２０１７年度　修士論文

Supporting telepresence communication by using augmented reality

指導教員　田中二郎 教授

早稲田大学大学院情報生産システム研究科

情報生産システム工学専攻

インタラクティブプログラミング研究

44161036-6　　LI RUI

ABSTRACT

Currently, remote communication can be executed in real time through mobile phones, personal computers, etc. However, it is difficult for the user's 3D information to be conveyed to party place, and the discussion cannot be performed smoothly. So we must support telepresence communication to deal with remote problems, we concentrate on the part of the party place. Using augmented reality technology, remote users can see virtual 3D model of local user. Local users can watch the steam video in party place at the same time. With Augmented Reality and 3D Reconstruction in this system, we can make telecom more realistic than 2D video calling. We can correctly deliver the user's conversation and expression. The remote viewing user can see the 3D view of the local people by moving to the right or left. The system consists of two parts. One is a local place and there are two RGB-D cameras in the same coordinate system, the angle of the two RGB-D cameras is 90 degrees. RGBD data is processed on the computer and reconstructed as a 3D model. The 3D model is displayed as a color point cloud. Local people can see the party place scene through the projector. The other is party place. This section uses Epson AR glasses, webcams and computers. There is a Unity application for receiving and rendering point clouds on Android OS. When install and start the application on Epson AR glasses Controller, we can see the real-time 3D video of the local place called 3D model. This 3D model can be moved, scaled and rotated. In the case of a web camera, capture the image of the party place using the camera, process the captured video data with the computer, and then the steam data will be exported.

Keywords: Telepresence communication, Augmented Reality, RGB-D Camera

ACKNOWLEDGMENTS

I would like to express my gratitude to all people who have encouraged and helped me with my research during this two years. First, I'd like to thank my supervisor Professor Jiro Tanaka. The door to Prof. Tanaka office was always open whenever I ran into a trouble spot or had a question about my research or writing. He steered me in the right direction whenever he thought I needed it. He not only gives me a lot of advice about my research and provide me the device that I can do experiment with, but also gives me a lot of help in life.

I would like to thank the other members of my laboratory for providing me with unfailing support and continuous encouragement throughout my days of study and through the process of researching and writing this thesis. When I did this research, they also give me a lot of suggestion. Without their help, I cannot go this far. This accomplishment would not have been possible without them.

Then I want to thank my parents who support my study in Japan. They always encourage and support me for my study and life.

At last I want to thank all those who helped me in study or life. Best wishes for them for their kindness. Thank you.

Contents

Chapter 1

Introduction

1.1. Background

Nowadays, there are many families are in a long-distance situation using remote communication. Remote communication can be done regardless of long-distance and even overseas. It can be performed in real time in various places by using a mobile phone or a PC, and so on. For cellular phones with a camera and apps that support remote communication like Skype, there are video calls that can convey audio and video to the conversation partner in real-time.

The phone call and even a video call and some instant messaging applications are thought as a good way to deal with long-distance problem. However, it should be improved from several aspects:

1. The senior people(Parents) need telepresence communication and more concern from us.

2. We often talk with looking at the listener's face, using a phone call can’t confirm whether he is listening or whether the talk is correctly transmitted.

3. When we use a video call, the 3D information e.g. the 3D motion of myself is difficult

to be transmitted to our parents. In this way, the discussion may not proceed smoothly.

1.2. Telepresence

1.2.1 Overview

Telepresence refers to a set of technologies which allow a person to feel as if they were present, to give the appearance of being present, or to have an effect, via telerobotics, at a place other than their true location.[1]

Telepresence requires that the users' senses be provided with such stimuli as to give the feeling of being in that other location. Additionally, users may be given the ability to affect the remote location. In this case, the user's position, movements, actions, voice, etc. may be sensed, transmitted and duplicated in the remote location to bring about this effect. Therefore information may be traveling in both directions between the user and the remote location[2].

A popular application is found in telepresence videoconferencing, the highest possible level of videotelephony. Telepresence via video deploys greater technic al sophistication and improved fidelity of both sight and sound than in traditional videoconferencing. Technical advancements in mobile collaboration have also extended the capabilities of videoconferencing beyond the boardroom for use with hand-held mobile devices, enabling collaboration independent of location[3].

Telepresence has been described as the human experience of being fully present at a live real-world location remote from one's own physical location. Someone experiencing video telepresence would therefore be able to behave, and receive stimuli, as though part of a meeting at the remote site. The aforementioned would result in interactive participation of group activities that would bring benefits to a wide range of users[4].

1.2.2 Techniques

Virtual presence (virtual reality), it is An Online Video Web Conference in an Office Telepresence refers to a user interacting with another live, real place, and is distinct from the virtual presence, where the user is given the impression of being in a simulated environment. Telepresence and virtual presence rely on similar user-interface equipment, and they share the common feature that the relevant portions of the user's experience at some point in the process will be transmitted in an abstract (usually digital) representation. The main functional difference is the entity on the other end: a real environment in the case of telepresence, vs. a computer in the case of immersive virtual reality[5].

About the Applications, in Connecting communities, Telepresence can be used to establish a sense of shared presence or shared space among geographically separated members of a group[6], Many other applications in situations where humans are exposed to hazardous situations are readily recognized as suitable candidates for telepresence. Mining, bomb disposal, military operations, the rescue of victims from fire, toxic atmospheres, deep sea exploration, or even hostage situations, are some examples. Telepresence also plays a critical role in the exploration of other worlds, such as with the Mars Exploration Rovers, which are teleoperated from Earth.

In Pipeline inspection, Small diameter pipes otherwise inaccessible for examination can now be viewed using pipeline video inspection.

In Remote surgery, the possibility of being able to project the knowledge and the physical skill of a surgeon over long distances has many attractions. Thus, again there is considerable research underway in the subject[7]. (Locally controlled robots are currently being used for joint replacement surgery as they are more precise in milling bone to receive the joints). The armed forces have an obvious interest in the combination of telepresence, teleoperation, and telerobotics can potentially save the lives of battle casualties by allowing them prompt attention in mobile operating theatres by remote surgeons[8].

Recently, teleconferencing has been used in medicine (telemedicine or telematics), mainly employing audio-visual exchange, for the performance of real-time remote surgical operations – as demonstrated in Regensburg, Germany in 2002. In addition to audio-visual data, the transfer of haptic (tactile) information has also been demonstrated in telemedicine[9].

1.3 Augmented Reality

1.3.1 Overview

Augmented reality (AR) is a live direct or indirect view of a physical, real-world environment whose elements are "augmented" by computer-generated or extracted real-world sensory input such as sound, video, graphics, haptics or GPS data[10]. It is related to a more general concept called computer-mediated reality, in which environment is modified (diminished or augmented) by a computer. Augmented reality can enhance one’s perception of reality, while virtual reality, in contrast, replaces the real world with a totally virtual one. Augmentation techniques are typically performed in real-time. Real time and semantic context with environmental elements, such as supplemental information will be overlaid. The information about the surrounding environment of users becomes digitally interactive with the help of advanced AR technology such as adding object recognition and computer vision[11]. Information about the objects and surrounding environment can be overlaid on the real world. This kind of visual information can be real or virtual. Real sensed or measured information such as electromagnetic radio waves overlaid in exact alignment with where they are in space[12]. Augmented reality combines the digital world with a person's perceived real world. There are currently two general definitions of augmented reality. One is proposed by Ronald Azuma from University of North Carolina in 1997. He believes that augmented reality includes three aspects: combining virtual object with reality, real time interaction and three dimensions[13]. Another is put forward in 1994 by Paul Milgram Fumio Kishino, which is named Milgram's Reality-Virtual continuum[14]. They took the real environment and the virtual environment as the two sides of the continuous system and the middle of system is called hybrid reality. Augmented reality is located between the real environment and mixed reality while augmented virtual is located between the virtual environment and mixed reality.

1.3.2 Techniques

Augmented reality is closely related to hardware, software and application level. In terms of hardware, an AR platform combines processors, displays, sensors, and input devices. In terms of software, the key to the AR system is how to integrate the augmented object with the real world. At the application level, it was used initially for the military and then extended to daily life.

Hardware components for augmented reality includes: input devices sensors display and processor. Modern mobile computing devices such as tablet computers and smartphones contain these elements which usually include a camera and MEMS sensors such as GPS, solid state compass and accelerometer to make them suitable AR platforms[15]. Current AR hardware includes optical projection systems, monitors, mobile devices, head mounted displays, head up displays and computers. How realistically virtual object with the real world are integrated is a key measure of AR systems. The software should extract real world coordinates from camera images, which are independent from the camera. Usually those approaches include two parts. The first stage is to detect interest points, optical markers in the camera images. This step can use feature detection methods like thresholding, edge detection, blob detection, corner detection or other image processing approaches. Next stage restores a real world coordinate system from the data obtained in the previous stage. Some methods assume objects with known geometry or markers are present in the scene. In some of those cases 3D structure of the scene should be calculated in advance. If part of the scene is unknown simultaneous localization and mapping can map relative positions. Without information about scene geometry, structure from motion methods like bundle adjustment are adopted. Mathematical approach adopted in the second stage include geometric algebra, projective geometry, kalman and particle robust statistics, rotation representation with exponential map and nonlinear optimization[16]. In Architecture, AR can aid in visualizing building projects. Computer-generated images of a structure can be superimposed into a real life local view of a property before the physical building is constructed there; this was demonstrated publicly by Trimble Navigation in 2004. AR can also be employed within an architect's workspace, rendering animated 3D visualizations of their 2D drawings. Architecture sight-seeing can be enhanced with AR applications, allowing users viewing a building's exterior to virtually see through its walls, viewing its interior objects and layout[17][18][19]. And in Spatial immersion and interaction, Augmented reality applications, running on handheld devices utilized as virtual reality headsets, can also digitalize human presence in space and provide a computer generated model of them, in a virtual space where they can interact and perform various actions. Such capabilities are demonstrated by "Project Anywhere", developed by a postgraduate student at ETH Zurich, which was dubbed as an "out-of-body experience"[20][21].

Chapter 2

Research Goal and Approach

2.1 Problem

The problem we are targeting is helping the users in separated place get a feeling of being party together during a telepresence communication. By the common long-distance communication, the conversation and the face express cannot be transmitted to the other side smoothly and realistic. For example, based on the two-dimensional object, when the users who are in party place watch the object directly, they cannot see the left side or right side of the object, and we want to blend the three-dimensional captured and reconstructed object into AR system, so if the users in party place move to left slightly then the right side of the local place person can be seen soon, and if the users in party place move to right slightly then the left side of the local place person can be seen soon too. And still the problem of party place, the scene of party place cannot be transmitted to the local place in real-time while we need a two-way communication for this two side. In this case, we need some devices which can capture and process the image of party place and then the local place can receive the data. Although we figure out the problem of the 3D model which can be seen in the party place, the user in party place may want to fix the size of the 3D model, so if the user who uses the AR glasses in party place could zoom in or zoom out the 3D model by rotating and enlarging or shrinking, the communication and interaction of this system will be better and more efficient.

2.2 Research Goal

In this research, we solve the problems of existing remote communication with video, It is aimed to make it possible to communicate with long-distance people be realistic and smooth by using the RGB-D camera and Augmented Reality and RGB-D cameras. In this case, Local place people can see the video steam of the party place and party place people can see the three-dimensional image of local place person, and this is a 2-way communication system.

2.3 Approach

In this research, in order to solve the problem of remote communication with video, when make remote communication with video, we propose to present the three-dimensional model of the local place person to the party place.

By displaying the three-dimensional model, it is possible to express three-dimensional motion that cannot be expressed in two dimensions, if the viewer moves slightly to the right, the user can see the model's left side.

Although this is natural in the real world, in mainstream video calls now, even if the user moves to the right, it is only looking at the display diagonally, there is no change in the displayed image and it will be difficult to see from the diagonal direction.

Express this on the display of the computer like the real world It is difficult to do, because it is difficult to move the display in the direction of the user's face.

In this research, this problem was solved by using an augmented reality glassed, and the glasses will be used for displaying three-dimensional model.

And in the same time, we will use the web camera to capture the scene of party place, especially the party place people’s faces and bodies, so the communication system will be in two-way and interactive.

Chapter 3

System Design

In this chapter, we introduce the designs of local place and party place and the explain how to realize such sensation.

3.1 System Overview

The system which is shown in Figure 3.1 will consist of two parts. One is Local Place, there are two RGB-D cameras placed in the same coordinate system and two RGB-D cameras' angle is 90 degrees. The RGBD data will be processed in computers and then will be reconstructed as a 3D model. The 3D model will be in the form of colored point clouds. Person in local place can see the scene of Party Place by the projector. The other is Party Place, there is a Unity application for receiving and rendering point clouds in Android OS. Once install and start the application, we can see the 3D model which can be moved, scaled and rotated. Using the camera to capture the view of Party Place and its steaming data will be sent to the local place.

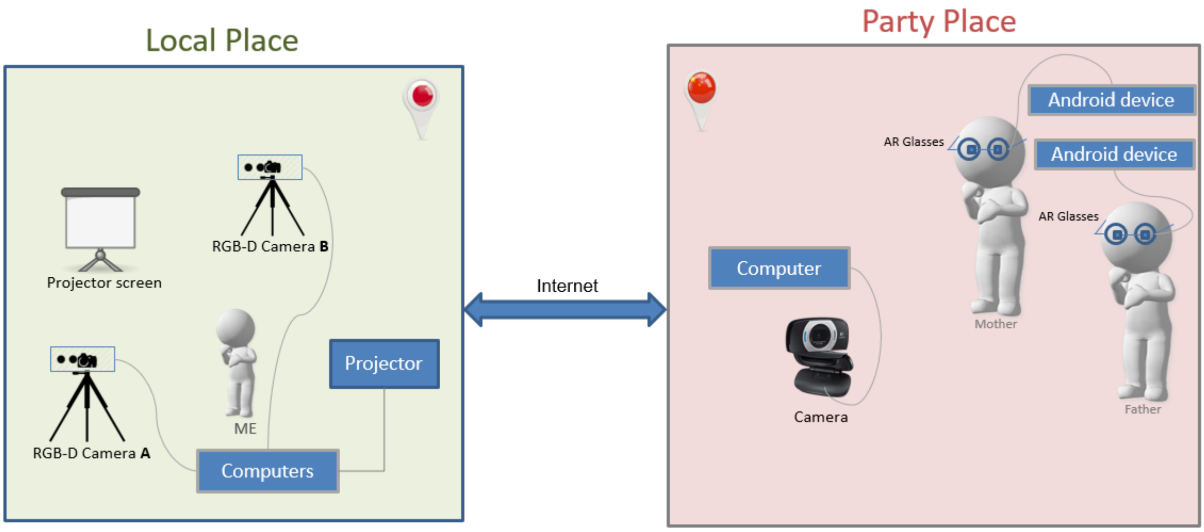


Figure 3.1 System Overview

3.2 Local Place

3.2.1 Local Place Setup Overview

Figure 3.2.1 shows the overview of the local setup. It includes the RGB-D Cameras, a projector and its screen and 3 desktop PCs. The local user sits at the table to use our system. In Local Place, there are two RGB-D cameras placed in the same coordinate system and two RGB-D cameras' angle is 90 degrees. The RGBD data will be processed in computers and then will be reconstructed as a 3D model. The 3D model will be in the form of colored point clouds. Person in local place can see the scene of party place by the projector.

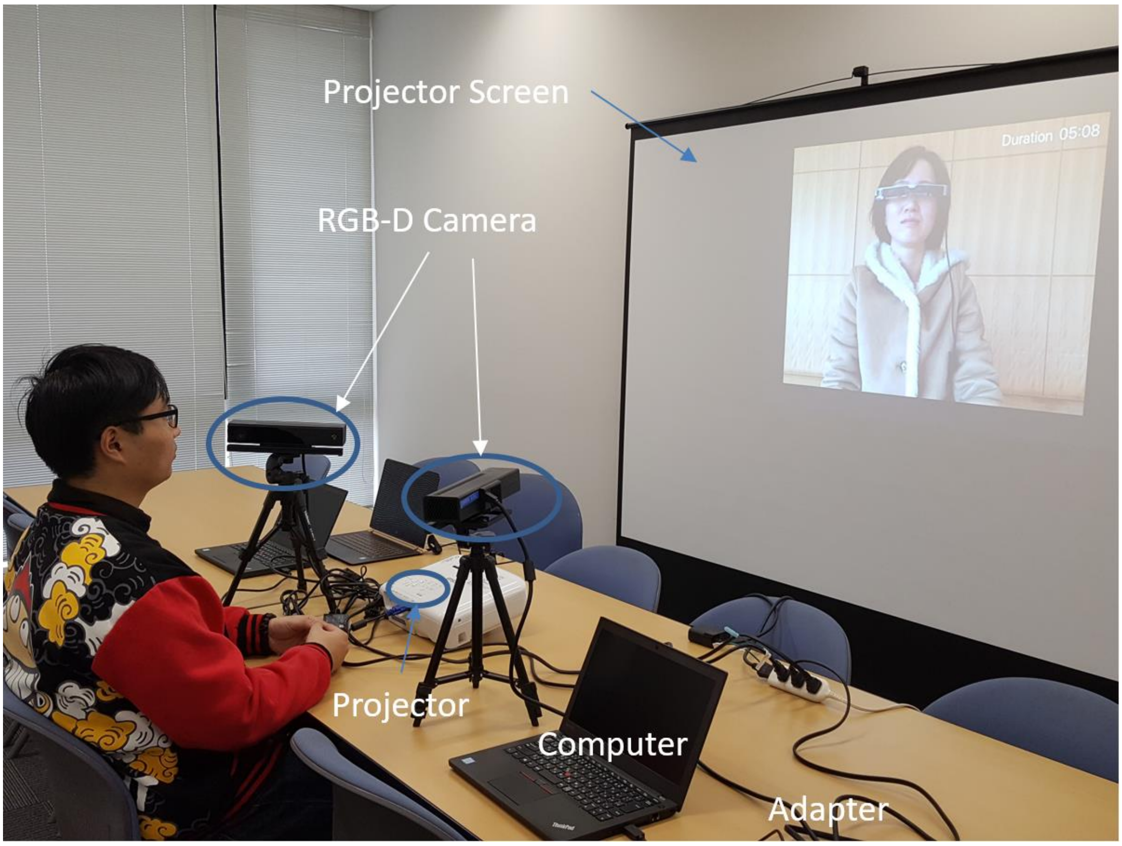


Figure 3.2 Local place setup overview

3.2.2 RGB-D Camera

In this paper, we use Microsoft Kinect v2 be a RGB-D camera. Kinect (codenamed Project Natal during development) is a line of motion sensing input devices by Microsoft for Xbox 360 and Xbox One video game consoles and Microsoft Windows PCs. Based around a webcam-style add-on peripheral, it enables users to control and interact with their console/computer without the need for a game controller, through a natural user interface using gestures and spoken commands. [22]

The Kinect V2 has greater accuracy with three times the fidelity over its predecessor and can track without visible light by using an active IR sensor. It has a 60% wider field of vision that can detect a user up to 3 feet from the sensor, compared to six feet for the original Kinect, and can track up to 6 skeletons at once. It can also detect a player's heart rate, facial expression, the position and orientation of 25 individual joints (including thumbs), the weight put on each limb, speed of player movements, and track gestures performed with a standard controller. The color camera captures 1080p video that can be displayed in the same resolution as the viewing screen, allowing for a broad range of scenarios. In addition to improving video communications and video analytics applications, this provides a stable input on which to build interactive applications. Kinect's microphone is used to provide voice commands for actions such as navigation, starting games, and waking the console from sleep mode. [23] [24] The Kinect Sensor and its Adapter are shown in Figure 3.3.

Figure 3.3 Kinect Sensor(left) and its Adapter(right)

3.2.3 Projector

This projection screen will be installed at local place to display the video steam of party place.

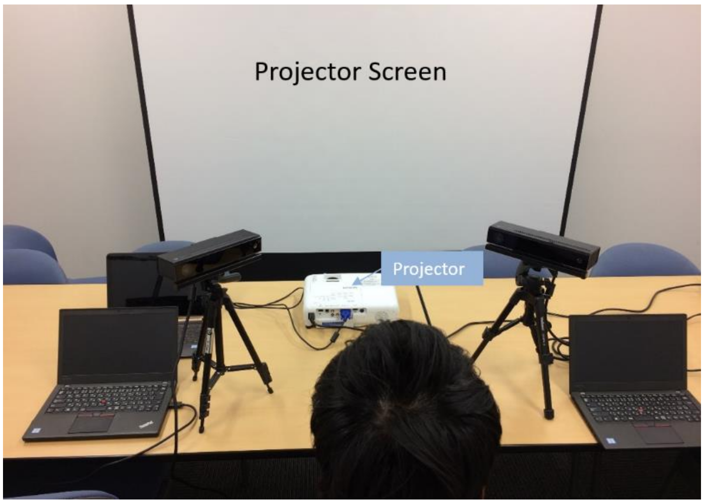


Figure 3.4 Projector Screen

3.2.4 Computers

In this part, we use projector to display the video steam of the party place and we need a computer for process the live video steam making. Table 3.1 and Figure 3.5 shows the information and photo of the PC we used.

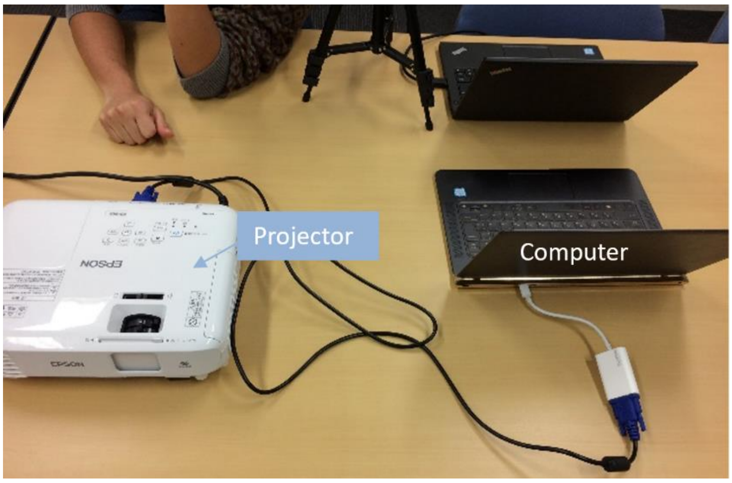


Figure 3.5 Projector and Computer

And for 3D capturing and 3D reconstruction, we used two high level PC, Table 3.2 shows the information of the PC we used.

Table 3.1 Local Laptop PC-1

Laptop PC	
Operating System	Microsoft Windows 10
Graphics Card	Intel(R) HD Graphics 520
CPU	Intel Core i5-6500 2.5GHz
RAM	4.00 GB
Internet Connection	Wi-Fi 802.11n

Table 3.2 Local Laptop PC-2

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

3.3 Point Clouds:

A point cloud is a set of data points in some coordinate system. In a three-dimensional coordinate system, these points are usually defined by X, Y, and Z coordinates, and often are intended to represent the external surface of an object. In computer vision and pattern recognition, point set registration, also known as point matching, is the process of finding a spatial transformation that aligns two point sets. The purpose of finding such a transformation includes merging multiple data sets into a globally consistent model and mapping a new measurement to a known data set to identify features or to estimate its pose. A point set may be raw data from 3D scanning or an array of rangefinders. For use in image processing and feature-based image registration, a point set may be a set of features obtained by feature extraction from an image, for example, corner detection. Point set registration is used in optical character recognition, augmented reality and aligning data from magnetic resonance imaging with computer-aided tomography scans. [25] As the Figure 3.6.1 and 3.6.2 shows, we can see the Point Clouds of the local place person who are captured by RGB-D camera. The left side of the figure shows the point clouds and the right side of the figure shows the color image and the Skeleton information of the local place user.

Figure 3.6.1 Point Clouds and Color image

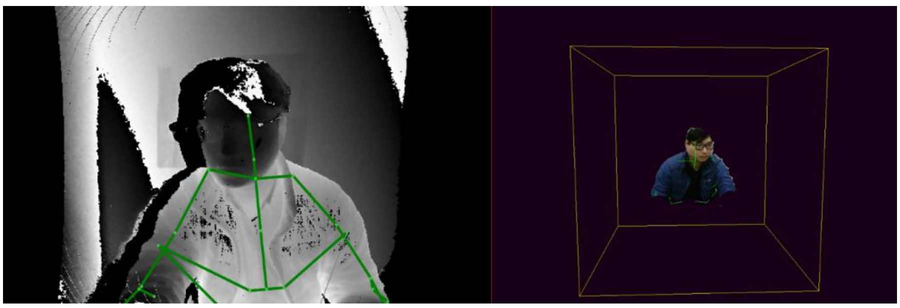
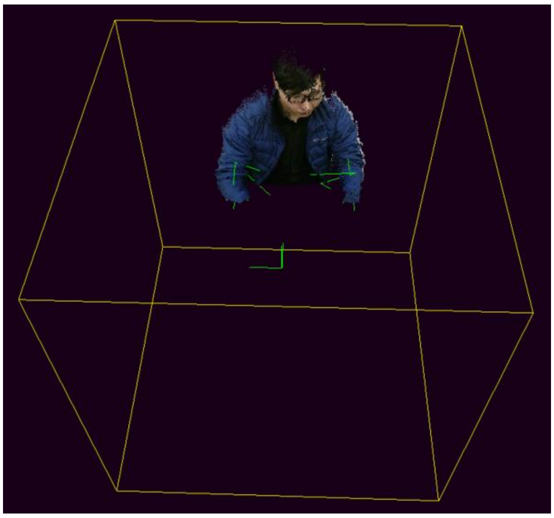
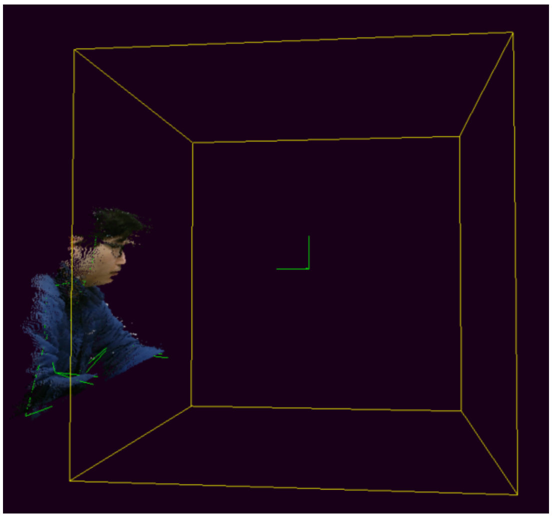


Figure 3.6.2 Point Clouds and Depth image

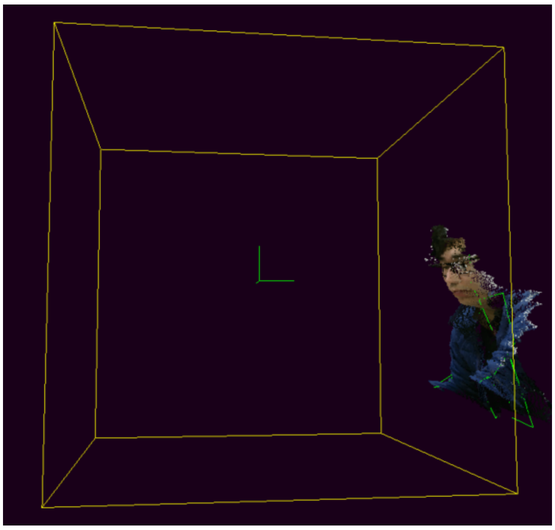
By rotating the 3D model, we can see the left, right, above and below side of the point clouds. As the Figure 3.7 shows. The green right-angle line is the location of RGB-D camera, the other green lines are the skeleton information.



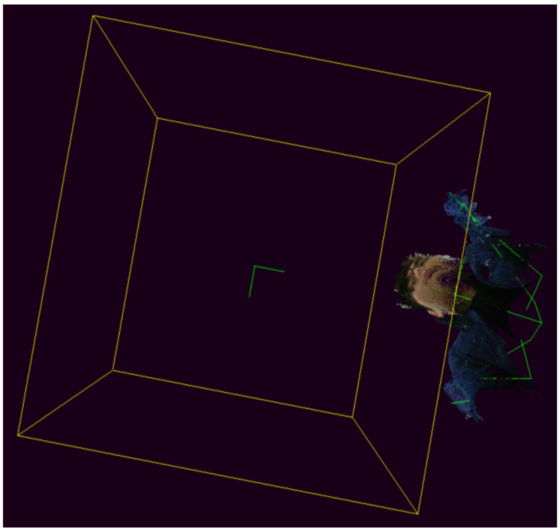
(a) Above



(b) Right



(c) Left



(d) Below

Figure 3.7 The above(a), right(b), left(c) and below(d) side of 3D Point Clouds.

3.4 Party Place

3.4.1 Party Place Setup Overview

In the party place as the Figure 3.8.1 shows, we use augmented reality see-through glasses and its Android OS device and web camera. Using the web camera, the scene of party place especially the face and the movement of the people who are in party place can be captured and then the video steam will be sent to local place. And there a Unity application in Android OS, once deploy the application, the 3D image of the person who is in local place can be seen in the smart glasses.

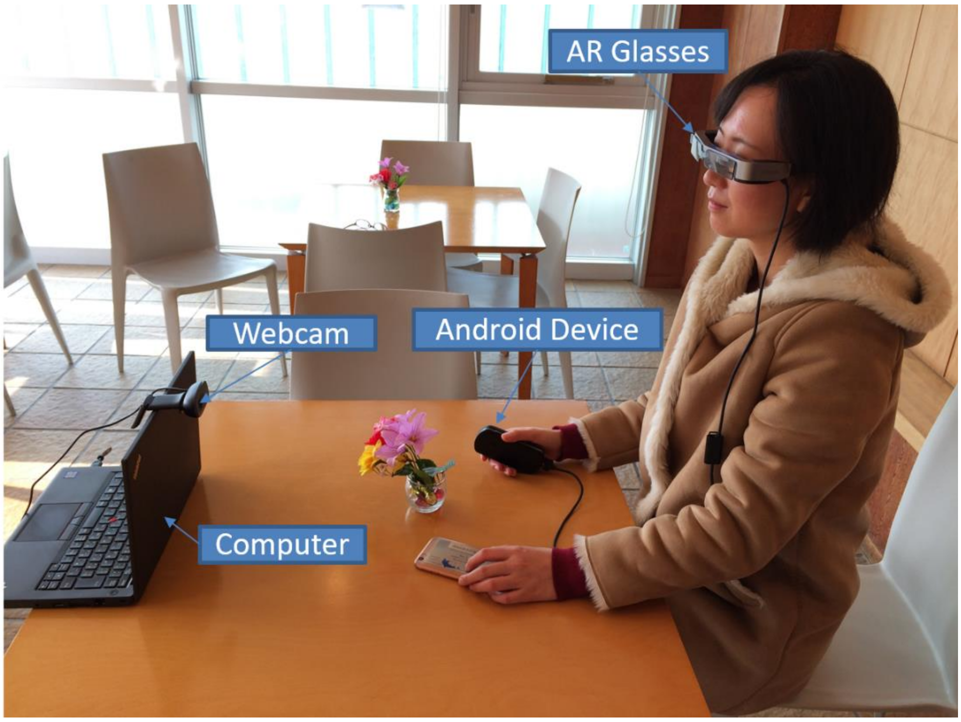


Figure 3.8.1 Party Place Setup

3.4.2 Smart Glasses

The Party Place user wears an augmented reality smart glasses-EPSON Moverio BT-300 just like Figure 3.8.2 shows which is light and supports an HD binocular displays. It packs with a motion-tracking sensor to detect the user's facing direction and a wireless module to exchange information with the local side via the Internet. It presents a semitransparent display on top of the physical world while allows the user to view the physical world clearly. It provides an audio output with an earphone.

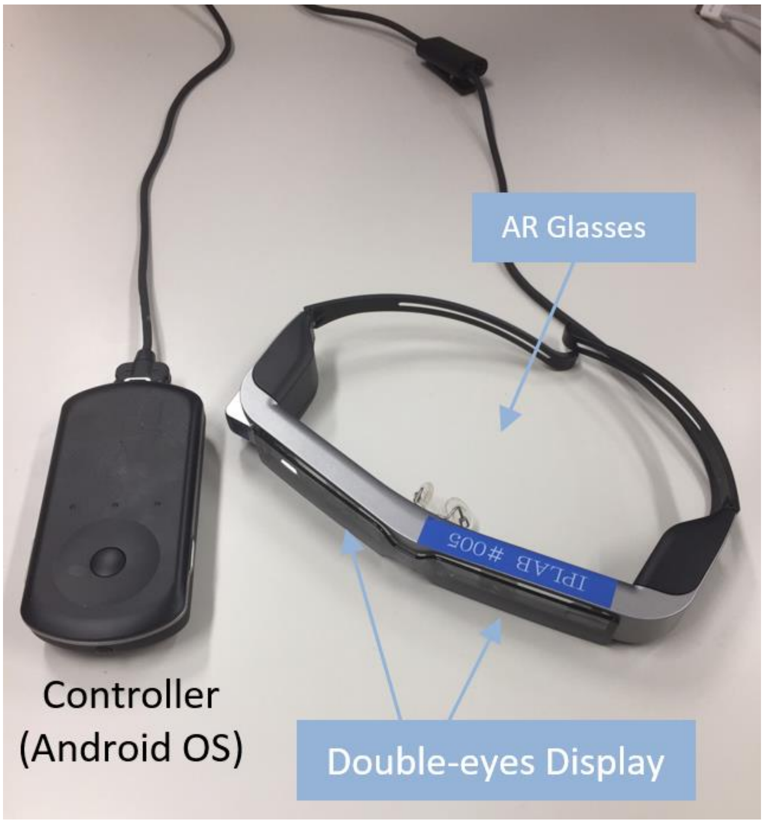


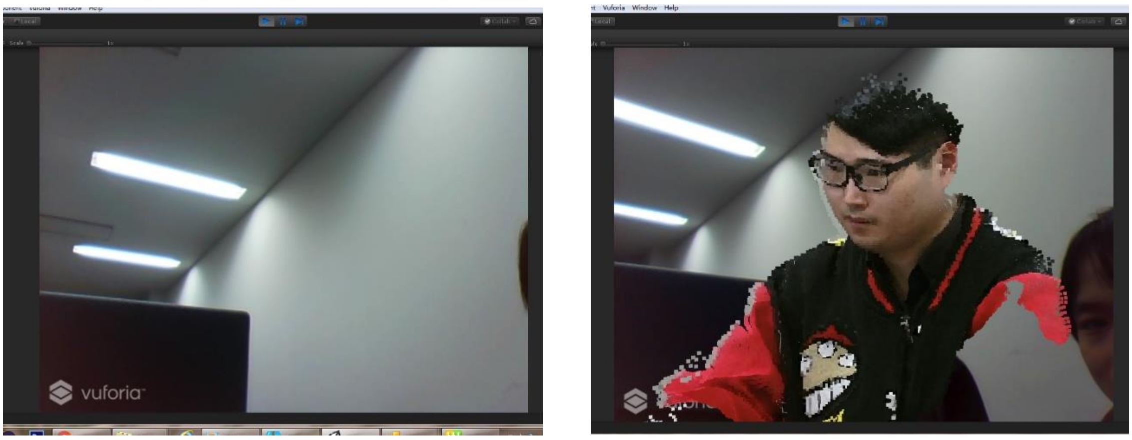
Figure 3.8.2 AR Glasses Set

As the Figure 3.8.3 shows, the user in party place is wearing the AR glasses and holding a controller in his hand. The user can control the model by operating the controller which is a mobile device based on Android OS.

Figure 3.8.3 Party Place user wearing a AR glasses

3.4.3 Point clouds fusion

In Figure 3.8.4, we will show the comparison of two conditions which are both in AR system but one is live 3D model displaying and other is not.



a.Without live 3D model   b. Live 3D model displaying

Figure 3.8.4 Comparison

3.4.4 Live 3D model fusion

In this part, based on the captured live 3D model, we use the Vuforia Augmented Reality SDK to display the live 3D image in a AR system, Vuforia is an Augmented Reality Software Development Kit (SDK) for mobile devices that enables the creation of Augmented Reality applications. It uses Computer Vision technology to recognize and track planar images (Image Targets) and simple 3D objects, such as boxes, in real-time. This image registration capability enables developers to position and orient virtual objects, such as 3D models and other media, in relation to real-world images when these are viewed through the camera of a mobile device. The virtual object then tracks the position and orientation of the image in real-time so that the viewer's perspective on the object corresponds with their perspective on the Image Target so that it appears that the virtual object is a part of the real-world scene. Additional features of the SDK include localized Occlusion Detection using 'Virtual Buttons', runtime image target selection, and the ability to create and reconfigure target sets programmatically at runtime. The SDK supports both native development for iOS and Android while also enabling the development of AR applications in Unity that are easily portable to both platforms.[26] In this case, all the 3D information of the point clouds will be compiled in the Unity 3D, we just set up the server and the Unity implementation on a same computer. As the Figure 3.9 shows, we can see that the point clouds are displayed in an AR system, the live dynamic 3D image only will be displayed with any background.

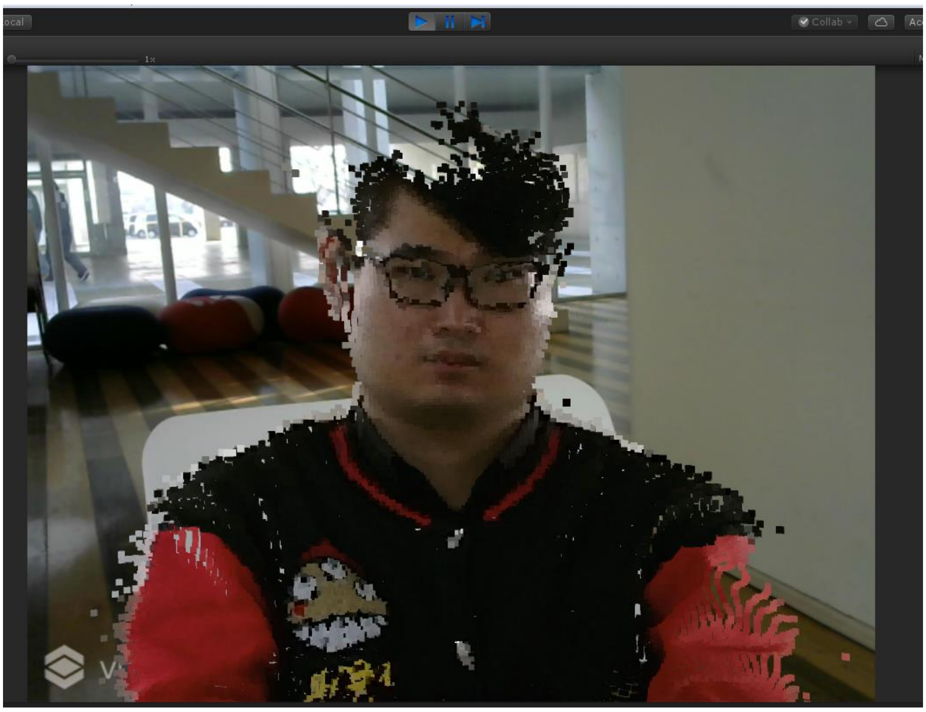


Figure 3.9 Live 3D image fusion

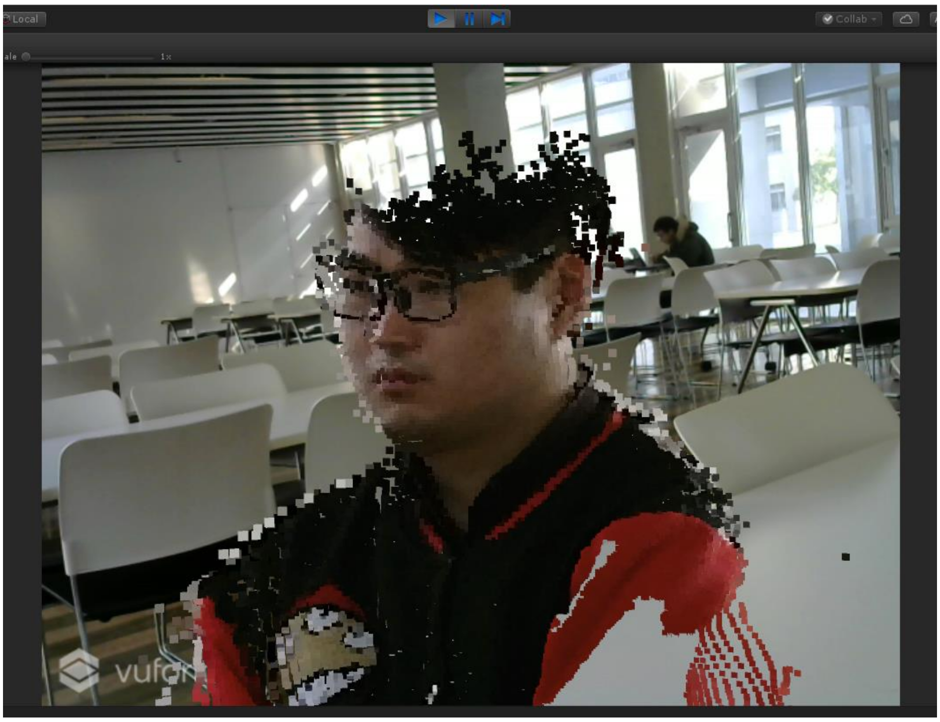


Figure 3.9.1 Live 3D image (Left Side)

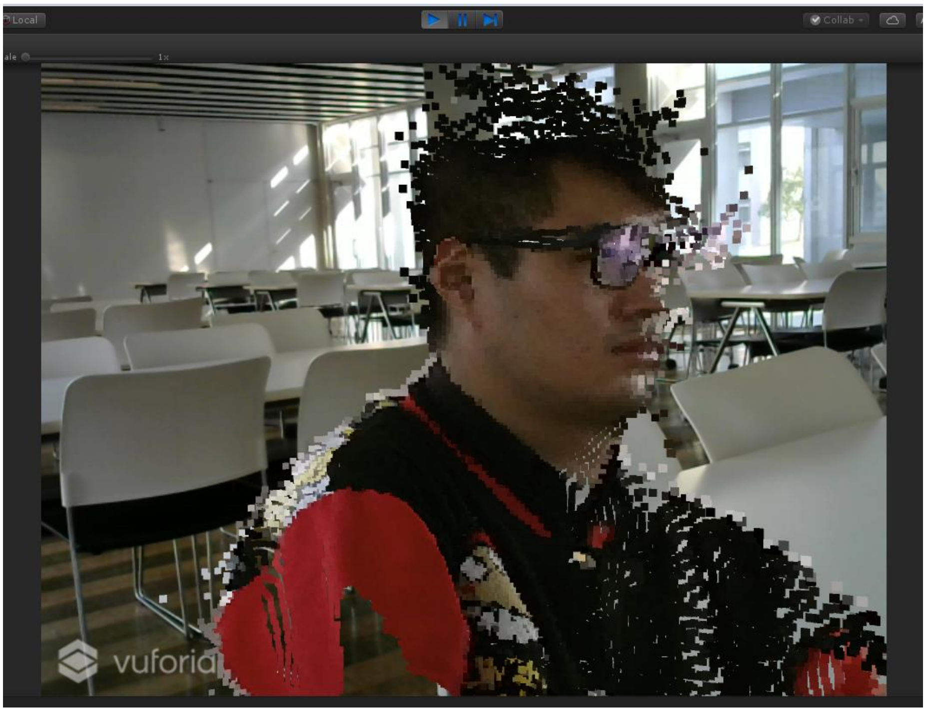


Figure 3.9.2 Live 3D image (Right Side)

We just replace the 3D model to a new location which can show the left side and the right side by rotating the 3D model using the Epson Glasses' controller. As the Figure 3.9.1 and Figure 3.9.2 shows, we can see the left side and right side of the 3D model.

3.4.5 Web Camera

We use web camera to capture the scene of the party place, this is a normal 2D camera which can take the live video and the data can be sent to the computer to be processed by connecting to the PC, Figure 3.10 shows the camera we use.

Figure 3.10 Web Camera

3.4.6 Computer

The laptop PC on the party place side is used to analyze data from web camera and process then send the live video to local place side. Table 3.3 shows the information of the PC we used.

Table3.3 Remote Laptop PC

Laptop PC	
Operating System	Microsoft Windows 10
Graphics Card	Intel(R) HD Graphics 520
CPU	Intel Core i5-6500 2.5GHz
RAM	4.00 GB
Internet Connection	Wi-Fi 802.11n

Chapter 4

System Implementation

4.1 Development Environment

The implementation will mostly consist of two parts, one is 3D model reconstruction, and other is AR part.

For 3D model reconstruction, we use Microsoft Kinect SDK and Visual Studio(based on C++), and Open CV. We develop the AR system using Unity 5 and development language is C#, and we use Android SDK and use EPSON Moverio controller as terminal. As for Unity AR programming, we use Vuforia as SDK. Windows 10 is used as server environment and MySQL is used as database. Sockets will be used to implement the communication between the application and the server.

4.2 System Hardware Overview

Our system's hardware includes two parts: the local place side and the party place side. Figure 5.2 shows the system hardware and information overview.

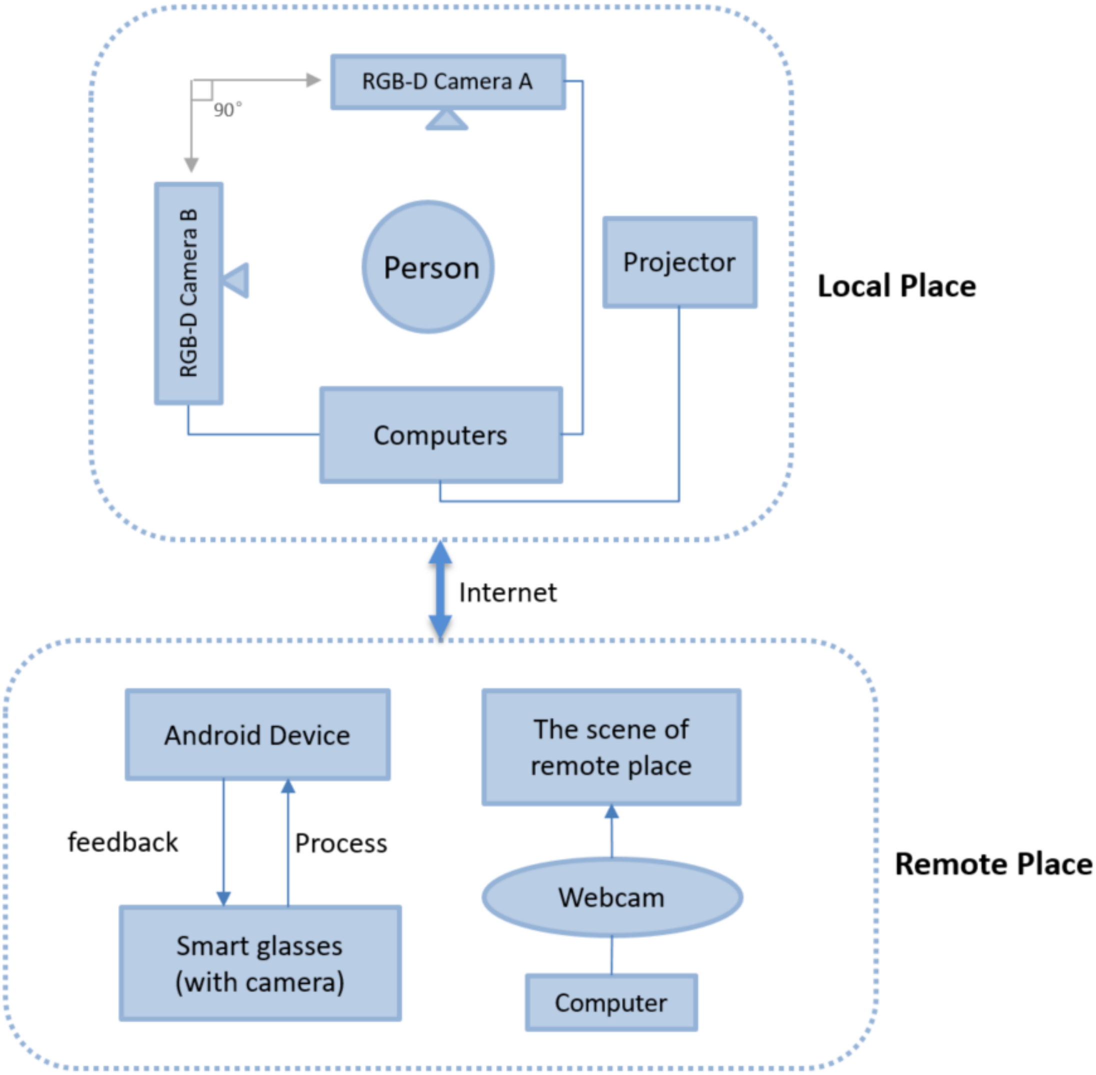


Figure4.1 Hardware Overview

4.3 RGB-D Data

Figure 5.3 shows the overview of the local setup. It includes the RGB-D Cameras, a projector and its screen and 3 desktop PCs. The local user sits at the table to use our system. In Local Place, there are two RGB-D cameras placed in the same coordinate system and two RGB-D cameras' angle is 90 degrees. The RGBD data will be processed in computers and then will be reconstructed as a 3D model. The 3D model will be in the form of colored point clouds. Person in local place can see the scene of Remote place by the projector. [27]

The SDK provides a sophisticated software library and tools to help developers use the rich form of Kinect-based natural input, which senses and reacts to real-world events. The Kinect and the software library interact with your application, as shown in Figure 4.2.

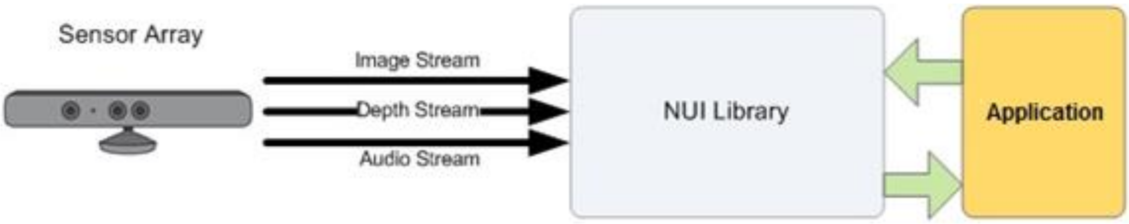


Figure 4.2 Hardware and Software Interaction with an Application

The components of the SDK will be shown in Figure 4.3.

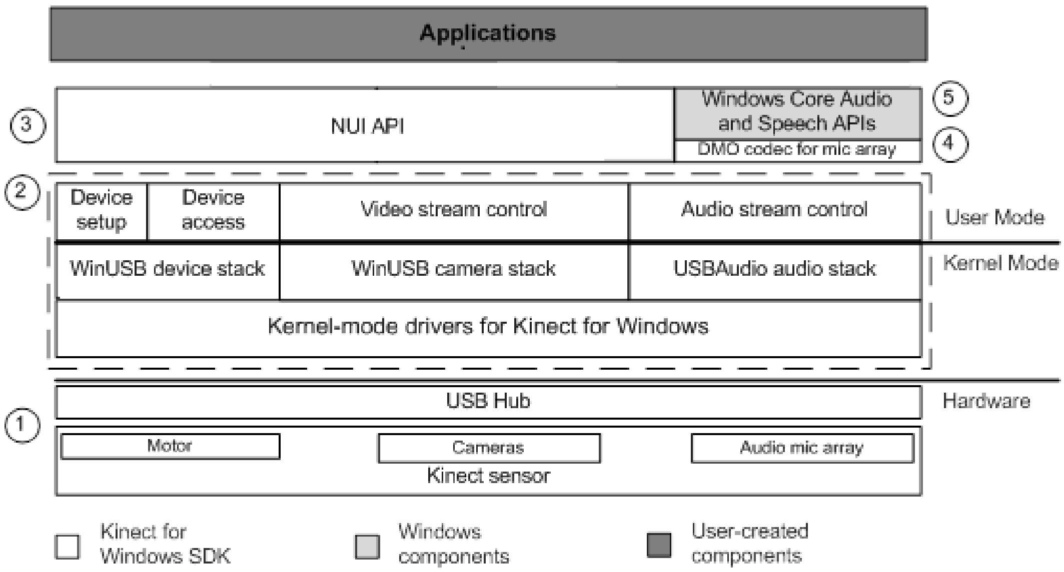


Figure 4.3 SDK Architecture

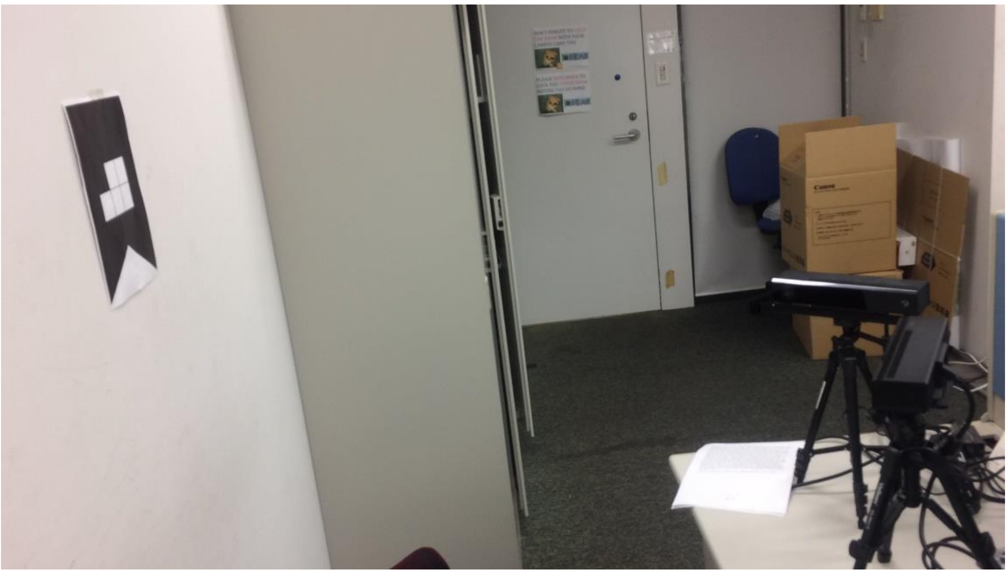


Figure 4.3.1 The usage scenario of RGB-D camera

In our research, we use two Kinects and one marker to realize the capturing for the local place person, just as Figure 4.3.1 shows, the angle of two Kinects is set up as 90 degrees, and the wall behind the desk is stick with a marker, we use this marker to calibrate the localization of this two Kinects and the localization of the person who will be captured.

4.4 Epson Moverio Development

We use unity 3D to blend the live 3D model, as we know they are point clouds of local place’s person, and then we use Epson moverio(AR glasses) to display the point clouds. So we can see the live 3D image of local place’s person in the AR glasses.

By using the Epson original APIs as well as the standard Android APIs, the BT-300 can use specific functions that are not built-in to the standard Android system. This product uses a Si-OLED (Organic EL Pane) for display panel. By using the Epson original APIs as well as the standard Android APIs, the BT-300 can use specific functions that are not built-in to the standard Android system. We will explain the development of the Epson moverio step by step:

1. Download Android Studio
2. Download the JDK
3. Install Android Studio
4. Platform-tools and SDK Platform
5. USB driver settings
6. Connecting the BT-300 to a computer (Figure 4.6.1)
7. Use the SDK provided by Epson

In our research, for the step 7, just as the Figure 4.6.2 shows, we downloaded the Epson moverio SDK and Unity plugin package from the homepage of the technical support for Epson company.

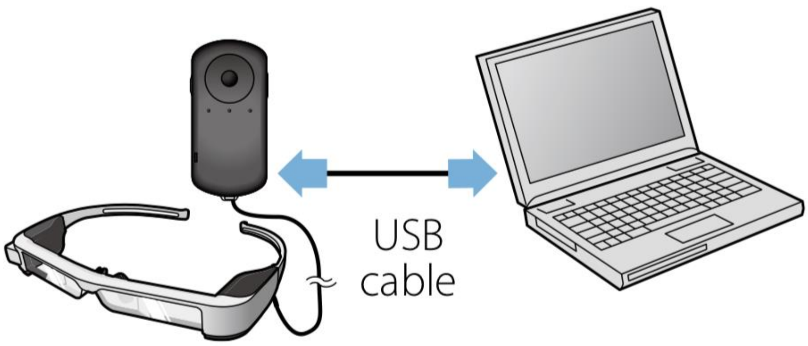


Figure 4.6.1 Connection of the devices

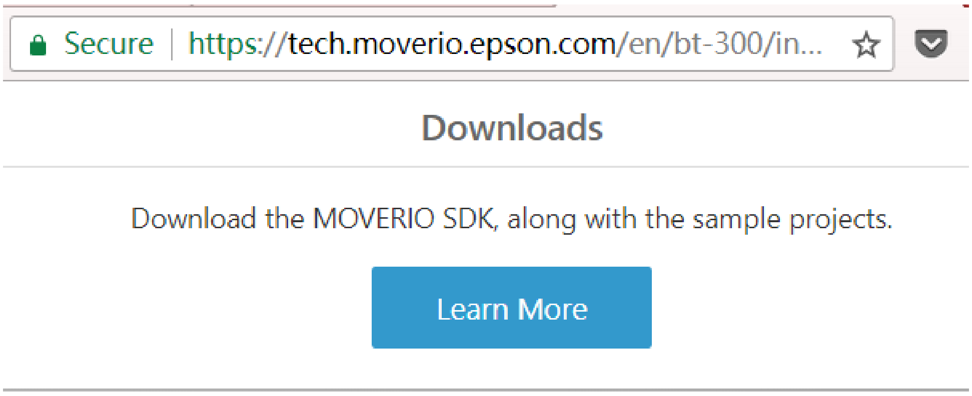


Figure 4.6.2 Downloads of SDK and Unity plugin

At last we import the Unity plug-in Assets to the Unity 5.5.0 and then run the project, the configuration of the Epson Moverio will be completed.

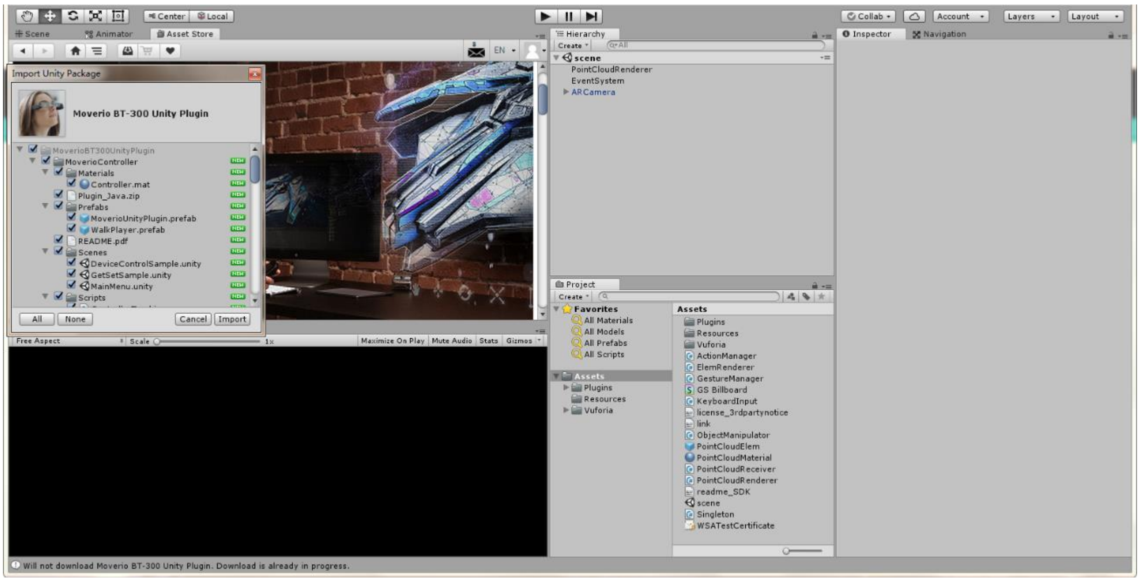


Figure 4.6.3 Epson Moverio SDK and its Unity plugin

I just let three members of our lab to help me. One of them wear the AR glasses and use the glasses to view other two members, when the viewer run the Unity App in the controller of the AR glasses, just as he will see my live 3D dynamic image (Point Clouds) and other two lab members.

Figure 4.6.1 Viewpoint of party place

4.5 ICP Algorithm

If the correct correspondences are known, the correct relative rotation/translation can be calculated in closed form just as the Figure 4.4 shows.

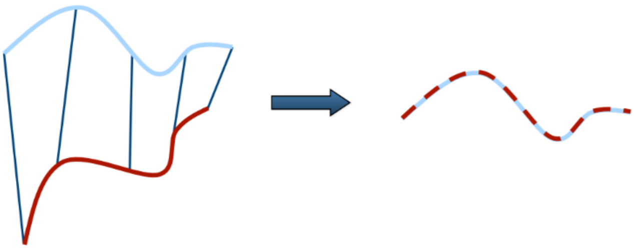


Figure 4.8.1

If the correct correspondences are not known, it is generally impossible to determine the optimal relative rotation/translation in one step, just as the Figure 4.5 and Figure 4.6 shows.

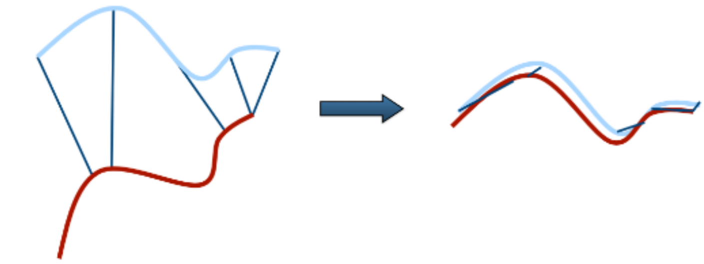


Figure 4.8.2

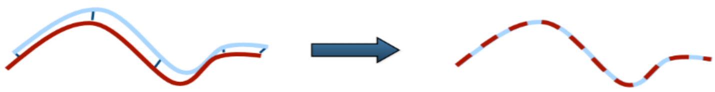


Figure 4.8.3

In this paper, in order to achieve the perfect fit of two 3D live dynamic image which captured by two RGB-D cameras, so we use ICP algorithm to make the two 3D images overlap, and the reconstructed 3D model will be more realistic and vivid.

4.6 Network Communications:

In this part, we use network socket to realize the communications. A network socket is an internal endpoint for sending or receiving data at a single node in a computer network. Socket originated in Unix, and one of the basic Unix / Linux philosophy is "everything is a file," you can use "open -> read write write / read -> close close" mode to operate. My understanding is that Socket is an implementation of the mode, socket is a special kind of file, and some socket functions are operations on it (read / write IO, open, close). Socket function corresponds to the ordinary file open operation. Ordinary file open operation returns a file descriptor, and socket () is used to create a socket descriptor, it uniquely identifies a socket. The socket descriptor and file descriptor, follow-up operation is used to it, as an argument, through which to read and write operations.

Just as different values can be passed to fopen to open different files. Create a socket, you can also specify different parameters to create a different socket descriptor, socket function of the three parameters were: Domain, Type, and Protocol.

Client-server model is that users do a series of operations on the database, which requires master-slave synchronization will be written to a log (binary form), and then the main database binary log distributed to the slave database through the network interface, from the database through the execution of the sequence of operations in the log, the data is written to the slave database summary, so as to achieve the master-slave data consistency.

Part of the above specific analysis:

1.When the data on the master changes, the events (insert, update, delete) changes in accordance with the order to write to the binlog. (Statement data change format)

2.When the slave is connected to the master, the master machine opens the binlog dump thread for the slave.

3.When the master binlog changes, binlog dump thread will notify the slave, and the corresponding binlog content sent to the slave.

Chapter 6

Related Work

6.1 Research about 3D model making and AR

The paper written by Jozaki et al.[28] used 2 RGB-D cameras to capture the half body's 3D information and made the 3D model be seen in a HMD. In this research, they developed a remote communication support system for 3D display of the figure of the other party by using augmented reality. By using this system, user can see the 3D model of the remote partner, making it easier to convey information to assist communication such as gestures and hand gestures. They have been developing on the assumption of a one-to-one call, but we plan to make many-to-many calls so that we can use it even during conference calls in the future.

6.2 Research about RGB-D camera

The paper written by Peter Henry et al.[29] investigated how RGB-D cameras can be used for building dense 3D modeling of indoor environments. In this paper, they investigated how such cameras can be used for building dense 3D maps of indoor environments. Such maps have applications in robot navigation, manipulation, semantic mapping, and telepresence. They present RGB-D Mapping, a full 3D mapping system that utilizes a novel joint optimization algorithm combining visual features and shape-based alignment. Visual and depth information are also combined for view-based loop-closure detection, followed by pose optimization to achieve globally consistent maps. They evaluate RGB-D Mapping on two large indoor environments and show that it effectively combines the visual and shape information available from RGB-D cameras.

6.3 Research about telepresence and AR glasses

The paper written by Sean Fanello et al.[30] made a system named Holoportation for Virtual 3D Teleportation in Real-time. They present an end-to-end system for augmented and virtual reality telepresence, called Holoportation. The system demonstrates high-quality, real-time 3D reconstructions of an entire space, including people, furniture and objects, using a set of new depth cameras. These 3D models can also be transmitted in real-time to remote users. This allows users wearing virtual or augmented reality displays to see, hear and interact with remote participants in 3D, almost as if they were present in the same physical space. From an audio-visual perspective, communicating and interacting with remote users edges closer to face-to-face communication. This paper describes the Holoportation technical system in full, its key interactive capabilities, the application scenarios it enables, and an initial qualitative study of using this new communication medium.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this research, by using augmented reality technique, the remote users can see local user's virtual 3D dynamic image in AR glasses, the local user can see the remote place's streaming video in the same time. Using augmented reality technique and 3D reconstruction in this system can make remote communication be more realistic compared with phone call and 2D video call. The users' talk and facial expression can be correctly transmitted. And remote viewing users can see the local person's 3D view by moving to the right or left.

The system consists of two parts. One is Local Place, there are two RGB-D cameras placed in the same coordinate system and two RGB-D cameras' angle is 90 degrees. The RGBD data will be processed in computers and then will be reconstructed as a 3D model. The 3D model will be in the form of colored point clouds. Person in local place can see the scene of Remote place by the projector. We use Epson AR glasses and web camera and one computer in party place. there is a Unity application for receiving and rendering point clouds in Android OS. We can see the live 3D dynamic image of local place's person in AR glasses. The 3D model can be moved, scaled and rotated. We use the camera to capture the view of remote place and the captured video data can be processed in the computer and then export the steaming data.

In the same time, the live steaming data can be sent to the local place, in the computer of the local place, the live steaming video can be displayed and the person in local place can see the scene of party place by the live video steam.

7.2 Future Work

In the future, we plan to add two or more RGB-D cameras in local place, which can make the captured image more solid and realistic. At the party place, we need to add real-time voice communication modules, which can make the communication more realistic and fluent. Based on the importance of cross-platform, we will also port the Android applications which exported from the Unity to other platforms so that the system can be compatible with more mobile devices. Finally, we want to integrate and package all the modules into one application. This will make the whole system more concise and easier to use.

Reference

[1] Telepresence, teleoperation & telerobotics, *IDC Technologies*, www.idc-online.com

[2] Ian Yeoman, *Business & Economics 2050 Tomorrow's Tourism*, pp. 102, 2012.

[3] René Visser, Michaël Gründeman, Nick Bakker, Education, EXIN Green IT Foundation, pp. 69, 1970.

[4] Video Conferencing: A guide to making a Telepresence Business Case Matt Walker, 2007.

[5] W Zeng, Using Immersive Virtual Reality to Create Presence in Online Shopping *Australasian Conference on Information Systems*, Zeng & Richardson, Hobart, Australia, 2017.

[6] WAS Buxton, Telepresence: Integrating Shared Task and Person Spaces.

[7] Sandor, Jozsef; Haidegger, Tamas; Benyo, Zoltan (2012). "Surgery in Space: The Future of Robotic Telesurgery". *Surgical Endoscopy*. 26 (1): pp. 681-690.

[8] Murphy, Challacombe; Khan; Dasgupta (2006). "Robotic Technology in Urology". *Postgraduate Medical Journal*. 82 (973): pp. 743-747.

[9] Westwood et al. (Eds.), Medicine meets Virtual Reality 12: Studies in Health Technology & Informatics. W. Kahled, S. Reichling, O.T. Bruhns, H. Böse, M. Baumann, S. Egersdörfer, D. Klein, A. Tunayer, H. Freimuth, A. Lorenz, A. Pessavento & H. Ermert,

Palpation Imaging using a Haptic System for Virtual Reality applications in Medicine. pp. 147-153, IOS Press, 2004.

[10] Schuettel, Patrick, The Concise Fintech Compendium. *Fribourg: School of Management Fribourg/Switzerland,* 2017.

[11] S.M.Abbas, S.Hassan, and J.Yun, “Augmented reality based teaching pendant for industrial robot,” in *Control, Automation and Systems(ICCAS),2012 12th International Conference on.* IEEE, 2012, pp. 2210-2213.

[12] M. GHINEA, G. C. DEAC, and C. N. GEORGESCU, “Improving the quality of image in virtual reality applications for industry,” *emergence,* vol. 9, p. 10.

[13] R. T. Azuma, “A survey of augmented reality,” *Presence: Teleoperators and virtual environments,* vol. 6, no. 4, pp. 355-385, 1997.

[14] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE TRANSACTIONS on Information and Systems,* vol. 77, no. 12, pp. 1321-1329, 1994.

[15] R. Metz, “Augmented reality is finally getting real,” Technology Review, vol. 2, 2012.

[16] B. Boser, “Indoor floorplan with wifi coverage map android application,” 2013.

[17] Divecha, Devina. Augmented Reality (AR) used in architecture and design. *Design MENA* 8 September, 2011.

[18] Architectural dreams in augmented reality. *University News, University of Western Australia 5 March*, 2012.

[19] Outdoor AR. TV One News, 8 March 2004.

[20] Davis, Nicola (January 7, 2015). "Project Anywhere: digital route to an out-of-body experience". *The Guardian. Retrieved September 21*, 2016.

[21] "Project Anywhere: an out-of-body experience of a new kind", *Euronews, 2015-02-25. Retrieved September 21*, 2016.

[22] Microsoft. June 1, 2009. Archived from the original on January 21, 2012. Retrieved June 2, 2009.

[23] Microsoft's new Kinect is official: larger field of view, HD camera, wake with voice.

[24] "Kinect is always listening on Xbox One, but privacy is a 'top priority' for Microsoft". Polygon. Retrieved May 22, 2013.

[25] Rusinkiewicz, S. and Levoy, M. 2000. QSplat: *a multiresolution point rendering system for large meshes.* In Siggraph 2000. ACM , New York, NY, 343–352.

[26] *Vuforia website,* https://www.vuforia.com

[27] *Kinect's SDK architecture*, https://msdn.microsoft.com/en-us/lirary/jj131023

[28] Ryo Jozaki, Jiro Tanaka, School of Informatics, University of Tsukuba, Faculty of Engineering, Information and Systems, University of Tsukuba, *AR Chat: Remote Communication Support System by Augmented Reality,* The 74th National Convention of IPSJ, March 2012

[29] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, Dieter Fox: RGB-D mapping: *Using Kinect-style depth cameras for dense 3D modeling of indoor environments*, SAGE journals- Volume: 31 issue: 5, page(s): 647-663, Article first published online: February 10, 2012; Issue published: April 1, 2012

[30] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim: *Holoportation: Virtual 3D Teleportation in Real-time* in: UIST '16 Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, October 16 - 19, 2016 Pages 741-754