Bachelor's thesis

Information Technology

Internet Technology

2015

Xing Huang

VIRTUAL REALITY/ AUGMENTED REALITY TECHNOLOGY: THE NEXT CHAPTER OF HUMAN-COMPUTER INTERACTION

Xing Huang

VIRTUAL REALITY/ AUGMENTED REALITY TECHNOLOGY: THE NEXT CHAPTER OF HUMAN-COMPUTER INTERACTION

No matter how many different size and shape the computer has, the basic components of computers are still the same. If we use the user perspective to look for the development of computer history, we can surprisingly find that it is the input output device that leads the development of the industry development, in one word, human-computer interaction changes the development of computer history.

Human computer interaction has been gone through three stages, the first stage relies on the input command to interact with the computer and the main input device is the keyboard. The second stage, which occurs presently, is the graphical user interface and the main input device is the mouse. The third stage is the touch screen which is used in every smartphone and tablet and the main interaction at this stage is finger. There is one clear point in such history, the basic evolution direction is to let people have more freedom, bring a more natural way to gain information, people want to fully use the human five senses to feel the world, thus virtual reality and augmented reality has become a hot topic.

This thesis gives a brief history of human-computer interaction and an introduction of virtual reality and augmented reality, discusses their development and how they work, and provides reasoning why virtual reality and augmented reality will be the next stage of human-computer interaction.

KEYWORDS:

Human-Computer interaction, Virtual Reality, Augmented Reality

ACKONWLEDGEMENTS

After four years study in Finland, with my thesis done, my Bachelor's degree study has come to an end. Special thanks go to my parents who made a wise choice to let me study in such a great place. Without their support, I could not have had the opportunity to broaden my horizon and experience a totally different culture.

Thanks are due to my girlfriend and friends for their companionship. It was great to travel with my girlfriend to enjoy the beauty of other European countries and celebrate the New Year together. Studying with my friends helped me to solve problems we faced and we helped each other if we had difficulties. I could not go back to China for the Chinese New Year, but thanks to my friends so that we could spend the Chinese New Year together.

I would like to thank my teachers in TUAS who taught me IT, strengthened my Math and English skills. I gained useful IT knowledge which is really helpful for my further study and future job search.

CONTENTS

PICTURES

LIST OF ABBREVIATIONS (OR) SYMBOLS

ACM	Association for Computing Machinery
HCI	Human Computer Interaction
HMI	Human Machine Interface
VR	Virtual Reality
AR	Augmented Reality
Oculus Rift	Virtual Reality wearable device
Google Glass	Augmented Reality wearable device
HoloLens	Augmented Reality wearable device
NLS	oN-Line System, a computer collaboration system
GUI	Graphical User Interface
PDA	Personal Digital Assistant
MS-DOS	Microsoft Disk Operating System
WIMP	Windows, Icons, Menus, Pointer
CRT	Cathode Ray Tube
LCD	Liquid Crystal Display
PPI	Pixels Per Inch
IMU	Inertial measurement unit
FPS	First Person Shooting
Fps	frame per second
HMD	Head Mount Display
IR	Infrared Radiation
HDMI	High-Definition Multimedia Interface
DVI	Digital Video Interactive
FOV	Field of View
LED	light-emitting diode
PTAM	Parallel Tracking and Mapping
ISMAR	International Symposium on Mixed and Augmented Reality

CG	Computer Graphics
LCoS	Liquid Crystal on Silicon
PBS	Polarizing Beam Splitter
HPU	Holographic Processing Unit
SLAM	Simultaneous Localization And Mapping

1 INTRODUCTION

With the rapid development of modern technology, technology is everywhere. The computer is the most representative product, as within a few decades, now there are many different types of computers, for example, the huge server located in the room, the personal computer on the table, the laptop on the knees, the smartphone and tablet in our hands, even the wearable device on our wrist or on our head. The relationship between the computer and human has gone through a fundamental change. Human–Computer Interaction (HCI) is a field of study that helps people control the machines more easily, so that computers can be used by a range of users, from the minority specialist to the majority of people around the world. ACM defines HCI as "A discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them." (Hewett et al., 2009).

At first, we could only type the command to let the computer work, then after the development of the graphical user interface, with the birth of the mouse, we started to click the icon to work. In the recent years, the technology of touch screen has freed our hand and we can touch the screen to use our device. People still want to find more ways of interaction with the computer. As one of the main types of Human–Computer Interaction Virtual Reality has become a hot topic in the recent years. Oculus Rift, Google Glass, and HoloLens represent the most advanced technology in the field of Virtual Reality / Augmented Reality (AR). The communication between human beings and the real world is quite normal, the virtual world is over isolated, so is there a communication platform between reality and virtuality which helps people find more friends?

People just repeat tedious everyday life activities and feel bored. Is there a way that can make the real world full of variety? AR technology can solve the above problems, it is a technology that can improve and fulfill new demands and it will be the next big thing.

This thesis will introduce the history and development of human–computer interaction and Virtual Reality / Augmented Reality. It will analyze three most representive technologies, Oculus Rift, Google Glass and HoloLens. Based on these, it will discuss the advantages and disadvantages of those technologies, and will conclude on how much possibility there is that VR and AR technology will be the next stage of HCI.

The thesis contains the following chapters:

Chapter 1 introduces the background of the thesis.

Chapter 2 describes the development of Human–Computer Interaction.

Chapter 3 introduces Virtual Reality and analyses some of the VR technologies, like Oculus Rift.

Chapter 4 introduces Augmented Reality and analyses Google Glass and HoloLens.

Chapter 5 is the conclusion of the thesis.

2 HUMAN–COMPUTER INTERACTION

2.1 Definition of HCI

Human-Computer Interaction or Human-Machine Interaction is the study that researches the relationship between the system and user. The system can be any type of machines, or it can also be the computerized system and software. This can be seen as a communication language between the computer and human by using a certain interaction to finish the task that designed to be done. The interface of HCI usually means the visible part user can see. User controls on the interface can be small like a home button on the smartphone, big like a panel on the plane craft. The design of a HCI interface should include the understanding of the system to create usability and user-friendliness.

2.2 Detailed interpretation of HCI

The keyword for HCI function is "user friendliness". The HCI function mainly relies on the input, output equipment and related software. The equipment used in HCI are keyboard, mouse, touch screen, speech recognition devices and any other pattern recognition devices. The software that is related to that equipment is the system that offers the HCI function, the use of it is to control the function of the related system and understand every command and request that is sent from the HCI equipment.

At the very early stage, the HCI equipment was the button, like the typewriter. People pressed buttons to create letters. Later, with the birth of the computer, HCI equipment changed to the keyboard and monitor. Operators typed the command on the keyboard, the system ran the process and showed the result on the screen. In the late 1960s, with the rapid development of science technology, more and more people realized the potential of the computer. They thought the computer would become their personal computer and it would no longer be used only by companies or the governments but it could also serve the normal people. At that time although the screen on the computer could only show the green color terminal interface, some people had looked far ahead into

the future. One of the important visionaries was an American inventor, Douglas Engelbart. He gave a demonstration on December 9, 1968, of a new computer system called NLS, the oN-Line System, and the prototype of the computer mouse was a three-button palm-sized contraption (Edwards, 2008). This is not only a breakthrough on computer history, but also a big step in HCI development. For the next ten years, many companies started to find out the potential of the computer, the birth of the Graphical User Interface cooperating with the mouse and keyboard, has now become a standard setup for the personal computer. The invention of the computer mouse brought the operation of the computer into a new era.

People are always trying everything new and the same applies to HCI. People want to use the GUI more directly and easily. Another new concept shows up: touch screen, a display that is sensitive to human touch or other kinds of medium. E.A. Johnson is considered as the first one to develop touch screen in 1965 (Johnson, E.A. 1965). So far after half-century of development, people get in touch with touch screens everywhere, ATM machines, PDA, monitors, computers, and phones. After the launch of iPhone in 2007, touch screen technology ushered in a blowout and now we are using it all the time. Touch screen technology is a more natural interaction because the human hand is the most powerful tool in the human history. Using the finger to control the computer that we want to use is a great development in HCI.

With the development of the pattern recognition, the input of speech recognition has been used into the interaction between the operator and computer. This is a big step because people can work on the human natural language, not only follow the binary digits. When we drive a car, we do not need to type on the phone to send messages, we can use our voice to send messages to others, which is really convenient. This can be seen as an intellectualized HCI.

2.3   History of HCI

In 1959, based on how to reduce fatigue when operating the computer, an American scholar Brian Shackel published an article about the ergonomics for computer design which considered the first concept of HCI (Shackel 1959). One

year later, Liklider first coined the term Human-Computer Close Symbiosis (Licklider 1960), which is regarded as the enlightenment view of HCI. Later in 1970, two HCI research centers were established, HUSAT at Loughbocough University in the UK, and Palo Alto by Xerox, in America. The next three years there are four monographs related to HCI which gave a direction to HCI development. Then in the next decade, HCI study developed its own theoretical system and categories, on the one hand, independent of ergonomics, and placing more emphasis on the humanities, like cognitive psychology, praxeology and sociology; on the other hand, it expanded from the human-computer interface, underlined interaction from computer to human reaction. Thus the Interface changed to Interaction.

From late 1990s until now, with the rapid development and popularization of the high speed processor, multimedia technology and Internet Web technology, the research has focused on intelligent interaction, multimedia interaction, virtual interaction and Man-machine coordinated interaction, which puts people first. The development of HCI is a change from people adapting to the computer to the computer adapting to people.

2.4   The HCI model

The HCI model is a concept model structure that describes the interactive mechanism in HCI system. So far scientists have come up with many models, like User Model, Interaction Model, HMI Model, Evaluation Model and so on. Those models describe the characteristics of humans and computers in HCI from different angles. Knowing the HCI model is the base of developing a HCI system. It is also a must understand knowledge to design a successful HCI product.

In the model study of HCI, an important model in the early stage is Norman's Model of Interaction which can be considered a logical simplification of an execution-evaluation cycle (Norman, 1988). In this model, Norman divided the HCI process into two stages, execution and evaluation. These stages usually have the following seven steps: Forming the goal; Forming the intention; Specifying an action; Executing the action; Perceiving the state of the world;

Interpreting the state of the world; Evaluating the outcome. Each step is the action of the user. First, the user forms a goal. This is a task for a user. The language of the task needs to be formulated in the different field, but the task might not be that clear, so it needs to be translated into a clearer intention and the actual action to achieve the goal. After execute the action, the user feels the new shape of the system and gives the interpretation of his expectation. At last, if the status of the system reflects the user's goal, the interaction is successful, otherwise the user needs to rebuild a new goal and repeat this cycle. This model is given in Picture 1.
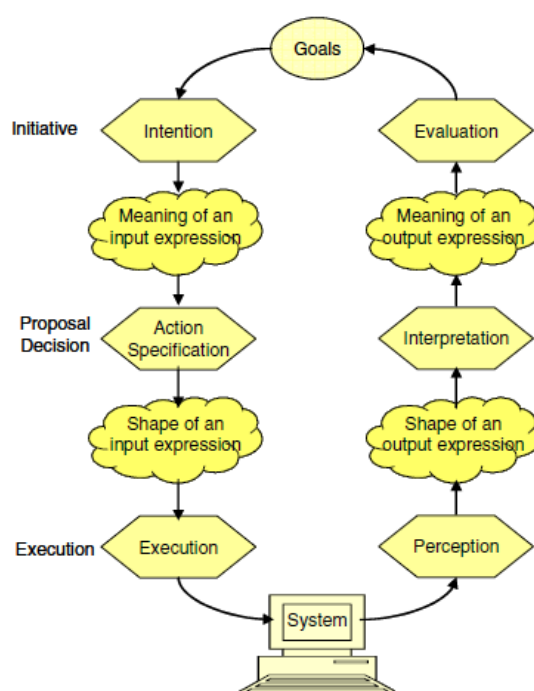
Picture 1. Norman's Model of Interaction

This execution-evaluation cycle model is an effective way to understand HCI. It has a clear thinking and intuition, and it offers a universal frame for the past experience and analysis work. However, it fully pays attention to the user's view of the interaction and only regards the computer system as the interface of interaction and ignore the conduct of the computer system by interface communication. In one sense, it is not a quite completed HCI model.

In 1991, Abowd and Beale improved this model, and made some extension and announced the framework of interaction (Abowd and Beale 1991). This model reflects the characteristics of the user and the system in an interactive system

at the same time, which makes the interactive process more complete. This model divided interaction into four main parts: system, user, input and output. Each part has its own language. Besides the user task language and the system core language, it also includes input and output language and these languages express a concept of application area from their own angle. The system language is called the core language which describes computing features in the application area. The user language is called task language and it describes attributes related to the user' intention. This model reflects the general characteristics better in interaction. The interactive process shows information moving through these four parts and its conversion describes the system operation. Input and output together is the human-computer interface. One interactive cycle has four stages: articulation, performance, presentation and observation. The first two stages are responsible for understanding the intention of the user. The next two stages are responsible for the explanation and evaluation of system output. Picture 2 shows the Interaction Framework model.
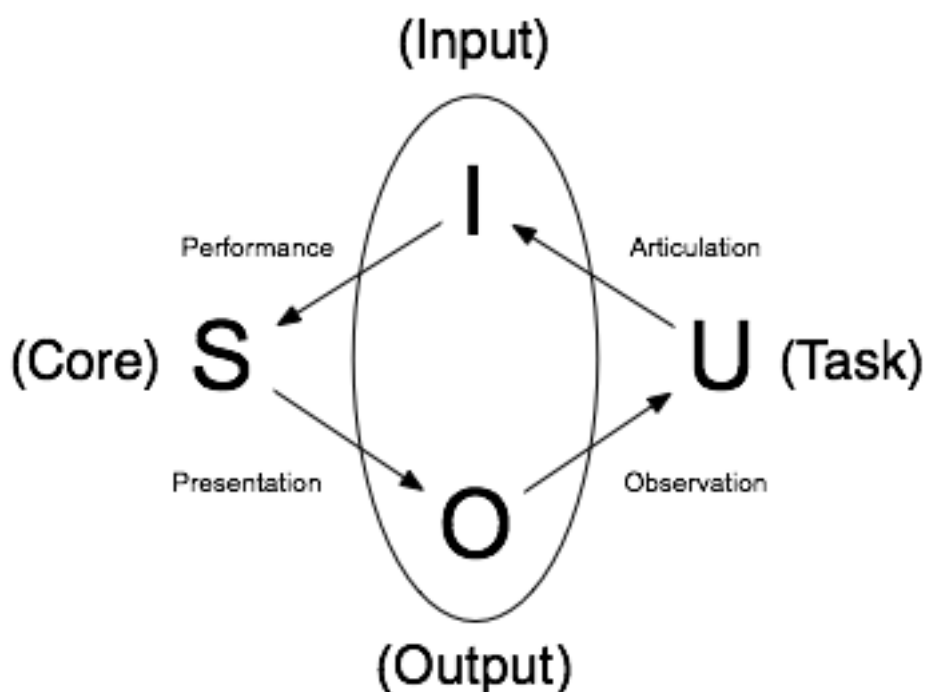
Picture 2. Interaction Framework

2.5 Interaction Styles

In fact, the process of HCI is the process of input and output. The user inputs the command through the HCI interface, the computer shows the output result to the user after processing. The form of inputs and outputs between humans and computers are diverse, so are the forms of interaction.

2.5.1 Data interaction

Data interaction is a way that people input data to communicate with the computer. It is one important way of HCI. The general interactive process works like this: first the system sends a prompt to the user and reminds the user of input and how to input; next the user inputs data into the computer using input equipment; then the system will respond to the input and give feedback information, shown on output equipment such as the screen. In the meantime the system will check the input. If it is not correct, it will prompt the user to input again. The data here can be any kind of information symbols, like digit, symbol, color, image and so on. Here are some kinds of data interaction:

- Command line: the user types in commands for the program, MS-DOS and UNIX use this style;

- Question and answer: the system asks questions, and the user answers. This is simple, but monotonous;

- Menus list as much data on the screen that the user can select, like the PDA operation interface and settings on mobile phone;

- Form filling: the user types data into the required fields, like the database applications;

- Function keys offers special keys, like the computer games;

- Graphical direct manipulation: the user can point, click, drag, type. Most Windows systems use this style, known as WIMP which stands for "windows, icons, menus, pointer".

2.5.2 Graphical interaction

Scientific research shows that humans transfer information mainly by speech, text and graphics, Humans absorb more than 70%information through vision, thus the research into graphical interaction is really important and the application areas of graphical interaction are unprecedentedly widespread, like face recognition, handwritten interaction, digital ink, etc.

Graphical interaction, in simple terms, is that, according to the people's behavior, the computer makes, understands and reacts. In this respect, making the computer have visual perception is the first problem to be solved. So far there are three layers of the machine vision system:

1. Image processing: processing the image to improve the visual effect, transferring the input image to another image that has the required features;

2. Image recognition: detecting and measuring the specific target in the image. Image recognition gains the objective information to build the description of the image, transfers the image to data;

3. Image perception is based on image recognition, further researches the interactive relationship in the image and draws comprehension of image content as well as the explanation of original scene, transfers the image to an interpretation.

2.5.3 Speech interaction

Speech is accepted as the most natural, convenient way of information communication. There is 75% of communication in human daily life occurring through speech. The auditory pathway has many advantages. For example, the auditory signal detection speed is faster than the visual signal detection speed; people are sensitive to the change of the voice through the time; providing auditory information and visual information can make people feel a more real and sense of presence. So, the auditory pathway is the most important information pathway in interaction between people and the computer.

Speech interaction is a technology researching how to interact with the computer by using a natural language or a synthetic language. It involves a lot of different subjects, and it not only needs to study speech recognition and speech synthetic, but also research the mechanism of interaction. Generally, there are two methods of speech interaction: one is based on speech recognition and understanding and, mainly relies on the audio signal to interact; the other is speech interaction combined with other interaction styles. In this system, speech interaction is only a part of the interactive system. Some well-known speech interactions are voice assistants in smartphones, like Siri, Google Now and Cortana and people can use them to set an alarm, or open an application, or send a message.

2.5.4 Behavioral interaction

In daily communication, besides using speech interaction, body language is also helps people express their thought. This is called human behavioral interaction. This interaction can not only enhance communication ability, sometimes it can even play a role that speech interaction can not play, like a fashion show, dancing, or sketch comedy, etc. The computer locates and recognizes the user, captures the body language and facial characteristics to understand the human action and behavior and gives a response.

Behavioral interaction brings a brand new interaction style. The computer can predict what the user wants by interpreting user behavior. For example, the computer captures the user's eye to decide if they need a phone call or to surf the Internet. Now on the market, there are some of the experimental projects are trying new forms of interaction (Summers 2013):

Leap Motion is a tiny metallic bar in front of the screen. People can wave their hand to swipe the screen, or rotate image zoom in, or zoom out.

MYO, is like an arm band. It will measure the muscle activity when user uses that arm. Apart from Leap Motion, this is Room Free. There is no need to sit in front of the screen, the user can move to anywhere;

Kinect, Kinect offers a much healthier way of playing the video games. Several motion sensors and cameras will analyze user's motion, so that the player can enjoy the game.

Finally there are three VR/AR devices, Oculus Rift, Google Glass and HoloLens, which will be the focus of the next chapter.

3 VIRTUAL REALITY

3.1 Background

Mobile was the next big thing and although we are in the mobile era, the age of mobile has passed. So we can see large tech companies trying to discover new products that will change the game. These companies do not want to be the next Nokia. The tricky aspect is that no one knows what will be the next. There is only a vision of what can be done to make technology serve humans better. Thus, Google is running a secret lab called Google X which develops some modern or even future tech products, like Google Glass, an augmented reality glass, Project Loon which provides internet Service by balloons, driverless car, etc. Apple and Samsung are investigating wearable devices and smart home. Facebook acquired Oculus Rift to develop virtual reality. All these companies want to take the vantage point in the next round competition and they all believe that virtual reality will mostly be the next thing. If we look back into the past, the computing center changed from the personal computer to mobile, and it not only the size that becomes smaller, the interaction also changes, from the keyboard and mouse to the touch screen. This change makes people get more close to their mobile device. Virtual Reality will bring the user into an immersive environment and it might subvert the entire industry.

3.2 Virtual Reality

Virtual Reality is a computer-simulated three dimensional virtual space, offers virtual sight, sound, touch, and makes the user feel being in a real world, and observe the virtual space with no limit. There is seven different concepts describing virtual reality: simulation, interaction, artificiality, immersion, telepresence, full-body immersion, and network communication (Heim 1993). From a technical point of view, virtual reality has three basic features: Immersion, Interaction, and Imagination. These three "I" features emphasize that the human takes the leading role in the virtual reality system. In the beginning, people could only observe the result from the outside computer system but now they can immerse in an environment that the computer creates. People used to use the keyboard and mouse to communicate with the computer

system but now people can use multiple sensors to interact with a multi-information environment. In one word, in the future's virtual system, the aim of human is to let the computer system satisfy people's need, not to force the people make do with the computer system.

3.3   History of Virtual Reality

Virtual Reality is not an advanced technology. In fact, the first prototype of VR device was created in 1968, around half century from now, but it was until now that the word Virtual Reality came into the public view and became a popular topic. This is not strange, after all, at that age, even the computer mouse was just created in 1968, when there was no personal computer, only the computer. Similarly, people might have seen the concept of Virtual Reality as crazy at that time, but this was due to technological limitations.

The prototype The Sword of Damocles was created by the father of VR, Ivan Sutherland. At that time, he only called it as a head-mounted three-dimensional display (Sutherland,1968). This became the first prototype of VR because it defined some key features of VR. First, it has a stereo display. The prototype had two CRT displays to show the image from different optic angles. This is important because it creates stereovision. Then it had a virtual image. created by a computer program which did real-time computing. Next, it had two ways to measure the head position, one was a mechanical linkage, and another was an ultrasonic wave. Next was model generation. In model generation there was a simple cube in the display, but it built up based on the space coordinates. When the user wore this device and turned the head around or changed the visual angle, the cube also changed. Picture 3 shows the system.
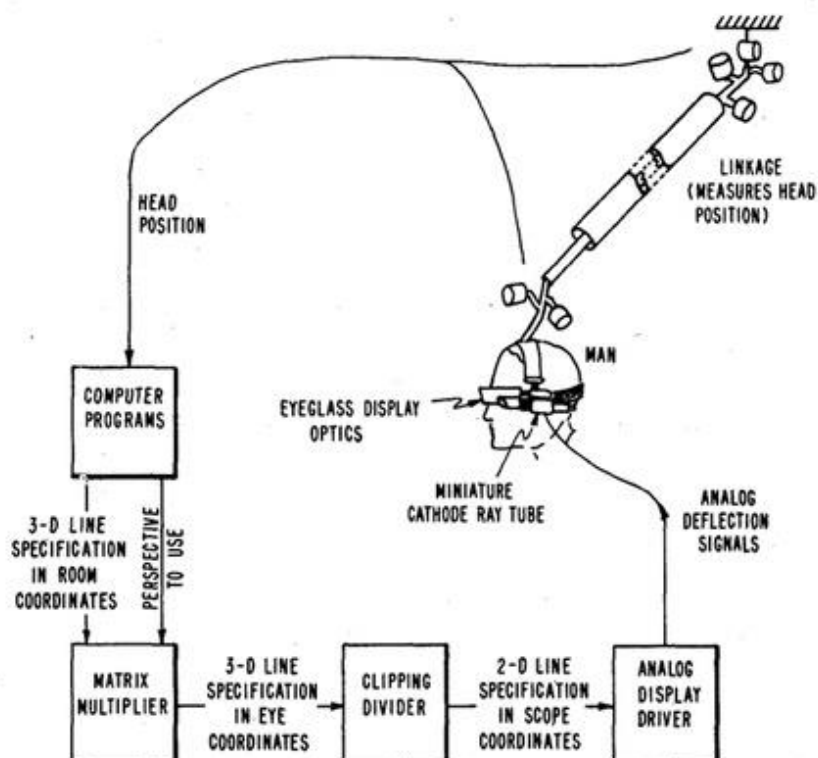
Picture 3. The parts of the three-dimensional display system

We have to admit that, in today's view, The Sword of Damocles is quite simple, and heavy, but was considered as an innovation in 1968 and it was definitely a piece of work much beyond this lifetime.

The most important change in the 1980s was that the parts of a VR device could be purchased individually. We could buy portable LCD displays, the graphics card could support more complicated pictures. A company developed a 6 degrees of freedom head tracking equipment, compared with the past, the accuracy had improved greatly. There were also some gloves that have the joint motion sensors so that the user could have more interaction with the computer. One big moment at that time was that another pioneer of VR, Jaron Lanier, produced the first truly commercial VR product the EyePhone, a virtual reality head-mounted display. After that, VR started to be used in many different fields.

Back to the present, VR technology has a qualitative change than before. The display in the past was not good. So far, the pixel per inch on smartphone has over 500 ppi, brightness and color have also greatly improved. The graphical

card can support a large 3D game. The special Inertial Measurement Unit (IMU) used to measure the position of head, it was not only tiny, but also very accuracy. There are also more choices of interaction with the virtual environment, such as, motion control, gesture recognition and speech recognition, etc., we you can find some relatively mature products on the market.

The personal computer, from the prototype to popularization was developed in a few decades, the Internet took a dozen years to connect the whole world, the smartphone and mobile Internet only took less than a decade to take over our daily life. On these grounds, people widely consider that the new interaction on the VR/AR technology will be the next thing.

3.4   Use of Virtual Reality

The use of VR was quite broad. In 1993, according to statistics of VR research projects around the world, there were over 800 projects The result showed that entertainment, education and art take the lead, then is military and aviation, followed by medical science, then robots and commerce.

In entertainment, the rich sensibility and 3D display environment make VR the best tool for developing games. Players want to be in the game, from the very first word game, to two dimensional game, three dimensional game and Internet 3D game. The interaction, the reality and immersion become better and better especially in games like role play, FPS, car racing. VR can fully satisfy the requirements, so now on the market, high tech companies focus on VR games, even on the game platform Steam, there are some games especially designed for VR devices., It is not hard to demonstrate that VR in game play has a bright future.

In the educational field, the interactivity and vitality of VR technology is used to explain abstract concepts in solid geometry, physics, chemistry and geography VR is a powerful tool, and it can save money, without limits of space. In some special training institution such as driving or flight schools, VR can avoid any kind of physical danger. There is no need to worry about a car accident or air crash. In art, VR has a spot sense of participation that can transfer static art like

painting or sculpture to dynamic art and people can better understand the thought of the artist. At the same time, VR can enhance the artistic expression; a virtual musician can play many kinds of instruments, so inconvenient people like in the other city can enjoy the concert.

In military and aviation, VR provides vivid simulated training, like the battlefield, or zero gravity environment for astronaut training.

In medical science, VR has a very important purpose. In the virtual environment, we can build a virtual human body, students can understand and observe the organ structure easily which is much more efficient than reading a textbook. In medical schools, students can do autopsy and surgery practice, by using the VR technology, thus eliminating the need to prepare the real body and training expenses. If the VR system is good enough, the advantages are great, surgeons can simulate the surgery in advance, find the best program, and the remote surgery provides some necessary help.

Emergency drills and taking preventive measures are the key in some dangerous industries, firefighting, electric power, oil exploitation, in ensure minimum loss after an accident. Regular emergency exercises are a traditional way, but they cost a lot of money, a mass of manpower, and material resources. If we simulate a man-made accident in a VR environment, this can save money, enhance the frequency of the exercise and ensure the safety. This also works in any other maintenance service. There are more, more fields and more people can benefit from the VR technology.

3.5 Barriers of Virtual Reality

From the above, we can conclude that VR will be a breakthrough in our future life, full of potential, but VR so far is more likely to be used in the commerce or some special fields. There is a long way to go if VR technology goes into the consumer market. There is still a technical limit which is how to provide a true reality environment, and some other problems still do not have a good solution.

Firstly, there is not a real way to step into the virtual world. There is a joke in Oculus Rift developer circles that every time when the staff asks the user to stand up and walk around, the user usually do not dare to move, because at

present stage, most of the VR HMD are connected to the computer, this dramatically limits the interactive range. The VR device just covers our eyes, simply changes our eye sight. in our virtual view, it could be a whole world. We can move to anywhere we want, but the current state of VR did not cover all of our sight range. It is also awkward that user still needs to use the mouse and keyboard to control. So in fact we are still sitting in front of the computer. Some developers are try to build a special room for the VR device, but the usual problem is that some of interactions include squatting, hiding, jumping or climbing and these interactions can not be fulfilled.

Second is the input problem, which is the core of interaction. The majority of VR devices can only capture the user's head motion, but not the rest of the body, for example the hand action cannot be simulated. Input is the most important and significant experience for users, otherwise users will confuse where their hands are. Traditional computer peripherals are the keyboard and the mouse, but it is obvious that when the user is covered in the VR environment, it is difficult to use those smoothly. Gamepad seems to be a compromise, but for the traditional PC user, this is still somewhat strange. Some developers try to develop peripherals like a game gun or a sword, but the delayed action is the big problem. In the VR fields, eye tracking and motion tracking is the most efficient way, so development in this area is needed.

Third is lack of uniform standards. There is no doubt that VR technology is still in the primary stage. Each developer has its own way to show the demonstration on a VR device and there is not a uniform standard although developers show much interests in it. As a new platform, the key to success is to attract people's interest and game players are the core audience of VR. but so far game players mainly gather around PCs, home video game consoles, now even mobile devices. So converting people from a mature market to a quite strange market or making these markets compatible is a big problem. Another shortcoming is the lack of talents to develop VR game.

Fourth, it is easier to make people feel tired. Some people may have this experience, when they watch a movie or play a game, when the camera moves

fast, it will bring different focus, so human's eye needs to refocus, this might cause some people feel dizzy and sick, especially to female users.

There are others barriers, too. Compared with the first prototype of VR device, the VR device so far is tinier, but it is still quite heavy and not that comfortable. It even looks silly. The price is still high, the consumers will not buy it, and there is a lack of a special platform. Some technology seems to be cool, but it will lose attention as time goes by, so VR has a long way to go. There will be one day when it becomes a new way to change the interaction between human and computer and achieve a more natural interaction.

3.6   Challenges in Virtual Reality

On the market, most of the VR devices are head-mounted display, because HMD can bring user a high level of immersion, but at the same time, this high immersion will cause people to feel dizzy. In fact, this is called motion sickness., The reason cause this problem is that human eyes feel the movement of images which is uncoordinated with the body feeling the movement.

In the human body, feeling the movement of the body is accomplished by relying on the vestibular system. It is like the accelerometer and gyroscope, detect acceleration and angular velocity. When playing a 3D game, like the FPS game, the eyes will tell the brain that our body is moving, but because we are just sitting or standing still, and only use the game pad or keyboard to control the virtual character move, the vestibular system will tell the brain we are not moving, and then the conflict will occur. On the other hand, when a player shakes the head, and suddenly stops, unlike the eyes, the vestibular system will not stop immediately, because of the inertia, it will tell the brain we are still moving, so the signals will be conflicting again.

There are around 30% to 40% people sick of the equipment, 60% to 70% people sick of the content. For the hardware, the resolution of the display screen is not enough, 2K or even 4K resolution might fulfill the need of a HMD, the refresh rate of the display needs to be enhanced, 60 HZ still has space to increase. Recently HTC announced a VR device whose refresh rate is 90 frames per second, and the users said that the problem of dizziness has been

solved (Yin-Poole 2015). In the 3D environment, when we move, there is no force feedback, so if some related accessories are added, the problem can be solved. For the software, on the market most of the games have incorporated the software to the VR equipment. The user interface and the play instruction are not suitable for VR devices, so this part should be redesigned. Then, there are two necessary conditions for the frame per second (fps) in the game. First, it should be stable, second, fps must be more than 30, otherwise there will be a problem. If player turns the head from 90 degrees in 1 second, but the fps is not stable and slow for this movement, the brain will think it was done, the eyes will tell it is not done and the player will feel dizzy. Field of View should get close to human eyes, if the FOV is too large, there will be a stretch of the edge of the display, if the FOV too small, when there is movement, the image will change fast. Both of these are not acceptable.

3.7   Oculus Rift

3.7.1   Development history

In 2012, the Oculus Rift project showed up on Kickstarter, a global crowdfunding website. Oculus Rift is a VR HMD designed for game play. It will connect with the VR world to make the player feel real. The target of it on the website was to raise two hundred and fifty thousand dollars, but it seems that people showed a great interest in it. In the end it was funded more than 2.4 million dollars (Kickstarter, 2012). In late 2012, Oculus released the first development kits and after a few months, there were dozens of games supported this VR device: Until now, there are more than 560 apps or games available to use (Riftenabled.com 2015). Early in 2014, Facebook spent around two billion dollars acquire Oculus Rift and from this move, it is not difficult to reach the conclusion that the top tech companies see the huge potential of virtual reality. In July 2014, the second version of the Oculus development kits released, and maybe in 2015, the consumer version will be released.

3.7.2   Oculus Rift teardown

The development kits 2 include a headset, several cables, one camera, two lenses and several different plugs for different countries. The motherboard has

many sensors, like the accelerometer, gyroscope and magnetometer, it also has a built in latency tester, IR sensor, the camera is used to capture the IR sensor and the display is 5.7 inches, per eye with a resolution of 960 X 1080 (iFixit, 2014). Picture 4 shows the Oculus Rift DK 2.

Picture 4. Oculus Rift DK 2

### 3.7.3 How the Oculus Rift works

Basically, Oculus Rift is an external monitor supported HDMI and DVI input, in the meantime, it is a USB device that can detect head motion. In order to achieve 3D image, the screen displays two images, the left eye sees the left half of the screen, and the right eye sees the right half of the screen. In development kit 1, the resolution of screen is not that high, so in development kit 2, Oculus increased the resolution to Full HD and the user gets a better experience. A set of lenses are placed on the top of the screen, and this is the key to fulfill the stereoscopic 3D image which is used to focus and reshape the picture. The reason why Oculus Rift is the star in VR devices is that it can nearly restore the human's field of view, almost 180 degrees forward facing horizontal, 135 degrees vertical (Wikipedia, 2015), Oculus Rift has a 100-degree vertical FOV (Oculus VR, 2014) which enhances the immersion of user. Two convex

lenses are the secret, based on the simple optical theory, which magnify the screen. As shown in Picture 5, by using the convex lens, it can magnify the original length from x to x'. However, there is a consequence, at the same time the convex lens magnifies the image, the distortions will exist, so on software layout, Oculus Rift gives a solution to correct the distortions by creating the same size but opposite distortion. Thus the man-made distortion can counteract the optical distortion and the user can see the normal image through the screen.
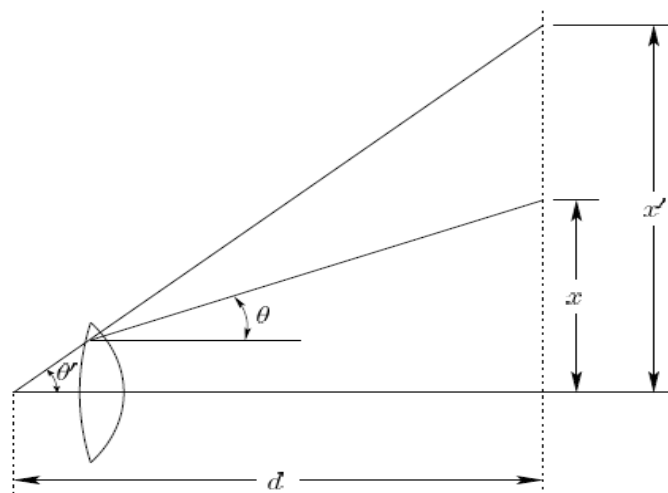
Picture 5. Magnifying glass theory

Head tracking is important in VR devices because in the real world, people can turn their head around to see the world. In the first prototype of VR device, the mechanical linkage was used for head tracking, nowadays thanks for the technology development, it can be replaced by much tinier sensors, gyroscope, accelerometer and magnetometer. In the second development kits of Oculus Rift, one more external camera helps to track the headset position more precisely. First, the sensors create a 3D position vector for which the coordinate system uses these axis definitions: X is positive to the right, Y is positive in the up direction and Z is positive heading backwards. The opposite directions in each axis have the following rotations: **Pitch** is rotation around X, positive when pitching up, **Yaw** is rotation around Y, positive when turning left and **Roll** is rotation around Z, positive when tilting to the left in the XY plane (OCULUS VR, LLC, 2014). Picture 6 shows the head tracking sensors system.
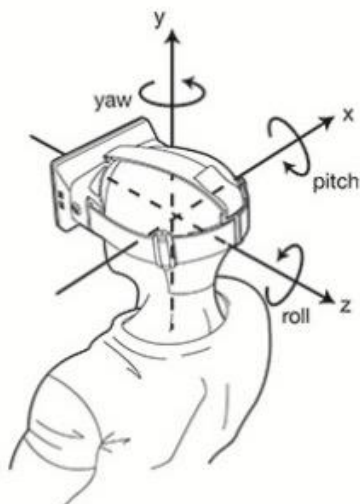
Picture 6. The Rift coordinate system

In the first development kits, there is no IR camera, which means that it can track head rotation very well. However, when we move our whole body, such as when sitting or jumping and swinging, the virtual head on display would not move. This decreased the immersion and reality. Therefore in the second development kits, some changes were made and an IR camera was added. On the headset there are several IR LEDs.   Capturing the set of IR LED opens new ways of game play, is more comfortable and more immersive. Players can hide in a corner, swing their body to dodge bullet, and much more. Picture 7 is the position tacking system.
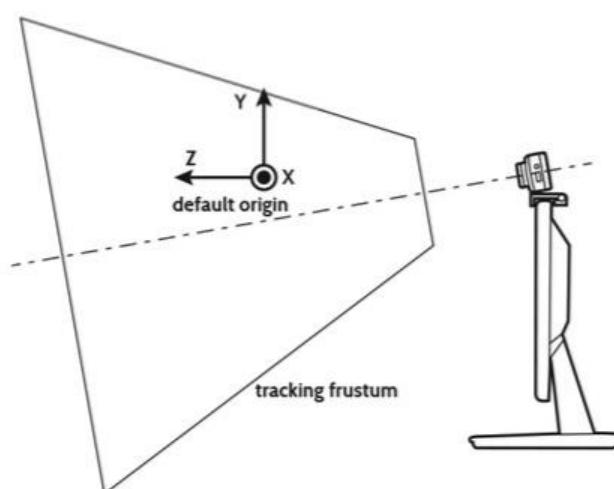
Picture 7. Position tracking camera

3.7.4 How to solve the latency

To enjoy the game, players can bear many things. 3D games might make some people feel sick and VR devices might aggravate this feeling. The feeling of dizziness that Oculus Rift causes comes from the latency of the whole system. The sensors, the screen technology, the GPU speed are all the factors cause the problem. In general, latency occurs when user moves his head, and the image shows the right one on the screen, there is a time sensors will make response, image transmission and display response. Even with today's technology, latency can not be eliminated completely. It can only be controlled to be as low as possible. With the power of the processor, on the market some VR HMD decrease latency to 100 millisecond or 40 millisecond, but it is not enough, research shows that 20ms is the watershed, 15ms will remove the simulator sickness (Abrash, 2012).

Oculus Rift divided the latency time into six parts:

1. The user inputs information,

2. The USB cable transforms the command from Rift to computer,

3. The game engine translates command to GPU,

4. The CPU gives order to write a new image,

5. The display transforms pixels to an image,

6. The image is fully generated.

Oculus is trying hard to cut down every step's time. The first step is to minimize the input latency which means finishing the transfer from head motion to digital command as fast as possible. Differently from the prototype of the Oculus Rift, now the current version has a built-in tracker integrated with gyroscope, accelerometer and magnetometer and time has decreased from 15ms to 1ms. Then the signal transfer from Rift to computer will make 1 to 2ms latency and the problem is with the cable, so unless the USB cable is reinvented, we can not reduce the latency time.

Next, game developers will take responsibility for the latency The game fps plays a particularly important role. At present, most games are around 60fps, which means that every image transfer to GPU will cost about 16.67ms, so if developers improve the speed 1 times, the latency will be the half as before. The GPU transfers the command to pixel on display through USB cable. Some pixel transform fast, for example, black to white only takes less than 10ms, some pixels need more time. To save the time, every single pixel starts to transform once it has received the command from GPU. The Rift begins to write from bottom to top, so when the command transfer to the top, the pixels at the bottom will have finished the transformation. The whole process takes 20 to 30ms, based on a 60fps game, the whole latency takes 40ms which is still too much for a VR device.

It is seems that latency time is mainly wasted on the screen. Oculus Rift used to use LCD display, but OLED technology can have an extremely fast refresh rate, lasting only a single millisecond. So using OLED display might solve the problem. However there is only one manufacturer offering OLED technology, that is Samsung, but Samsung has not sold that technology to any third party. However, on second development kits, the display use OLED technology. From the teardown of Oculus Rift can see that it just uses the Galaxy Note 3 display panel directly on the screen, a 5.7 inches Super AMOLED 1080p screen, and Oculus slightly overclocks the display panel, and changes the refresh rate from 60Hz to 75 Hz to decrease the latency time on screen. Besides that, the internal tracker has a really high data sampling capability, 1000 times per second, which not only reduces the latency, but also has a capability to predict the head position that user may move to. If the head movement is fast, it will not be able to stop immediately as it can only slow down from one point. Thus if players move their head, at one period of time, the tracker can predict the next position of the head, after the head slows down, the prediction moves down to 0. This process not truly changes the latency time, but it allows players see the image much faster. This can change the latency subjectivity for about ten milliseconds.

All those efforts have successfully shortened the latency into 20ms, which makes the Oculus Rift the best VR HMD so far, except latency. However there are still other things that need to be resolved before the product is delivered to

the consumer, like the accuracy of the tracking, the quality of the image, the image resolution. However with the support of Facebook, the enthusiasm of the developers and the development of the technology, we can hope that Oculus Rift will offer the best VR experiences to the player one day.

4 AUGMENTED REALITY

4.1 Definition

Augmented Reality is an advanced technology based on Virtual Reality, Augmented Reality is a technology that calculates the position and angle of the camera and adds related images on to the relevant real world in real time. The goal of this technology is to put the virtual world and real world on screen together and make interaction. The concept of the Augmented Reality was first mentioned in the 1990s, Tom Caudell and his colleagues from Boeing designed a system that overlaid some text tips and routing paths on machinist's head-mounted display in real time. The information could help the machinist teardown the machine easily (Lee, 2012). So far there are two general definitions of Augmented Reality, the first one is Milgram's Reality-Virtuality Continuum mentioned by Paul Milgran and Fumio Kishino in 1994 (Milgram and Kishino 1994). They point out that real environment and virtual environment on two sides of a line, as Picture 8 shows below, the middle space between them is called Mixed Reality, the space near the real environment is called Augmented Reality and the other side is called Augmented Virtuality.

Picture 8. Milgram's Reality-Virtuality Continuum

The other definition is more well-known.   According to Ronald Azuma (Azuma, 1997), Augmented Reality contains the following three points: It combines real and virtual; it is interactive in real time; and it is registered in 3D.

Augmented Reality is in a way combined virtual technology to re-observe the real world. AR is able to provide information we cannot gain from the real world, at a deeper level. This information can make the world more diverse. A strange street, in normal eyes, is just a street, but from the sight of AR, it will cover a different 'information coat', or another so called ' The Long Tail'. We will know

what the name of this street is, what kind of shops it has, which are some of the best shops. People can know everything they want. These are already some of the characteristics in some AR product at present. With the development of technology, AR could be more advanced and makes the interaction expand from accurate location to the whole real environment, from the simple communication between human and screen to integrating humans into the surrounding environment.

4.2   Types of AR systems

A complete AR system is combined with the real-time hardware components and relevant software. There are three main types of AR systems:

1. *Monitor-based* AR system: It is based on the computer monitor, a camera will take the image from the real world, input it to the computer, then compound the real world image and virtual image simulated by the computer graphics system, finally it shows on the screen and the user will see the final AR image. It is simple, but it cannot give the user much immersion. Picture 9 shows how the monitor-based system works.
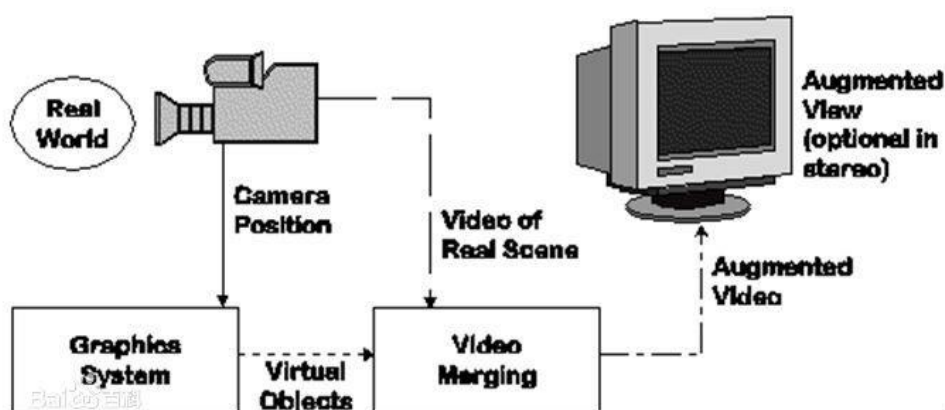
Picture 9. Monitor-based system

2. *Optical See-through* AR system: It is based on the optical principles and uses a head-mounted display. Unlike the Monitor-based system, the user can see the real world in this system. There are optical combiners in front of the user's eyes, which can be imagined as a glass, so the real world can be seen. The head tracker will track user's head position, then

the scene generator will generate related images and show them on the monitors and the user can see them on the optical combiners. Picture 10 shows how the optical see-through system works.
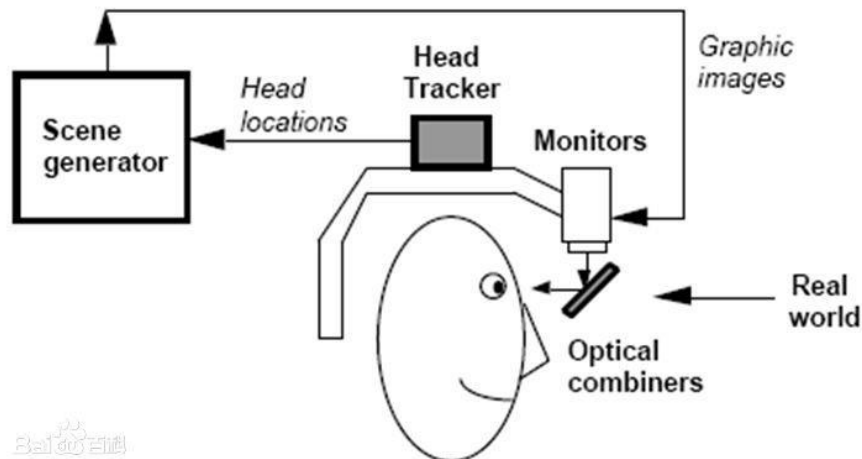
Picture 10. Optical see-through system

3. *The Video see-through* AR system, also needs HMD, and it is based on the video synthetic technology, quite similar to the optical see-through system. The only difference is that user sees the real world through the monitor. Video cameras will help to record the real world information before eyes, the video of real world will be combined with the virtual images and then the combined video will be shown on the monitor. Picture 11 shows how the video see-through system works.
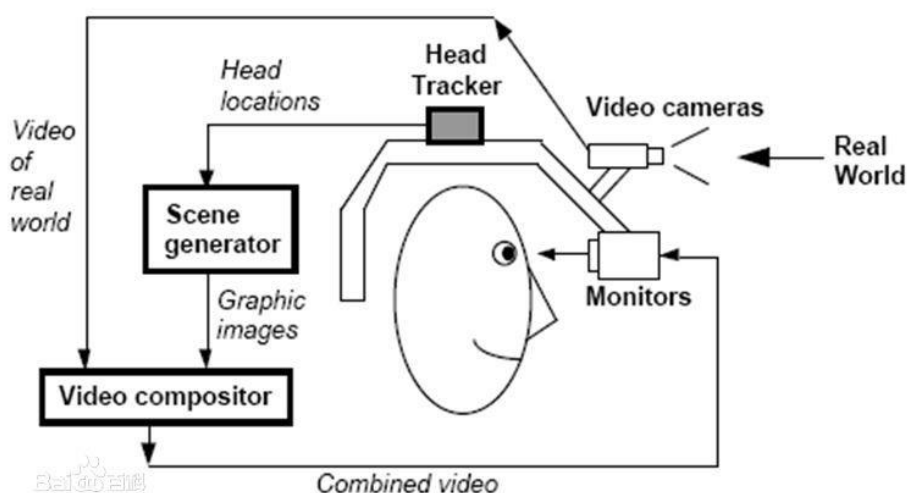
Picture 11. Video See-through system

To compare optical and video see-through systems, they both have advantages and shortcomings. This is a short analysis of them. The optical system is simple, because the system only needs to process one virtual image, the real image just directly seen by user's eyes, when the head tracker tracks the head position and shows the relevant images. the latency is short. However, the video system needs more time to handle both real and virtual images, so the delay is longer than in the optical system. Besides that, because of the extra video cameras, the video system is heavier than that of the optical system, and because the video compositor needs to process and combine the real and the virtual images, so it is more complicated than the optical system.

The resolution used to be a problem of the video system in the past, because the display technology was not advanced enough and human eyes can see the pixels clearly on the screen. However, now with the development of the screen technology, the PPI of the screen has developed far beyond the limit of human eyes, so this is no longer a problem. Another shortcoming that related to eyes is the eye offset. In other words, the video camera is not exactly at the same position as human eyes and might be a little higher on the HMD, so this might cause some misjudgments of the real environment. For example, if an engineer is trying to maintain a machine, the misjudgment of the position will most likely create a mistake. The solution to this problem is, use the mirrors to change the path of the light which will make the system slightly more complicated.

The optical system has a vital drawback. In order to reflect the information on the monitor at the optical combiners, the design of the combiners cannot be fully transparent, so the amount of the light will be reduced and the users may feel they wear a pair of sunglasses. On the other hand, the video system does not have such a problem.

One of the greatest strengths of the video system is that it has a wide FOV. The vertical range of human eyes is about 135 degrees, but the video camera can have 360 degrees of view if there are several cameras working at the same time. The optical might occurs the distortion so that the image will be rather unreal and strange, but with the video camera, the real world image is digital, by using the software, we can correct the distortion, and finally an undistorted

image can be displayed on the screen, and the user will have a better experience.

4.3   Use of Augmented Reality

The use of Augmented Reality has many application fields similarly to the Virtual Reality, such as sophisticated weapons, the development of aircrafts, virtual training, and entertainment. Because it has the characteristic of augmenting the real world, it has applications in fields like medicine, the maintenance of precise instruments, military and engineering. AR technology has more obvious strengths than VR technology.

Medical field:

Application in the medical field offers huge benefits. As mentioned in Chapter 3, the VR technology can be used in this field. Students can learn the human body and do the surgery. However AR technology can do better, students still need to observe the true human body and surgeons need to have an operation in real life. At this time, AR technology can play a supporting role, for example it can indicate which places on a human body should start use the scalpel to do the incision or which place has the tumor. This assistance can help the surgeon perform his surgery more precisely and enhance the success rate of the operation.

Military field:

AR technology can help the troops obtain location in real time, and the geographic data on a strange battlefield. AR also provides basic navigation and airplane information, target location in the military aircrafts.

Museums and historic sites:

When we visiting a museum, we can see many cultural relics, some of those relics are not complete, and if there are many visitors, it is hard to see the explanation of that relic, some of the outdoor historic sites also fragmentary. AR technology can bring a whole new experience, visitors can see the reconstruction of the imperfect relics, what its original shape was.

Entertainment:

AR games can improve the interaction between players and the real environment. The mobile game Ingress is probably the best example of the AR games. This game has a simple science fiction background. There are two camps that player should choose from at first, then player needs to find and collect the items on the map, the item can be seen as the experience, then find the 'Portal ', attack or defend. The map in the game is based on the Google Maps, so the location of the item and the Portal is real, they can be a museum or a specific building. The player needs to go to that place in the real world to collect the item and get close to the Portal. This is a worldwide game and there are some clubs offline gather the player to hold some events, play this game together, so this is a really cool thing and much more interactive with the player. Picture 12 is the user game interface of Ingress, and, the map in the background is of New York City.
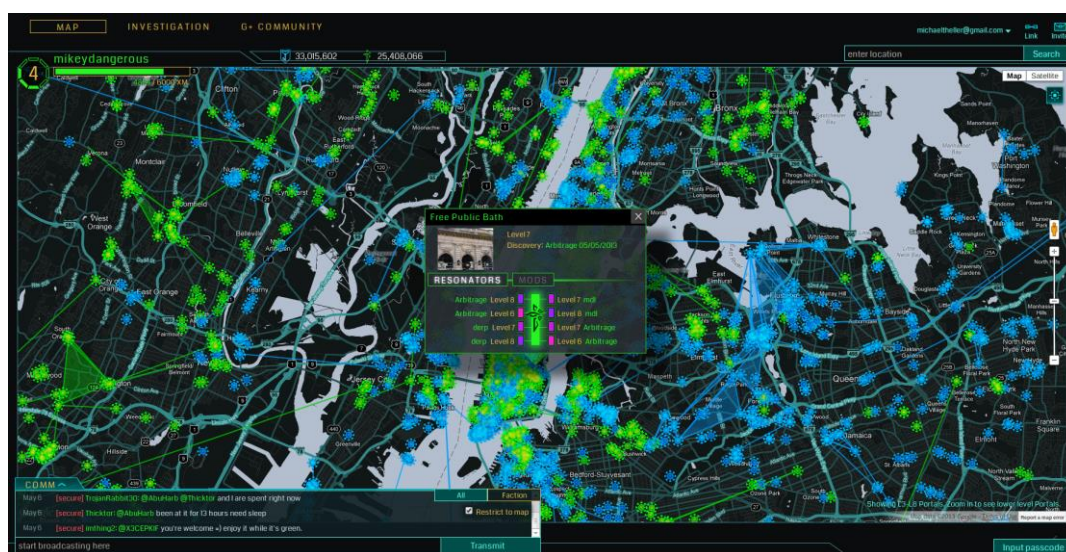
Picture 12. Ingress game

Navigation:

The traditional way of navigation on the mobile phone is good, but not interactive, for people in a strange place, sometimes cannot be sure if this is the right place shown on the map. Using AR technology in t navigation, those problems can be solved perfectly. Nokia City Lens is the best software to describe this. Through the camera and GPS on a mobile phone, the camera

can display the scene around you, like shops, restaurants, hotels, even bus stops. There are several labels on the screen showing the exact location of those buildings, and their approximate distance. When we click on one of the labels, it will show the detailed information, like the name of the restaurant, the special food of it and the feedback from people. When the place is a little far from your location, the software can give turn-by-turn navigation and ways of transportation to arrive at the destination. This is a powerful tool when travelling to an unfamiliar place, more convenient and interactive with the people. Picture 13 below is the user interface of this software.

Picture 13. Nokia City Lens

Other fields:

In industrial maintenance, the HMD can offer supplementary information to the user, including the virtual panel, the internal structure of the device and the blueprint of the item. In the television relay, such as basketball games, it can overlay the basketball player's data, like points, assists, rebounds, so the audience can have more information. In video chat communication, the user can wear some funny virtual hat or glasses to make the video chat more interesting.

4.4   Realistic way to achieve AR

There are about three ways to achieve AR: GPS and sensor, marker recognition and PTAM.

GPS and sensor are mostly used on the smartphone. At present, nearly all the smartphones have those sensors The GPS sensor can get the longitude and latitude of the location and some of the smartphone even have a barometer sensor to get the height, the electronic compass to get the orientation. The accelerometer sensor can get the tile angle and lastly, using the camera on the phone to get the surrounding image and the related information can be shown on the screen. This is the simple way to achieve AR technology. Nokia City Lens is a good representive of an AR application. Therefore, we can conclude the AR is not a high-end technology that normal people can not used. On the contrary AR technology is already in our daily life.

Marker recognition stores the Marker images in advance, uses the camera to capture the image, then uses the image recognition technology to recognize the Marker image in the database, also gets the position and direction of the Marker, finally overlays related information and displays it on screen. This technology is based on an open source project called ARToolKit. It is a computer library written in C/C++ language and helps developers to create an AR application more easily. In order to develop an AR application, the most difficult part is to overlay the virtual images on the real images in real time and make a precise alignment to the real world. ARToolKit uses computer graphic technology to calculate the relative position between the camera and Marker, so developers can overlay the virtual image on the Marker successfully. ARToolKit includes tracking library and full source code and developers can adjust the interface in different platforms.

One example of Marker recognition is a mobile application called Blippar. Blippar cooperates with some famous brands to help them show user friendly advertisements. When we open the application on the smartphone, the application will scan the Marker 'B'. If there is no 'B', it is still ok if the advertisement has cooperation with Blippar, after a few seconds, an AR advertisement will show on the screen and user can make a   simple interaction

with the advertisement, can click on it to see more information about the product. Nevertheless, this way to achieve AR has many limitations, because it should define the Marker beforehand, and it cannot gain information outside the Marker, so it is only suitable for games, books and advertisements. Picture 14 is how Blippar works.

Picture 14. Blippar application

Image analysis and recognition uses the camera to analyse the objects, space and the surrounding environment and then overlays related information. The most famous related project is Parallel Tracking and Mapping, shortened as PTAM. It is a camera tracking system for AR, first showed on ISMAR 2007. Georg Klein and David Murray from the University of Oxford published a thesis called Parallel Tracking and Mapping for Small AR Workspaces (Klein and Murray, 2007), it is an open source project whose main purpose is to capture the feature points using the camera, then detect a plane, next build a virtual 3D coordinates on that plane, and create the composite image and CG. The unique feature is parallel processing which is used to detect the 3D plane and create composite the images, so it can save a lot of time to solve some large and complicated operation. This could be one method to achieve AR, because there is no need to find the special Marker. AR can be achieved by analyzing the image and building real images on coordinates and compose AR more naturally. However this method also has some problems, like the large number of data to be processed, differentiated and analyzed on the plane and the

vertical plane. The technical difficulty of PTAM is relatively large, so PTAM AR still needs to be developed. Picture 15 shows how PTAM works.
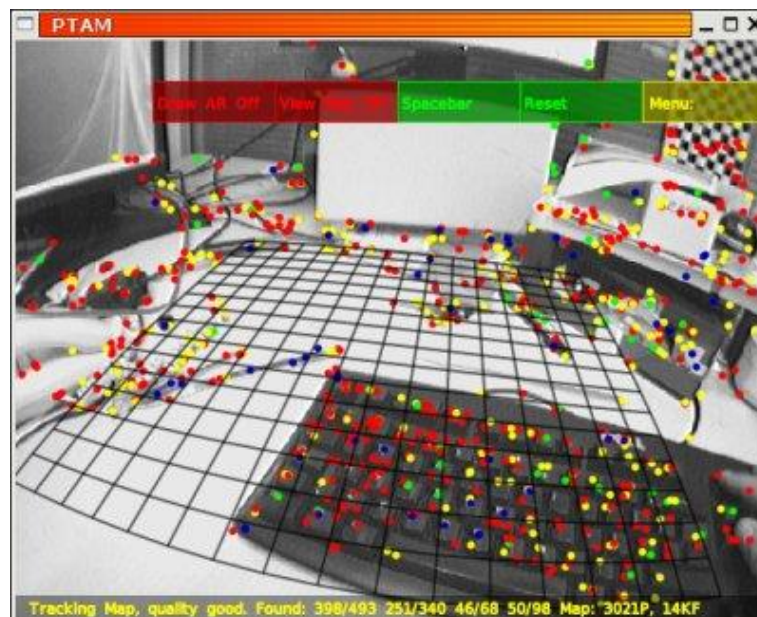

Picture 15. PTAM

4.5 Challenges in Augmented Reality

The first challenge is the same as in VR. In order to fully immerse the user, it is best to have the image in real time, but with different problems. Latency is always a problem because AR has a high demand for real time synchronization. As the user moves in the real world, the virtual information should be overlaid as soon as possible, so the user can get the correct information. Minimized latency will enhance the immersion of the user.

Then there is the computing capability. Unlike the VR which only needs to handle the virtual signal, AR has to analyze the real world environment and at the same time manage the virtual signal. There is vast amount of data in AR, so the hardware should be powerful. The hardware is a head-mounted device, so it should be tiny, light, at the same time. The power should be as low as possible, otherwise the battery life and the heat will confuse the user. Based on Moore's law, a small but powerful hardware is not a problem in the future. The capture and tracking system need work accuracy, so that they can give the user the right information, the software should also be optimized to filter the information, remove useless data and keep the useful data.

Another challenge that needs to be addressed is the upstream bandwidth. In order to show the virtual information on the real world, the AR device needs to send the real world data onto the nearby server, and the nearby server will give back the information. Because of the character of the AR technology, there is a high demand for the upstream bandwidth, but nowadays the wireless internet device has a large downstream bandwidth. According to a recent study, shown in Picture 16 below, the downstream bandwidth is 3 to 10 times larger than the upstream bandwidth (Broadband Backhaul, 2014). So for purpose of better use of AR HMD in the future, designing a new wireless Internet infrastructure is necessary.
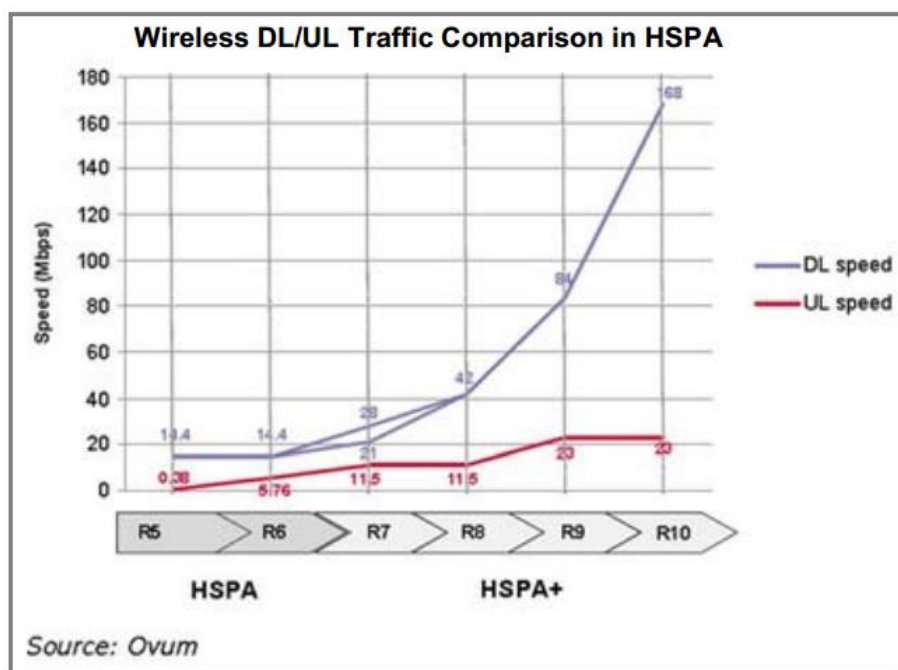
Picture 16. Wireless DL/UL traffic comparison

The last but not the least concern of AR device is privacy. AR devices are very attractive and, a lot of people want to use them, but the privacy still needs to be considered. Like Google Glass, the camera on it allows it to take pictures or videos without being noticed, which means that a user can record a movie at the cinema, and when walking on the street, the AR technology might find out the users name, age and other information through liking in Facebook. Although these actions looks trendy, they actually violate privacy and that is why after Google Glass was launched, some companies forbade their employees

wearing Google Glass on some special occasion. So these kinds of devices need a serious discussion on privacy.
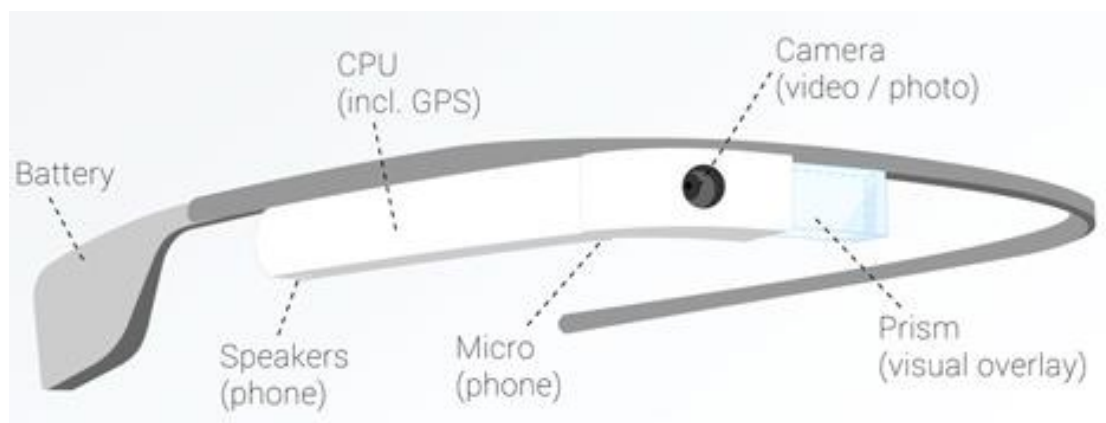
4.6 Google Glass

4.6.1 Brief introduction

Google Glass, first presented in 2012 by Google Glass (Goldman, 2012), was seen as the first truly AR hardware device in the world. The aim was to offer ubiquitous computing to consumers, Google Glass has some functions that the smartphone has, like taking pictures, having a video chat, sending messages, surfing the internet and giving a road direction. So far it has two explorer editions at the price of 1500 dollars, but in January 2015, Google announced that it discontinued the explorer edition of Google Glass. The public has no idea if this move means that Google gave up this project or not. Although Google Glass seems dead, it brings several labels into `the public eyes: Augmented Reality, head-mounted display, wearable device and wearable computer. Augmented Reality has become a hot topic in the last two years and people start to look into new ways of interaction with the computer.

### 4.6.2 Composition of Google Glass

The main components of Google Glass are shown in Picture 17 below. It has five megapixels camera, can record a 720p video, on the side of the glass there is a track pad, the user can swipe forward, backward, down and tap. Inside the trackpad is the motherboard, with CPU, gyroscope, accelerometer, compass and some sensors. There is a one-day battery on the back, the sound system uses the bone conduction transducer to play the sound, and Bluetooth is used to connect to the mobile phone. At the front, there is a display, the size of it is like a 25-inch high definition screen from eight feet away, and there is a prism used to reflect the visual images from the projector to the user 's eyes (Support.google.com, 2015). The weight is so light so it just likes normal glasses.

Picture 17. Composition of Google Glass

4.6.3 How Google Glass Works

Here is a simple explanation on the display technology behind the Glass, how it shows virtual information to user's eyes. There is a projector on the Glass. Unlike the normal projector technology, Google Glass uses Liquid Crystal on Silicon, a display based on reflection which has a better light utilization, so it can achieve a high resolution and color performance, and it is power saving and cheap, thus it is the best choice for the mobile projector. First, the LED light shines through the polarizing beam splitter onto the LCoS panel, then it reflects on the PBS again and refracts at a 45-degree angle to the collimator at the end of the prism, next reflects on the partially reflecting mirror and refracts at another 45-degree angle into user's eye. That is how the user sees the virtual overlaid image. Depending on how the user wears the Google Glass, normally, the virtual image will show on the top right corner of the visual field, and, at the same time, the real world lights just shine through the partially reflecting mirror into the user's eye (Google Inc., 2013). Besides the display, the communication between the user and the computer is also important in the AR device. Voice communication is one of the interactions on Google Glass, but there is a problem, if we want to use Google Glass outside in a noisy environment, to keep the voice communication clear, Google uses the bone conduction transducer. Bone conduction uses human cranium as the medium of sound, and conducts the sound directly into the human brain, so it has the best effect of sound conduction, and s the Glass more convenient. Picture 18 below shows how Google Glass display works.
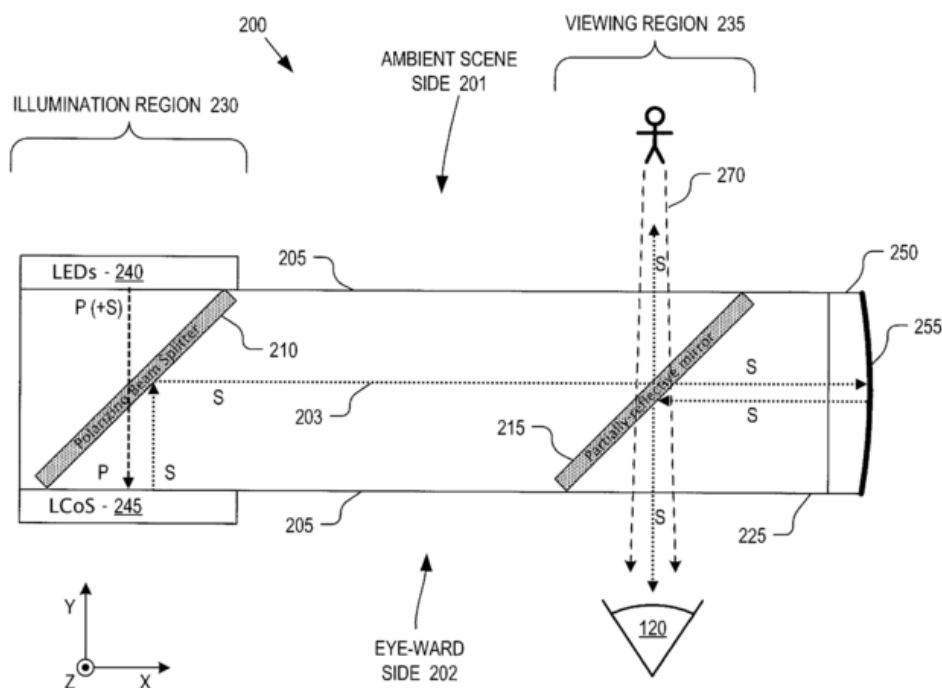
Picture 18. Google Glass Explorer Optics Schematic

4.6.4 Features of Google Glass

Google Glass has the features that a smartphone also has. It runs the customized Android operating system, so it is just like a wearable smartphone. Voice recognition is one of the main interactions of Google Glass. When the user says "OK, Glass", a menu will show with some icons indicating actions such as take pictures, record a video, use Google map, call somebody, send messages or e-mail. This is very convenient when the user does not have a free hand to hold the phone. For example, if we want to take a picture, we just need to say "OK, Glass, take a picture", and Google Glass will take a picture, or there is a picture button on the top of the glass and we can tap that button to take a picture. The same applies to recording a video. This is one of the cool features which make Google Glass special because the picture can be taken as a first person looking, the special camera angle can record your personal life in a unique way and we can have both of the hands free.

Another killer application is the Google Map on the glass. HMD is the best device to show the map application to the user, because it is hand free. Thus, it does not matter if the user is walking, riding or driving, the map application on

HMD is extremely convenient. When wearing a Google Glass, we just need to say "OK Google, get the direction to some place", and we can choose the way we go there, the turn by turn navigation will show on the right side of the eye, so when we are riding a bike, we do not have to worry about accidents because both hands are on the wheel.

Other basic build-in features include sending or receiving e-mail and text messages, Google search, looking at the weather. Google Glass has the potential to become more functional and powerful if the developers are willing to develop new applications on this platform. Many third party developers and companies have already made some applications like some news apps, The New York Times, social network apps like Facebook and Twitter, note apps like Evernote, and facial recognition, exercise, translation.

In conclusion, Google Glass is a small but powerful device. It includes a lot of high-tech hardware like Bluetooth, Wi-Fi, gyroscope, accelerometer, compass, camera, microphone, speaker, and a tiny screen. The shape is just one pair of normal glasses. It uses the most natural way of human communication: voice recognition to interact with the computer, all we need to say is "OK, Glass". The convenient navigation will lead us to our destination, we will never get lost. Real-time acquisition is like an intelligent assistant, there is no need to worry about forgetting important things, because it can remind the user the flight information, hotel information, meeting time. It also supports both Android and iOS operating system, works as an extra device, answers phone calls. The design of the Google Glass is fashionable and it even appears in some fashion shows. This is what the future technology looks like and is the core of human computer interaction, ubiquitous computing. The computer will integrate into the surrounding environment, disappear from the eyesight, people can gain and manage the information in any place, any time, any way.

4.6.5 Main problems of Google Glass

It is a fact that Google has officially announced that it will stop selling the explorer edition of Google Glass. After around two years, here are some possible main problems of Google Glass.

The first one is the privacy. It is not only Google Glass that has this problem. All the wearable smart devices have the same concern, especially those devices that have cameras can take pictures or record videos without any notice. Because of that, Google Glass has been forbidden in many public places, like bars and restaurants. At the meantime, because of the pirated movies, a lot of cinemas in America also prohibit Google Glass and other wearable video equipment.

Second is the safety and security problem. When wearing the Google Glass, due to the virtual image on the top right corner of the visual sight, in order to see the information, the user needs to focus on the corner. This might cause distraction, especially for the drivers, and accidents. On the other hand, Google Glass is connected to the mobile phone through Bluetooth, so hackers can steal the smartphone password by cracking the Glass, and steal other information inside the mobile phone. However there are other ways to improve security, for example by using fingerprints to unlock the phone.

The third problem is the lack of "killer application" and the target audience is not clear. So far on Google Glass, some of the core applications just work like on the mobile phone. There is no application that motivates the users to use Google Glass to achieve that aim. Google Glass acts too much like a phone on the head. It becomes an unnecessary product unless there is one or several "killer applications" that makes the user want to keep wearing the Google Glass. Except that, the market positioning of Google Glass is unclear. Google wants Google Glass to be a necessary item in our normal life too much. Just like the shape of Google Glass too much like as a normal pair of glasses, this strategy is not good for a potential game changer product. At present, the smartphone is still the main market and people can now do anything on the smartphone, thus it is better for Google Glass to be used in specific fields.

As the first true augmented reality wearable device, Google Glass takes too much pressure. Although it is no longer on the market for now, it is hard to say whether Google Glass is success or failure. Google Glass is a quite advanced product, there is a long way for it to become popularized, but it represents the thought of the next way of computing, a much more interactive product for

humans. It is undeniable that Google Glass has many flaws, but with the saturation of the mobile computing market, it needs a new product to step into the next revolution. Google Glass is more like a symbol that bring AR technology into the public eye, encourages more and more companies to develop this technology together, and makes AR become the next stage of human-computer interaction.

4.7   HoloLens

4.7.1   Announcement of HoloLens

On the 22nd of January 2015, Microsoft announced a fantastic Augmented Reality HMD called HoloLens (Hempel 2015), a short name of Hologram Lens. A lot of people became shocked by watching the demonstration which felt like moving scenes in science fiction movies to the real world. Because HoloLens is an Augmented Reality device, there is a short comparison to the Google Glass. Unlike the Google Glass which can only overlay a two-dimensional layer for the user, from the demonstration of HoloLens, we can see that it can also overlay three dimensional images, just like the same as hologram technology, so the user will feel more vivid by using HoloLens. From this angle it is more powerful than Google Glass. Then the HoloLens supports gestures, which means that the user can tap, swipe, pinch, zoom in and out through the air. This greatly strengthens the interaction between user and computer. Last but not the least, in our opinion, HoloLens will not have the same end as Google Glass because Microsoft Pictures out that this kind of technology should first be used in inside room, like the home and office, so people will not feel strange that one guy wears a big HMD and make some weird gestures. Afterwards this behaviour will be accepted and popularized by people and it is not hard to see people wear it outside. Picture 19 below is the prototype of Microsoft HoloLens.

Picture 19. Microsoft HoloLens

4.7.2   How Microsoft HoloLens works

At present, Microsoft just announced HoloLens but it is still needs time for the development version, there are no technical details from the official announcement so far, so the next points is only represent our point of view that HoloLens probably works like this. First, although Microsoft called it a Mixed Reality Product, HoloLens is basically an Augmented Reality product. It has an independent computing unit, CPU and GPU from Intel and one Holographic Processing Unit (HPU) and it is an application-specific integrated circuit customized by Microsoft.

It is clear that HoloLens is not a VR device, totally different from Oclulus Rift whose feature is to immerse the user into the computer simulated three dimensional world but it cannot help people better understand the real world. HoloLens is not Google Glass. It is much more advanced than Google Glass, HoloLens has the ability of three-dimensional perception, after scanning, HoloLens can model the surrounding world, but Google Glass can only see the two-dimensional pixel value. Next HoloLens is capable of 3D rendering and stronger interaction, not only voice recognition but it can also use gestures to control the device. HoloLens is not an AR device based on camera for mobile phon. As mentioned before, these applications can only detect the panel of the special image, but HoloLens can detect 3D scenes from any angle.

Going back to the definition of Augmented Reality, in order to fulfill the AR, it must first understand the reality through the camera data to gain the 3D depth information. So why can the camera on HoloLens perceive the depth? It is all because of the technology accumulation on another Microsoft product, Kinect. There is a CMOS infrared sensor on Kinect and also on HoloLens. There might be 4 cameras two on each side, and those four cameras with the sensors can cover a wide horizontal and vertical angle. The CMOS infrared sensor can detect surroundings by black and white spectrum. Pure black means infinitely far, pure white means infinitely near and the grey zone between them means the distance between the item and sensor. The sensor collects every single point in the visual field, simulates a depth map in real time. The depth map is used for Stereo Vision technology. The aim of Stereo Vision is to rebuilt 3D scenes by using the camera, usually two cameras to gain the distance data. Unlike the human eyes, the computer stereo vision is quite backward. One of the main challenges of AR on the HMD is to better manage the stereo vision. Picture 20 is the sketch map of the Stereo Vision.
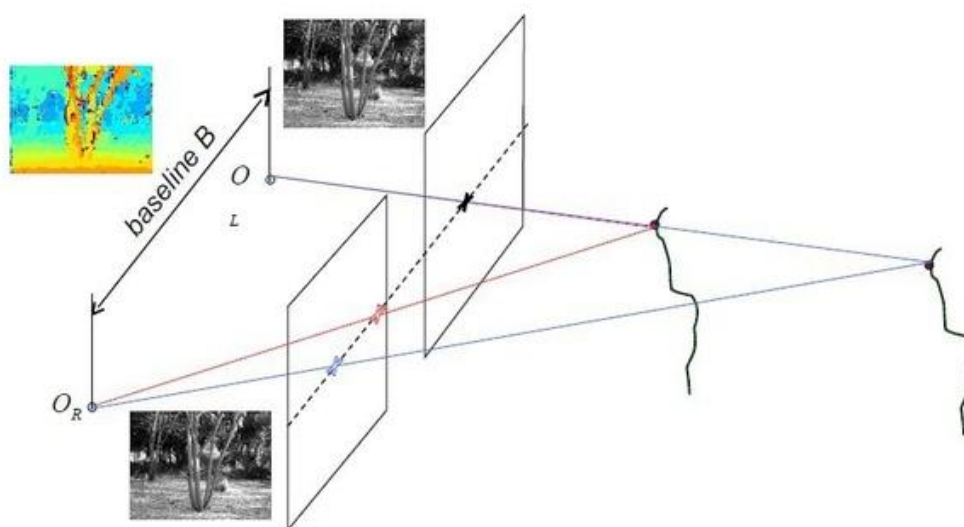
Picture 20. Stereo Vision technology

The first step of the Stereo Vision is to correct camera distortion. The lens on the camera has the distortion, so in order to gain the precise data, the camera needs to be corrected., The common way of correction is to take a few pictures of the chessboard, then calculate the matrix parameter of the camera and the distortion can be corrected. The next step is image rectification. Because two

cameras stand in different positions, there is a deviation of two cameras. The left camera can see the leftmost image and the right camera can see the rightmost image. The aim of image rectification is to get the same part of the image. Next is image correspondence. Left and right images should be matched and the last step is to use the reprojectImageTo3D function in OpenCV to build the depth map (Docs.opencv.org, 2014). However, getting only one depth map is not enough. It only represents the image at one moment. To get the whole 3D scenes, there is a series of depth map that need to be analyzed. Picture 21 shows the depth map.


Picture 21. Depth map

4.7.3 How to rebuilt 3D scene from depth map

The answer to rebuilding a 3D scene is SLAM, Simultaneous Localization And Mapping. It is a system used in robots, driverless cars, unmanned aerial vehicles and some other computing systems.   The aim is that the robot starts in an unknown environment, using the SLAM system to locate its position and rebuilt the 3D map in real time, at last arrive at the destination. The SLAM system deals with a very philosophical problem, what the world looks like and where I am. Based on SLAM, Microsoft created its own algorithm called

KinectFushion (Izadi et al. 2011) and published two theses. After using SLAM on Kinect, there is a high possibility of SLAM used on HoloLens. The user wears HoloLens and moves in a room, the system will gain different angles of depth maps, iterate in real time and accumulate those depth maps to calculate an accurate 3D model in a room.

KinectFushion rebuilds a 3D scene from the depth map in four main stages. The first stage is called depth map conversion. It converts the depth map from the image into space coordinates, saves them as floating points and calculates the vertex coordinate and normal vector on the space. The second stage is called camera tracking and it calculates the camera pose, including positions and orientation under global coordinates, by using the iterative alignment algorithm to follow the values, thus the system will always know the change of the camera from the original position. The third stage is called volumetric integration and it integrates the depth data from stage two into a 3D space. The fundamental element is not the triangle but square. In the demonstration of HoloLens, there is one scene that a user wears HoloLens in his room to play the MineCraft game. Thus, some of the HoloLens applications run on stage three. The fourth stage is called ray casting and it is a common way of 3D rendering, because HoloLens is AR HMD, the screen is transparent, so there is no need to render the room because the human eye has already rendered it. Picture 22 depicts   the stages of KinectFushion.
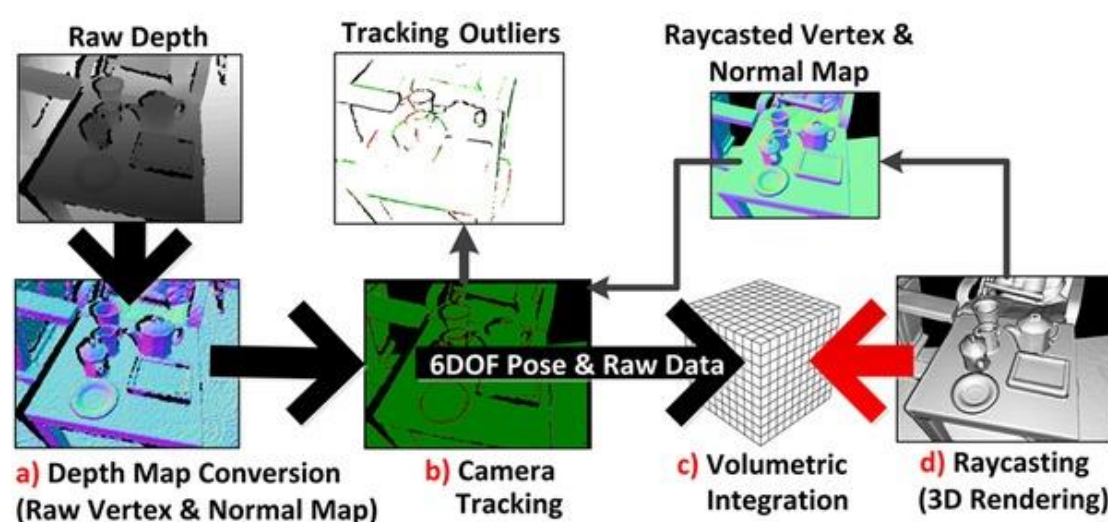
Picture 22. KinectFushion

4.7.4 How gesture recognition works on HoloLens

There is also that HoloLens will use the same gesture recognition technology on Kinect. Gesture is one part of communication in human daily life, use this technology on computer, people can free their hand, use a more natural way to interact with the computer. In the field of HCI, gestures represent a simple command, have arbitrary meanings, are based on conventions and might be misunderstood. To achieve gesture recognition, Microsoft created a Skeleton Tracking system.

On Kinect, the Skeleton Tracking system will analyse the depth map data to extract the human body silhouette, then create the coordinates of human joint. There are around 20 joints on the human body that need to be tracked, and these 20 joints simulate a skeleton system. By recording and comparing the skeleton with the database, the computer does the related task.

As on HoloLens, the camera only needs to capture the hand gestures and does the same thing to achieve the gesture recognition. However, as mentioned before, unlike the command, the gestures cannot be exactly the same every time, so this kind of gesture recognition highly relies on the machine learning. The recognition engine will record, handle and reuse the data, so as the time goes by, the precision of gesture recognition will be enhanced.

4.7.5 Strengths and drawbacks of HoloLens

From the demonstration of HoloLens, this AR HMD can be used in many different scenes. First is the game play, users can play a game like Minecraft based on their real living room, some buildings on the sofa, some on the table and some on the floor. It is a quite interesting experiences. The second demonstration is the video chat, by using Skype, a father can help his daughter fix the tap, the father can teach and show the virtual array on how to operate, so the daughter can fix the problem easily. The third demonstration is in the office, an architect is designing the shape of a fancy motorcycle, she can see the 3D model of the motorcycle and adjust a little as she wants, this is a much more direct way that architect can change his design. The final demonstration

is cooperating with NASA to explore the land on Mars, user will be on the surface of the Mars, the color, the landform, just like the real.

From the four demonstrations HoloLens, we can see that it has great potential to be the game changer as it can be used in many different fields. Compare with another AR HMD Google Glass, HoloLens is much more advanced.

Although AR HMD has much potential, Google Glass still failed and left some unsolved question that so far HoloLens cannot give the solution, either.

As the pioneer of the software products around the world, Microsoft will fully support developing applications related to HoloLens, but as HoloLens is a brand new highly integrated product, the development of third party applications is a huge challenge. Because at first HoloLens is an independent product, so for the developer, it is a new platform. In addition, there are a lot of different sensors on HoloLens, so developers will not only think about the quality of the application, but also would want to make sure the interaction between human and machine. Finding the balance between virtualilty and reality, and creating an immersive experience are not easy to achieve. Secondly, refer to another product Kinect, there are also few developers who can make a high quality game.

Compared with Google Glass, HoloLens is quite large, as a wearable device, it should be worn outside. So there is a portability problem, Besides that, it is still unknown how the battery life is, whether there is a heating problem and whether it can satisfied shortsighted users. Last but not the least is the price of HoloLens and popularization of it. For example, in the case of Google Glass, the high prices and a high profile promotion are the main factors that it failed. At this stage people still define AR and VR as high-end electronic products and the market is for developers and fans.

There is not doubt that the research and development cost on HoloLens is far more than that of Google Glass. Thus it is hard to make it popular at least in the first iteration. What Microsoft should consider is how to make HoloLens more useful and on this basis try to persuade more consumers and potential users to buy this product.

5 CONCLUSION

In conclusion, as a bold try of the next stage of human-computer interaction, virtual reality and augmented reality now draw people's attention. Several high tech companies start to develop these kinds of devices one after another. From the analysis, we can see that the future of HCI is to let people more free, bring back a natural way to gain information, let people rely on the simpler things. At this point, many people may say usability, but in fact it is not good, because usability is from the angle of the product. What is more important is to put people first.

There are two ways to achieve the freedom that people expect. The first, one is the virtual way, people put their mind into a vast virtual space and release it. A pure digital spirit can be built, people get their freedom and this is what virtual reality technology wants to do. This is easy to achieve, so many companies develop virtual reality and there are some great products on the market. However, this direction is not that ideal, no matter how real the virtual environment is. After all it is a lie and people will find a way to go back to the real world. Thus virtual reality is more popular in games and design industries. In the long term, augmented reality will become a ubiquitous technology, It breaks the limitation of space and time, combines the digital content and real world, people's body and spirit can get fully used. Technology strengthens our body and serves our mind, people can live in a free and real world, hence the augmented reality can be used in many different industries.

Both virtual reality and augmented reality have years of history. Virtual reality even has more than 40 years history. There are two main reasons why they are not popular at present. Firstly, from a scientific and technological point of view, there are many limitations that need to be solved, like the battery life, power of the processor, motion-sickness problem with the HMD. The other reason is the rigid demand from the people whether they really need this technology in their daily life, or it is only needed in some the fields.

All in all, as a revolutionary technology, although there are some barriers between virtual reality and augmented reality, Microsoft, Google, Oculus Rift

and other companies, those great development teams, never give up finding the next way of HCI. Even if those advanced products disappear in the history, they still have the unforgettable value, to lead the industry step into a new time. From the PC to the smartphone, we saw the new interaction come, from smartphone to the wearable device, we will look forward to the next way of interaction.

REFERENCES

Abowd, G. D., and Beale, R. (1991). 'Users systems and interfaces: a unifying framework for interaction. In: D. Diaper and N. Hammond.   Eds: HCI'91: People and Computers VI, pages 73-87.   Cambridge: Cambridge University Press.

Abrash, M. (2012). [Blog] Available at: http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr/

Azuma, R A Survey of Augmented Reality Presence: Teleoperators and Virtual Environments, pp. 355–385, August 1997.

Broadband Backhaul. (2014). 1st ed. [ebook] Ceragon Networks Ltd. Available at: http://www.ceragon.com/images/Reasource_Center/Technical_Briefs/Ceragon_Technical_Brief_Asymmetric_Transport.pdf.

Docs.opencv.org, (2014). Camera Calibration and 3D Reconstruction — OpenCV 2.4.11.0 documentation.   [online]   Available   at: http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html [Accessed 7 Apr. 2015].

Edwards, B. (2008). The computer mouse turns 40. [online] Macworld. Available at: http://www.macworld.com/article/1137400/mouse40.html

Google Inc., (2013). Lightweight eyepiece for head mounted display. US20130070338 A1.

Google Inc., (2013). Heads-up display including eye tracking. US 20130207887 A1.

Goldman, David (April 4, 2012). "Google unveils 'Project Glass' virtual-reality glasses". Money (CNN).

Hewett; Baecker; Card; Carey; Gasen; Mantei; Perlman; Strong; Verplank. (2009) "ACM SIGCHI Curricula for Human-Computer Interaction". ACM SIGCHI.

Heim, M. (1993). The metaphysics of virtual reality. New York: Oxford University Press.

Hempel, J. (2015). Project HoloLens: Our Exclusive Hands-On With Microsoft's Holographic Goggles | WIRED. [online] WIRED. Available at: http://www.wired.com/2015/01/microsoft-hands-on/.

iFixit, (2014). Oculus Rift Development Kit 2 Teardown. [online] Available at: https://www.ifixit.com/Teardown/Oculus+Rift+Development+Kit+2+Teardown/2763

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., ... & Fitzgibbon, A. (2011). KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the 24th annual ACM symposium on User interface software and technology (pp. 559-568). ACM.

Johnson, E.A. (1965). "Touch Display - A novel input/output device for computers". Electronics Letters 1 (8): 219–220.

Kickstarter, (2012). Oculus Rift: Step Into the Game. [online] Available at: https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game.

Klein, G. and Murray, D. (2007). 1st ed. [ebook] Available at: http://www.robots.ox.ac.uk/~gk/publications/KleinMurray2007ISMAR.pdf .

Lee, Kangdon (March 2012). "Augmented Reality in Education and Training". Techtrends: Linking Research & Practice To Improve Learning 56 (2). Retrieved 2014-05-15.

Licklider, J. (1960). Man-Computer Symbiosis. IRE Transactions on Human Factors in Electronics, HFE-1(1), pp.4-11.

Norman, D. (1988). The psychology of everyday things. New York: Basic Books.

Oculus VR, (2014). The All New Oculus Rift Development Kit 2 (DK2) Virtual Reality Headset. [online] Available at: https://www.oculus.com/dk2/.

OCULUS VR, LLC, (2014). 1st ed. [e-book] OCULUS VR, LLC. Available at: http://static.oculus.com/sdk-downloads/documents/Oculus_Developer_Guide_0.4.4.pdf.

P. Milgram and A. F. Kishino, Taxonomy of Mixed Reality Visual Displays IEICE Transactions on Information and Systems, E77-D(12), pp. 1321-1329, 1994

Riftenabled.com, (2015). RiftEnabled™ – Oculus Rift enabled list of games & demos. [online] Available at: http://www.riftenabled.com/admin/apps/.

Shackel, B. (1959) Ergonomics for a computer. Design 120, 36-39.

Sutherland, I. E. (1968). "A head-mounted three dimensional display". Proceedings of AFIPS 68, pp. 757-764

Summers, N. (2013). 10 Experimental Projects that Completely Rethink Computer Interfaces. [online] The Next Web. Available at: http://thenextweb.com/insider/2013/05/28/instinctive-innovation-10-experimental-projects-that-completely-rethink-computer-interfaces/

Wikipedia, (2015). Field of view. [online] Available at: http://en.wikipedia.org/wiki/Field_of_view.

Support.google.com, (2015). Tech specs - Google Glass Help. [online] Available at: https://support.google.com/glass/answer/3064128?hl=en&ref_topic=3063354.

Yin-Poole, W. (2015). Watch Valve's HTC Vive VR Rezzed developer session. [online] Eurogamer.net. Available at: http://www.eurogamer.net/articles/2015-03-13-watch-valves-htc-vive-vr-rezzed-developer-session.