Title	The development of shared virtual reality tourism system with emotional connections
Sub Title	
Author	王, 思媛(Wang, Siyuan) 小木, 哲朗(Ogi, Tetsurō)
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2021
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2021年度システムエンジニアリング学 第337号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002021-0011

Master's Dissertation 2021

The Development of Shared Virtual Reality Tourism System with Emotional Connections

Siyuan Wang
（Student ID Number：81934610）

Supervisor　Tetsuro Ogi

September 2021

Graduate School of System Design and Management,
Keio University
Major in System Design and Management

SUMMARY OF MASTER'S DISSERTATION

Student Identification Number	81934610	Name	Siyuan Wang

Title

The Development of Shared Virtual Reality Tourism System with Emotional Connections

Abstract

Tourism is travel for pleasure during the time of leisure and it is a common and effective way to build and strengthen the relationship between people. However, living with physical impairment or an ongoing infectious pandemic can have a large impact on physical tourism. A technology that has advanced and became more affordable is Virtual Reality (VR). VR tourism provides a substitute for physical travel. While most existing systems provide only a single-user experience, this shared VR tourism system aims to introduce socialization features in VR tourism and explore the effect on maintaining emotional connections.

In this master thesis, we present the design of the shared VR tourism system that allows multiple users to immerse themselves in the travel experience together with HTC Vive Headset and interact through non-verbal communications realized by eye-tracking and lip tracking technology to build emotional connection effectively. And the performance was evaluated based on 3 variables: co-presence, social presence, and satisfaction. The results showed that this system can improve the emotional connection in the VR tourism experience.

Key Word(5 words)

Shared Virtual Reality, Tourism, Eye tracking, photorealistic avatar, non-verbal communication

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

1.1.1 Overview of Virtual Reality (VR)

Virtual Reality (VR) is the computer-generated simulation of a 3D environment, which seems very real to the person experiencing it, using special electronic equipment. VR is the technology that immerses you in a simulated world. The objective is to achieve a strong sense of presence (being there) in the virtual environment [1].

The display devices are the technological equipment used to visualize the VR content. They are classified according to the level of immersion they can provide, which is the perception of being physically present in a non-physical world. Immersive systems can be characterized by the sensorimotor contingencies (SCs) that they support. SCs refers to the actions that we know to carry out in order to perceive, for example, moving your head and bending down to see underneath something [2]. Since immersion is objective, one system is more immersive than another if it is superior in at least one parameter while others remain equal. There are three categories of immersive display systems, non-immersive, semi-immersive and immersive.

- Non-immersive systems are simpler devices that use a single screen, such as a desktop to display the virtual environment. A typical desktop display system is to display a virtual environment in a web browser with WebVR. WebVR is a JavaScript API that allows web applications to present content in VR. And user can interact with the virtual environment by using mice or some VR devices [3].

- Semi-immersive systems use a set of large projections to display the virtual environment on walls, enveloping the viewer, such as the cave automatic virtual environment (CAVE), or the

Powerwall screen. Semi-immersive display system depends on a consistent projector system with motion capture [4].

Figure 1-1 Cave automatic virtual environment (CAVE)

- Immersive devices, such as HMDs, are the full immersive system that isolates the user from the external world. Head-mounted displays (HMD) are usually in the form of a goggle or helmet that holds the display close to the user's eyes and adjusts the focus through the optical path to project the image and colors to the eyes. The separated left and right images for each eye ensure the stereo vision and the user can watch a three-dimensional scene in three dimensions. Immersive display system provides a complete simulated experience, including a stereoscopic view, which response to the user’s head movement. With a VR headset, users can participate in a fully immersive experience [5].

Figure 1-2 Immersive Display Device (HTC Vive)

During the past two decades, VR has usually been displayed through desktop PCs or semi-immersive display systems, such as CAVE. However, the increasing alongside the popularity and performance of the new generation of HMD devices, especially the affordable mobile VR HMD devices, is boosting the application of using immersive display system in research [6]. Specifically, VR display in this research refers to HMD-based display.

1.2 Overview of VR Tourism

1.2.1 Definition

Travel is the movement between different geographic locations for various purposes. One of the motivations to travel is to build and strengthen relationships between people. The shared experience of travel brings people together. However, people are incapable to travel in real life for various reasons, such as physical disability and time conflicts. Especially, with the recent travel restriction order, traveling in real life is unsuggested during the current pandemic period。

To provide people the experience of traveling without physical travel, Virtual Reality (VR) technology was introduced to immerse them into the simulated tourism world with wearing an HMD. And the experience is called VR tourism, which is the perfect alternative for physical travel in pandemic time. A virtual tour is a simulation of an existing location, usually composed of a sequence of multimedia elements such as videos, images and sound effects [7]. However, in this research, virtual tour was specified to 3D VR tourism, which is the realistic 3D revisualization of virtual environment. VR tourism allows user to explore physical places without physical travel. The typical VR tourism consisting of panoramic video, panoramic images and 3D models of the real world.

1.2.2 Needs of VR tourism

- Before Covid-19

  VR tourism has been applied in several industries, such as real-estate, hotel booking and tourist sightseeing. William and Hobson(1995) stated, “From a marketing perspective, VR has the potential to revolutionize the promotion and selling of tourism ”. The primary potential of VR tourism is its ability to provide extensive sensory information to prospective tourists [8].

And this makes VR an ideal tool for providing rich data to potential tourists who have problem with physical travel. Cheong even discussed VR as a possible threat to the travel industry [9].

VR has always been seen as a promising technology. However, due to the expensive start expense, the users of VR stayed in a limited number for years. Virtual reality is having an increasing impact on the travel and tourism industry, with this increasing alongside the popularity of devices, especially the affordable mobile VR devices.

- Post Covid-19

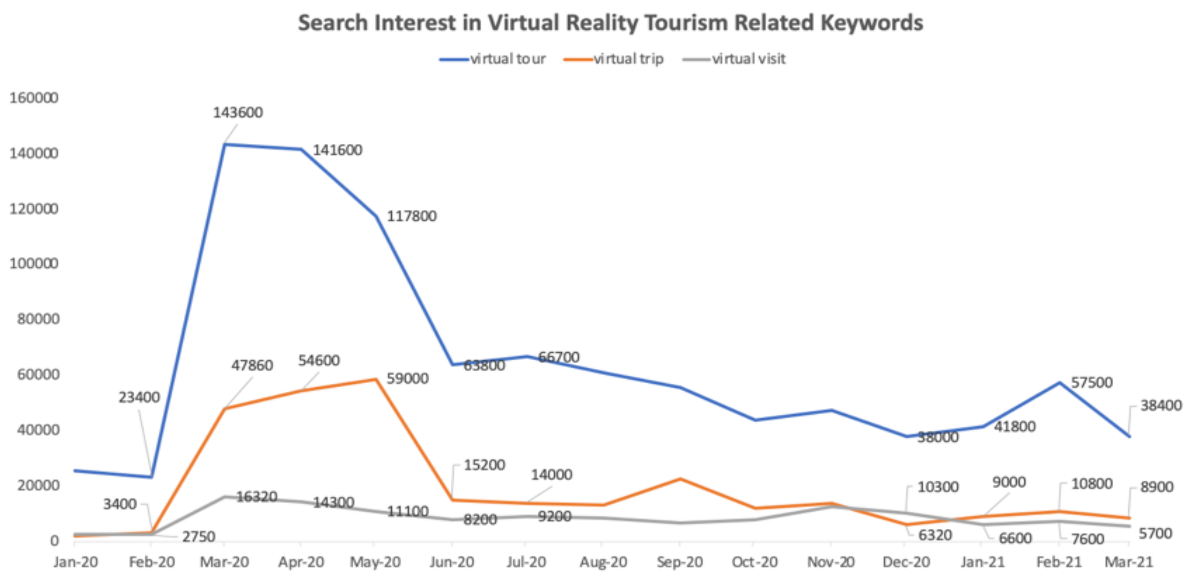


Figure 1-3 Online Search Interest in Virtual Reality Tourism Keywords

Covid-19 made a huge impact to the world in the past year. Social distancing and quarantine life became the new commons during the Covid-19 pandemic. Covid-19 and the rise of the remote economy generated new needs and trends for consumers and businesses, driving funding and M&A deals in areas such as virtual collaboration tools, entertainment at home and AR/VR (remote) tourism [10]. Investors directs their attention to VR tourism industry.

With worldwide lockdowns and domestic restriction bringing travel to a standstill, tourism industry faces the biggest challenge of a generation, this industry is eager to find a substitute for physical travel, even just a temporary solution.

According to the data from online search interests of VR tourism related keywords which was showed in Figure 1-3, the public interest of VR tourism also raised after Covid-19 happened. Specifically, the search of VR tourism related keywords increased rapidly in March 2020, right around the outbreak of Covid-19. Comparing with February 2020, the volume of search of VR tourism related keywords nearly quadrupled that of Feb's count. Additionally, the monthly search of VR tourism remained at least double that of the search volumes in the same months in 2019 even after the peek past.

In the new pandemic period, people would prefer less risky experiences and would see VR tourism as a substitute for the conventional travel. In May of 2020, UNWTO Secretary-General recognized virtual reality among solutions for accelerating the recovery of tourism from Covid-19. "We now have an opportunity to rethink tourism and do things better. Virtual reality, artificial intelligence, and big data will all have a part to play in our joint response to COVID-19, and in building resilience for the future" [11].

1.2.3 Current VR tourism

The exploring of VR tourism has already been adopted by the tourism industry for years. VR provides the ability to visit several destinations in a short time. Additionally, user can have a basic understanding of their destination from the VR. Three examples of current VR tourism system are listed below:

- Google released the Google Earth to allow user to explore the whole earth with their mice. User can check the 3D landscape of their destination easily. With the updated Google earth VR, user can have an immersive tour experience form the first-person perspective, such as fly over a city and walk along the street [12].

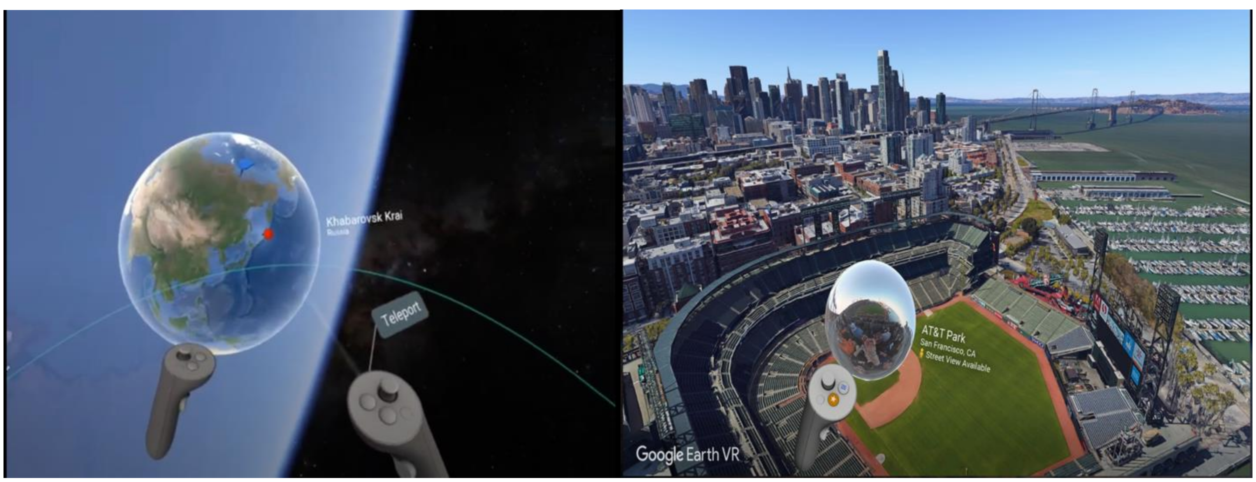


Figure 1-4 Google Earth VR

- VR technology also been applied to museum visiting. With VR, user can walk around and explore those remarkable works in the virtual museum. VR museum also enriched the visiting tour. While most exhibits in reality are prohibited to touch, even be locked in the glass box, VR museum allows user to explore the artwork freely in the way he wants. A typical example of VR museum is the “The VR Museum of Fine Art”, where user can walk around, duck under and even “touch” the virtual artwork with VR device [13]. Additionally, informative holographic plaques are set near the virtual exhibits to help user who wants to know more about this artwork, just like the way in real life.

Figure 1-5 The Virtual Reality Museum of Fine Art

Recently, VTEC Laboratories Inc. released a virtual tourism web service named “Milapse Trek”. This service is based on the latest technology if 360-degree video. With collaboration with travel agency, this service provides a digital experience of walking around the city and some famous sightseeing spots. User can control the speed of video playing, change the direction at the intersection, and walk around the city freely. It is easy to operate the 360-degree videos on the smartphone or PC to explore the view of the city [14].

Figure 1-6 Yokohama Chinatown in Milapse Trek

VR technology in tourism industry is not only used by the tourist sites, but also by hotels. VR technology is utilized to create a digital first-person point of view of the user, which allows user to have a “Try before buy it” experience. User can explore the real hotel room with a set of panoramic images provided by the hotel. Lee introduced many innovation examples of using VR technology in hotel industrial. For example, Marriott hotel's virtual honeymoon from London to Hawaii incorporate VR to enhance customers' immersive experiences [15].

1.3 Motion Tracking Technologies in VR

One of the most unique things about VR is the sense of presence, allowing user to immerse themselves into the virtual world, be able to look around, move and interact with the virtual environment.

Motion tracking is the process of digitizing your movements for user in computer software, and it is a key component of creating a truly immersive virtual reality experience. Current commercial VR systems such as HTC Vice and Oculus use additional sensors in the game setup to capture the user's pose and movement to achieve interaction.

In the following contents, several main motion-tracking technologies used in VR were introduced.

1.3.1 Body tracking technology

The most commonly used motion tracking technology in VR is the body tracking by headset and VR controllers. And it is essential to create an immersive VR experience. There are two broad categories of body tracking technology:

- Optical tracking

Optical tracking refers to tracking body motion by using an imaging device, such as Kinect from Microsoft. In optical tracking system, the user being tracked has optical markers, usually some reflective dots on certain points of their body or the equipment such as the HMD or the VR controller. Once the camera captured these certain dots, motion information will be tracked. The most successful existing commercial body-tracking system, the Kinect, is able to achieve non-mark tracking through software methods combined with special infrared and depth sensing camera [16]. But the accuracy of this system still cannot compare to the mark-based body tracking system yet.

- Non-optical tracking

Non-optical tracking refers to tracking body motion with several equipment, that contains micro-electromechanical sensors, that are often attached to the body of user. Those sensors can measure the lateral, rotational and compass orientations, which means that low-latency, precision motion data can be provided. A typical non-optical tracking method is to use the VR controller and HMD. we default that HMD represents the head, and controller represents the hand. With this basic VR equipment setup, the fundamental body motion of user can be tracked.

With the improvement of sensors, non-optical tracking systems are expected to become more compact and accurate. Also, combining non-optical tracking methods with optical methods (e.g. infrared tracking) can be powerful for motion tracking.

1.3.2 Hand tracking technology

Similar to the body-tracking, hand tracking in VR also can be divided into 2 categories:

- Front-camera-based hand tracking (Optical tracking)

In this method, an optical tracking device (camera) is used to capture the movement. A classic example of camera-based hand tracking is the Lap Motion. It creates high-resolution real-time scans of objects in close proximity to its camera. In VR situation, this device can be attached to the front of HMD and provide a precisely tracked movement of the user’s hands. With this simulation of hand movement, interaction in VR environment can be realized. Besides the Leap Motion, HTC and Oculus also released their own camera-based hand tracking solutions by using the front camera embedded in their HMDs to help developer to track finger movement and gestures with the front camera [17].

- Controller-based hand tracking (Non-optical tracking)

As the name implied, VR controller is representing the hand. The location movement of hand will be synchronized with the hand controller, while the typical hand gestures can be simulated based on the input from controller. Each trigger and button on the hand controller corresponds to a certain movement of hand.

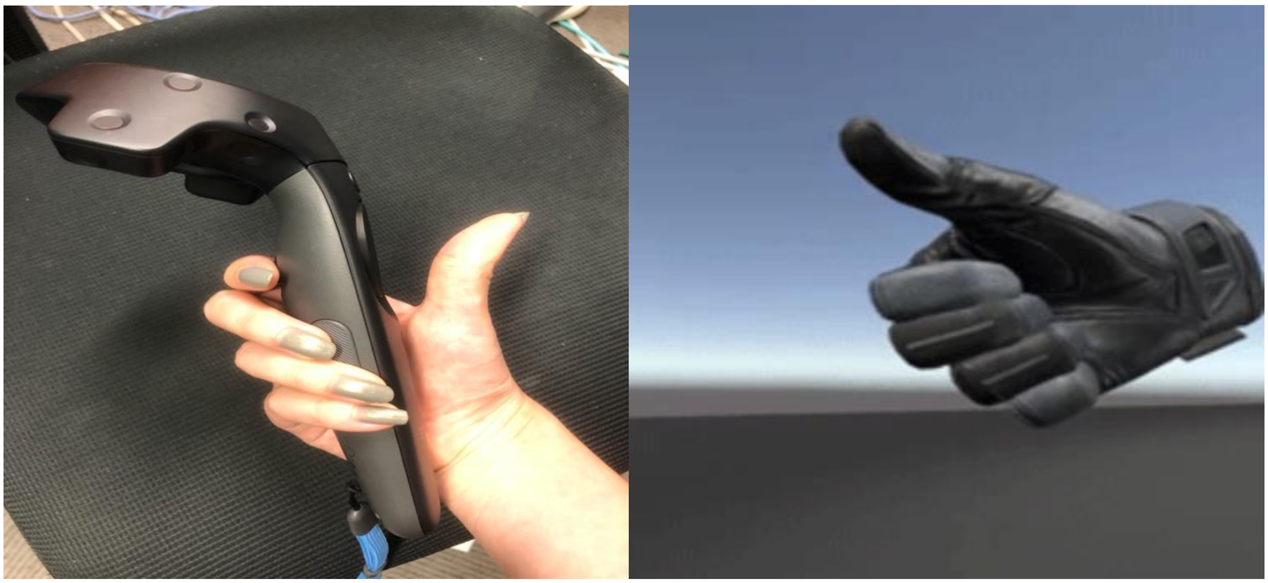


Figure 1-7 Controller-based hand tracking (HTC Vive) [19]

1.3.3 Facial motion tracking and facial occlusion problem in VR

Facial motion tracking is the process of electronically converting the movement of a person's face into a digital database using cameras or laser scanners. Facial motion tracking is related to body motion tracking but is more challenging due to the tracking requirement to detect and track tiny expression of eyes and lips. In recent years, face tracking and facial movement reconstruction based on machine learning has been applied in various fields, access control, attendance tracking, security and surveillance, and others.

Facial expression of emotion is a major channel in daily communications to transmit and enhance information that cannot be provided by speech. They provide a mean for conveying thought and emotion through visual cues that may not be easy to articulate verbally. The facial expression and eye contact is considered as the most natural form of interaction between human [20]. While lots of studies have been conducted on automatically inferring human expressions from images and video sequences most of them focused on non-occluded situation and to our knowledge, few of them focused on user's expression in virtual reality environment. One of the most challenging obstacles for getting accurate facial expression recognition is the partial occlusion in the face. When some parts of face being obstructed by sunglasses, hands and hair, the difficulty of extracting facial features from those occluded area was increased [21,22]. The traditional method of reconstruct facial expression is to use a fixed front-facing camera and track facial movement information based on facial landmarks.

Figure 1-8 Face tracking with 68 landmarks [23]

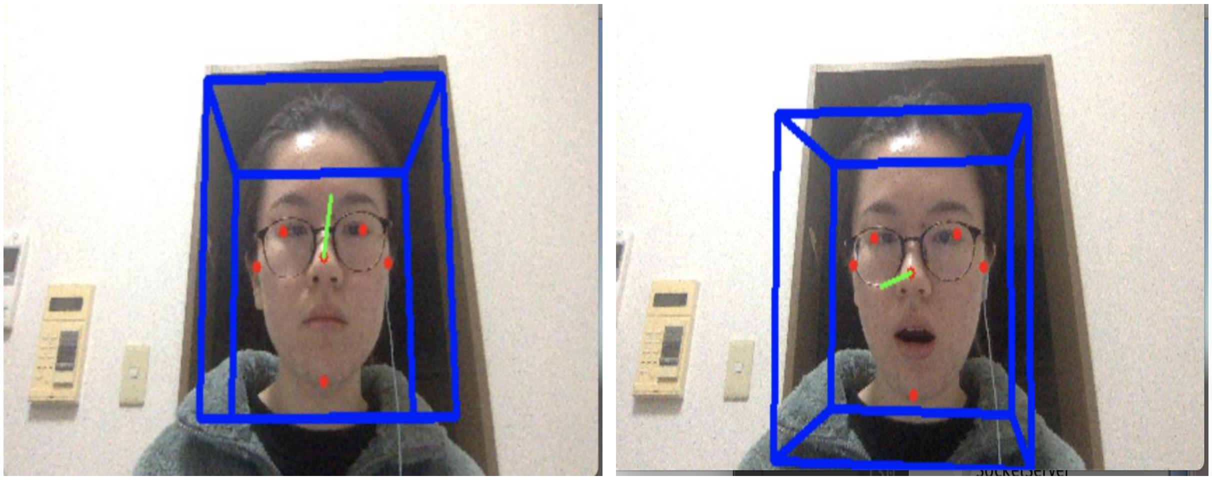


Figure 1-9 The result of face tracking

However, in VR situation, there are 2 major obstacles that restrict getting facial movement information from the user's face.

- Partial face occlusion

While user wearing the head-mounted display, about half of the face was covered. Landmark of eye and eyebrow are lost. This type of occlusion is mostly damaging since it results in the whole features of relevance to judging facial expression being obscured.

In Virtual Reality, a large part of the face is occluded by a HMD which cause the facial expression hard to be tracked, which restricted social interaction in VR [24].

Figure 1-10 Facial occlusion by HMD

- Uncertainty of movement

The biggest difference between VR and traditional display method is that user can move freely in virtual environment. They can rotate their head and turn their body around, which is difficult to track with a fixed camera. In traditional display method, usually user will be in front of the screen and staying in the capturable area of the camera. However, user can move around and rotate their head in VR. Once they turned back, the camera will be lost tracking of the face.

Figure 1-11 Rotating head

With the rapid development of computing technology, virtual reality (VR) is becoming mature. However, facial expression recognition in virtual reality environment is still difficult. Capturing facial expressions and eye gaze while wearing HMDs is one of the urgent problems need to be solved in VR technology.

1.3.4 Eye tracking

Eye tracking is the process of measuring either the point of gaze or the motion of an eye relative to the head. An eye tracker is a device for measuring eye positions and eye movement. Traditional eye tracking technology use external camera to track the eye motion and it is part of the face tracking process. For precise tracking result, professional eye tracker also been used to track gaze direction, pupil position and other accurate eye movement [25].

In order to solve the face occlusion problem in VR, many VR headsets are now beginning to incorporate eye-tracking technology into their HMD. The embedded eye tracker works as a sensor that monitors the user’s visual attention in VR. Take VR eye tracker from Tobii as an example, the infrared cameras are mounted behind the lens and point at eyes and surrounding areas to avoid interference with the VR display. Eye tracking technology enhance VR experience with more natural interactions through gaze, eye tracking contributes to more immersive and user-friendly experiences in VR [26]. The details of Tobii VR eye tracker would be introduced in the latter contents.

Figure 1-12 Embedded eye tracker for HTC VR HMD

1.3.5 Lip tracking

Lip tracking is a component of facial motion tracking and it focus on the capture of mouth area. Common lip tracking method use a fixed camera to capture movement of the mouth and mirror the movement to the target object. Additionally, speech or audio can also be used to track lip movement sometimes.

Traditionally, facial expression tracking, and recognition generally involves tracking 68 well-defined landmarks to locate and reflect facial motion [27]. The whole face is detected, and each part of face are labeled as different numbered facial landmark features. Chen et.al presented an optimized lip contouring algorithm to allow more accurate and detailed contouring output and accelerate the machine learning process by adding an image gradient term to detect image edge at low contrast areas [28]. A commercial application of the traditional optical lip-tracking method is the HTC Vive facial tracker. This additional device consisted of cameras that can capture movements in the lower half of the face and can be attached with the Vive HMD.

To simulate the mouth movement with speech, the speech-driven facial expression technologies can be classified into: (1) face model-based and (2) non-model based. Face model-based approaches parameterize the facial expression into blend shapes and predict these parameters from the audio input [29]. Cudeiro used DeepSpeech voice recognition method to produce an intermediate representation of the audio signal [30]. Then they were regressed into the parameters of the FLAME face model [31]. Non-model-based approaches map the audio signals directly to the 3D vertices of the face mesh or 2D point locations of the mouth [29]. In Karras et al.'s research, an Long short-term memory (LSTM) is used to learn this mapping [32], and the final photorealistic results are generated in Suwajanakorn's study [33].

1.4 Problem Statement

Though the use of VR in tourism is not new and many tourist endeavors use technology to make traveling more convenient and effective, there are many facets of VR use that could be improved. Specifically, there is appeals for far more efforts to be put on improving the tourists' experiences [34]. According to William, the 3 key elements of VR are:

(1) Visualization, where people can see their surroundings inside the virtual world.

(2) Immersion, the objective level of sensory fidelity a VR system provides.

(3) Interactivity, degree of control over the experience [8].

Presence is another term used to characterize VR, which is user's subjective psychological response to a VR system, the subject sense of "being there" [35]. In order to improve the level of presence, the interaction behavior in VR needs to be enhanced.

Moreover, due to human's nature of socialization needs, share VR needs to be introduced to simulate the interactions in reality.

Since most existing VR tourism research are like “Google earth VR” and “Milapse Trek”, only focus on single user’s experience, which reveals the problem of lacking socialization feature among users existing in VR tourism. Moreover, the immersion of social interaction (communication) in current VR tourism is inadequate.

According to the introduction above, the most urgent problem in existing VR tourism is the lack of socialization between users. VR tourism needs socialization feature and more natural interactions to improve the sense of presence in the system.

In summary, the areas covered by this thesis are:

- Multi-user, specifically closely coupled (sharable), VR theory and interaction.
- The use of VR in tourist contexts such as tourist attraction visiting experience and tourism video.
- Practice of social interaction in VR that may provide for more complete reflection of how traveling happens in shared virtual tourism environment with the consideration of how the photorealistic avatar representative and tourist contexts affect VR traveling.

1.5 Research Purpose

Considering the problems of current VR tourism systems above, we are looking for a VR tourism system which can combine adding socialization features and expanding interactions in VR with existing motion tracking technologies. And also, emotional connection enrichment is being expected. In this research, we will try to realize these expectations by proposing a shared VR tourism system with emotional connection.

- Multi-user, specifically closely coupled (sharable), VR theory and interaction.
- The use of VR in tourist contexts such as tourist attraction visiting experience and tourism video.
- The use of eye-tracking and audio-driven lip-syncing technology in shared VR.
- Practice of social interaction in VR that may provide for more complete reflection of how traveling happens in shared virtual tourism environment with the consideration of how the photorealistic avatar representations and tourist contexts affect VR traveling.

2. Related Work

2.1 Shared VR System and Avatar Representation

In order to upgrade the interaction in VR into a more natural level, shared VR was introduced to simulate social interactions in Reality. Shared VR was born on the ground of collaborative virtual environments [36]. Shared VR refers to a 3D virtual spaces where multiple users can interact with others through HMDs [37]. Comparing to other shared digital space (such as internet-based 3D virtual world and games), which mainly depends on mediated communication on computer's screen. Shared VR is capable to provide more natural embodied interaction that is similar to our interpersonal interaction in reality.

One of the most active commercial shared VR systems is VRChat, a massively multiplayer online virtual platform where players can interact with others as customized 3D character [38]. VRChat support both VR headset and web browser. Players from both platforms can socialize in the shared virtual environment, talk to each other through voice chat and body language. For more natural conversation, VRChat also allows user to modify their character individually and add auto-facial expression simulation features. Overall, VR Chat is a virtual social platform where people can have online chat experience with virtual avatar representation.

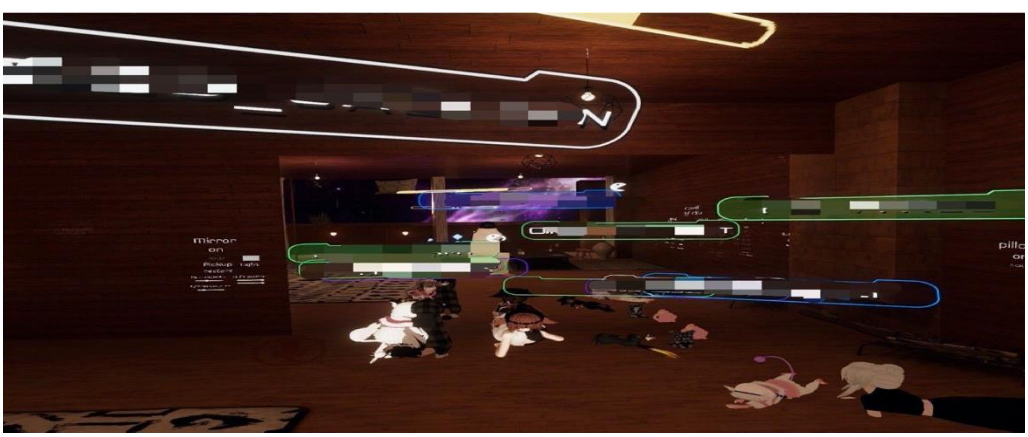


Figure 2-1 VR Chat

One of the most active commercial shared virtual worlds is called Second Life, where people can have an internet-based virtual life includes socialize, network and create their own virtual space [39]. Second life even includes traveling feature in the virtual world. With the expanding of population in Second life's virtual world, many tourist attractions created the replica of their own tourist site, such as Paris' Eiffel Tower. Much like the real life, social component is an important part in Second life. Users can travel to those virtual tourist sites and interact with other players there.

Shared VR is used in diverse areas of teaching and training, including project management, business and healthcare [40]. Shared VR provided a virtual space where industry design teams can meet together and collaborate in various fields [41]. In addition to education, Grasset's research discussed about the possibilities of using AR and VR technology for collaborative task [42]. In Li's study about improving virtual museum visiting experience by allowing user to interact with the virtual artifacts with VR technology. Li also stated that users were interested in additional interaction that requires higher level of collaboration beyond seeing each other's artifacts' moving. They prefer the unintentional collaboration and sharing information in the VE [43].

Due to the technological features of social VR that introduce a strong illusion of virtual body ownership through the avatar representation [44,45]. The typical representations are the simulated audience which generates effective response to user but has limitation of application situation. William addressed that the significant impact of game on players are most likely comes from the interaction between players and avatars [46]. Therefore, the presence and representation of another person can have significant impact on the user no matter the technology.

Current methods of creating avatar representation were classified by the degree of user modeling into three main types:

- Puppeteered

Puppet is often used in online games, as well as multiplayer world like Second life. These avatars are controller by some simple operation, such as pressing a key to trigger an animation.

- Reconstructed

Real-time reconstructed avatar promises a faithful replication of the user's appearance in reality. It was applied in various scenarios, such as online meeting [47].

- Tracked

Real-time tracked avatar requires elaborated sensor technology like full-body motion tracking technology, face tracking technology and an authentic body model to matching with the input from sensors. The accuracy and realism of tracked avatar depends on the extent of sensory coverage, which depends on the degree of fitting the avatar to the controlling user [48,49].

Marc's experiment reveals the positive effect of co-located social companions on the quality of experience of virtual worlds, likely to improve co-presence and possibility of interaction [37]. His research also proved the positive effect of using photorealistic avatar on improving presence.

2.2 Non-verbal Communication in VR

Comparing to other shared digital space (such as internet-based 3D virtual world and games), which mainly depends on mediated communication on computer's screen. Shared VR is capable to provide more natural embodied interaction that is similar to our conversation in reality. The characteristic of VR that differs from other shared digital space is that VR support verbal and non-verbal communicatiom, which provides an immersive experience similar to face-to-face interaction in reality [50,51].

Based on that, Maloney specifically indicated the importance of using non-verbal communication on developing more natural embodied interaction in VR system [52]. Non-verbal interaction includes hand gestures, gaze and facial expression. In addition, non-verbal interaction is also sited to be an effective method to build and maintain emotional connection. Further, a feature comparison of several social VR applications reveals that despite the potential to induce authentic social experiences, there are still many open challenges and opportunities for designing and utilizing expressive non-verbal communication features [53, 54, 55, 56].

More specifically, the feature of VR that verbal and non-verbal communication were supported creates an experience that is similar to face-to-face interaction [57].

Figure 2-2 Example of non-verbal communication in VR (body language)

2.3 Facial Expression and Emotion Recognition in VR

Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. As mentioned above, the facial occlusion problem is always the main obstacle for research of facial expression in VR. Fortunately, with the development of facial tracking technology and sensor, the number of research on facial expression with partial facial occlusion was increased. Scheirer et al. used face mounted piezoelectric sensor to classify confused and interested face by sensing specific facial movement [58]. Suzuki et al. proposed a facial expression mapping technique that mirroring user's facial expression into avatar by using an embedded optical sensor and machine learning [59]. Hickson developed a method to classify facial expressions by using the eye tracking data. And they discussed about the relationship between eye movement and emotions [60].

Emotions are a mental state that are experienced by a human and is associated with the feelings and a degree of pleasure or displeasure [61]. Also, they are states of feeling that result in psychological changes that affect human action or behavior [62]. And emotions are a very important trend topic in fields like tourism and marketing recently. With the "emotional tourism" increasing, using emotions analysis becomes the new trend in tourism marketing [63]. The effect of emotion on sense of presence have so far been investigated but with unclear result. Diemer believed that the unclarity was due in part to the excessive virtual environments and subsets of emotions considered in the study. Another possible reason is the ignorance of the 2 common dimensions of emotion, valence and intensity. Hence, they studied the effects of emotional valence and intensity on presence in VR within a single experimental design, using controlled VEs. According to the result, they confirmed that there is indeed a relation between emotion and presence in VR, specifically that emotional intensity is the major factor in the formation of presence [64].

There are many ways that can be used to simulate emotions, such as through watching a movie, looking at a still image or have a virtual tour. To recognize those emotions, many technologies were utilized. For example, eye tracking technology was used in the research of Soleymani et al. The author recorded the EEG and gaze data (pupillary response and gaze distance) to get the response of user while watching video clips. Based on the result, the three classes of participants'' response were defined on both arousal and valence aspect [65]. Another example is to use heartbeats for emotion recognition. Choi's research examined the validity of Heart rate variability as a tool to evaluate emotions using the International Affective Picture System (IAPS). For experimentation, five photos of each of the categories of "happy", "unhappy" and "neutral" from the IAPS were presented to the participants and the subjects were required to fill out the Self-assessment Manikin after each picture [66].

3. The Development of Shared VR Tourism System with Emotional Connection

3.1 System Concept

The system provides a shared virtual environment where multiple users can meet. Then when users are enjoying a common virtual tourism experience, the system allows users to interact with each other through voice chat and non-verbal communication (body language, facial expression etc.). By all these social interactions, the emotion of user can be exchanged and emotional connections between users can be enhanced. For example, when a user finds an interesting object in the virtual scene, he/she can share this finding with the other user right away and they can have a mutual talk about this object.

The function flow is shown in figure 3-1. Both User A and User B will enter this VR tourism system with an HMD. They can explore the VR tour environment together, communicate through body language, facial expression and voice chat. And all those information are transferred through the Internet.

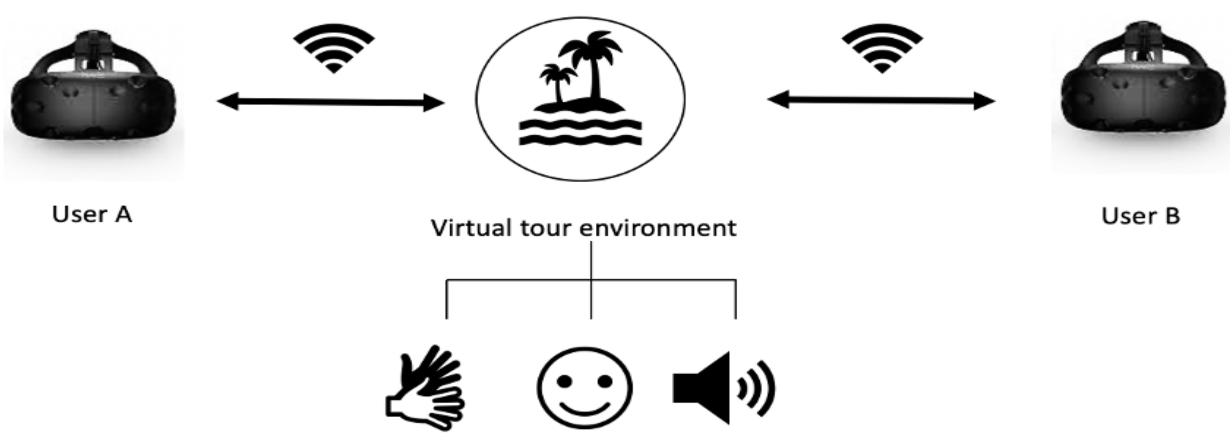


Figure 3-1 Function Flow of the System

The framework of the shared VR tourism system is shown in figure 3-2. The system consists of 2 major parts: VR tourism system and the shared VR system Networking. VR tourism system is responsible to the build of virtual tour environment, handle local interaction logic and handle data input form the local VR devices. And received data from the other user will be used to simulate an avatar that represents the remote user B. Then the shared VR system networking refers to the networking framework that transfer and synchronize data through Internet.

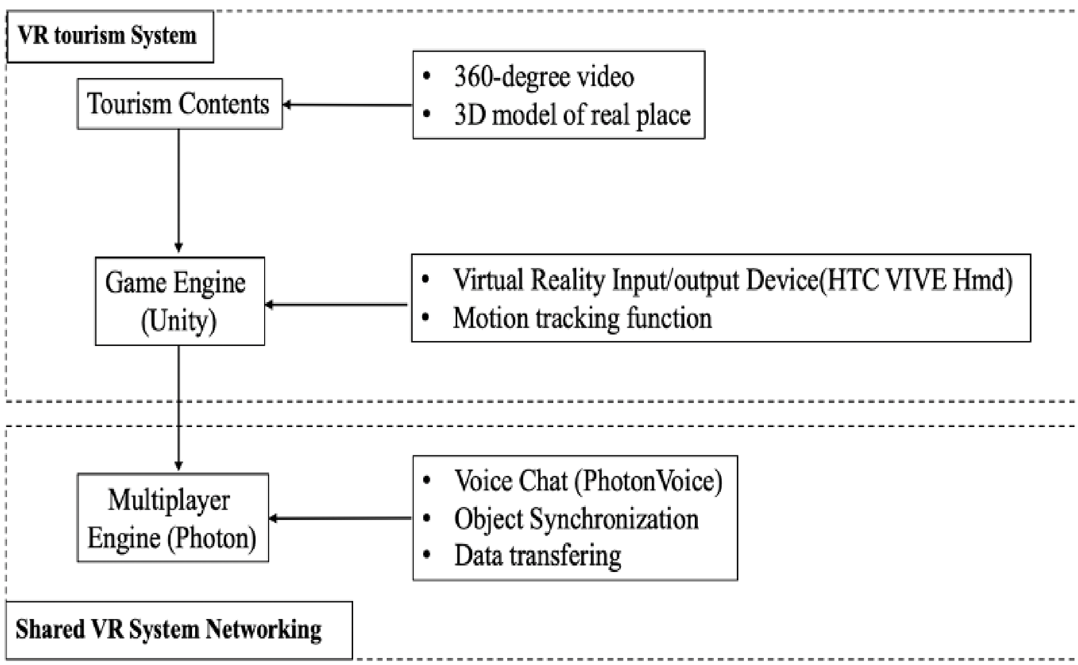


Figure 3-2 The Framework of Shared VR Tourism System with Emotional Connection

3.2 Development Environment Setup

3.2.1 Hardware setup

HTC Vive Pro Eye set

- Hand controllers
- Embedded Eye tracker
- Embedded Microphone

Figure 3-3 HTC Vive Pro Eye

3.2.2 Software setup

- Character creator 3
- Blender
- Autodesk Maya
- Oculus Lipsync
- Unity 2019.04.28
- Visual Studio 2017

- Photon Unity Networking (PUN)
- Photon Voice
- OpenVR Unity XR
- Tobii XR SDK
- HTC Vive SRanipal SDK

Unity is a cross-platform game engine to create and operate interactive 3D and 2D content, as well as virtual reality experience. OpenVR Unity XR plugin and Photon Unity networking plugin was integrated in unity to create the social VR available development environment [67].

Before building the shared VR system, we need to prepare 3D face model of each user in advance. Character creator 3 is a software that can generate a realistic 3D avatar from a single picture. With this software, a 3D human model in fbx format and photorealistic textures can be exported.

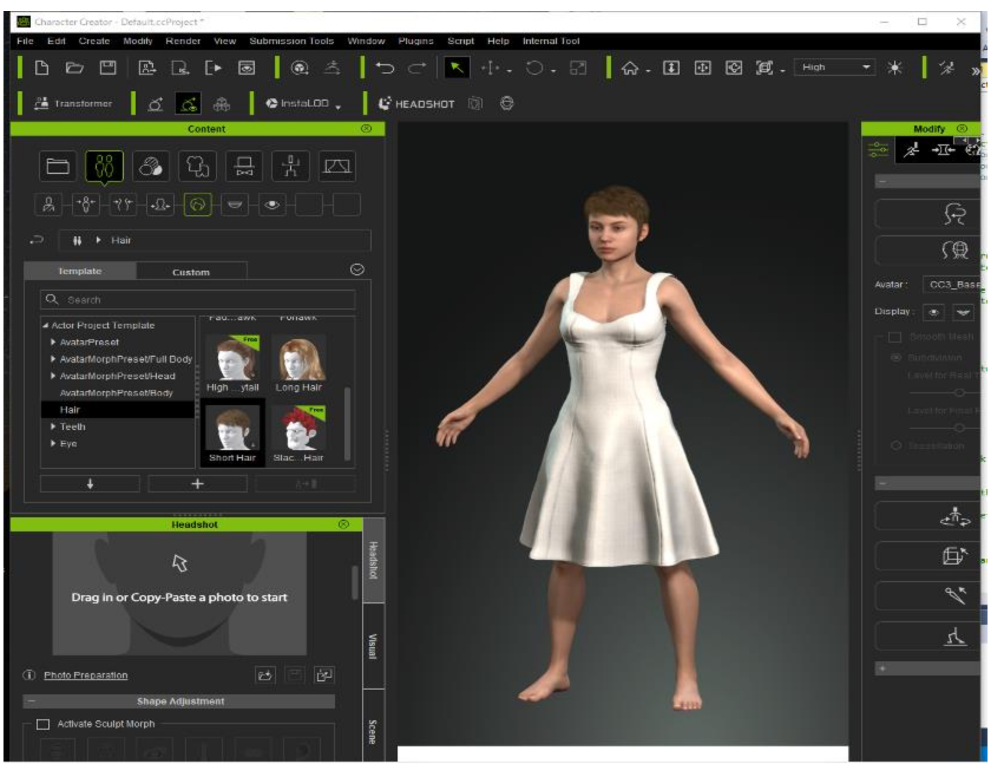


Figure 3-4 The user interface of Character Creator 3

In order to simulate facial expressions of user in our system, we need to set up a software environment which allows motion information data from the motion tracking system being fully utilized. Hence, eye tracking SDK(Tobii XR SDK and HTC Vive SRanipal SDK) and lip syncing plug-in (Oculus Lipsync) were used in the system.

3.3 VR Tourism Environment

3.2.1 VR tourism content

The background environments are critical to create an immersive virtual experience. In this VR tourism system, user can experience 2 types of VR tourism content.

- 360° panoramic video

  The 360-degree video is shot with an omnidirectional camera. It provides a spherical video with vivid sound effect. The view of every direction that was recorded is displayed to the user. User can change the perspective by rotating their head, but the position of user is fixed in this scene.

  In this system, several panoramic videos were utilized. Self-made 360-degree videos were captured by RICOH THETA V, a panoramic camera that is capable of 360-degree video and photos. Also, 360-degree video of oversea tourist attractions was cited from Internet [68].

Figure 3-5 Examples of 360-degree video

(A -Yamashita Park; B - Raichi farm; C – Yagami riverside; D – Crystal shower fall)

- 3D environment modeling

3D environment modeling is the generation of realistic environment based on a real-world observation. By combining certain sound effect, user can immerse themselves into the 3D environment modeling. In the 3D environment, user is free to move around, look close to objects and interact with the environment. In this system, the used 3D environment model is made with Autodesk Maya and Blender.

Maya is a 3D computer graphics application that is used to create assets for interactive 3D applications. Maya is used on film, television and architecture. Maya is a great choice to build the 3D virtual environment in this study.

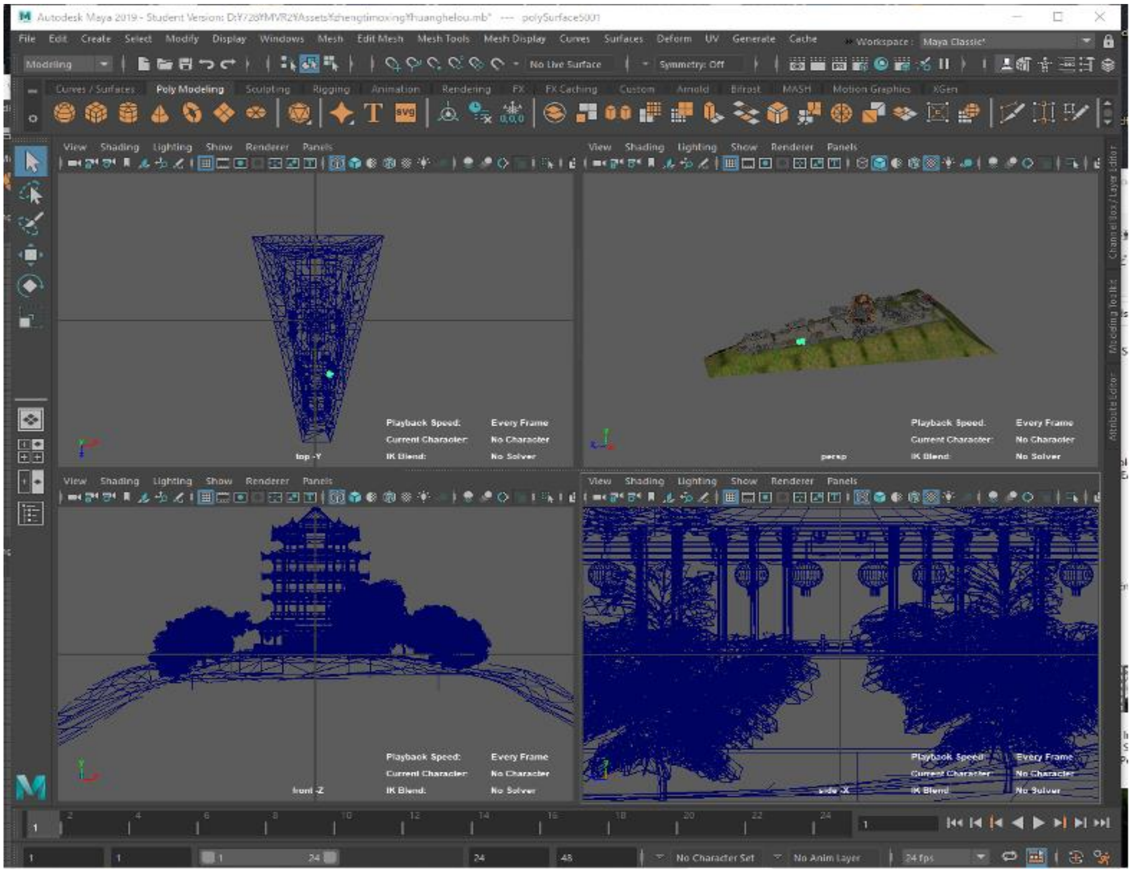


Figure 3-6 The user interface of Autodesk Maya

Blender is another digital design software to create 3D models, motion graphics and virtual reality components. Its features include 3D modelling, texturing, rigging and skinning. Blender has a strong rendering system which make it a grate choice to create scene assets.

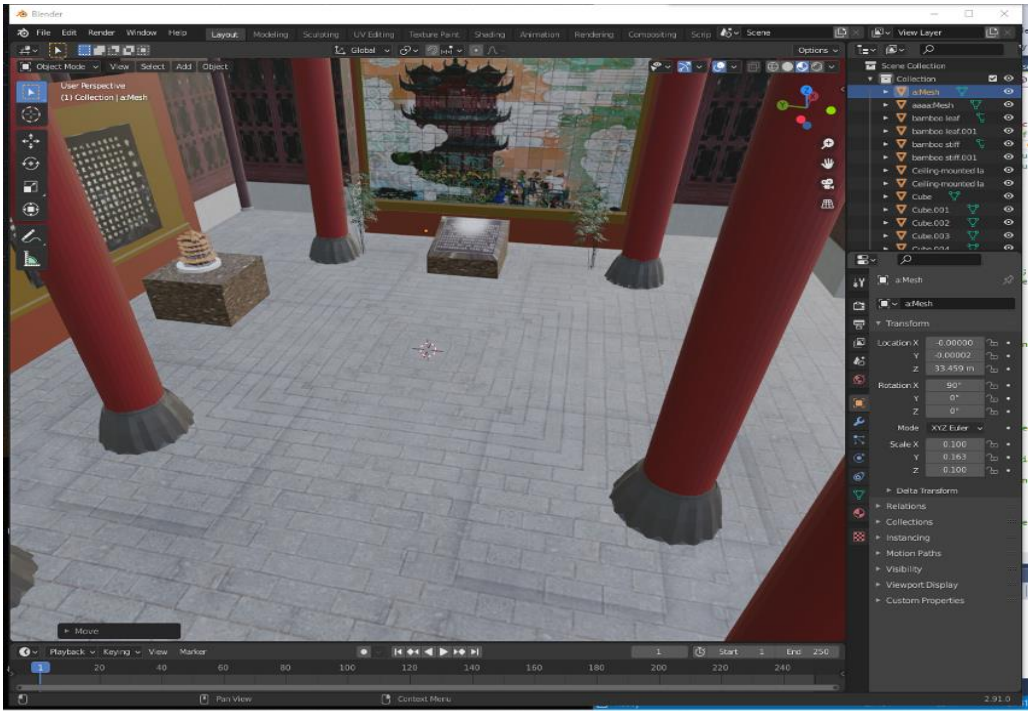


Figure 3-7 The user interface of Blender

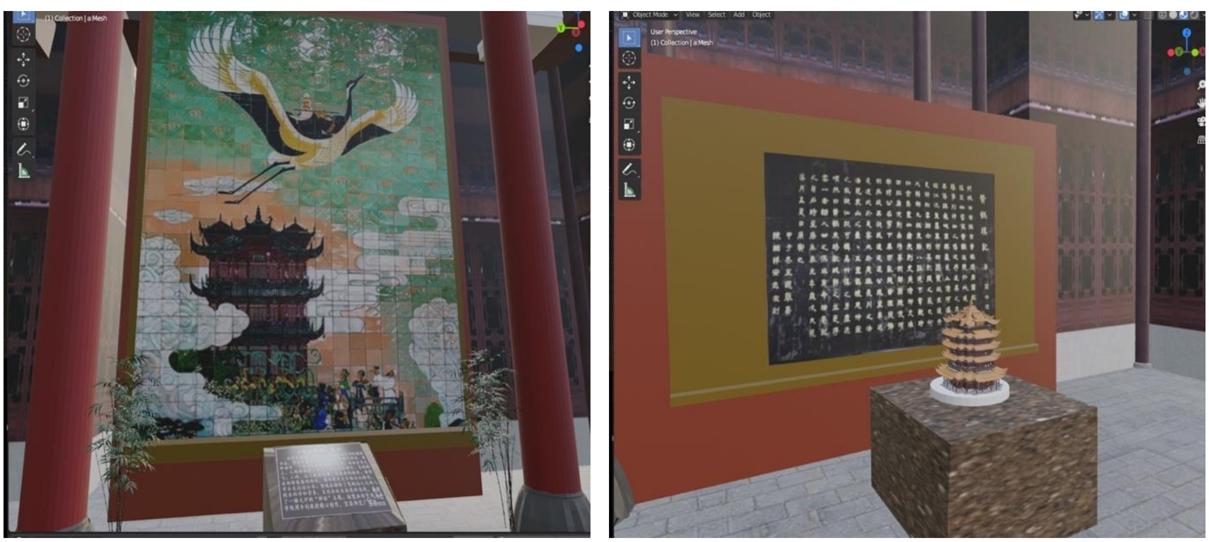
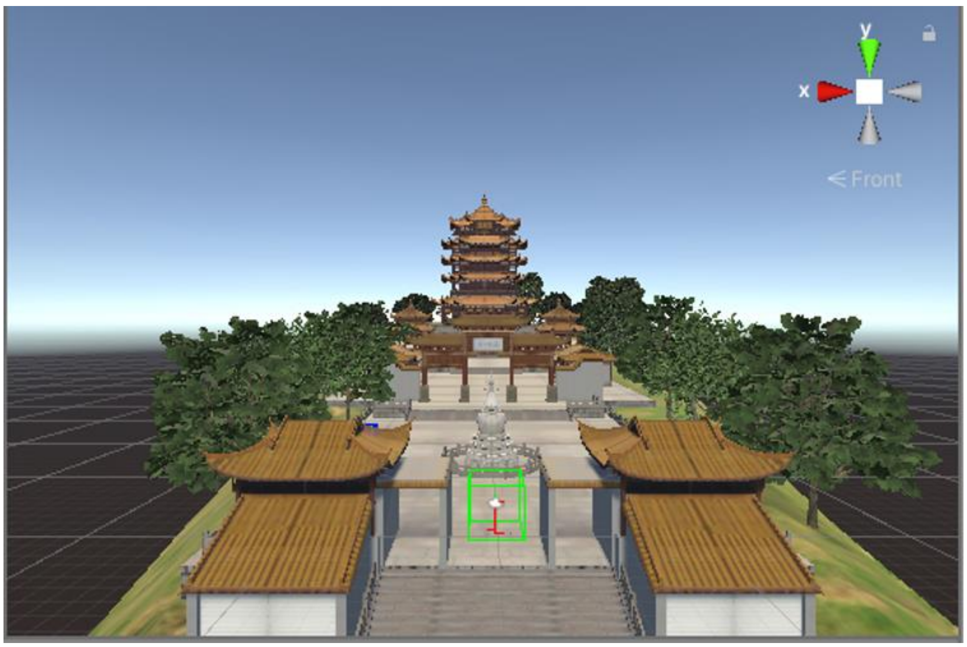


Figure 3-8 3D environmental model

In this study, the 3D environmental model is created based on a famous historical tourist attraction “Yellow Crane Tower” in Wuhan, China. It is considered one of the Four Great Towers of China. Inside the scenic area, there are the tower itself, the tower’s bell, and several statues. We have tried our best to reconstruct this tourism spot in the virtual reality environment.

In the virtual Yellow Crane Tower, the main architectures and environment were modeled. And the first floor of the tower structure was recreated, the indoor view of the tower, includes some murals and paintings, can be seen in the virtual space.

To improve the immersion of the 3D model-based environment, a corresponding background music was added into the whole experience as well. Since the model is based on the Yellow Crane Tower in Wuhan, China. We used the music of an ancient Chinese Chime Bell (Bianzhong) playing performance, which was the music that can be heard in Yellow Crane Tower [69].



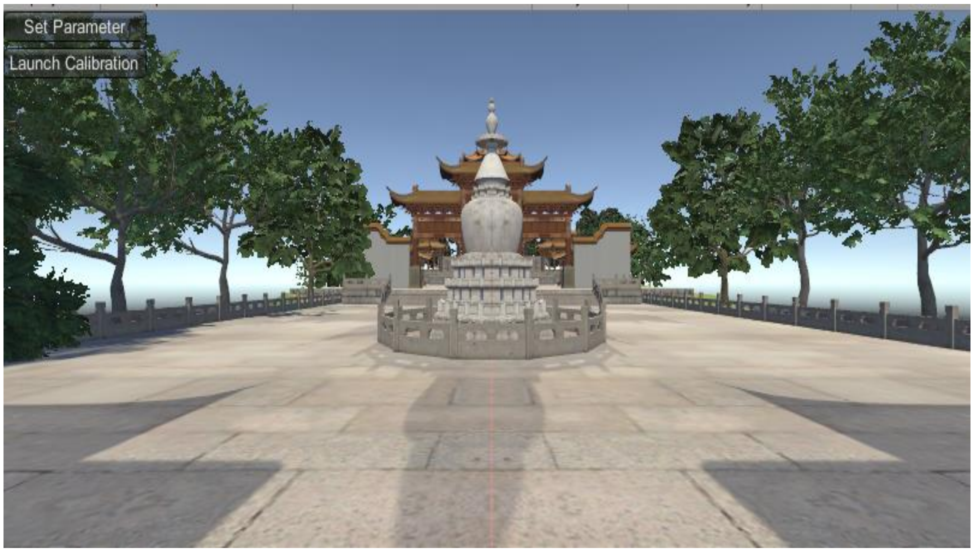


Figure 3-9 The view of Yellow Crane Tower in Unity editor

3.4 Expansion of Interaction with Emotional Connection

3.4.1 The use of Photorealistic avatar

(1) Character Creator 3

To create the customized 3D avatar for each user, Character creator 3 provides an efficient solution to create 3D face model from a single picture of the user. Since adequate partial embodiment increased co-presence [47], a 3D avatar consists of photorealistic head and 2 hands are used to represents the user in this research.

(2) Blender

Then the generated 3D face model was imported into Blender where details of the model is adjusted, and blend shapes are created. The 3D model is called meshes in Unity and meshes contain vertices and multiple triangle arrays. By modifying those vertices, the shape of meshes will change. Thereby, the shape of 3D model changes as well. Blend shape (shape key) is a morph target animation that realized by changes of mesh that is stored as a series of vertices. In each key frame of this blend shape, the vertices are updated with the current value. The blend shape is applied to create facial expression. For example, if the face is first modelled with a neutral expression (base blend) and a “target deformation” of open mouth (shown in Figure 3-11). Once the blend shape is triggered, the face will morph between the base shape and the morph target.

Figure 3-10 The process of creating a photorealistic avatar

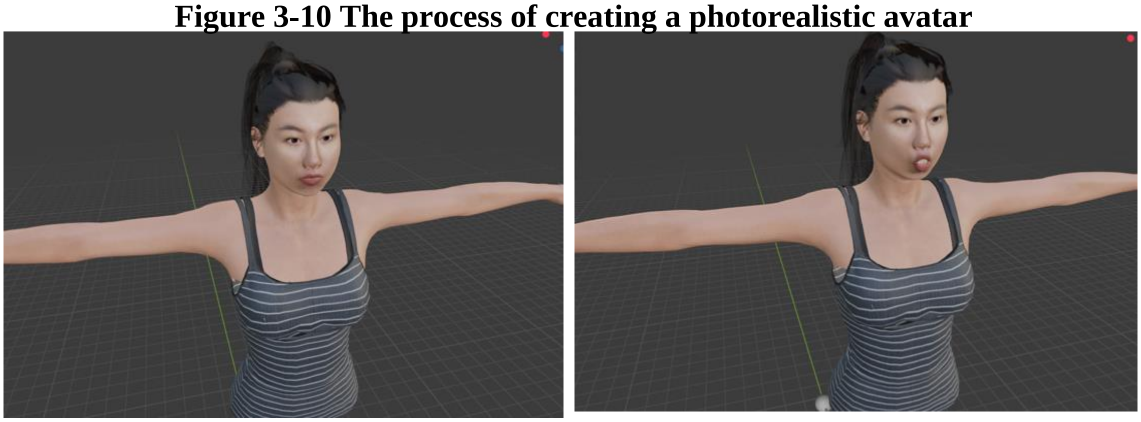


Figure 3-11 Examples of facial motion (blend shape)

(Left image – neutral; Right image – open mouth)

(3) Unity

After the 3D avatar was imported into Unity editor, several animation clips were created to mimic the movement of user, such as Left eye blink, laugh and other characteristic movement.

Unity's Animation features includes animation clips, blend shape system of facial animations, full control of animation weights through script at runtime and much more. In order to create facial animation, several animation clips and animation controllers are needed.

Animation clip is the animation data that can be used to animate game object or simple movement of character. Animation clips can be imported from external sources, such as asset store, Maya. Also, Unity editor allow user to create and edit animation clips internally. In this research, combination of external animations and internal animation clips were used. Since several blend shapes of facial expression has been created in Blender, they were imported and used in Animation Window to create new animation clips. As result, each face model has 23 animation clips for facial expression reconstruction purpose (see the list of animation clips in Figure 3-12), each of the animation clip is corresponding to a particular facial motion.

Figure 3-12 List of animation clips

Animation controller allows user to arrange and control a set of animations. Animation controller manages various animation states and the transition between animation states in State Machine, which is similar to a workflow control chart. The change of animation states is controlled with parameters, that can be accessed and assigned values from the scripts.

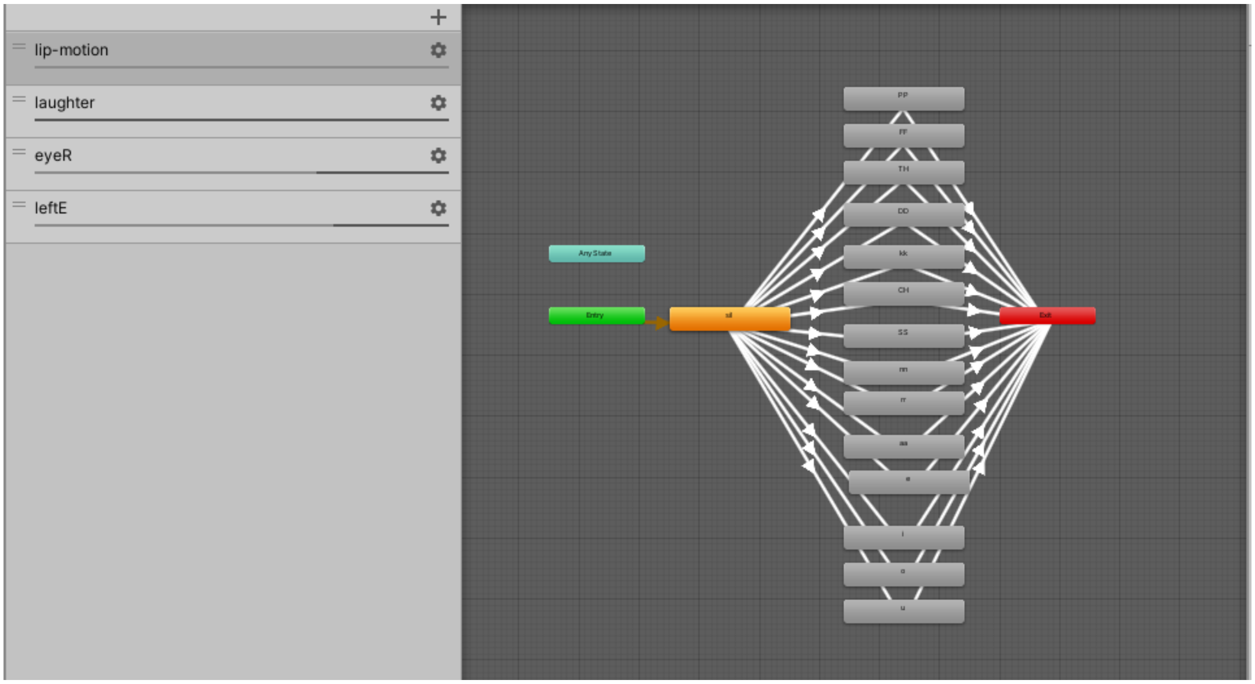


Figure 3-13 Example of State Machine (Lip-motion Layer)

3.4.2 Facial expression reconstruction

(1) Eye-tracking technology

To track eye movement, there are several features involved. And the main eye features that is relevant with emotion includes pupil diameter, pupil position, fixation duration, the distance between sclera and iris, motion speed of the eye and pupillary response [70].

Tobii, the eye tracker manufacturer, provides high precision eye tracking and more detailed data analysis as well as heat map and eye tracking for HTC VIVE Pro Eye. Eye trackers from Tobii Pro use infrared illuminators to generate reflection patterns on the corneas of the subject's eyes.

These reflection patterns, together with other visual data about the subject, are collected by image sensors. Sophisticated image processing algorithms identify relevant features, including the eyes and the corneal reflection patterns. Complex mathematics calculate the 3–D position of each eyeball and the gaze direction from the eye [71].

With the embedded eye tracker inside the HTC Vive Pro Eye, serval eye data can be collected. Tobii XR SDK converts the eye data into readable signals in Unity. Then those signals drive animations of the user's avatar to simulate the facial expression around eye.

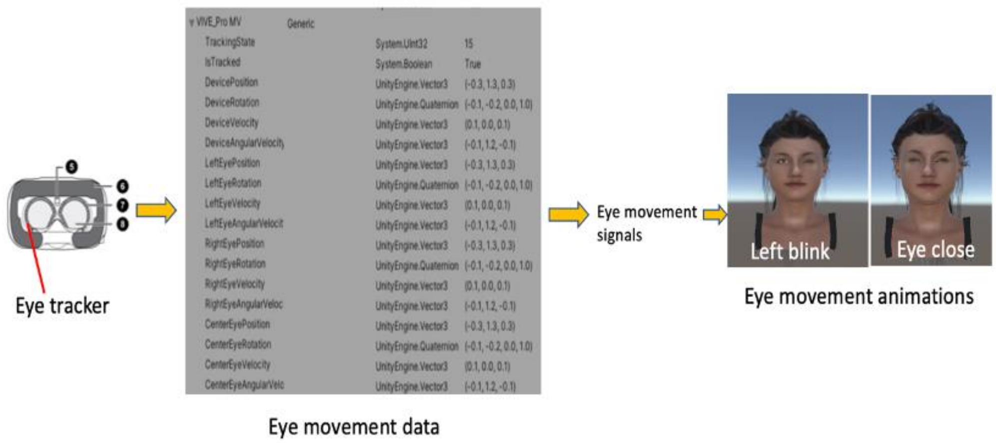


Figure 3-14 The process of simulating eye movement

(2) Audio-driven lip-syncing technology

Oculus Lipsync is a simple solution to sync mouth movements of the user to the speech sound from the real-time audio input. It syncs user's speech movement of the live microphone input in real-time by mapping human speech to a set of visemes, which are the visual analog of phonemes [72].

Oculus Lipsync is powered by deep learning and it allows facial expression of a virtual avatar to be driven by speech and enrich the social interactions. Real-time audio-driven laughter detection bring the project a step closer towards enabling expressive facial animation, rich presence and non-verbal communication in virtual reality. Both the viseme prediction and laughter detection models rely on a Temporal Convolution Network (TCN) applied to the input stream of 32-dimentional LogMel feature vectors extracted from the raw audio signal. The viseme prediction network first predict 44 phonemes that are mapped to the 15 viseme mouth shape outputs. On the other hand, the laughter detection network directly predicts the probability of laughter occurring as a floating-point value in range 0.0 to 1.0 and mapping it with the laughing viseme mouth shape [73]. The reference of viseme mouth shape is shown in Figure 3-15.

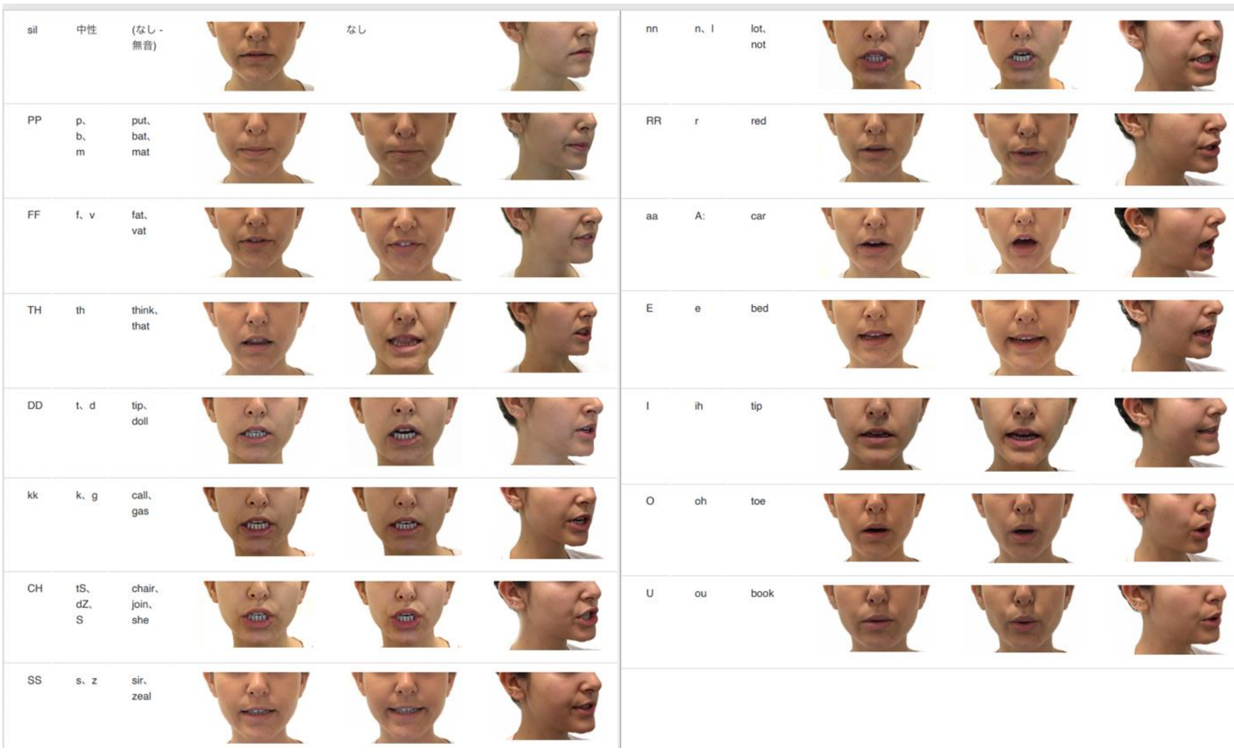


Figure 3-15 The reference of visemes

In this system, real-time audio input from the microphone is analyzed and categorized into a set of viseme signals. We use these signals to control the animations of visemes we created before for a vivid speaking motion simulation.

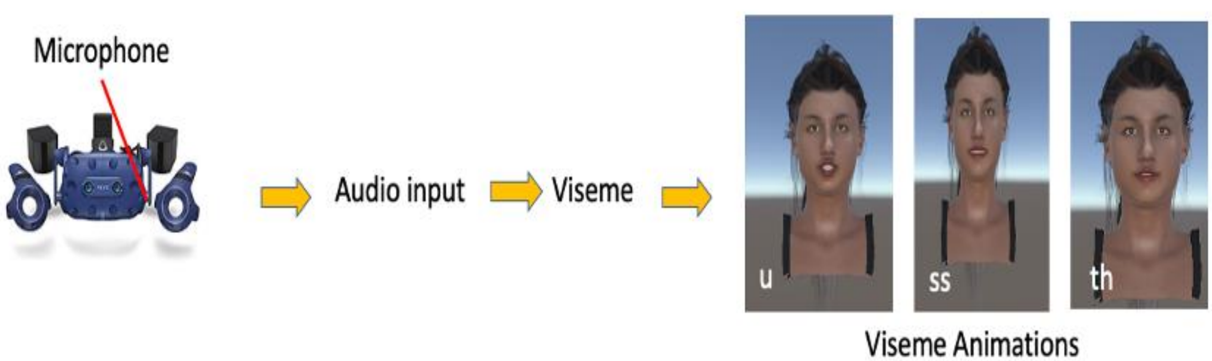


Figure 3-16 The process of simulating lip movemen

3.5 Network Setup

Photon Unity Networking (PUN) is a Unity package for multiplayer games. It contains various features to realize communication over the network through Photon server. Photon Voice is a Unity package that allows client state of audio can be automatically synchronized with another PUN's client state. PUN and Photon Voice have been used for the implementation of the multiplayer features and the voice chat, respectively. Both of the two Unity packages are based on Photon Engine [74].

In order to synchronize user's movement through Internet, the components attached to each user's avatar are shown in the Figure 3-17. Photon also provides the implementation of voice chat. Since verbal communication is the necessary component for social VR, Photon voice provided an easy and reliable approach to communicate users. To realize the synchronization, each player has network-instantiated PUN prefab (the avatar representation in this system) where location and audio is 3D positioned. The location information is recorded with the components of PhotonView, PhotonTransformView and others. A single audio stream has a pair of components on each side of sender and receiver: the Recorder and the Speaker. When using PUN, all components and state is linked via Photon View ID.

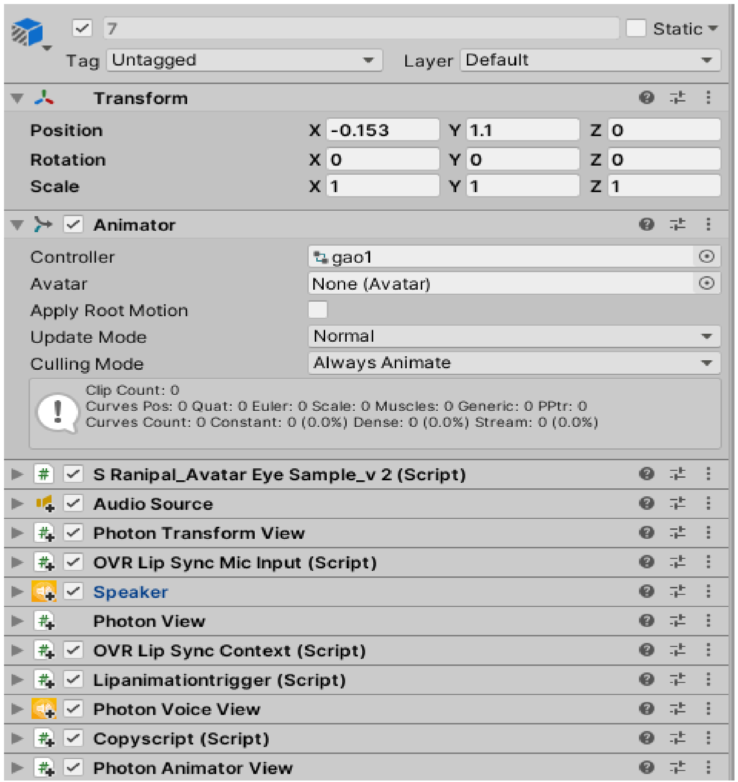


Figure 3-17 Inspector window view of user's avatar

Components "Photon View", "Photon Animator View" and "Photon Transform View" were used to synchronize the location and facial animations of the avatar. And "Photon Voice View" and "Speaker" component were attached for voice chat function. Besides components related to Photon, "OVR Lip Sync Mic Input", "Lipanimationtrigger" and some other components were used to manage device input and control animations, and they were the fundamental of realizing animation synchronization.

4. Evaluation of System

4.1 Experiment 1: The Verification of Facial Expression in Shared VR Tourism System

4.1.1 Objective

This experiment is designed to verify the feasibility of creating recognizable facial expression of emotions with the eye-tracking technology and audio-driven lip-syncing technology, and to verify if user can recognize these avatar's facial expressions.

4.1.2 Participant

The experiment involved 3 participants (all females) with the age range of 23-27 years old. All subjects had previous experience of using VR device but had little to no experience with social VR. The final sample is 3 participants.

4.1.3 Apparatus

This experiment used a set of HTC Vive Pro Eye connected to a desktop PC to record the facial expression. The facial expressions example that the participants watched was displayed on the screen of VR HMD.

The VR simulation was run via Unity. A basic virtual space was created that all VR contents displayed in.

4.1.4 Procedure

Participants first listen to the general instruction about the device and techniques used in the experiment. The experiment requires participants to observe and classify 7 facial expression on avatar which was generated with the eye-tracking and lip-syncing technology in advance. According to Ekman's theory of 7 universal facial expressions, the virtual facial expression consisted 7 emotions of joy, fear, sad, disgust, surprise, angry and contempt. These 7 facial expressions can be recognized, transcending language, regional, cultural and ethnic differences [75].

Figure 4-1 Example of universal facial expression of emotion

The facial expressions were displayed on the screen of VR HMD, and it were played in random order. While participants observing the facial animations, a dashboard of selections was displayed behind the avatar, see Figure 4-3. The options include angry, surprise, disgust, sad, contempt, fear, joy and not sure. After each facial animation being presented, participants were asked to choose an option from the dashboard to classify the facial expression they just observed. The facial expression will change every time the participant made choice. After all facial expressions were seen, the correct ratio was calculated.

Figure 4-2 Experimental scene in experiment 1

Figure 4-3 The operation interface in experiment 1

4.1.5 Result

Table 1 Result of different participant

	Joy	Fear	Sad	Surpris e	Disgust	Angry	Contem ption
Participant 1	○	○	○	○	○	○	×
Participant 2	○	×	○	×	○	○	×
Participant 3	○	○	○	○	○	○	○

The result of answer is showed in Table 1. Both participant 1 and 2 failed on recognizing the facial expression of contempt. Participant 2 also mistook the facial expression of surprise and fear. In total, the correction ratio of recognizing the facial expression is 0.875. According to the result, the function of generating facial expression with the motion-tracking technologies was verified. Participants can recognize the facial expression and the emotion they represented.

4.1.6 Discussion

Although the function of generating facial expression was verified, there are some limitations shown in the experiment. In the case of contempt, participants had the problem to distinguish it from the other emotional facial expressions. The possible reason is the fundamental limitation of audio-driven lip-syncing technology. For mouth movement with no voice signal, it is difficult to simulate the motion with limited resource. The accuracy of non-voice facial expression needs to be improved.

4.2 Experiment 2: The Validation of Shared VR Tourism System with Emotional Connections

4.2.1 Objective

This experiment is designed to compare the VR tourism system with multiplayer feature or not and validate the performance of the system from users' perspective.

4.2.2 Participant

Ten university students (two males and eight females) with an age range of 23 -27 years took part in the experiment. All participants were paired in groups of two (5 groups in total). 3 groups consisted of participants who were familiar with each other. Most participants had previous experience with VR but had little to no previous experience with social VR. The final sample is 10 participants (5 groups).

4.2.3 Apparatus

This experiment used 1 set of HTC Vive pro eye and a set of HTC Vive, each set consisting of 2 base stations, an HMD, two hand controllers, and connected to a desktop PC, seeing in Figure 4-4. To connect the Photon's server, the PCs were setup on Internet connected. The VR tourism content is the 3D modeling environment of Yellow Crane Tower.

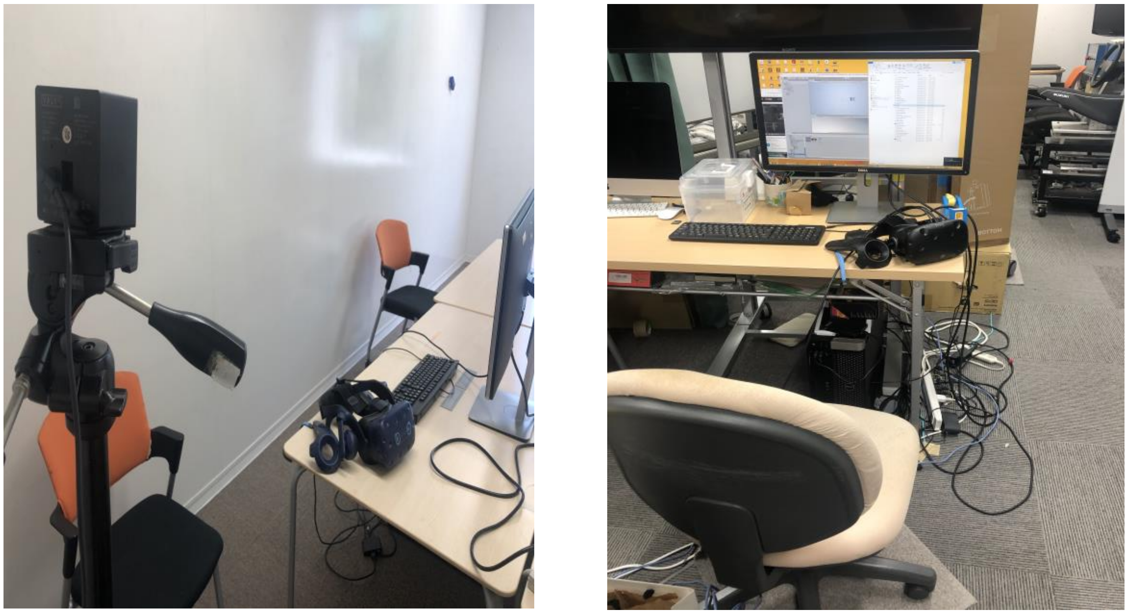


Figure 4-4 Experiment environment setup

(Right side is the HTC Vive set; Left image is the HTC Vive Pro Eye set)

4.1.4 Procedure

The experiment is aimed to compare the traditional single user VR tourism with the modified shared VR tourism system. The experiment procedure contained two trials, solo scenario and multiplayer scenario. These are described separately below. Prior to having the experiment, participants took some pictures for making the photorealistic avatar of themselves. After welcoming participants and explaining the study, participants were paired with a partner. The researcher confirmed that both participants were willing to share VR tour experience with each other.

Then participant was instructed to put on the HTC Vive Pro Eye and VR controllers for the first trial. The first trial used a solo scenario. Participant entered the VR tourism environment alone and explored the virtual environment by themselves, walk around or interact with the environmental object. Participants had up to 5 minutes to explore the tourism content. Then they can press the menu button to exit from the simulation. The scene of trial 1 is shown in Figure 4-5, 4-6 and 4-7

Figure 4-5 Experimental scene of solo scene

Figure 4-6 The view from participant in the solo scene

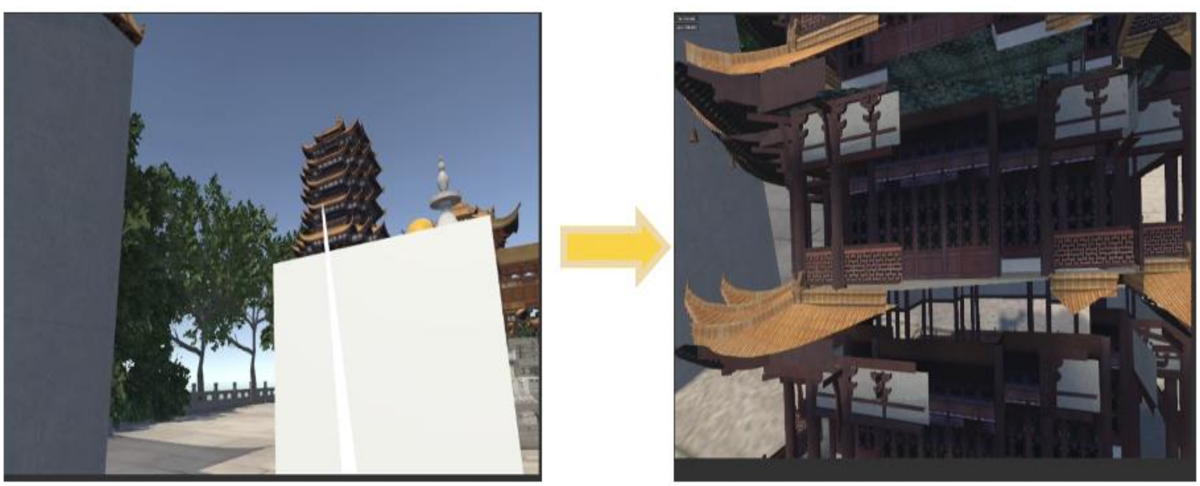


Figure 4-7 Interaction between participant and VR environment

The second trial is a multiplayer scenario. One participant wore the HTC Vive Pro Eye set, while the other one wore the HTC Vive set. Participants started off the trial by having a short function confirmation where 2 participants check their allocations and the connection status of each other. After the check was done, participants started the shared experience. Similar to the solo scenario, participant can explore the virtual environment, but they have a companion this time, they can interact with each other. They can see each other, talk to each other through voice chat, also see the movement respectively. The second scenario took about 10-20 minutes.

Figure 4-8 Experimental scene of multiplayer scene

Figure 4-9 The view from participants in multiplayer scene

After finishing each trial, participants were asked to fill out a questionnaire (see Appendix 1). The questionnaire started with 2 true or false questions about their past experiences of using VR and followed by 15 specific question related to the evaluation. In addition to discuss the overall experience of the experiment, short answer of feedbacks was collected with each participant separately. The total duration of the experiment was about 40 minutes per group, with two scenarios of tour experience and questionnaires filled out after each round.

Table 2 Question list

Variable	Questions	Reference
Co-presence	I was aware that other people were with me in the virtual room.	Questiones on the sense of being together [71]
	To what extent, if at all, did you have a sense of being with the other person?	
	To what extent did you have a sense that you where in the same place as the other person during the course of the experience?	
	When you think back about your last experience, do you remember this as more like talking to a computer or communicating with a real person?	
	To what extent was your experience in traveling in this VR system like that other real experience, with regard to your sense of doing something together?	
	Overall rate the degree to which you had a sense that there was another human being interacting with you, rather than just a machine?	
Social presence	I empathized with the other.	The networked minds maesure of social presence [72]
	I felt connected to the other.	
	I found it enjoyable to be with the other individual.	
	I sympathized with others.	
	What my partner felt affected what I felt.	
	What I felt affected what my partner felt.	
Satisfication	I think that I would like to use this system frequently.	Usability scale [73]
	I thought this system was easy to operate.	
	I enjoyed this experience a lot.	

4.1.5 Result

The experiment was evaluated with 3 dependent variables: co-presence, social presence, and satisfaction [76, 77, 78]. These variables were determined with 15 questions in the questionnaire as listed in Table 2. Co-presence refers to the sense of "being together" with other people in a shared simulated virtual environment. Meanwhile, social presence is more psychologic related and defined as "the feeling that others are involved in the communication process."

The basic descriptive of the three variables are shown in Table 3. To evaluate the performance of the system in 2 trials. We calculated and compared the mean score of each variables. The results showed a significant difference in the usability. It indicates that adding socialization feature in VR tourism have positive effect on improving satisifcation of using the system.

Table 3 Basic descriptive

Variable	Sample	Mean	SD
Co-presence	10	1.000	0.000
2) Co-presence	10	3.000	0.816
Social Presece	10	1.286	0.117
2) Social presence	10	2.855	0.305
Satisfaction	10	2.865	0.318
2) Satisfaction	10	3.416	0.332

As shown in Table 4, the result of correlation analysis of co-presence and social presence is $r = 0.692 > 0$, $p = 0.027 < 0.05$. The correlationship between co-presence and social presence was significant, and it revealed that co-presence and Social presence can affect each other.

Table 4 Correlation analysis [r(p)]

	2) Co-presence	2) Social presence	2) Usability (satisfaction)
2) Co-presence	1		
2) Social presence	0.692(0.027)	1	
2) **Satisfaction**	-0.421 (0.225)	-0.160(0.660)	1

4.1.6 Discussion

Participants provided several feedbacks after experiment. And we found some limitations according to the feedbacks.

- The avatar moved too fast with VR teleport locomotion, it was hard to have face-to-face communication and observe the facial motion.
- Participants have less experience with other facial expression of emotion except joy.
- The photorealistic face looks strange to participants.

4.3 Experiment 3: The Validation of Shared VR Tourism System with Emotional Connections

4.3.1 Objective

This experiment is designed to validate the performance of the system in different emotional states by providing user various VR tourism contents. The experiment environment setup is the same with experiment 1.

4.3.2 Participant

4 university students (all females) with an age range of 23 -27 years took part in the experiment. All participants were paired in groups of two (2 groups in total). All participants had previous experience with VR but had little to no previous experience with social VR. The final sample is 4 participants (2 groups).

4.3.3 Apparatus

This experiment used 1 set of HTC Vive pro eye and a set of HTC Vive, each set consisting of 2 base stations, an HMD, two hand controllers, and connected to a desktop PC. To connect the Photon's server, the PCs were setup on Internet connected. The VR tourism content is the 360-degree videos of (1) roller coaster scene, (2) Telpher scene, (3) riverside scene.

4.3.4 Procedure

Since all participants in experiment 3 had joined the previous experiment 2, the avatar data was reused for this experiment. After the initial preparation phase where participants listen to the instructions of the experiment, they were explained how to express their emotions with facial motion and voice.

For this experiment, participants were shown with three 360-degree videos. The roller coaster scene consisted of a roller coaster riding experience. The telpher scene showed the experience of taking the ropeway and enjoying the view of ski resort form the air. And the riverside walking scene is the walking alongside the river in a sunny day.

Figure 4-10 Experimental scene in experiment 3

In this experiment, participants were seated in the fixed location during the whole experiment to ensure they can see each other. Then, they were asked to view three 360-degree videos that help the participants to keep in a certain emotional state. The optional videos include: (1) roller coaster riding; (2) Telpher riding; (3) riverside walking. This experiment includes 2 trials, one for using avatar without facial expression function, another for using avatar with certain facial expression of emotion that is triggered by participants' movement. Each video lasts for about 3 minutes. During the whole experiment, the avatar representations were fixed in the seat, but participant can rotate their head to change the angle of view. In the first trial, facial expression function was disabled and the facial expression on the avatar was set to default facial expression (neutral) for the trial 1.

Figure 4-11 The view of participant in trial 1

The second trial examined the facial expression function under different tourist scene. Some examples of facial expression in trial 2 were shown in Figure 4- 10. The participant can not only see the avatar representation of their partner, but the facial expression was also synchronized in real-time. The participants were told to observe the reaction and look at the avatar as much as possible during the experience. Addition to that, participants were asked to feel the emotion of their partner and consider if the mutual empathy existing between them.

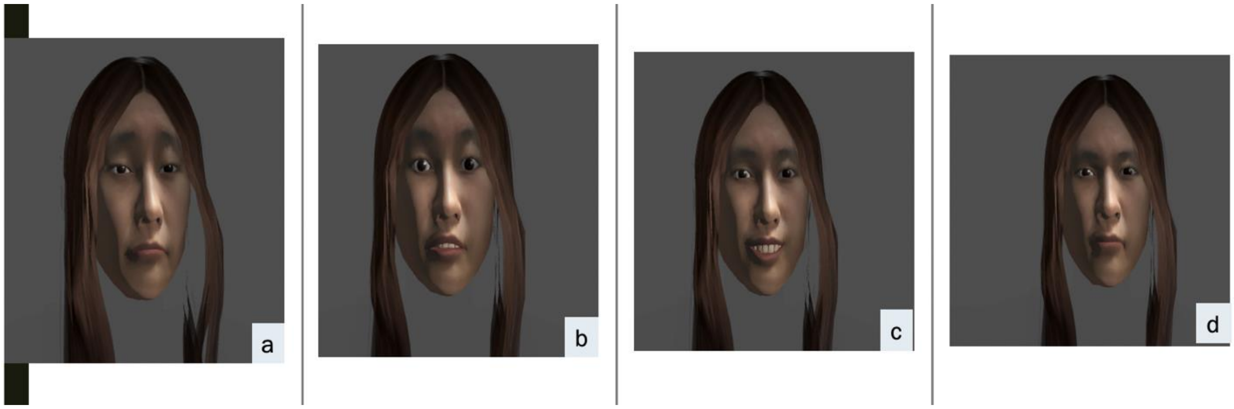


Figure 4-12 Typical facial expressions

(a – sad; b – surprise; c – joy; d – angry)

After each trial, the same questionnaire from experiment 1 was used to evaluate the performance of the system and the effect of facial expression on social presence and co-presence.

4.3.5 Result

Table 5 Paired T-test of three variables. * p<0.05 ** p<0.01

Paired T-test						
Pair No.		Mean	SD	MD	t	p
1	social presence	1.83	0.28	-1.73	-7.343	0.005**
	2) social presence	3.56	0.44			
2	Co-presense	2.33	0.49	-1.32	-6.718	0.007**
	2) Co-presense	3.65	0.25			
3	Satisfaction	2.67	0.27	-0.81	-5.166	0.014*
	2) Satisfaction	3.48	0.08			

To compare the results of experiment with facial expression function and without it, a paired T-test was conducted. The pairs of variables are shown in Table 5. To evaluate the statistical difference of the three variables in situation with no facial expression and situation with facial expression function, we performed a paired T-test with paring the same variable from 2 trials. The results showed there are significant difference on all three pairs of variables. Both social presence and co-presence showed significant difference in 0.01 level and significant mean difference between the 2 trials. It indicates that facial expression has positive effect on improving social presence and co-presence. In the pair of usability, the significant difference reveals that facial expression is effective to improve user's satisfaction.

4.3.6 Discussion

In this experiment, we examined the shared VR tourism system with different VR tourism content and different emotional states. The result of experiment provided evidence to suggest that introducing facial expression can make social interaction in VR tourism a similar experience to physical tour with emotional contagion.

The interview after each experiment revealed some common themes experienced by the participants. And we found some potential limitations according to the feedbacks.

- The avatar was akward and it was difficult to look at it for a long time.
- The vary of emotion is frequent and slight, but only significant and typical emotional facial expression was reflected.
- The duration of each video was too short. It was difficult to enjoy the traveling experience and oberseve the reaction of the paterner at the same time in such a short time.

5. Conclusion

5.1 Summary

In this research, concerning the problem of lacking social interactions in VR tourism, we have presented our shared VR tourism system with emotional connection. The shared VR tourism system with emotional connections is a system for integration of social features and VR tourism system. The system includes 2 principal parts: VR tourism content and multiplayer interaction. The system has been tested with success on the feasibility test, for synchronize real-time photorealistic facial expression with eye-tracking technology and audio-driven lip-syncing technology and communicating between users through the facial expression. Based on the structure of our system, we enriched the way of interaction in VR.

This paper reports the result of an experimental study conducted to examine the effect of adding social features, in addition to the non-verbal communication, on the performance of VR tourism system, social presence and co-presence between remote participants. We conducted experiments to prove the feasibility of the realizing recognizable facial expression with the existing technologies and the performance of using this facial expression in our system. We also conducted extra experiment to research the emotional connection and the emotional response of facial expression on avatar in VR tourism environment.

All in all, the research has successfully built a novel shared VR tourism system that allows user to socialize in various ways in VR and help to build and maintain emotional connection for remote users. With the more natural interaction in VR environment, the immersion and sense of presence of using this VR system increased.

5.2 Limitation and Future Work

Due to the limitations of the existing hardware and time constraints the experiments were conducted in small scale. If it were to examine the large group of participants, the result would be more comprehensive and accurate. If more time were available, there would be some large-scale experiments conducting with consideration of variety of nation, age and balance of gender. The way of expressing emotion may vary from different cultural background and the emotion response to an identical content varies from people. For example, some people feel scared and they scream or cry while riding a roller coaster. But others may feel excited and laugh during the ride.

This paper showed the exploration of improving current Virtual Reality tourism system with enriched social interaction through novel motion – tracking technology. We decided to focus on the facial expression which is the representation of non-verbal interpersonal interaction. And reconstructing facial expression by combining eye tracking and audio driven lip-syncing method. Because of the limitation of speech recognition, the analysis of mouth movement is not accurate as expectation which leads to the limitation of missing some untypical facial expression. More work should be carried out to improve the efficacy of the facial expression synchronization. A future system would integrate the optical lip-tracking technology to lip-syncing function, which this system did not include. Instead of audio-driven lip synchronization, optical lip tracking will provide a better result of tracking, and the realizing of various shape of mouth movement is possible. With the accuracy of generated facial expression, it will allow us to further investigate the novel approach to share emotion with others and the emotional connection in virtual environment.

Another possible limitation could arise in this study is the use of partial body avatar (head and hands). Although previous research clarified that partial embodiment is adequate for increasing co-presence in virtual reality, our participants pointed out that the current 3D avatar have somewhat reduced the presence of the experience. It was difficult to feel sympathy with this avatar. In the future research, much work needs to be conducted in improving the quality of user interaction.

6. Acknowledgement

Time always flies so fast that before I realized, the last 2 years, when I was a student in Keio University, will come to a happy end. During the master's program, I have been helped by many people around me. When my thesis is about to be completed, and when I am about to graduate, I would like to thank all the people who have offered invaluable support in the past 2 years.

First and foremost, I would like to express my deep and sincere gratitude to Professor Ogi for giving me the opportunity to do research and providing invaluable guidance throughout this research. He has taught me the methodology to carry out the research and to present the research as clearly and logical as possible. It was a great honor to learn and grow under his guidance.
I would also like to thank all my committee members for your valuable comments and suggestions that helped me to revising the graduate thesis. Additionally, I want to thank all the professors in SDM, Keio University. Thank you for the insightful lectures.

Secondly, I want to thank everyone who has participated my experiment. Thank you all for attending and helping me in this research transcending the tough pandemic situation. There is no chance that I can finish it without your support.

Finally, my deepest gratitude goes to my family for their continuous support and understanding for my past 24 years. Although we haven't seen each other face to face for almost 2 years, their help and support have accompanied me through the difficult period of the thesis and moments of my life.

7. Reference

[1] Burdea, G. C., & Coiffet, P. (2003). Virtual reality technology (Vol. 1): Hoboken, NJ: John Wiley & Sons.

[2] Slater, M. Place illusion and plausibility can lead to realistic behavior in immersive virtual environments.Philos. Trans. R. Soc. B Biol. Sci. 2009, 364, 3549–3557.

[3] What is WebVR? From: https://webvr.info/

[4] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Robert V. Kenyon, and John C. Hart. 1992. The CAVE: audio visual experience automatic virtual environment. Commun. ACM 35, 6 (June 1992), 64–72. DOI: https://doi.org/10.1145/129888.129892

[5] Shibata, Takashi (1 April 2002). "Head mounted display". Displays. 23 (1–2): 57–64. doi:10.1016/S0141-9382(02)00010-0. ISSN 0141-9382.

[6] Vecchiato, G.; Jelic, A.; Tieri, G.; Maglione, A.G.; De Matteis, F.; Babiloni, F. Neurophysiological correlates of embodiment and motivational factors during the perception of virtual architectural environments. Cogn. Process. 2015, 16, 425–429.

[7] Cheong, R. (1995). The virtual threat to travel and tourism. *Tourism Management, 16*(6), 417-422. doi: 10.1016/0261-5177(95)00049-T.

[8] Williams, P., & Hobson, J. S. P. (1995). Virtual reality and tourism: fact or fantasy? *Tourism Management, 16*(6), 423-427. doi: 10.1016/0261-5177(95)00050-X.

[9] Cheong, R. (1995). "The virtual threat to travel and tourism". *Tourism Management, 16*(6), 417-422. doi: 10.1016/0261-5177(95)00049-T.

[10] Kevin, S and Shubham, S. "The next normal arrives: Trends that will define 2021—and beyond". Retrieved July 8, 2021, from https://www.mckinsey.com/featured-insights/leadership/the-next-normal-arrives-trends-that-will-define-2021-and-beyond#

[11] “THE FUTURE IS NOW! UNWTO RECOGNIZES WORLD’S BEST INNOVATORS FACING UP TO COVID-19”2020, May 07. Retrieved July 8, 2021, from https://www.unwto.org/news/the-future-is-now-unwto-recognizes-world%E2%80%99s-best-innovators-facing-up-to-covid-19

[12] “Google Earth VR. “Retrieved July 8, 2021, from https://arvr.google.com/earth/

[13] “The VR Museum of Fine Art”, Retrieved July 8, 2021, from https://store.steampowered.com/app/515020/The_VR_Museum_of_Fine_Art/?l=japanese

[14] “Milapse Trek”, Retrieved July 8, 2021, from https://trek.milapse.com/namba

[15] “Virtual Reality the new trend for Travel”(2016, December 24). Retrieved July 8, 2021, from https://www.virtualhoneymoon.com/blog/?tag=marriott

[16] Kinect. Retrieved July 8, 2021, from https://developer.microsoft.com/en-us/windows/kinect/

[17] VIVE Hand Tracking SDK . Retrieved July 8, 2021, from https://developer.vive.com/resources/vive-sense/sdk/vive-hand-tracking-sdk/

[18] The Leap Motion. Retrieved July 8, 2021, from https://www.ultraleap.com/

[19] HTC Vive. Retrieved July 8, 2021, from https://www.vive.com/us/

[20] C. Runa Miranda and V. Costa Orvalho, ”Assessing Facial Expressions in Virtual Reality Environments,” in 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Rome, Italy, 2017.

[21] L. Zhang, B. Verma, D. Tjondronegor and V. Chandran,” Facial Expression Analysis under Partial Occlusion: A Survey,” ACM Computing Surveys (CSUR), vol. 51, no. 2, 2018.

[22] H. K. Ekene and R. Stiefelhagen, “Why Is Facial Occlusion a Challenging Problem?,” in International Conference on Biometrics, Berlin, Heidelberg, 2009.

[23] Face detection and attributes. Retrieved July 8, 2021, from https://docs.microsoft.com/en-us/azure/cognitive-services/face/concepts/face-detection

[24] S. Hickson, N. Dufour, A. Sud, V. Kwatra and I. Essa, "Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras," in IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 2019

[25] Robert J. K. Jacob; Keith S. Karn (2003). "Eye Tracking in Human–Computer Interaction and Usability Research: Ready to Deliver the Promises". In Hyona; Radach; Deubel (eds.). *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Oxford, England: Elsevier Science BV.

[26] "Tobii Pro VR Integration – based on HTC Vive." Retrieved July 8, 2021, from https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-vr-integration-product-description.pdf/?v=1.7

[27] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. J. Mach. Learn. Res. 10 (12/1/2009), 1755–1758.

[28] Hasan, Md & Shoeb, Ahmed. (2014). Human Lips-Contour Recognition and Tracing. International Journal of Advanced Research in Artificial Intelligence. 3. 10.14569/IJARAI.2014.030107.

[29] Habibie, I., Xu, W., Mehta, D., Liu, L., Seidel, H. P., Pons-Moll, G., ... & Theobalt, C. (2021). Learning Speech-driven 3D Conversational Gestures from Video. arXiv preprint arXiv:2102.06837.

[30] Daniel Cudeiro, Timo Bolkart, Cassidy Laidlaw, Anurag Ranjan, and Michael Black. Capture, learning, and synthesis of 3D speaking styles. Computer Vision and Pattern Recognition (CVPR), 2019. 3

[31] Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, 36(6), Nov. 2017. Two first authors contributed equally.

[32] Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. Audio-driven facial animation by joint endto-end learning of pose and emotion. ACM Trans. Graph., 36(4), July 2017. 1, 3

[33] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: Learning lip sync from audio. ACM Trans. Graph., 2017. 3, 4

[34] Tavakoli, R., & Mura, P. (2015). ‘Journeys in Second Life’ – Iranian Muslim women's behaviour in virtual tourist destinations. *Tourism Management, 46*, 398-407. doi: 10.1016/j.tourman.2014.07.015.

[35] Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol. 2018 Aug;109(3):431-433. doi: 10.1111/bjop.12305. Epub 2018 May 21. PMID: 29781508.

[36] S. Benford, J. Bowers, L. E. Fahlen, C. Greenhalgh, and D. Snowdon. ´ User embodiment in collaborative virtual environments. In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 242–249, 1995.

[37] L, Marc & K, Florian et al., Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality. IEEE transactions on visualization and computer graphics. PP. 10.1109/TVCG.2019.2899250.

[38] VR Chat. Retrieved July 8, 2021, from https://hello.vrchat.com/

[39] Second Life. Retrieved July 8, 2021, from https://secondlife.com/

[40] Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications, 79, 5501-5527. https://doi.org/10.1007/ s11042-019-08348-9.

[41] Hong, S. W., El Antably, A., & Kalay, Y. E. (2019). Architectural design creativity in multi-user virtual environment: A comparative analysis between remote collaboration media. Environment and Planning B: Urban Analytics and City Science, 46(5), 826-844. https://doi.org/10. 1177/2399808317733267.

[42] Grasset, Raphael & Duenser, Andreas & Billinghurst, Mark. (2008). Design of a Mixed-Reality Book: Is It Still a Real Book?. 10.1109/ISMAR.2008.4637333.

[43] Li Y, Ch'Ng E, Cai S, See S (2018) Multiuser interaction with hybrid VR and AR for cultural heritage objects. In: Proceedings of the 2018 3rd digital heritage international congress, digital heritage 2018—held jointly with the 2018 24th international conference on virtual systems and multimedia, VSMM 2018. https://doi. org/10.1109/DigitalHeritage.2018.8810126

[44] R. Schroeder. Being There Together: Social interaction in shared virtual environments. Oxford University Press, 2010.

[45] Steed and R. Schroeder. Collaboration in Immersive and Nonimmersive Virtual Environments. In Immersed in Media, pages 263–282. Springer, 2015.

[46] Dmitri Williams, Nick Yee, Scott E. Caplan, Who plays, how much, and why? Debunking the stereotypical gamer profile, *Journal of Computer-Mediated Communication*, Volume 13, Issue 4, 1 July 2008, Pages 993–1018, https://doi.org/10.1111/j.1083-6101.2008.00428.x

[47] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social VR. In 2017 IEEE Symposium on 3D User Interfaces (3DUI), pages 233–234, 2017.

[48] V. Kasapakis, E. Dzardanova, and C. Paschalidis. Conceptual and technical aspects of full-body motion support in virtual and mixed reality. In L. T. De Paolis and P. Bourdot, editors, International Conference on Augmented Reality, Virtual Reality, and Computer Graphics, pages 668–682. Springer International Publishing, 2018.

[49] B. Spanlang, J.-M. Normand, D. Borland, K. Kilteni, E. Giannopoulos, A. Pomes, M. Gonz ´ alez-Franco, D. Perez-Marcos, J. Arroyo-Palacios, ´ X. N. Muncunill, and M. Slater. How to build an embodiment lab: Achieving body representation illusions in virtual reality. Frontiers in Robotics and AI, 1:9, 2014.

[50] F. De Simone, J. Li, H. G. Debarba, A. El Ali, S. N. Gunkel, and P. Cesar. Watching videos together in social virtual reality: An experimental study on user's qoe. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 890–891. IEEE, 2019.

[51] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12, 2018.

[52] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. 2020. "Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality. Proc. ACM Hum.-Comput. Interact. 4, CSCW2, Article 175 (October 2020), 25 pages. DOI:https://doi.org/10.1145/3415246

[53] T. J. Tanenbaum, N. Hartoonian, and J. Bryan. " how do i make this thing smile?" an inventory of expressive nonverbal communication in commercial social virtual reality platforms. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13, 2020.

[54] F. De Simone, J. Li, H. G. Debarba, A. El Ali, S. N. Gunkel, and P. Cesar. Watching videos together in social virtual reality: An experimental study on user's qoe. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 890–891. IEEE, 2019. 28

[55] A. Kolesnichenko, J. McVeigh-Schultz, and K. Isbister. Understanding emerging design practices for avatar systems in the commercial social vr ecology. In Proceedings of the 2019 on Designing Interactive Systems Conference, pp. 241–252, 2019.

[56] F. Moustafa and A. Steed. A longitudinal study of small group interaction in social virtual reality. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, pp. 1–10, 2018.30

[57] S. Zamanifard and G. Freeman. "the togetherness that we crave" experiencing social vr in long distance relationships. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, pp. 438–442, 2019.

[58] Scheirer, J., Fernandez, R., and Picard, R. W. Expression glasses: A wearable device for facial expression recognition, 1999

[59] Suzuki, K., Nakamura, F., Otsuka, J., Masai, K., Itoh, Y., Sugiura, Y., & Sugimoto, M. (2016). Facial Expression Mapping inside Head Mounted Display by Embedded Optical Sensors. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*.

[60] Hickson, S., Dufour, N., Sud, A., Kwatra, V., & Essa, I. (2019). Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1626-1635.

[61] Cabanac, M. What is emotion? Behav. Process. 2002, 60, 69–83.

[62] Daniel, L. Psychology, 2nd ed.; Worth: New York, NY, USA, 2011.

[63] Hadinejad, Arghavan & Moyle, Brent & Scott, Noel & Kralj, Anna. (2018). Emotional responses to tourism advertisements: the application of FaceReader™. Tourism Recreation Research. 44. 1-5. 10.1080/02508281.2018.1505228.

[64] Diemer, Julia & Alpers, Georg & Peperkorn, Henrik & Youssef, Shiban & Mühlberger, Andreas. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. Frontiers in Psychology. 6. 10.3389/fpsyg.2015.00026.

[65] Soleymani, M.; Pantic, M.; Pun, T. Multimodal Emotion Recognition in Response to Videos. IEEE Trans. Affect. Comput. 2011, 3, 211–223

[66] Choi, Kwang-Ho & Kim, Junbeom & Kwon, O. & Kim, Min & Ryu, Yeon & Park, Ji-Eun. (2017). Is heart rate variability (HRV) an adequate tool for evaluating human emotions? – A focus on the use of the International Affective Picture System (IAPS). Psychiatry Research. 251. 10.1016/j.psychres.2017.02.025.

[67] Unity 3D. Available online: https://unity3d.com/ (accessed on 5 July 2021).

[68] Crystal Shower Fall. Retrieved July 5, 2021, from https://www.atmosphaeres.com/video/449/Behind+The+Waterfall+Curtain

[69] "Bianzhong of Marquis Yi - Traditional Chinese Bells." Retrieved July 5, 2021, from https://www.youtube.com/watch?v=KrLjz7D8jwQ

[70] Lim JZ, Mountstephens J, Teo J. Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. Sensors. 2020; 20(8):2384. https://doi.org/10.3390/s20082384

[71] "Tobii Pro VR Integration – based on HTC Vive." Retrieved July 8, 2021, from https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-vr-integrationproduct-description.pdf/?v=1.7

[72] Oculus Lipsync. Available online: https://developer.oculus.com/downloads/package/oculus-lipsync-unity/(accessed on 5 July 2021).

[73] "Oculus LipSync - Lip Animation from Audio." Retrieved July 8, 2021, from https://anenbergb.com/projects/lipsync/#powered-by-deep-learning

[74] Photon Engine: Multiplayer Game Development Made Easy. Available online: https://www.photonengine. com/ (accessed on 5 July 2021).

[75] Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion. Segerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture, 27, 46.

[76] Cagatay Basdogan, Chih-Hao Ho, Mandayam A. Srinivasan, and Mel Slater. 2000. An experimental study on the role of touch in shared virtual environments. ACM Trans. Comput.-Hum. Interact. 7, 4 (Dec. 2000), 443–460.

[77] Biocca, F., Harms, C., & Gregg, J. (2001, May). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In 4th annual international workshop on presence, Philadelphia, PA (pp. 1-9).

[78] Brooke, John. (1995). SUS: A quick and dirty usability scale. Usability Eval. Industry.189.

8. Appendix

Appendix 1 Questionnaire for Experiment 1

Evaluation of VR Tour System Using Experience

This is a questionnaire for the research of "The Development of Shared VR Tourism System with Emotional Connection".

Please indicate if ("yes/no")

1. Have you used VR before?

○Yes

○No

2. Do you have experience of using social VR before?

○Yes

○No

Please rate the following questions on a 5-point Likert Scale (1 = "strongly disagree", 5 = "strongly agree")

3. I was aware that other people were with me in the virtual room.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

4. To what extent, if at all, did you have a sense of being with the other person?

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

5. To what extent did you have a sense that you where in the same place as the other person during the course of the experience?

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

6. When you think back about your last experience, do you remember this as more like talking to a computer or communicating with a real person?

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

7. Think about a previous time when you travel together with another person. To what extent was your experience in traveling in this VR system like that other real experience, with regard to your sense of doing something together?

Not at all	○1	○2	○3	○4	○5	Very real

8. Overall rate the degree to which you had a sense that there was another human being interacting with you, rather than just a machine?

Not at all	○1	○2	○3	○4	○5	Very real

9. I empathized with the other.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

10. I felt connected to the other.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

11. I found it enjoyable to be with the other individual.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

12. I sympathized with others.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

13. What my partner felt affected what I felt.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

14. What I felt affected what my partner felt.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

15. I think that I would like to use this system frequently.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

16. I thought this system was easy to operate.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree

17. I enjoyed this experience a lot.

Strongly disagree	○1	○2	○3	○4	○5	Strongly agree