第 26 回日本バーチャルリアリティ学会大会論文集（2021 年 9 月）

A Shared Virtual Reality Tourism System with Emotional Connections

Siyuan Wang[1)], Tetsuro Ogi[2)]
1) Graduate School of System Design and Management, Keio University
2) Graduate School of System Design and Management, Keio University

Abstract： Virtual Reality (VR) tourism provides a substitute of physical travel. While most existing systems provide only single-user experience, our shared VR tourism system aims to introduce socialization feature in VR tourism and explore the effect on maintaining emotional connections. In this paper, we present the design our shared VR tourism system that built with HTC Vive Pro Eye that multiple users to immerse themselves in the travel experience together and interact through non-verbal communications realized by eye-tracking and lip tracking technology to build emotional connection effectively. And we evaluate the performance based on 3 variables: co-presence, social presence and satisfaction. Our results that this system can improve the emotional connection in VR tourism experience.

Keywords： Virtual Reality, VR tourism, eye tracking, lip tracking, non-verbal communication

1. Introduction

Travel is the movement between different geographic locations for various purpose. 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, travelling in real life is unsuggested during current pandemic period.

To provide people the experience of traveling without physical travel, Virtual Reality (VR) technology is 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.

In order to upgrade the VR experience into a more natural level, interactivity is seen as one key element [1]. Natural embodied interaction behavior needs to be enhanced in VR. Additionally, due to human's nature of socialization needs, shared VR needs to be introduced to simulate the interpersonal interactions in reality. Shared VR refers to a 3D virtual spaces where multiple users can interact with others through HMDs [2].

However, most existing VR tourism research and programs 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.

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 [3]. The positive effect of using no-verbal communication, such as gestures, gaze and facial expression, on developing more natural embodied interaction in VR system was proved [4]. Facial expressions are the facial changes in response to a person's internal emotional states, intentions, or social communications. With the development of facial tracking technologies, the number of research on facial expression was increased. However, facial tracking in VR has 2 main obstacles: (1) Facial occlusion with HMD, over half of user's face were covered and unrecognizable while wearing the HMD. (2) Random movement, the facial motion is hard to track with traditional camera while user is moving, rotating constantly during the whole experience. Hence, facial expression with partial facial occlusion was researched on to solve the problems. 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 [5]. 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 [6].

Considering the problems of current VR tourism systems above, we have developed a shared VR tourism system with emotional connections that can combine adding socialization features and expanding interactions in VR with existing motion tracking technologies. Our system provides a shared virtual environment where multiplayer can meet with using VR headset

(Fig.1). Users can have more natural embodied interaction under the problem of facial occlusion in VR. User can interact with each other through voice chat and non-verbal communication (body language and facial expression). The emotion of user can be exchanged, and emotional connection can be enhanced. Additionally, improvement of social presence of VR tourism is expected.

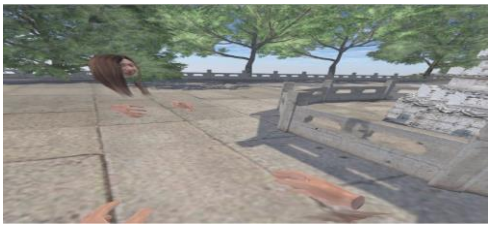

Fig. 1 The running scene of the system

2. System design

The shared VR tourism system with emotional connections allows users to enter the VR tourism system with their HMD. Users can explore the VR tour environment together, communicate through verbal communication and non-verbal communication. And all the information are transferred through the Internet.

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.

2.1.1 360-degree panoramic video

These videos are captured by a panoramic camera and user can see the view of every direction that was recorded with vivid sound effect. 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.

2.1.2 3D-modeling environment

3D environment modeling is the generation of realistic environment based on a real-world observation. 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.

2.2 Human-like Avatar Representation

First, we used Character creator 3 to generate 3D photorealistic avatar of user from a single picture. Then the generated 3D face model was imported into Blender where details of the model was adjusted, and blend shapes were created to simulate facial motion. In Unity, blend shapes were used to create animation clips that can simulate facial expression.

2.3 Facial expression (Motion-tracking technologies)

In this system, facial expression was divided into 2 facial motion area, eye area and mouth area. And the expression around each area is generated respectively.

2.3.1 Eye-tracking

With the embedded eye tracker inside the HTC Vive Pro Eye, several 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, such as blinking, eyebrow frowning.

2.3.2 Audio-driven lip syncing

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

2.4 System networking

Photon Unity Networking (PUN) and Photon Voice have been used for the implementation of the multiplayer features and the voice chat, respectively. In order to synchronize users' moving status through Internet, several scripts were attached to the avatar representation and motion data were transferred via Photon server to each client. Since verbal communication is the necessary component for social VR, Photon voice is used to create real-time voice chat that allows user hear each other in the shared environment.

3. Experiment

Several experiments were conducted to validate the performance of the system and explore the effect of using facial expression in VR tourism.

A pilot experiment was conducted to verify if user can recognize avatar's facial expressions that were generated with the motion-tracking technologies used in this system. Participants were asked to observe and classify the 7 universal facial expressions on avatar which were recorded in advance . And the correction ratio was 0.875, it indicated that participants can recognize the facial expression and the emotion they represented. However, this experiment also revealed that some facial expression that requires no voice signals is difficult to generated with lip syncing.

3.1 Experiment 1: The validation of shared VR tourism system with emotional connections

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.

3.1.1 Participants

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).

3.1.2 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 3D modeling environment of Yellow Crane Tower.

3.1.3 Procedure

The whole experiment was divided into 2 trials. The first trial used a solo scenario. Participant can explore the virtual environment by themselves, walk around or interact with the environmental object. The first trial lasted about 5 minutes. The second trial is a multiplayer scenario. Similar to the solo scenario, participant can explore the virtual environment, but they have an 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.

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.

3.1.4 Result

The experiment was evaluated with 3 dependent variables: co-presence, social presence, and satisfaction[7,8,9]. These variables were determined with 17 questions in the questionnaire. 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."

Table 1 Basic descriptive of three variables

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

"2)" represents the second trial of experiment

The basic desccriptive of the three variable are shown in Table 1. 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 2 Correlation analysis of three variables {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

r = Pearson correlation; p = probability

As shown in Table 2, 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.

3.1.5 Feedbacks

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.

3.2 Experiment 2: The validation of shared VR tourism system with emotional connections

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.

2.2.1 Participants

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).

2.2.2 Procedure

In this experiment, participants were seated in the fixed location 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 (see Fig.2). Each video lasts for about 3 minutes. After each trial, the same questionnaire from experiment 1 was used to evaluate the performance of the system.

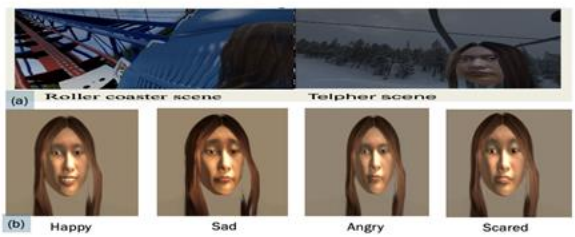

Fig. 2 (a) The running scene with different VR tourism content.
(b) Generated facial expressions in experiment.

3.2.3 Result

The pairs of variable are shown in Table 3. 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 variable. 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 have 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.

Table 3 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			

4. Conclusion

In this research, concerning the problem of lacking social interactions in VR tourism, we proposed and successfully built a novel shared VR tourism system with emotional connections by using eye-tracking technology and audio-driven lip-syncing technology. Based on the structure of our system, we enriched the way of interaction in VR. We also conducted experiments to prove the feasibility of the realizing recognizable facial expression with the existing technologies and the performance of using this facial expression function in our system.

This paper shows 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.

In the future, implanting other motion-tracking technology into this system to improve the accuracy of facial expression is in plan. 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. Moreover, the small sample of experiments may lead the inaccuracy of result. Another experiment with a bigger scale of participants needs to be conducted in the future. Additionally, When analysis the experiment result, characteristics of sample should also be considered, such as the effect of nationality, age on generating facial expression.

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 social presence of using this VR system increased.

5. Reference

[1] Williams, P., & Hobson, J. S. P. (1995). Virtual reality and tourism: fact or fantasy? *Tourism Management, 16*(6), 423-427.

[2] 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.

[3] 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.

[4] Divine, M & Guo, F et al., 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.

[5] Suzuki, K., et al (2016). Facial Expression Mapping inside Head Mounted Display by Embedded Optical Sensors. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*.

[6] Hickson, S & Dufour, N., et al (2019). Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1626-1635.

[7] 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).

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

[9] Slater M. Immersion and the illusion of presence in virtual reality. Br J Psychol. 2018 Aug;109(3):431-4