Evaluation of Low-end Virtual Reality Content of Cultural Heritage: A Preliminary Study with Eye Movement

Jeremy Tzi-dong Ng
The University of Hong Kong
Hong Kong
jntd@connect.hku.hk

Weichen Liu
University of California, San Diego
San Diego, California, U.S.A.
wel008@ucsd.edu

Xiao Hu
The University of Hong Kong
Hong Kong
xiaoxhu@hku.hk

Tzyy-Ping Jung
University of California, San Diego
San Diego, California, U.S.A.
jung@sccn.ucsd.edu

ABSTRACT

The affordances of virtual reality (VR) have made it widely adopted for presenting the content of cultural heritage in digital libraries. In recent years, non-specialists including university students were involved in creating low-end VR content using low-cost equipment and software. Among various design options, the question as to which ones could be more effective in presenting cultural heritage remains. This study aims to evaluate and compare the effectiveness of user-created VR content of cultural heritage with different designs, through collecting and analyzing self-report and eye movement data from end users. Results show that the presence of text annotations in the VR content helped users understand the cultural heritage being presented, whereas users' visual attention was largely attracted to the text annotations and additional images when the VR content contained such visual information. This preliminary study also explores the feasibility of using the eye-tracking method to analyze user interactions with VR content of cultural heritage. The results provide empirical evidence on the effects of different designs of user-created VR content on end users' understanding of cultural heritage.

CCS CONCEPTS

•Human-centered computing ~ Human computer interaction (HCI) ~ HCI design and evaluation methods ~ User studies

KEYWORDS

Virtual reality; cultural heritage; eye-tracking; user evaluation

ACM Reference format:

Jeremy Tzi-dong Ng, Weichen Liu, Xiao Hu, and Tzyy-Ping Jung. 2020. Evaluation of Low-end Virtual Reality Content of Cultural Heritage: A Preliminary Study with Eye Movement. In *Proceedings of Joint Conference on Digital Library, Virtual Event, August 1-5, 2020 (JCDL'20), August 1–5, 2020, Virtual Event, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3383583.3398603

1 Introduction

Virtual reality (VR) is defined as an immersive and interactive virtual environment created by exploiting existing technologies (e.g., 3D photogrammetry) [1]. For example, VR has been employed in digital libraries for displaying cultural content [2]. The affordances of VR have made it adopted in facilitating and enhancing cultural heritage learning [3]. Learning about cultural heritage involves interests, engagement and affective responses [6], well facilitated by the use of VR [3][9]. Despite being widely used, its effectiveness in presenting learnable content has mostly been measured by self-report that could be imprecise or biased [4]. To date, most VR content has been created by specialists, which limits the accessibility of VR to the general audience [5]. VR content created by non-expert users enables their experience of authentic activities and creative interpretation of reality [7]. However, the efficacy of user-created VR content has rarely been formally investigated. In a previous study [8], low-end VR content of cultural heritage, defined as comprising a spherical panorama and basic additional features such as text annotations, was created by undergraduate students as an assignment of a repurposed Digital Library (DL) course. This "low-tech barrier" approach was found effective in both learning about cultural heritage and development of literacy skills [8]. Notwithstanding the importance of existing work, in this study we switch the focus to the end users of the created VR content, investigating and comparing the user effectiveness of alternative designs of such VR content. Another goal of this preliminary study is to explore the feasibility of using the eye-tracking approach to examine end users' interaction with user-created VR content. Eye-tracking, as an objective and unobtrusive means of user data collection, can complement the traditional self-report data and help avoid possible response bias [4]. In this study, an exploratory user experiment was conducted to answer the following specific research questions (RQs):

RQ1: How do user-created low-end virtual reality content with different designs affect end users' understanding of cultural heritage?
RQ2: How do different designs affect users' visual attention in the VR environment?

The first research question aims to explore how the user-created, low-end VR content with different designs (e.g., with or without text annotations) affects end users' understanding of cultural heritage, whereas the second question sets out to investigate which parts of the VR content attract users' visual attention. The first question will be answered by eliciting end users' self-reported perceptions while answers to the second question will be based on end users' eye movement behaviors. Our findings will provide empirical evidence on 1) the effectiveness of design variations of low-end VR content in presenting cultural heritage, and 2) the feasibility of exploiting eye movement data in studying user interactions in low-end VR environment. Findings also offer insights to DLs displaying VR content of cultural heritage.

JCDL '20, August 1–5, 2020, Virtual Event, China
 ACM ISBN 978-1-4503-7585-6/20/08…$15.00
https://doi.org/10.1145/3383583.3398603

2 Related Work

2.1 VR for Cultural Heritage

Aligned with the constructivist approach, VR can visualize abstract concepts and offer direct experience in the virtual environment [4]. A fundamental feature of VR that distinguishes itself from other types of digital media is its ability to deliver the experience of "immersiveness", which refers to the suspension of being aware of the real world with the feeling of being situated in the virtual environment [9]. In addition to communicating knowledge and information about cultural heritage, VR stands out for its capacity in promoting interests and engagement and generating emotional responses [3]. In particular, the audience's experience from their affective interpretation of a heritage site is shown to be impactful on their perceived significance of cultural heritage [10]. VR is thus recommended as an innovative medium for presenting cultural heritage information [1], when compared to the conventional approach of mere verbalization and assimilation of facts and information [11]. VR is therefore widely adopted by memory institutions (e.g., museums, libraries) to preserve and present socio-culturally and historically significant cultural heritage [12]. Recently, digital libraries (DL) have exploited VR technologies to display historic sites and monuments [13], where usability testing and user evaluations have yielded encouraging results [14][15]. Accessible VR experience also allows the general public to learn about and appreciate cultural heritage with minimal risks of damaging it and disrupting its environment [16].

2.2 VR Content Creation

The task of VR production has been predominately in the hands of programmers due to non-specialists' lack of skills [17], but it has been recognized that the potential of VR would not be fully realized if the content authoring work excludes novice users [5]. To bridge the gap, low-cost equipment and software have recently been introduced for creating and viewing low-end VR content [18]. The commonly agreed components of low-end VR content include spherical panoramas with a 360-degree field-of-view and additional features enabled by authoring tools (e.g., computer graphics application, online platforms) such as text annotations, images and audio [19][20]. These give rise to important design decisions that have usually relied on the accumulation of experience of creators who are specialists. There have been few studies on end users' evaluation of low-end VR content, for judging which combinations and alternatives (e.g., text only, text plus image) are more effective in delivering the content [21].

2.3 Eye-tracking in VR Environment

Eye-tracking, as an objective and unobtrusive means of data collection [24], has been adopted in the VR environment in recent years (e.g., [22][23]). Nevertheless, previous studies focus on presenting a pipeline of data collection or data architecture rather than interpreting eye movement data in a designated context (e.g., cultural heritage). Given the affordances of VR in facilitating the presentation and apprehension of cultural heritage information [1] and the novelty of eye-tracking in the VR environment, this study also explores the feasibility of applying the eye-tracking method for analyzing user interactions with VR content of cultural heritage.

3 User Experiment

3.1 Main Hypothesis and Stimuli

Given that texts and images are basic features of low-end VR content [19][20], four student-created VR stories of four different built heritage sites from Hong Kong were evaluated, each with two alternative designs. Two of the VR stories had designs with or without text annotations, while the other two had designs with or without additional images of (parts of) the heritage itself, showing details, alternative angle (e.g., the bird's-eye view), or the past appearance of the heritage site. There were thus eight stimuli in four conditions for each participant. These VR stories were presented via a VR application built by using Unity3D. The presence and absence of text and image features are the conditions, while users' visual attention based on their eye movements and self-reported understanding of the cultural heritage are the measures of user effectiveness. The following hypotheses were tested:

H1: VR content with or without text annotations has different effects on users' understanding of the cultural heritage.
H2: VR content with or without additional image has different effects on users' understanding of the cultural heritage.
H3: VR content with or without text annotations has different effects on users' visual attention based on their eye movements.
H4: VR content with or without additional image has different effects on users' visual attention based on their eye movements.

3.2 Participants and Procedures

Eight participants (four females) aged from 22 to 46 (mean = 24.9) were recruited for this pilot user experiment, five of whom were bespectacled. They reported a low level of familiarity with cultural heritage from Hong Kong. The experiment was conducted individually in a quiet and spacious room and each session lasted for about 30 minutes. Upon giving informed consent, participants filled in a pre-experiment questionnaire on demographic information and their previous experience with VR. They were then instructed to wear a VR headset, HTC Vive Pro Eye, that enabled them to see the VR content. After calibrating the eye tracker embedded in the headset, the participants were presented with each VR stimulus and were asked to freely explore it in the VR environment, each for 1.5 minutes. The four VR stories were presented to each participant in a counter-balanced order according to the Latin square design, while the conditions with or without text/image were alternating in the sequence of stimuli. After each stimulus, a short oral questionnaire was conducted to probe participants' understanding of the cultural heritage presented. After all stimuli, a post-experiment questionnaire with open-ended questions was conducted to elicit participants' perceptions towards the VR content with different designs.

3.3 Data Collection and Analysis

In the questionnaire after each stimulus, participants were asked to rate the extent to which the VR content enabled them to understand *what* the cultural heritage is and *why* it is cultural heritage. In the post-questionnaire, they were asked to rate their understanding of cultural heritage in general based on their view of VR content with text annotations and images respectively. Ratings were on a 7-point Likert scale of 1 (Strongly Disagree) to 7 (Strongly Agree). Participants' eye movements were recorded with the HTC Vive Pro Eye headset that is

a state-of-the-art device for tracking eye movements in the VR environment. It tracks users' eye movement by combining the head movement with the 2-dimensional coordinates of users' eye gaze detected by its built-in Tobii eye-tracking system. From the raw eye movement data, eye fixations are calculated by the disperse-based algorithm[1] [24]. The target areas, as the "regions of interests" (ROI), include the cultural heritage, surrounding urban constructions, tamed nature, text annotations, additional images, and others. Statistics based on these eye fixations were computed and compared across the four design conditions. Figure 1 displays the sample heatmap of eye fixations on the VR content of one cultural heritage site in this study, while the corresponding ROIs are presented in Figure 2.

The quantitative analysis focused on the effects of the conditions (i.e., presence and absence of text, or additional image) on the user effectiveness measures including eye fixations in different ROIs, and their understanding of the cultural heritage presented. Statistical tests, including t-tests (for numeric data) and Mann-Whitney U tests (for ordinal data), were used to test the hypotheses.

Figure 1: Sample heatmap of eye fixations on the VR content of Tin Hau Temple, Hong Kong (with text annotations)

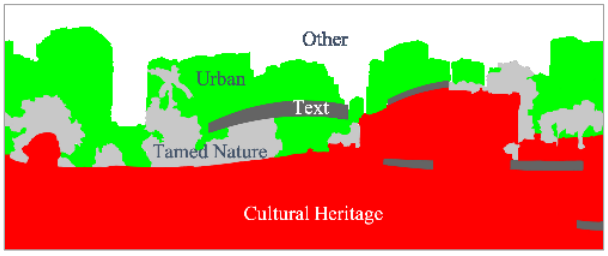

Figure 2. Region of Interests (ROIs) of eye fixation on the VR content of Tin Hau Temple, Hong Kong

4 Results and Discussion

4.1 RQ1: Understanding of cultural heritage

Hypotheses 1 and 2 are to answer RQ1. They can be tested by comparing participants' responses to the questions on their understanding of the cultural heritage after each stimulus. Table 1 shows the statistics as well as the results of Mann-Whitney U (MWU) tests. Hypothesis 1 is supported but Hypothesis 2 is not. For the comparison between the conditions with and without text annotations, users reported better understanding ($p < 0.05$) of what the heritage was and why it was regarded as a heritage when the VR content contained text annotations. From the open-ended post-questionnaire responses, a user commented that the absence of text annotations made it difficult to identify the heritage building that might not inherently "stand out" in the VR content from other different items (e.g., Christmas decorations, shopping mall) (Participant #4). In contrast, there was no significant difference ($p > 0.05$) on users' understanding of the presented cultural heritage between the conditions with and without additional images. For instance, after viewing an additional image showing the bird's-eye view of the heritage building in a VR scene, a participant commented that he/she "*felt [like being] on top of a mountain*" without mentioning anything about the heritage itself (Participant #7). This may suggest that the additional image did not help the user understand more about the heritage, but implies his/her perceived immersiveness when viewing the VR content. For another VR scene of archaeological remains of a military site without additional image, a user mentioned his guess on it being the "*old wall of a city*" (Participant #3), evidencing the user's ability to recognize what the cultural heritage is even without an additional image.

Table 1. Statistics of understanding of the heritage across different conditions

Condition	Understand what cultural heritage is Median	Understand why it is a cultural heritage Median
With Text	6.00	5.00
Without Text	4.00	3.00
p of MWU test	0.022*	0.029*
With Image	4.00	3.00
Without Image	3.00	2.00
p of MWU test	0.472	0.064

Note: N = 16; *: significant at $p < 0.05$ level; **: significant at $p < 0.01$ level.

4.2 RQ2: Visual attention on VR content

Hypotheses 3 and 4 are to answer RQ2. Table 2 shows the aggregated number of eye fixations in different ROIs across different conditions. Chi-square independent tests indicate that the fixation distributions (excluding "Text/Image") were significantly different between the conditions with and without text annotations ($x^2 = 11.56$, df = 3, $p = 0.009$**), but not significantly different between the conditions with and without additional images ($x^2 = 7.16$, df = 3, $p = 0.067$). In other words, Hypotheses 3 is supported but Hypothesis 4 is not.

Paired-samples t tests reveal that with text annotation or additional image, users' visual attention to the heritage decreased significantly. The additional images seem to have seriously distracted users' attention from the heritage itself, by reducing the number of fixations on the heritage to nearly a third of that in the original condition (i.e., without additional images). In connection to the results that users did not have a better understanding of the heritage when there were additional images (Table 1), adding images on top of the spherical VR photos may not be an effective design even though the images were about the heritage itself.

The results also show that, with text annotation, users had significantly fewer fixations on the surrounding urban structures such as residential buildings. This is desirable as the urban structures are usually irrelevant to the heritage and thus are regarded as distractions. Nonetheless, some users expressed a need for "more elaboration" than only text labels (Participant #6), for the VR content to be "more informative" (Participant #7). In the cases of having additional images,

[1] The disperse and duration threshold were set as 5 degree and 250ms respectively.

users looked at the surrounding nature significantly less than when there were no additional images. The same result is observed for the "Other" category. Considering users had twice as many fixations on the images as those on the text, it seems the images drew a lot of attention of the users, so that they generally paid less attention to all other parts of the VR scenes. Furthermore, a user perceived that adding an image to a VR scene might be "unnecessary" when there is abundant contextual information (e.g., the heritage itself being a temple) for users to "figure out what it is" (Participant #3). In sum, while text annotation might have helped users understand more about the heritage (Table 1), additional images might not have provided additional useful information about the heritage, particularly for easily recognizable heritage, and might have been distracting.

Table 2. Number of eye fixations on different ROIs across different conditions

Condition	Heritage	Nature	Urban	Text/Image	Other	Total
With Text	963	78	148	311	66	1566
Without Text	1351	108	216	--	49	1724
p of t-test	< 0.01**	0.11	0.03*		0.16	
With Image	467	117	2	798	34	1418
Without Image	1247	268	15	--	59	1589
p of t-test	< 0.01**	0.02*	N/A[a]		0.01*	

Note: N = 16; *: significant at $p < 0.05$ level; **: significant at $p < 0.01$ level; [a]: two few data points for a valid test.

5 Conclusion and Future Work

To evaluate the effectiveness of user-created virtual reality content in presenting cultural heritage, an exploratory user experiment was conducted to compare different design options of VR content for end users' understanding of cultural heritage. As a preliminary study, self-report and eye movement data of eight users were collected and analyzed. Results show that the presence of text annotations in VR content helped users understand the cultural heritage, whereas users' visual attention was largely attracted to the text annotations and additional images when the VR content contained such designs. These results validate the efficacy of user-created VR content with text annotations in facilitating end users' understanding of cultural heritage, which could be informative to DLs that display VR content of cultural heritage. Results also demonstrate the feasibility of applying the eye-tracking method to examine users' interactions with low-end VR content of cultural heritage. A limitation of this study is the small sample size and caution needs to be taken when interpreting the findings. Future work can recruit larger samples, refine VR content design, and optimize the eye-tracking approach in the VR environment.

ACKNOWLEDGMENTS

This study is partially supported by a faculty research fund from the Faculty of Education, University of Hong Kong. We thank Dr. Ying Wu's assistance in the user experiment.

REFERENCES

[1] Marcello Carrozzino and Massimo Bergamasco, 2010. Beyond virtual museums: Experiencing immersive virtual reality in real museums. *Journal of Cultural Heritage* 11, 4, 452-458.

[2] Cuiyuan Zhang, 2018. Research on Digital Library Based on Virtual Reality Technology. In 2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018), 101-106.

[3] David Fonseca, Isidro Navarro, Isabela de Renteria, Fernando Moreira, Álvaro Ferrer and Oriol de Reina, 2018. Assessment of wearable virtual reality technology for visiting World Heritage buildings: an educational approach. *Journal of Educational Computing Research 56*, 6, 940-973.

[4] Hsiu-Mei Huang, Ulrich Rauch and Shu-Sheng Liaw, 2010. Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education 55*, 3, 1171- 1182.

[5] Ralf Dörner, Marcelo Kallmann, and Yazhou Huang, 2015. Content creation and authoring challenges for virtual environments: from user interfaces to autonomous virtual characters. In *Virtual Realities*, 187-212. Springer, Cham.

[6] Hicela Ivon and Dubravka Kuscevic, 2013. School and the cultural-heritage environment: pedagogical, creative and artistic aspects. *CEPS Journal 3*, 2, 29-50

[7] Nagore Barrena, Andrés Navarro, Sara García and David Oyarzun, 2016. Cooltour: Vr and ar authoring tool to create cultural experiences. In *Intelligent Interactive Multimedia Systems and Services 2016*, 483-489. Springer, Cham.

[8] Xiao Hu, Jeremy Ng and Jin Ha Lee, 2019. VR creation experience in cultural heritage education: A preliminary exploration. In *Proceedings of the Annual Meeting of the Association for Information Science and Technology (ASIS&T) 2019*.

[9] Mel Slater, 2017. Implicit learning through embodiment in immersive virtual reality. In *Virtual, augmented, and mixed realities in education*, 19-33. Springer, Singapore.

[10] Savvas Makridis, Spyridon Alexiou and Maria Vrasida, 2017. The Role of Experience in Shaping Student Perception of the Significance of Cultural Heritage. In *Tourism, Culture and Heritage in a Smart Economy*, 467-482. Springer, Cham.

[11] Jiménez Pérez, Roque; Cuenca López, José María and Ferreras Listán, Mario (2010). Heritage education: Exploring the conceptions of teachers and administrators from the perspective of experimental and social science teaching. *Teaching and teacher education* 26, 6, 1319-1331.

[12] Loris Barbieri, Fabio Bruno and Maurizio Muzzupappa. (2017). Virtual museum system evaluation through user studies. *Journal of Cultural Heritage*, 26, 101-108.

[13] Xiao Hu, Jeremy Ng, and Shengping Xia. (2018). User-Centered evaluation of metadata schema for nonmovable cultural heritage: Murals and stone cave temples. *Journal of the Association for Information Science and Technology, 69*, 12, 1476-1487.

[14] Xiao Hu. (2018). Usability evaluation of E-Dunhuang cultural heritage digital library. *Data and information management, 2*, 2, 57-69.

[15] Xiao Hu, Eric M. Y. Ho, and Chen Qiao, 2017. Digitizing Dunhuang cultural heritage: A user evaluation of Mogao cave panorama digital library. *Journal of Data and Information Science, 2*, 3, 49-67.

[16] Hyuk-Jin Lee, 2017. Phenomenological classification of cultural heritage: role of virtual reality. *Virtual Archaeology Review* 8, 16, 69-74.

[17] Zayd Hendricks, Gary Marsden and Edwin Blake, 2003. A meta-authoring tool for specifying interactions in virtual reality environments. In *Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa*, ACM, 171-180.

[18] Sébastien Kuntz, Richard Kulpa and Jérôme Royan (2018). The Democratization of VR-AR. In Virtual Reality and Augmented Reality: Myths and Realities, 73-122.

[19] Hugo Coelho, Miguel Melo, José Martins and Maximino Bessa , 2019. Collaborative immersive authoring tool for real-time creation of multisensory VR experiences. *Multimedia Tools and Applications*, 1-21.

[20] Michael Nebeling and Maximilian Speicher, 2018. The trouble with augmented reality/virtual reality authoring tools. In *Int'l Symposium on Mixed and Augmented Reality Adjunct*, 333-337.

[21] Maria Economou and Laia Pujol, 2008. Educational tool or expensive toy? Evaluating VR evaluation and its relevance for virtual heritage. In *New heritage: New media and cultural heritage*, 242-260.

[22] Steven Hickson, Nick Dufour, Avneesh Sud, Vivek Kwatra, & Irfan Essa. (2019). Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. *In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1626-1635.

[23] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, & Matthias Nießner (2018). FaceVR: Real-time gaze-aware facial reenactment in virtual reality. ACM Transactions on Graphics (TOG), 37, 2, 25.

[24] Dario Salvucci and Joseph Goldberg, 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, 71-78.