A Case Study on the Use of Volumetric Video in Augmented Reality for Cultural Heritage

Emin Zerman
emin.zerman@scss.tcd.ie
V-SENSE, School of Computer Science, Trinity College
Dublin
Dublin, Ireland

Néill O'Dwyer
odwyernc@tcd.ie
V-SENSE, School of Computer Science, Trinity College
Dublin
Dublin, Ireland

Gareth W. Young
youngga@tcd.ie
V-SENSE, School of Computer Science, Trinity College
Dublin
Dublin, Ireland

Aljosa Smolic
smolica@scss.tcd.ie
V-SENSE, School of Computer Science, Trinity College
Dublin
Dublin, Ireland

ABSTRACT

Interest in the use of augmented reality (AR) is currently growing alongside advances in visual display and 3D reconstruction technologies. With the integration of volumetric video (VV), AR can be enhanced with representations of live-action actors. In this paper, we present an exploratory case-study of a museologically focused AR application that evaluated various parameters affecting the overall experience for visitors of the Long Room in the Old Library at Trinity College Dublin. Employing a volumetric 3D representation of Jonathan Swift, the benefits of applying VV in AR for a cultural heritage use-case scenario were explored. Moreover, we compared two AR platforms for this prototype application: a tablet and an HMD. For this purpose, we collected post-task opinions of the application and processed quantitative and qualitative data. The results highlighted differences between the two platforms and showed how the developed VV AR application was initially received by the users.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; • **Computing methodologies** → *Mixed / augmented reality*.

KEYWORDS

volumetric video, augmented reality, user experience, cultural heritage

1 INTRODUCTION

Interest in augmented reality (AR) is currently growing [14]. By harnessing integrated emergent sensor technologies, AR applications can now be easily run on commercially available devices, including tablets, smartphones, and head-mounted displays (HMDs) [6], such as HoloLens [12] and Magic Leap [11]. These AR applications can then be used for various purposes including communication, education, and entertainment.

Each technique for AR content creation [9] has its advantages and disadvantages. Computer-generated (CG) characters either feel "uncanny" and "unnatural" to viewers [13, 24] or require high-budget productions [17]. Motion capture offers some compromise; however, the requirement to wear markers creates another hurdle for content creators and actors. Volumetric video [21] (also known as 3D video or free-viewpoint video), instead, is generated using real-life, live-action footage, and 3D reconstruction techniques [7, 15, 16, 20], and it has its own limitations (see Section 4). VV reconstructs a 3D scene by placing multiple cameras around an object (or a scene) [7, 16, 20]. Using several cameras ranging from 12 [16] to more than 100 [7], the scene is simultaneously captured on all devices. Then a sequence of 3D models is generated via 3D reconstruction techniques [7, 16, 20] and stored as a sequence of textured triangular meshes or 3D point clouds. Hence, the reconstructed VV can be seen from any direction and any angle, see Figure 1.

Previous research has focused on the effect of different platforms [2, 3, 23]; however, these studies do not explore real-world case studies. Other comparisons of platforms [5, 10, 19, 22] (including AR in cultural heritage [5, 22]) have generally shown that tablets are easier to interact with and HMDs more immersive.

In this paper, a cultural heritage VV AR application was evaluated in which the users were presented with a 3D model of Dean Jonathan Swift. In collaboration with Trinity College Dublin (TCD), we used an application previously developed for this purpose [15] and provided to us by the Long Room Library. In this exploratory case-study, volunteers first used the AR application within the Long Room of the Old Library. Afterward, to explore various parameters affecting the users' overall experience, the participants were asked to fill in a questionnaire. The main contributions of this paper are
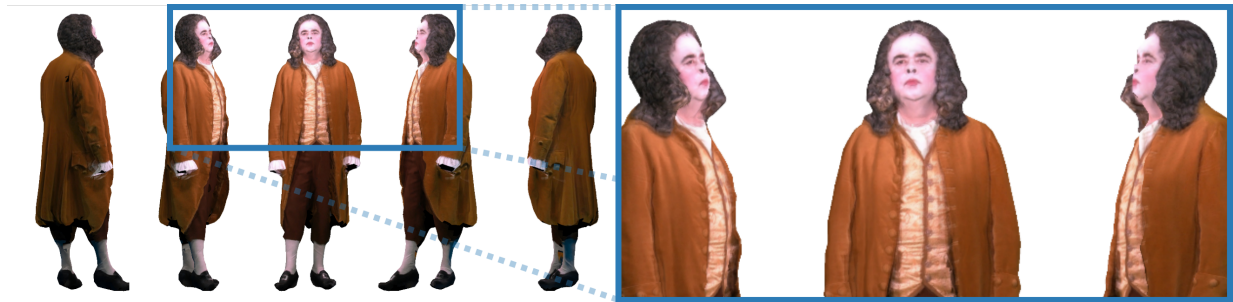
Figure 1: Volumetric video allows viewers to view its contents from all angles. Here we see the VV that was used in the presented study.

(i) a comparison of the effects of AR devices on perceived quality: tablet vs HMD, and *(ii)* a formative quantitative and qualitative analysis of parameters affecting the user experience for a cultural heritage AR application.

2 EXPERIMENT METHODOLOGY

2.1 Experiment Setup & Stimulus

To enrich visitor experiences [4], an experiment was conducted in the Long Room of the Old Library at TCD, adjacent to the bust of Dean Jonathan Swift, outside of regular opening hours; thus controlling for environmental continuity. An Apple iPad (version 6) and a first-generation Microsoft HoloLens were used as the tablet and the HMD interfaces, respectively. Participants held the tablet in their hands while wearing earphones and wore the HMD, adjusted accordingly.

The application [15] used during this case study was designed to be educational and lighthearted, a narrative innovation when compared to similar technologies used to enhance museum experiences. The participants were presented with the VV, a 3D capture of an actor (see Figure 1), represented as a sequence of textured 3D meshes. The character retells an anecdotal story of his lived experience while a student at TCD. Thus, the visual stimulus was presented with the corresponding monologue. Since the HMD had a limited storage capacity, only one volumetric video could be used as the stimulus. The stimulus was one minute long, and it was presented on a loop.

2.2 Experiment Procedure & Participants

We intentionally targeted participants who could be considered as naïve to both VV and AR technologies but were familiar with the current Long Room experience. Thus, the participants were recruited from the TCD university library and the faculty of arts and humanities via email. From within this group, administrators, students, and library employees attended. A total of 17 volunteers were recruited and participated in the study (between the ages of 25 and 65).

Following a briefing, participants were presented with the application using both the tablet and HMD in a counterbalanced (randomized) order and allowed to ask questions. Simulating normal use, the participants stood approximately 1-2 meters away from the bust of Jonathan Swift, either holding a tablet or donning the

HMD, see Figure 2. Participants were then asked to watch the volumetric video content on the device as if they were a visitor to this museum. That is, they were allowed to move around freely around their initial location as long as they kept their distance 1-2 meters away from the bust. Most participants stood stationary facing Swift directly or moving slowly around him in an arc of approximately 120 degrees. Only some participants moved around Swift to explore the VV fully. During the interaction, there was no time limitation imposed, but participants were asked to spend at least 3 minutes (3 loops of the one-minute stimulus) with each device to familiarize themselves with both the platform and the content. Immediately after using both devices, participants were given a questionnaire, see Table 1, which they filled out in private. Either a 5-point Likert scale with intensity labels or a two-option scale was used. These questions were selected for a high-level evaluation, indicating areas of interest for future, more focused analyses. The first question was targeted to identify general user experience and an overall appreciation for the VV for the prototype application. Considering realism and visual fidelity, the following questions were designed to understand the overall quality of the visual representation irrespective of the device and storytelling qualities. This overall quality comprises of the aspects such as image quality & resolution as well as how real it looks. Individual aspects mentioned in these questions, e.g., realism [18] and image quality [25], will be further analyzed in future work. Other questions focused on audio clarity, technical errors, and user preference for the devices. Participants were also asked *"Why did you give this score?"* to expand their answers further; allowing us to qualitatively analyze their individual experiences.

3 ANALYSIS

3.1 Quantitative Analysis

3.1.1 Quality of the Application. To measure a global rating for the presented materials, the participants were asked to rate the overall quality of the application, regardless of the device, as shown in Table 1. This included their delight or annoyance with the narrative, mise-en-scène (the stage design and arrangement of the actor), personal tastes, and the individual attributes and previous experiences of the user. Thus, we considered responses as the general perception of the user experience for this application. The overall
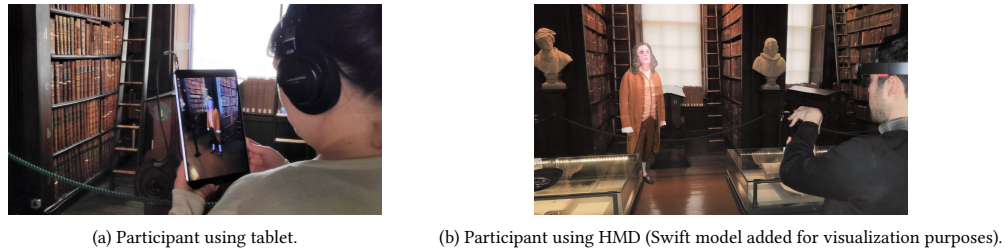
(a) Participant using tablet.

(b) Participant using HMD (Swift model added for visualization purposes).

Figure 2: Sample figures showing the experiment setup and the devices used.

Table 1: Questions in the post-task questionnaire.

Question	Scale
· How would you rate the quality of the application, in terms of a character disclosing historical information?	Likert-5
· How would you rate the quality of this representation of Jonathan Swift? Please rate considering the representation itself, e.g. realism, image quality, and independent of factors like the device & environment.	Likert-5
· How would you rate the quality of the representation of Jonathan Swift on the iPad?	Likert-5
· How would you rate the quality of the representation of Jonathan Swift on the HoloLens?	Likert-5
· How would you rate the clarity of the audio?	Likert-5
· Please rate the disturbance caused by a synchronization problem between audio and video.	Likert-5
· Please rate the disturbance if the representation was shaky or moved/floated around the screen.	Likert-5
· Which of the devices felt more immersive i.e. as if you were engaging with Jonathan Swift in person?	2-Options

user experience for this application was rated as "Good" by all participants ($n = 17$; $\mu = 4.03$; $CI = 0.69$), as shown in Figure 3.a.

3.1.2 Visual Quality of the Representation & Audio Clarity. Data were collected concerning the visual quality of the volumetric character, the visual quality of the representation on each device, and the audio clarity (see Figure 3). First, the participants rated the visual quality of the volumetric representation of Dean Jonathan Swift independent from the device. For this, the participants' overall rating of the quality of the representation was measured as being "Good" ($n = 17$; $\mu = 3.76$; $CI = 1.06$), as shown in Figure 3.a. Moreover, the participants rated the general clarity of the audio as "Excellent" ($n = 17$; $\mu = 4.59$; $CI = 0.52$). The participants were then asked to rate the visual quality of the representation on each device (see Figure 3.b). Participants rated both platforms as "Good": tablet ($n = 17$; $\mu = 3.88$; $CI = 0.80$); HMD ($n = 17$; $\mu = 3.65$; $CI = 1.02$). A Wilcoxon Signed Rank Test (a non-parametric paired difference test) revealed that there was no statistically significant difference in perceived quality between the two devices, $z = -1$, $p = 0.32$. Therefore, the median score could be considered as "Good" for both devices.

3.1.3 Intensity of Technical Errors. Since the use of AR technology for the display of VV is still in its infancy, a few technical disturbances were expected; such as audio-visual synchronization loss (lip-sync error) and the occasional loss of scene tracking – where the volumetric character appeared to float. To assess the impact of these anomalies on the users' perception of quality, the participants were asked to note these instabilities and rate the intensity of the

disturbance (see Figure 4.a). According to the responses, a majority of users did not note any loss in audio-visual synchronization ($n = 17$; $\mu = 3.88$; $CI = 1.14$); therefore, a median score of "Very Mild" was recorded. The participants' qualitative feedback on this topic suggested that this issue was more noticeable on the tablet than on the HMD. Regarding spatial destabilization, participants noted some loss of scene tracking ($n = 17$; $\mu = 3.71$; $CI = 0.87$), the median score for this question was recorded as being "Very Mild"; where ratings of "Very Mild" and "Mild" dominated the participants' feedback. Furthermore, users also verbally noted that they experienced spatial destabilization as "shakiness". This occurred more often on the tablet, and there were no references to spatial instability recorded on the HMD.

3.1.4 User Preference. To compare the two platforms, participants were asked to select which device was better than the other and in which particular aspects. Overall, the majority of the cohort (65% – see Figure 4.b) expressed a preference for the HMD, as it created an immersive experience (see the qualitative analysis for details).

3.2 Qualitative Analysis

The qualitative, open-ended responses collected from the participants were subjected to a content analysis [1, 8] to further analyze user experiences of VV AR in cultural heritage applications.

3.2.1 Visual Quality. Regarding the visual quality of Jonathan Swift's representation, participants revealed that they "...like it, but it needs to improve"; similarly, overall responses towards the
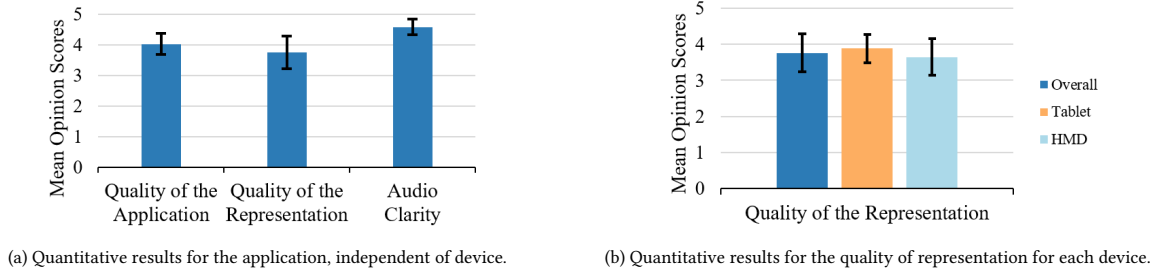
(a) Quantitative results for the application, independent of device.

(b) Quantitative results for the quality of representation for each device.

Figure 3: Participants' responses for quantitative quality scores with their confidence intervals.

(a) Participant ratings for technical limitations.
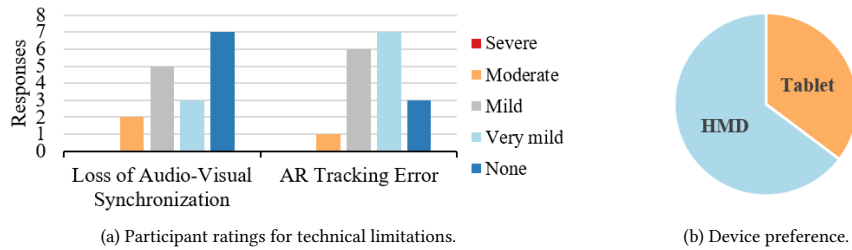

(b) Device preference.

Figure 4: Quantitative results for (a) technical limitations and (b) device preference for immersion.

character quality were mostly positive. When comparing the visual quality of both devices, the representation on the tablet was found to be of higher fidelity than that of the HMD, participants also suggested: "I think it's better on iPad but realistic on both" and "The iPad was a far better resolution, but I liked the headset". However, participants were also somewhat critical of the image quality, caused by "artifacts", that they experienced on both devices. These criticisms were elicited from phrases such as "pixelization", "rendering", and "tracking", and more articulated comments like "[The] Image quality is low. You walk around him you lost [sic] detail", "The model doesn't always fully render", and "Poor image rendition and poor tracking".

3.2.2 Technological Limitations. Participant comments highlighted the shortcomings of current state-of-the-art AR display and VV technology, particularly when concerned with pixelization, synchronization of playback, and real-world scene tracking. The face of the reconstructed model was considered especially weak – "contours of face, e.g. nose, not so successful" and "...the face was also a bit flat". Vis-à-vis the HMD, participants commented that "...the AR only showed parts of him at a time (narrow field of vision), but it's a prototype", and that "[it] was difficult to see the full body". They also stressed that with the HMD "there was some slight distortion & flickering of the image". The unstable visuals created by the occasional loss of scene tracking was noticed by some of the participants. They particularly noticed when "feet were floating on [the] image of Swift". Participants' commented that these shaky visuals were seen more often on the tablet than on the HMD. Nevertheless, both the quantitative results and qualitative feedback indicated that this "shakiness... detracted minimally from the experience". Loss of audio-visual synchronization was another noticeable disturbance on both platforms. However, some participants only

noted its occurrence on the tablet. These comments suggest that the tablet's higher resolution and image fidelity made the technical imperfections more noticeable.

3.2.3 Immersion. Concerning immersion, the AR application was successful in presenting the participants with an overall quality experience. Participants mentioned perceiving a "personal conversation" with a "very lifelike figure of Swift", and as one participant stated "It feels very compelling (it feels real enough for a good experience)". Another user suggested that "[Swift] was given a very 'characteristic' tone, that is, one feels they're in front of a person from the past". Comparing the two devices, the HMD was found to be more immersive than the tablet, for example, "Even though the quality of representation on HoloLens was not as good as iPad, the HoloLens was still a more immersive compelling experience", and "For audio, HoloLens felt more natural", although, the "HoloLens was a bit heavy". It was also mentioned that "Although the sound on the iPad was better, it filtered/blocked out the ambient environmental sounds. Left me feeling un-immersed in the Long Room itself [sic]".

Some criticisms were due to the representation being a pre-recorded volumetric video, so "[T]he model doesn't track you" and "[Swift] is looking at one point in space". It was suggested, therefore, that the users would feel more immersed if their location was tracked and Swift's gaze followed the interlocutor around the library. Another participant suggested that "Spatial sound" would improve the level of immersion they experienced, stating that "I think that audio is what we would use to find him in the space... The level needs to change as I move forward/away from Swift". This last comment hinted that although the participants felt immersion via the technology, there is room for more developments in volumetric capture and spatial audio to improve the overall quality of presence.

4 LIMITATIONS

Although it enables the capture of real-life 3D content, the VV technology has some limitations, as also pointed out by the participants. One limitation was the lack of interaction with the user. As current VV technology reconstructs a separate 3D model for each frame, this essentially constitutes a pre-recorded video that is played out to the user. Since the mesh structure is different for each frame, it is impractical to warp the head to follow the user in the current configuration; although, future advances in the technology will likely address this factor to create consistent mesh structures and help alleviate this limitation. The effects of this lack of interaction will be the topic of a more focused user study. Other limitations related to the reconstruction and visualization techniques are being actively researched and solved by the greater scientific community. Other limitations of this study included the lack of comparison between other CG models and VVs, analyzing the user responses considering their familiarity with the AR technology, and the limited participant pool.

5 CONCLUSION & FUTURE WORK

In this case study, we have presented a specific use-case experiment to evaluate a prototype, cultural heritage VV AR application for the Long Room in the Old Library at Trinity College Dublin. In this context, we collected user responses to measure the various parameters which affected the users' overall experience. The results show that the overall quality of the application was found to be "Good". However, for this particular application, further experiments are needed to validate the use of VV in AR at a larger scale. These results reveal some interesting preliminary insights and findings regarding the use of VV in AR for indoor museological and cultural heritage applications: *(i)* Experiencing VV AR for cultural heritage applications on a HMD was more immersive than on a tablet, and *(ii)* museum-goers wanted the AR content to interact with them personally. The qualitative aspects of the study suggest areas of interest that could potentially be used to expand the scale and tighten the focus of future user studies with immersive technology. These findings, therefore, stimulate further thought for user experience investigations. To fully address these limitations, we will conduct a larger field study during the Old Library's regular working hours, to explore the question: *"Is VV the best representation technique for CH oriented AR applications?"* Another aspect that may require further investigation is the long-term use of VV applications (to control for novelty factors) and exploring emergent AR devices since these parameters might prove to be more important as the proliferation of AR use increases.

ACKNOWLEDGMENTS

REFERENCES

[1] Anne Adams, Peter Lunt, and Paul Cairns. 2018. A qualititative approach to HCI research. In *Research Methods for Human-Computer Interaction* (2nd ed.), Paul Cairns and Anna Cox (Eds.). Cambridge University Press, Cambridge, UK, Chapter 7, 138–157.

[2] Evangelos Alexiou and Touradj Ebrahimi. 2018. Impact of visualization strategy for subjective quality assessment of point clouds. In *International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE.

[3] Evangelos Alexiou, Evgeniy Upenik, and Touradj Ebrahimi. 2017. Towards subjective quality assessment of point cloud imaging in augmented reality. In *19th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. https://doi.org/10.1109/MMSP.2017.8122237

[4] Ronald Azuma. 2016. Location-Based Mixed and Augmented Reality Storytelling. In *Fundamentals of Wearable Computers and Augmented Reality* (2nd ed.), Woodrow Barfield (Ed.). CRC Press, Boca Raton, FL, USA, Chapter 11, 259–276. https://doi.org/10.1201/b18703-15

[5] Chris Baber, Huw William Bristow, Sean-Lee Cheng, Anna Hedley, Yuri Kuriyama, Marc Lien, James Pollard, and Phil Sorrell. 2001. Augmenting Museums and Art Galleries. In *Interact*, Michitaka Hirose (Ed.). IOS Press, Amsterdam, 439–446.

[6] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications* 51, 1 (2011), 341–377.

[7] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-quality Streamable Free-viewpoint Video. *ACM Transactions on Graphics* 34, 4, Article 69 (July 2015), 13 pages. https://doi.org/10.1145/2766945

[8] Norman K Denzin and Yvonna S Lincoln. 2008. *Collecting and interpreting qualitative materials* (3 ed.). Sage.

[9] Borko Furht. 2011. *Handbook of augmented reality.* Springer Science & Business Media.

[10] Adam Greenfeld, Artur Lugmayr, and Wesley Lamont. 2018. Comparative Reality: Measuring User Experience and Emotion in Immersive Virtual Environments. In *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 204–209. https://doi.org/10.1109/AIVR.2018.00048

[11] Magic Leap. [n.d.]. Magic Leap One. https://www.magicleap.com/magic-leap-1. Accessed: 2020-01-24.

[12] Microsoft. [n.d.]. Microsoft HoloLens 2. https://www.microsoft.com/en-ie/hololens/. Accessed: 2020-01-24.

[13] MAsahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012).

[14] Ahmed K Noor. 2016. The HoloLens revolution. *Mechanical Engineering* 138, 10 (2016), 30–35.

[15] Néill O'Dwyer, Jan Ondřej, Rafael Pagés, Konstantinos Amplianitis, and Aljoša Smolić. 2018. Jonathan Swift: Augmented Reality Application for Trinity Library's Long Room. In *International Conference on Interactive Digital Storytelling*. Springer, 348–351.

[16] Rafael Pagés, Konstantinos Amplianitis, David Monaghan, Jan Ondřej, and Aljosa Smolić. 2018. Affordable content creation for free-viewpoint video and VR/AR applications. *Journal of Visual Communication and Image Representation* 53 (2018), 192–201. https://doi.org/10.1016/j.jvcir.2018.03.012

[17] T. S. Perry. 2014. Leaving the uncanny valley behind. *IEEE Spectrum* 51, 6 (June 2014), 48–53. https://doi.org/10.1109/MSPEC.2014.6821621

[18] Sandra Poeschl and Nicola Doering. 2013. The German VR Simulation Realism Scale-Psychometric Construction for Virtual Reality Applications with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine* 11 (2013), 33–37.

[19] Urs Riedlinger, Leif Oppermann, and Wolfgang Prinz. 2019. Tango vs. HoloLens: A Comparison of Collaborative Indoor AR Visualisations Using Hand-Held and Hands-Free Devices. *Multimodal Tech. Interact.* 3, 2 (2019). https://doi.org/10.3390/mti3020023

[20] O. Schreer, I. Feldmann, S. Renault, M. Zepp, M. Worchel, P. Eisert, and P. Kauff. 2019. Capture and 3D Video Processing of Volumetric Video. In *IEEE International Conference on Image Processing (ICIP)*. 4310–4314. https://doi.org/10.1109/ICIP.2019.8803576

[21] Aljoscha Smolic. 2011. 3D video and free viewpoint video—From capture to display. *Pattern Recognition* 44, 9 (2011), 1958 – 1968. https://doi.org/10.1016/j.patcog.2010.09.005 Computer Analysis of Images and Patterns.

[22] Vassilios Vlahakis, Nikolaos Ioannidis, John Karigiannis, Manolis Tsotros, Michael Gounaris, Didier Stricker, Tim Gleue, Patrick Daehne, and Luís Almeida. 2002. Archeoguide: An augmented reality guide for archaeological sites. *IEEE Computer Graphics and Applications* 22, 5 (Sep 2002), 55–60. https://doi.org/10.1109/MCG.2002.1028726

[23] Youguang Yu, Fuzheng Yang, Hantao Liu, and Wei Zhang. 2018. Perceptual Quality and Visual Experience Analysis for Polygon Mesh on Different Display Devices. *IEEE Access* 6 (2018), 42941–42949. https://doi.org/10.1109/ACCESS.2018.2859254

[24] Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–12.

[25] Emin Zerman, Cagri Ozcinar, Pan Gao, and Aljosa Smolic. 2020. Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. In *12th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE.