2021 Thirteenth International Conference on Quality of Multimedia Experience (QoMEX)

User Behaviour Analysis of Volumetric Video in Augmented Reality

Emin Zerman, Radhika Kulkarni, and Aljosa Smolic
V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
Emails: {zermane, kulkarnr, smolica}@tcd.ie

***Abstract*—Augmented reality (AR) is getting popular, and among other content creation techniques, volumetric video allows to bring dynamic real world content as captured by cameras into such applications. To develop efficient algorithms for compression and transmission of the volumetric media, it is important to understand how users will consume this new form of dynamic 3D content. In this paper, we analyse the user behaviour for the volumetric video consumption in AR. In particular, we study the distribution of users' viewpoints, relative locations, and average distances from the content. For this purpose, we built an Android AR application using the volumetric video and conducted a user study remotely. The results show that users spent most of their time looking at the frontal part of the volumetric video, and this indicates the importance of the face in visual attention. The collected user behaviour data are made public to support further research.**

***Index Terms*—user behaviour, user movement, augmented reality, volumetric video, user experience**

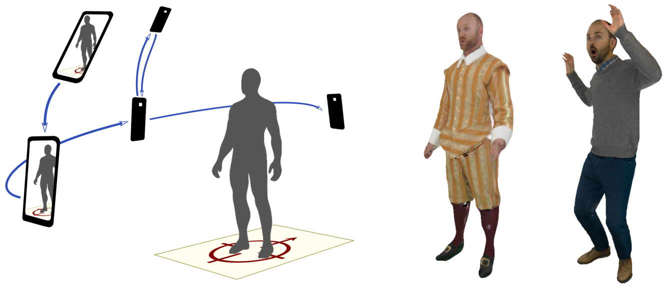
I. Introduction

Volumetric video (VV) is a new technique used to generate content for augmented reality (AR) and virtual reality (VR) applications. The VV algorithms build a dynamic 3D representation using real-life video from cameras surrounding the 3D object or scene [1]–[3]. The generated VVs are represented as point clouds or 3D textured mesh sequences [4], [5] and can be looked at from any viewpoint. The VVs are used in different marker-based [6] or markerless [7]–[11] AR applications. In almost all cases, these applications use VVs that are stored in the device; however, there is a growing interest in VV compression [12] and adaptive streaming [13]–[15], as real-time streaming is necessary for some applications, e.g., telepresence and remote collaboration.

Understanding how users interact with the VV in AR or VR is critical for optimising compression and adaptive streaming methods, such as viewport prediction [16] or rate-distortion or rate-utility estimations based on users' preferred distance [13]. User behaviour has been studied [17]–[19] and found critical in viewport prediction [20] for another immersive video technology: 360-degree video. For volumetric media, there are only a handful of studies that focus on understanding user interaction [21], AR viewport prediction [15], and navigation patterns for VR [14]. Previous studies reveal that participants preferred to stay in front of static point clouds and 1 metre away from them [21] and spent more time looking at the frontal view and faces of human models while determining the quality [22], [23]. In a recent social VR study [14], viewers' navigation patterns for VVs in VR were made publicly available, and it might be relevant for VV use in AR as well.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/27760.

978-1-6654-3589-5/21/$31.00 ©2021 IEEE



(a) Visualisation of the AR system (b) "Sir Frederick" (c) "Nico"

Fig. 1. A sample figure showing (a) how marker-based AR works with a smartphone where the VV is placed on top of a marker. The blue arrows depict a sample user movement around the marker to see the VV from different angles. Subfigures (b) and (c) show sample frames from selected VVs (b) "Sir Frederick" and (c) "Nico".

The only other work on VV use in AR discusses point cloud content delivery for AR, including compression, streaming, user movement patterns, and viewport prediction for streaming [15]. The collected user movement data are used in the viewport prediction step, and this helps reducing approximately 40% of data usage without quality reduction. Although this work collects user data, it is not a dedicated study on user behaviour analysis, e.g., it does not report user distance. To the best of our knowledge, the collected user trajectories are also not made publicly available. In our paper, different from [15], we focus on understanding user behaviour for dynamic 3D mesh-based VV use in AR. We also make the collected data publicly available along with the analysis codes.

In this paper, we provide an analysis of users' movement behaviour and their viewpoints while they are watching VVs in an AR setting. Using a marker-based AR application, device location and rotation data are collected from participants, see Fig. 1. We then analyse the users' preferences for viewing angle and viewing distance. The contributions of this paper are two-fold: *(i)* we provide a detailed analysis of user navigation in a marker-based AR scenario, and *(ii)* we make the collected user behaviour data available to help other researchers.

II. Mobile Augmented Reality Application

To collect user movement behaviour and viewpoints, we developed a marker-based mobile AR application for smartphones with Android operating system using Vuforia [24], Unity3D game engine [25] version 2018.4.16f1, and a bespoke VV player script. The application had a basic user interface with buttons to select and play the VV. The VV could be rotated by touching the screen and swiping using fingers.

To keep track of the user behaviour, 3D device position with respect to the marker and device rotation (yaw, pitch, and roll – in Euler angles) were logged into a text file in a formatted manner including a timestamp. Since the data collection was designed as a "remote" study (cf. Section III), the users were asked to submit the resulting logs via email using the file share functionality implemented using an add-on [26].

III. User Study

A. Volumetric Data

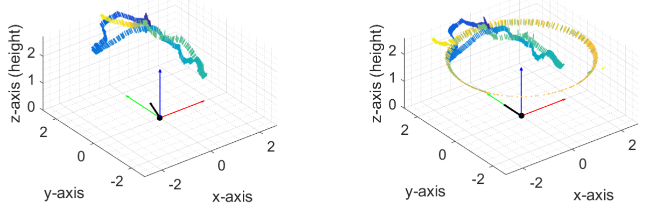
We used two different VVs kindly made available by Volograms [27] upon request: *"Sir Frederick"* and *"Nico"*, see Fig. 1. Both of these VVs depict a moving human. The selected VVs are represented as dynamic 3D meshes with texture information. *"Sir Frederick"* was a 60-seconds long VV, showing a man telling a story for the visitors of a castle [28], with ~25000 polygons and 1024×1024 pixel texture maps. *"Nico"* was a sample VV showing a surprised man which was 7-seconds long with ~16000 polygons and 1024×1024 pixel texture maps. Both VVs were 30 fps.

B. Setup & Procedure

Since the AR applications are generally designed to be consumed in any place desired by the users (e.g., their homes), in this study, we aimed to collect data for this scenario. Therefore, we designed a "remote" user study in which the participants use their own devices any time wherever they are.

For this user study, a marker image and an Android application package (APK) link were provided to the participants. Before starting, the participants were briefed about the aims of the project and informed that their movement will be recorded as a part of this user study. Following participants' informed consent, they were asked to download and install the APK, which took between 10-30 minutes. Participants were also asked to place the marker on a horizontal surface, either by printing the image or displaying it on a device such as a tablet. The exact environment and device settings were unknown.

The participants were then asked to start the application, select the content, and watch the VV in this AR application "however they wish" without any specific instructions. They were not limited in duration so that their movement behaviour was recorded in a natural way without forcing them into a time pressure or uncomfortable situation while they are consuming the VV. Once they finished watching, participants submitted the log file to the researchers via email. The device location in 3D space was measured in unit length that is relative to the marker size. The VV size was also determined by the size of the marker. Therefore, all of the distance units reported in this paper are relative measures. For this, the height of the human model in the VV was taken as 1 unit distance.



(a) Before - P#1 - "Sir Frederick" (b) After - P#1 - "Sir Frederick"

Fig. 2. Visualisation of the collected user movement data (a) before and (b) after preprocessing step (cf. Section IV-A). Preprocessing step ensures that the collected device location and orientation are always relative to a VV sequence that is facing +Y direction. The black arrow shows the initial VV orientation. The coloured arrows show the device position and orientation, and colours indicate time: blue is the beginning of user session and yellow is the end. It can be seen that this participant rotated the VV at one point while watching it, and this translated into a ring-like behaviour in (b).

This user study had been reviewed and approved by the institutional research ethics board prior to the study. The collected data and the analysis codes were made publicly available to support further studies in the field[1].

C. Participants

Responding to a call announced over the institution's email lists, 20 people participated in the data collection who did not receive compensation for participating. The data from these people (14 male, 6 female – mean age 30.2, std. 8.6) were used in the analyses. Similar to visual attention studies, for this user behaviour data collection, participants were selected among those who have not seen the contents beforehand.

IV. Analysis & Results

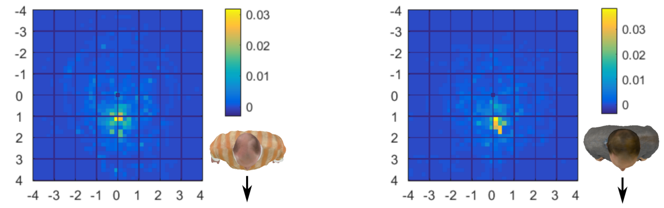
A. Preprocessing of the User Movement Data

As the user movement behaviour data was collected remotely, the marker was placed in varying orientations by the participants, see the arbitrary VV orientation indicated by the black arrow in Fig. 2.(a). To simplify the analysis, each point (i.e., the device location and orientation values) was rotated to align user movement data for different participants. First, the XYZ axes defined in computer graphics converted to the physical XYZ axes in which the XY plane defines the ground plane and Z axis is the height of the object. For this, two rotation matrices were calculated: RM_1 for the rotation between $RM_Y RM_X RM_Z[0,0,1]^\intercal$ and $[0,1,0]^\intercal$, and RM_2 for the rotation between $RM_1 RM_Y RM_X RM_Z[1,0,0]^\intercal$ and $[1,0,0]^\intercal$, where RM_X, RM_Y, and RM_Z are the rotation matrices for Euler angles. The user behaviour data points were rotated around the origin using the equation below:

$$\mathbf{y} = RM_2 RM_1 \mathbf{p} \quad (1)$$

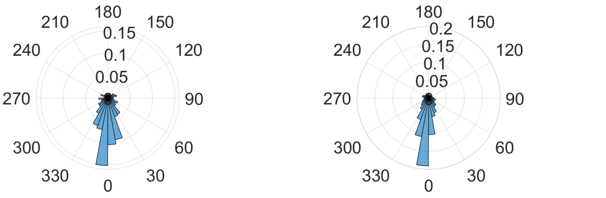
where $\mathbf{p} = [p_1, p_2, ..., p_n]^\intercal$ is a $6 \times N$ array of data points, $p_i = [x_i, y_i, z_i, u_i, v_i, w_i]$ is a data point for location (xyz) and

[1]The dataset and the analysis codes are available on https://v-sense.scss.tcd.ie/research/6dof/user-behaviour-analysis-of-volumetric-video/



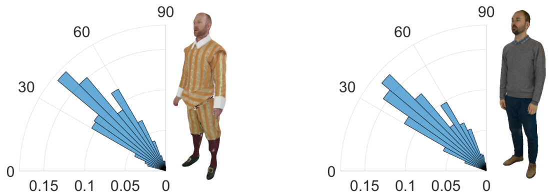
(b) “Sir Frederick” (b) “Nico”

Fig. 3. 2D histograms of users’ locations relative to the volumetric video on the ground plane for (a) “Sir Frederick” and (b) “Nico”. The volumetric video is placed in the centre (0,0), and the human in the volumetric video faces towards bottom of the page as shown on the side.



(a) “Sir Frederick” (b) “Nico”

Fig. 4. Distribution of users’ relative viewpoints in horizontal plane for (a) “Sir Frederick” and (b) “Nico”. The volumetric video is placed at the centre facing the bottom of the page, as depicted in Fig. 3.



(a) “Sir Frederick” (b) “Nico”

Fig. 5. Distribution of users’ relative viewpoints in vertical plane for (a) “Sir Frederick” and (b) “Nico”.

orientation (uvw) values in 3D space for i^{th} element, $\mathbf{y}$ is the output aligned $6 \times N$ array, N is the number of data points.

Fig. 2.(b) shows the “aligned” user behaviour data points after this preprocessing operation.

B. *Quantitative Analysis*

1) Distribution of viewpoints: Fig. 3 shows the 2D histograms of participants’ locations (or viewpoints) relative to the VV, and the results indicate that participants spent a great deal of their time looking towards the face and frontal body of the VV. The horizontal distances they spent most time in are ~1.1 units for “Sir Frederick” and ~1.5 units for “Nico”.

2) Distance from the volumetric video: In Table I, we report the average distance between each participant and the VV in terms of VV height. This information can be used in viewport prediction, visualisation, etc. We can observe that the average horizontal (on ground plane) and 3D space distances are around 1.66 and 2.37 times of the height, respectively.

3) Distribution of viewing angles: As shown in Fig. 4, participants spent around 44% of their time looking at the front of the VV ($\pm 20°$ difference from centre) and 72% of their time looking at a larger but still frontal arc of $\pm 60°$. Participants were also in agreement with one another in their vertical viewpoint distribution. Fig. 5 shows that participants spent around 44% and 74% of their time looking at VV with a slope of $[35°, 55°]$ and $[30°, 60°]$, respectively.

TABLE I

AVERAGE DISTANCES PER PARTICIPANT # ON 2D HORIZONTAL PLANE AND 3D SPACE FOR BOTH VV CONTENTS CONSIDERED. DISTANCE VALUES ARE NORMALISED: 1 UNIT DISTANCE IS THE HEIGHT OF THE VV. μ AND σ INDICATE AVERAGE AND STANDARD DEVIATION, RESPECTIVELY.

#	“Sir Frederick”		“Nico”		#	“Sir Frederick”		“Nico”	
	2D	3D	2D	3D		2D	3D	2D	3D
1	2.27	3.07	2.08	2.65	12	1.55	2.53	1.29	1.95
2	2.98	3.66	2.79	3.72	13	2.17	3.24	2.3	3.28
3	0.79	1.65	0.59	1.48	14	2.58	3.32	2.45	3.29
4	1.12	1.66	1.39	2.04	15	3.24	4.18	1.53	2.65
5	1.39	1.7	1.4	1.84	16	4.29	4.92	2.93	3.71
6	0.87	1.76	0.85	1.74	17	1.36	1.75	1.48	1.87
7	0.96	1.5	1.16	1.73	18	1.23	1.87	1.46	1.78
8	1.41	1.86	1.99	2.4	19	1.18	1.66	1.21	1.51
9	1.16	1.9	1.43	2.31	20	1.84	3	1.57	2.69
10	1.19	1.62	1.02	1.6	μ	1.73	2.43	1.59	2.3
11	0.94	1.79	0.96	1.84	σ	0.9	0.98	0.62	0.7

C. *Limitations*

This study focuses on entertainment, education, telepresence, and remote collaboration scenarios. As these scenarios focus almost always on capturing and displaying humans, only human models are selected and used in this study. Therefore, the results are only valid for two similar human models, and further studies are required for other types of contents.

In this study, user behaviour data was collected only from “remote” participants using smartphones to replicate daily use conditions for AR applications. The data collection will be extended including lab-based participants and head-mounted displays in future work, for further validation of our findings.

V. CONCLUSION & FUTURE WORK

Our analyses provide valuable insight into user behaviour in the sense how people consume the VV in a mobile AR application that uses a smartphone. Results show that the participants spent most of their time watching the VV from frontal view, which is in agreement with previous studies. Additionally, in this paper, the average distance of each participant from the VV is reported which can be used for VV rendering (i.e., minimum mesh resolution or point size), viewport prediction, or adaptive streaming rate-utility calculations.

In future work, we would like to update this mobile application to use it as a framework to conduct QoE studies for volumetric use in AR. People would approach AR visualisations for entertainment or commercials in a smartphone differently than a remote collaboration application presented with a head-mounted display. Therefore, we aim to conduct bigger comparative studies in the future to generalise our findings in this paper.

References

[1] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan, "High-quality streamable free-viewpoint video," *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 69:1–69:13, Jul. 2015.

[2] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolić, "Affordable content creation for free-viewpoint video and VR/AR applications," *Journal of Visual Communication and Image Representation*, vol. 53, pp. 192–201, 2018.

[3] O. Schreer, I. Feldmann, S. Renault, M. Zepp, M. Worchel, P. Eisert, and P. Kauff, "Capture and 3D video processing of volumetric video," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 4310–4314.

[4] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, "Textured mesh vs coloured point cloud: A subjective study for volumetric video compression," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, May 2020.

[5] K. Cao, Y. Xu, and P. Cosman, "Visual quality of compressed mesh and point cloud sequences," *IEEE Access*, vol. 8, pp. 171 203–171 217, 2020.

[6] S. Prince, A. D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato, "3D live: Real time captured content for mixed reality," in *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE, 2002.

[7] Metastage, "The Metastage App," https://metastage.com/the-metastage-app/, Accessed: 2021-01-27.

[8] Dimension Studio, "Madonna: 2019 Billboard Music Awards," https://www.dimensionstudio.co/work/madonna-billboard-music-awards-2019, Accessed: 2021-01-27.

[9] E. Zerman, N. O'Dwyer, G. W. Young, and A. Smolic, "A case study on the use of volumetric video in augmented reality for cultural heritage," in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 2020.

[10] Microsoft News Centre UK, "NHS staff are given some of the best training in the world thanks to technology," https://news.microsoft.com/en-gb/2020/07/29/nhs-staff-are-given-some-of-the-best-training-in-the-world-thanks-to-technology/, Jul 2020, Accessed: 2021-01-27.

[11] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou *et al.*, "Holoportation: Virtual 3D teleportation in real-time," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 2016, pp. 741–754.

[12] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging MPEG standards for point cloud compression," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar 2019.

[13] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 149–162, 2019.

[14] S. Subramanyam, I. Viola, A. Hanjalic, and P. Cesar, "User centered adaptive streaming of dynamic point clouds with low complexity tiling," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3669–3677.

[15] B. Han, Y. Liu, and F. Qian, "ViVo: Visibility-aware mobile volumetric video streaming," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020.

[16] B. Han, "Mobile immersive computing: Research challenges and the road ahead," *IEEE Communications Magazine*, vol. 57, no. 10, pp. 112–118, 2019.

[17] A. Singla, S. Fremerey, A. Raake, P. List, and B. Feiten, "AhG8: Measurement of user exploration behavior for omnidirectinal (360°) videos with a head mounted display," ISO/IEC JTC 1/SC 29/WG 11 JVET-H0050, ISO, Tech. Rep., Oct 2017, Macau, China.

[18] S. Fremerey, A. Singla, K. Meseberg, and A. Raake, "AVtrack360: An open dataset and software recording people's head rotations watching 360° videos on an HMD," in *Proceedings of the 9th ACM Multimedia Systems Conference*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 403—408.

[19] S. Rossi, C. Ozcinar, A. Smolic, and L. Toni, "Do users behave similarly in VR? Investigation of the user influence on the system design," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 2, 2020.

[20] X. Jiang, S. A. Naas, Y.-H. Chiang, S. Sigg, and Y. Ji, "SVP: Sinusoidal viewport prediction for 360-degree video streaming," *IEEE Access*, vol. 8, pp. 164 471–164 481, 2020.

[21] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.

[22] E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," in *Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019.

[23] H. Dutagaci, C. P. Cheung, and A. Godil, "A benchmark for best view selection of 3D objects," in *Proceedings of the ACM Workshop on 3D Object retrieval*, 2010, pp. 45–50.

[24] Vuforia, "Vuforia software development kit," https://developer.vuforia.com/, Accessed: 2021-01-27.

[25] Unity, "Unity3D Game Engine," https://unity.com/, Accessed: 2021-01-27.

[26] Y. Kula, "Native Share for Android & iOS," Unity Asset Store - https://assetstore.unity.com/packages/tools/integration/native-share-for-android-ios-112731, Accessed: 2021-01-27.

[27] Volograms, "Volograms homepage," https://www.volograms.com/, Accessed: 2021-01-27.

[28] Volograms, "Volumetric holograms in tourism and culture: how to engage visitors with real people in immersive media," https://medium.com/volograms/volumetric-holograms-in-tourism-and-culture-how-to-engage-visitors-with-real-people-in-immersive-2501f53fa971, Accessed: 2021-01-27.