A NEW CHALLENGE: BEHAVIOURAL ANALYSIS OF 6-DOF USER WHEN CONSUMING IMMERSIVE MEDIA

Silvia Rossi[⋆†] *Irene Viola*[†] *Laura Toni*[⋆] *Pablo Cesar*[† ‡]

[⋆] Department of Electronic & Electrical Engineering, UCL, London (UK)
[†] DIS, Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
[‡] INSY, TU Delft, Delft, The Netherlands

2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506525

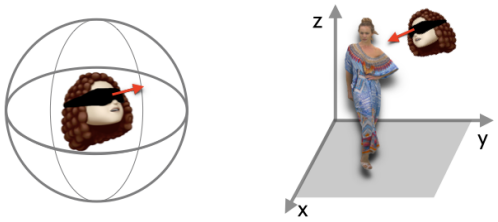
ABSTRACT

Thanks to recent advances in computer graphics, wearable technology, and connectivity, Virtual Reality (VR) has landed in our daily life. A key novelty in VR is the role of the user, which has turned from merely passive to entirely active. Thus, improving any aspect of the coding–delivery–rendering chain starts with the need for understanding user behaviour. To do so, we investigate the navigation trajectories of users within a 6-Degrees-of-Freedom (DoF) VR environment. Specifically, we investigate the main differences and similarities between 3 and 6-DoF navigation through existing methodologies adopted to study user behaviour in 3-DoF settings. Our simulation results, based on real navigation paths of users while displaying dynamic volumetric media in 6-DoF conditions, show the limitations of clustering algorithms for 3-DoF in assessing user similarity in 6-DoF. Given these observations, we state the need for developing new solutions for the analysis of 6-DoF trajectories.

Index Terms— Point Cloud, User Analysis, 6-DOF, Virtual Reality, Data Clustering

1. INTRODUCTION

Virtual Reality (VR) technology has revolutionised how users engage and interact with media content, going beyond the passive paradigm of traditional video technology, and offering higher degrees of immersiveness and interaction. In a generic VR scenario, a viewer can freely navigate the immersive scene, selecting the portion (named *viewport*) to be displayed based on her/his viewing direction. Depending on the enabled locomotion functionalities in the 3D space, VR environments can be classified as 3- or 6-Degrees-of-Freedom (DoF). In the first scenario, the viewer is virtually positioned at the centre of a sphere (Fig. 1 (a)) and, the immersive content media is typically a 360° environment projected into the virtual sphere. The media is displayed from an *inward* position and, the interaction is experienced only by rotating and translating the user head: the head is the only "interface" of interactivity. The setting of a 6-DoF scenario is rather different due to an extra level of interaction between user and content (Fig. 1 (b)). The user has now the freedom to change the viewing direction (rotating and translating the head as in 3-DoF) but also to change position inside the VR environment. The scene, populated by *volumetric objects* (*i.e.*, meshes or point clouds), is now observed from an *outward* position.



(a) 3-DoF (b) 6-DoF

Fig. 1. Viewing paradigm in 3- and 6-DoF VR.

Despite their differences, the user in both systems becomes the main driving force in deciding which media content (or portion thereof) is being displayed at any given time. Thus, to be able to deliver VR systems at a large scale, there is a compelling need to develop *user-centric* VR systems, which operate in a personalised manner (media processing is tailored specifically to the users' behaviour), to remain bandwidth-tolerant whilst meeting quality and latency criteria. To enable such user-centric systems, there is the need to understand users' interactivity models [1, 2]. User movement in VR environments has been analysed, for both 3-DoF [3–5] and 6-DoF [6–9] scenarios, in terms of total and averaged interaction time and angular velocity, among others. User navigation in 6-DoF scenarios was also studied in the past in the context of locomotion and display technology for CAVE environments [10, 11]. However, the focus has been mainly put on the analysis of completion time per task versus different setting conditions. While highly informative to summarise the interaction of users within a content, these metrics usually fail in providing other key information: which users

This work has been supported by Royal Society under grant IES—R1—180128 and by Cisco under Cisco Research Center Donation scheme.

navigate similarly within the content, and which are the dominant interaction behaviours among users. The importance of this information has been already proved in 3-DoF, and a spherical clustering algorithm and an information-theoretic approach have been proposed in [12, 13], respectively. This behavioural investigation has been instead overlooked in the emerging 6-DoF environment.

In this paper, we want to fill the gap of behavioural analysis in 6-DoF systems. The main research question we aim to address is how new physical settings and locomotion functionalities given to users can affect the analysis and understanding of their behaviour. In this first attempt of behavioural exploration for 6-DoF users, we assume the presence of a unique object of interest in a VR scene. In detail, we propose a comparative analysis of how a clustering algorithm defined for 3-DoF behaves when applied to 6-DoF trajectories. To do so, we explore how different distances (such as relative distance between user and content or between viewing direction) but also different metrics (for example euclidean versus geodesic distance) can be used to model consistent viewport overlap. Finally, we study how spherical clustering solutions fare when applied to the 6-DoF setting, using a publicly available dataset of navigation trajectories in 6-DoF [9]. Results indicate that 3-DoF clustering solutions are not able to capture similarities when users are placed at far distances between each other, suggesting that new solutions tailored for 6-DoF navigation are needed.

2. THE CHALLENGE: USER NAVIGATION IN A 6-DOF ENVIRONMENT

We are interested in analysing user behaviour, assuming that users interact similarly when they *observe the same volumetric content.* The user behaviour can be identified by the spatio-temporal sequences of the user's movements within the content, namely *navigation trajectories*. In the following, we compare the key characteristics of 3- and 6-DoF systems to highlight the main novelties of the latter in terms of navigation. It should be noted that, in a 3-DoF scenario, users are bounded to be viewing a portion of the omnidirectional content at any given time. This is not necessarily the case in a 6-DoF environment where one or multiple objects of interest are placed in the scene. For simplicity, we consider only one object of interest in an otherwise empty 3D scene of a 6-DoF system. Our analysis can be straightforwardly extended to multiple objects in the same scene.

In a 3-DoF scenario, the trajectory of a generic user i can be formally denoted by the sequence of the user's viewing direction over time $\{p_1^i, p_2^i, .., p_n^i\}$ where p_t^i is the center of the viewport projected on the immersive content (*i.e.*, spherical video) at timestamp t. The point p can be represented in spherical coordinates by $[\theta, \phi, r]$ where $\theta \in [0, 2\pi]$ is the azimuth angle, $\phi \in [0, \pi]$ the polar angle, and r is the distance between the point (viewport center projected on the immersive content) and the origin (user position). In a 3-DoF scenario, users are positioned at the centre of the spherical content; thus, r is constant during the interaction. As a consequence, the viewport centre alone is highly informative of the user behaviour, and can be used as a proxy of viewport overlap among users [12]. In particular, if the distance between viewport centres is low, the similarity between users corresponds to high viewport overlap.

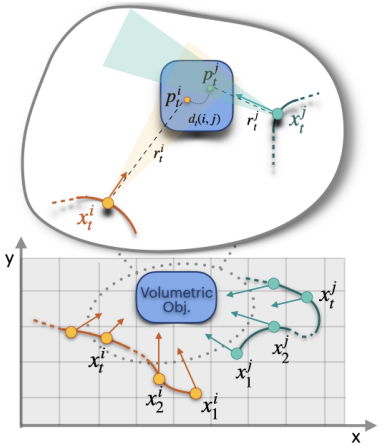
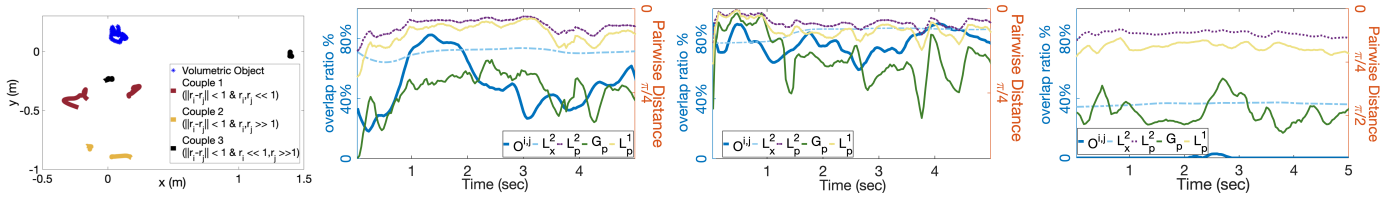


Fig. 2. An example of 6-DoF trajectories projected in a 2D domain for user i and j. In the circle, a snapshot at time t where coloured triangles represent viewing frustum per user.

In a 6-DoF setting, however, the distance between the user and immersive content can change over time due to the added degrees of freedom. In this scenario, the distance between viewport centres alone might not be sufficient to identify a common portion of the displayed point cloud. For instance, a small distance between viewport centres might suggest a high similarity between the corresponding users, which might not necessarily be true if they are at a very different relative distance from the volumetric content. Therefore, the distance r between the viewer and the object is now crucial to identify the actual displayed portion of the content. In the top part of Fig. 2, we have represented the user's *viewing frustum* by triangles, which indicate the area within the user's viewport. Given these users i and j at time t with $r_t^i \gg r_t^j$, the latter, who is very close to the object, will visualise a very focused and detailed part of it; conversely, user i is pointing to the same area but from further distance, thus she/he will experience the content differently. Thus, 6-DoF navigation trajectories cannot be merely represented only by time and viewport's center position, as the point of origin (*i.e.*, user position) is also needed. To take into account these differences, as shown in Fig.2, we define the spatio-temporal trajectory for a 6-DoF user i as $\{(x_1^i, p_1^i, r_1^i), (x_2^i, p_2^i, r_2^i), \ldots, (x_n^i, p_n^i, r_n^i)\}$. In addition to the viewport's center p_t^i projected on the displayed volumetric object, there are also x_t^i which represents the spatial coordinates (*i.e.*, [x,y,z]) of the user in the VR environment and r_t^i the distance between user and volumetric object.



(a) Users' position on the VR floor (b) Couple 1 (c) Couple 2 (d) Couple 3

Fig. 3. Comparison between significant couples of users navigating in PC3 (*Red and black*).

3. USER TRAJECTORY ANALYSIS IN 6-DOF

3.1. Methodology

We based our investigations on a publicly available dataset of dynamic Point Clouds (PC) presented in [9]. The dataset is a collection of navigation trajectories from 26 participants who experienced 4 different dynamic sequences named, *Long dress* (PC1), *Loot* (PC2), *Red and black* (PC3), *Soldier* (PC4).

We assume that two generic users i and j of the dataset are placed at given time t in positions x_t^i and x_t^j, respectively. Given the nature of the experiment, similarly to what is shown in Fig.2, a single object of interest was placed in the VR scene, and users were instructed to focus on the volumetric content for the duration of the session. Therefore, their viewport's centers can be projected at any given time on the volumetric object in p_t^i and p_t^j, respectively. We define $\mathcal{S}_t^i$ and $\mathcal{S}_t^j$ as the set of points of the volumetric content falling within the viewing frustum cast by users i and j. Then, we denote the overlap set by $\mathcal{S}_t^{i\cap j}$, defined as the portion of points displayed by both users. Equipped with the above notation, we can now introduce a key metric for the analysis: the *overlap ratio* $O^{i,j}$. The latter is defined as the cardinality of the overlap set, normalised by the cardinality of the set containing all points of the PC visualised by both users. Namely $O_t^{i,j} = |\mathcal{S}_t^{i\cap j}|/|\mathcal{S}_t^i \cup \mathcal{S}_t^i|$. The higher is the overlap ratio, the higher is the similarity between users, and vice versa. To verify if such overlap can be substituted with the distance between the two viewport centers $D(p_t^i, p_t^j)$, as shown in [12], we consider 4 different distance metrics to take into account the heterogeneous shape of the PCs: the euclidean distance between users' position in the space (L_x^2), and the distance between the viewport centres projected on the volumetric content in terms of euclidean (L_p^2), geodesic (G_p), and cityblock distance (L_p^1). In particular, geodesic distance is the shortest arc length connecting the points on a sphere, while cityblock evaluates the absolute differences between coordinates.

3.2. Distance as a proxy for overlap?

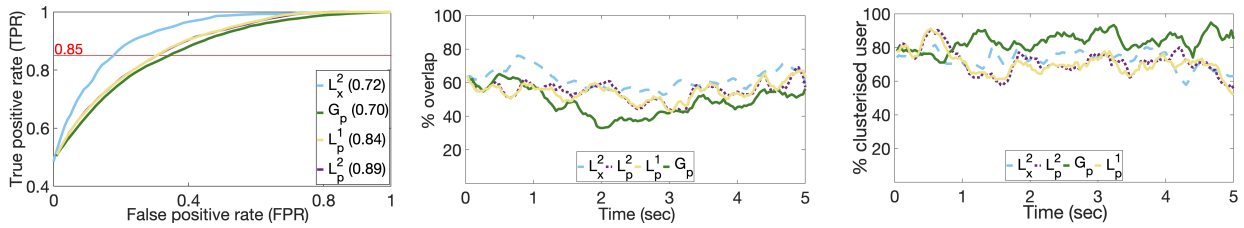
We conducted a first analysis of the relationship between viewport overlap and distance between the user and the volumetric object, by studying three couples of users with different behaviour. This difference lies mainly in the user position. In more details, we considered the following pair of users: **Couple 1:** users i and j sharing a similar position at a small distance from the object ($||r_i - r_j|| < 1\ ; r_i, r_j \ll 1$); **Couple 2:** users i and j sharing a similar position at a large distance from the object ($||r_i - r_j|| < 1\ ; r_i, r_j \gg 1$); **Couple 3:** user i (j) close to (far from) the object ($||r_i - r_j|| > 1\ ;\ r_i \ll 1,\ r_j \gg 1$). Figure 3 (a) depicts the spatial position over time of the selected users' couples (given by their HMD position) with respect to the centroid of the volumetric content in the sequence PC3. Fig. 3 (b-d) compare the viewport overlap over time (expressed in percentage) for each couple ($O_{i,j}$, blue solid line), which represents our ground truth information, versus their distance $D(i,j)$ for the four different distance metrics described in the previous subsection. When users share a similar position (Fig. 3 (b-c)), the correlation between pairwise overlap and distance metrics is quite evident (high overlap, low distance), especially when geodesic distance is considered and users are close to the object. Conversely, the euclidean distance between users is not so informative since is almost flat. In the context of the third couple, the overlap is negligible (given the quite different positions of the users from the object), but the distance metrics fail in capturing this behaviour. Finally, L_p^2 and L_p^1 work similarly in all cases, even though the overlap is substantially different (high in subfigure (b) and (c), very low in (d)). Only the geodesic distance between two viewport centres seems to be much higher compared with the previous couples.

3.3. Distance to assess users' similarity?

After showing that the distance metric does not perfectly replicate the overlap behaviour, we now show why this is a fundamental problem when studying user behaviour. We do so by looking at user similarities via clustering techniques. We use the clique-based clustering proposed in [12] to identify users that are attending the same portion of the content. The clustering algorithm identifies cliques of users all connected within a graph. This graph is built as follows: users are neighbours if their distance is below a given threshold. If the distance is a reliable proxy for the viewport overlap, this clustering technique ensures to identify the largest cluster of users with large viewport overlap. To implement this clustering, the first step is to identify the distance threshold value. As in [12], we empirically evaluate the Receiver Operating Characteristic (ROC) curves per each analysed distance metric and select the best value. In detail, we assumed that two users are attending the same portion of content if their

	PC 1				PC 2				PC 3				PC 4			
	L_x^2	L_p^2	L_p^1	G_p	L_x^2	L_p^2	L_p^1	G_p	L_x^2	L_p^2	L_p^1	G_p	L_x^2	L_p^2	L_p^1	G_p
Mean N. Tot Clusters	9.63	8.7	8.7	**6.2**	10.9	7.14	7.14	**6.76**	10.05	8.81	8.85	**6.49**	10.91	9.61	9.54	**7.19**
Mean N. Single Cluster (cl. = 1 user)	3.85	3.73	3.73	**1.90**	5.03	2.87	2.97	**2.20**	4.18	3.61	3.60	**1.92**	4.77	3.97	3.98	**2.23**
Mean Overlap within Cl. (cl. >2 user)	**62.84%**	59.73 %	59.59 %	49.31 %	**57.00 %**	40.19 %	40.01 %	42.05 %	**62.00 %**	55.04 %	54.62 %	48.48 %	**61.41 %**	54.51 %	55.19 %	46.95 %
Mean Clustered Population (cl. >2 user)	73.60 %	72.49 %	72.99 %	**85.95 %**	66.27 %	78.44 %	78.40 %	**78.96 %**	72.67 %	71.41 %	70.72 %	**83.41 %**	67.90 %	71.83 %	72.72 %	**84.22 %**

Table 1. Spherical clustering analysis over time, based on the different distance metrics and per each video content.



(a) ROC curves per distance metrics. In the legend, threshold values for the spherical clustering.

(b) Mean overlap over time (Clusters $>$ 2 users)

(c) Mean clustered user over time (Clusters $>$ 2 users)

Fig. 4. Spherical clustering results over time per sequence PC3 (*Red and black*).

viewports overlap by at least 80% of their total viewed area; we then computed the ROCs curves in Fig.4 (a) based on all users' navigation trajectories in the dataset. We selected threshold values to have a probability to correctly classify an event (*i.e.*, True Positive Rate (TPR)) equal to 0.85. In the figure, the selected values of threshold per metrics are shown in the legend. All the distance metrics achieve the selected TPR with False Negative Rate (FNR) values smaller than 0.4. Considering that FNR should ideally be minimised, the results confirm the validity of the chosen threshold. Using the selected values, we applied the spherical clustering at each content frame. To avoid misleading results with clusters composed of a single user, we only consider clusters composed of more than 2 users. At each frame, we evaluated the viewport overlap among all users within the same cluster and averaged across clusters. Fig. 4 (b) shows this mean as a function of the time frame for the four distance metrics under consideration. In Fig. 4 (c), instead, we measure how large clusters are on average. We depict this by plotting the percentage of users falling within each cluster (averaged over all clusters) as a function of time. We plot this for each of the four distance metrics considered. We observe that all metrics reach an average of viewport overlap within clusters between $40\% - 60\%$. Even if clusters based on L_x^2 seem to reach a higher overlap ratio within the same cluster, it is also relevant to notice that part of the user population is not covered, since they fall in small clusters (with less than 2 users). The percent of users took into account is indeed around 70 of the entire population (Fig. 4 (c)). On the contrary, clustering based on the geodesic distance between viewport centres (L_p^2) finds larger clusters but less meaningful ones as it leads to a smaller mean overlap ratio. A global view of the results is offered in Table 1, which provides results (averaged over time) for all the sequences in the dataset. Results in the table confirm the previously observed trend: clusters based on the geodesic distance between viewport centres (G_p) are able to identify consistent groups of users, while those based on the euclidean distance between users (L_x^2) perform better in terms of viewport overlap. Here, the first limitation of the metrics currently available to analyse users behaviour in 6-DoF: the lack of one metric that can provide highly populated clusters (as we would like to identify mainstream interactivity) with a large overlap ratio between users within clusters (as we need to identify representative clusters). Equally important, despite the metric used, the values of overlap ratio are below 63%. However, we recall that we set a distance threshold value corresponding to 80% overlap. Here the second limitation: current distance metrics are not a reliable proxy for the viewport overlap measure. As a consequence, this paper opens the door to a very new challenge on designing a proper metric to analyse users' behaviour in 6-DoF. The intuition is that this metric will need to consider both user positions (*i.e.*, x_i, x_j) and viewing directions (*i.e.*, p_i, p_j) to efficiently analyse 6-DoF users.

4. CONCLUSION

We have presented a first attempt of behavioural analysis of users while exploring a 6-DoF immersive content, focusing on studying user similarities. The core of the paper highlights the key differences in the interactivity models between 3-DoF and 6-DoF, showing that *i)* the definition of the trajectory is different, *ii)* current metrics fail in capturing similarity among users (in terms of overlap of the displayed content), *iii)* existing clustering methodologies used in 3-DoF cannot be reliably extended to 6-DoF due to the lack of proper metrics. As consequence, we highlight the need to develop new metrics and methodologies to be able to properly analyse user behaviour in 6-DoF. In future work, we will indeed investigate new metrics that better describe user similarity. We will also extend our analysis to other datasets in order to have a more complete overview of user behaviour in a 6-DoF environment.

5. REFERENCES

[1] 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 (TOMM)*, 2020.

[2] J. Van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6-DoF HTTP adaptive streaming through point cloud compression," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019.

[3] F. De Simone, J. Gutiérrez, and P. Le Callet, "Complexity measurement and characterization of 360-degree content," *Electronic Imaging*, 2019.

[4] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017.

[5] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, "Saliency in VR: How do people explore virtual environments?," in *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[6] W. Chen, A. Plancoulaine, N. Férey, D. Touraine, J. Nelson, and P. Bourdot, "6DoF navigation in virtual worlds: comparison of joystick-based and head-controlled paradigms," in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, 2013.

[7] S. Subramanyam, J. Li, I. Viola, and P. Cesar, "Comparing the Quality of Highly Realistic Digital Humans in 3DoF and 6DoF: A Volumetric Video Case Study," in *IEEE Conference on Virtual Reality and 3D User Interfaces*, 2020.

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

[9] 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.

[10] C. Swindells, B. A Po, I. Hajshirmohammadi, B. Corrie, J. Dill, B. Fisher, and K. Booth, "Comparing CAVE, wall, and desktop displays for navigation and wayfinding in complex 3D models," in *IEEE Proceedings Computer Graphics International*, 2004.

[11] C. Christou, A. Tzanavari, K. Herakleous, and C. Poullis, "Navigation in virtual reality: Comparison of gaze-directed and pointing motion control," in *18th mediterranean electrotechnical conference*, 2016.

[12] S. Rossi, F. De Simone, P. Frossard, and L. Toni, "Spherical clustering of users navigating 360 content," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.

[13] S. Rossi and L. Toni, "Understanding user navigation in immersive experience: an information-theoretic analysis," in *Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems*, 2020.