2022 14th International Conference on Quality of Multimedia Experience (QoMEX)

On the Impact of Spatial Rendering on Point Cloud Subjective Visual Quality Assessment

Davi Lazzarotto, Michela Testolina, Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
davi.nachtigalllazzarotto@epfl.ch, michela.testolina@epfl.ch, touradj.ebrahimi@epfl.ch

***Abstract*—Immersive imaging modalities have been receiving growing attention over the last years. In this context, point clouds demonstrated to be a competitive data representation format, mainly due to its compatibility with most acquisition devices such as LiDAR scans and depth cameras. On the other hand, the large associated data volume is a challenge to storage and transmission, resulting in the need for efficient point cloud compression methods. Multiple subjective studies have been conducted to assess the performance of such compression methods, mostly limiting the analysis to flat monitors or virtual and augmented reality headsets. In this paper, we investigate the impact of a novel eye-sensing light field display on several aspects of quality of experience, as well as on the subjective perception of compression artifacts. The two visualization devices have been observed to create distinct user experiences which lead to noticeable differences in subjective opinion. The advantages and disadvantages of each visualization strategy are underlined, based on rigorous statistical analysis. The subjectively annotated dataset is also released in order to foster future research.**

***Index Terms*—point clouds, immersive imaging, perceptual visual quality, subjective quality assessment, eye-sensing light field display**

(a) *CITIUSP* (b) *longdress* (c) *ipanemaCut*
(d) *fruits* (e) *mitch* (f) *wooden_dragon*

Fig. 1: Point cloud contents of the evaluated dataset.

I. Introduction

Novel technologies for displaying three-dimensional (3D) media are becoming increasingly popular, allowing users to interact with digital content in more natural ways than ever before. As a result, different imaging modalities have been assessed using a variety of such visualization devices, in order to enhance the quality of experience. Point clouds have been identified as an effective alternative, representing models as unconnected points in 3D space with associated color and additional attributes. In order to enable its usage in realistic use cases, compression methods are often needed, adding distortion artifacts that may potentially impair the subjective perception. The impact of these added distortions has already been evaluated in several visualization devices, such as flat monitors, virtual reality (VR) headsets, and augmented reality (AR) glasses [1]–[6].

In the last few years, autostereoscopic displays have been raising growing attention due to their capability of representing 3D images and videos in an immersive way without the need for additional devices such as glasses. In this context, Sony recently released the novel Spatial Reality Display, an eye-sensing light field display (ELFD) able to produce a spatial rendering of virtual objects and scenes, creating the perception of a volume protruding from the screen [7]. Such technology is based on a high-speed camera able to identify the position of the user's eyes, a real-time light field rendering system able to determine and render the correct view based on this position, and a novel eye-sensing-based lenticular lenses system. This technology has drawn growing interest in the fields of product designing, architecture, and entertainment [8], and promising results in the field of education [9] thanks to the high quality and realism of the rendered content.

While the interest in this type of visualization device is rapidly growing, to date, no previous effort was conducted to collect subjective visual quality scores on autostereoscopic displays. In this paper, we present the results of a subjective experiment assessing the quality of compressed point clouds, which was conducted on both the novel Sony's Spatial Reality Display and on a flat monitor. The scores obtained in these two different settings are compared in rigorous statistical analysis, yielding valuable insights for future research on autostereoscopic displays, which are likely to become increasingly pop-

This work was supported by the Swiss National Foundation for Scientific Research (SNSF) under the grant number 200021-178854.

978-1-6654-8794-8/22/$31.00

ular for the rendering of 3D immersive content. Additionally, the results of a survey on the quality of experience conducted following the subjective experiment are reported. The collected subjective scores as well as the results of the evaluation survey and the employed visualization framework are made publicly available to facilitate further research on the topic [1].

II. Previous work

The increased interest in point clouds as an immersive imaging modality has fostered the research on technologies for compression. Octrees are one of the several studied data structures for this purpose, being used to encode the geometry of both static and dynamic [10] point clouds by representing points as leaf nodes in a tree generated with a recursive partition of the space. Other methods [11] compress both geometry and color attributes by projecting the points into multiple views and generating two-dimensional maps with color and distance. These maps are then encoded using well-established video coding standards. Such technologies were adopted by MPEG in their coding standards, with the Geometry-based Point Cloud Compression (G-PCC) [12] leveraging octree-based coding for geometry and Video-based Point Cloud Compression (V-PCC) [13] being built from projection-based methods. For color compression, graph Fourier transforms [14] and region-adaptive hierarchical transforms [15] have been explored to encode attributes on a pre-defined topology, the latter being also added to the G-PCC compression standard.

Recent years have witnessed the rise of compression methods based on deep learning architectures. Autoencoders with three-dimensional convolutional layers were employed [16] to compress the geometry of points lying on a uniformly quantized grid represented as occupancy maps. Other authors leveraged architectures able to handle inputs directly in the point domain [17], without the need for prior voxelization. Efforts have also been devoted to compressing color attributes with deep learning, either with an extension of convolutional layers [18], learning-based operations that map colors into 2D maps [19] or learned volumetric functions [20].

The impact of compression artifacts on subjective perception was studied in several experiments. In [21], a comprehensive study was conducted in two laboratories using colored point clouds distorted with the two MPEG compression standards. Recently, a framework for crowdsourcing-based subjective visual quality assessment of models compressed with both conventional and learning-based codecs was proposed in [22]. While most of the effort conducted thus far on the visual quality assessment of compressed point clouds rendered the content on flat monitors, some studies evaluated subjective perception on alternative visualization devices. In [1], [2], subjects wearing VR headsets rated with 6 degrees of freedom point clouds compressed with G-PCC and V-PCC, respectively. The authors of [3] went further and conducted a similar evaluation with both flat monitors and VR headsets, identifying only small effects from the viewing devices on the final scores. [4] reported a novel methodology for point cloud subjective visual quality assessment on augmented reality glasses, targeting geometry-only point clouds corrupted by noise and compression-like artifacts. The same authors successively analyzed the impact of the different subjective methodologies for geometry-only point cloud quality assessment, i.e. interactive subjective quality assessment on a flat monitor and on a head-mounted display in an augmented reality scenario [5]. A larger confidence interval on the data collected using the AR approach was observed, probably caused by the variable level of interaction or by the increased level of discomfort caused by the head-mounted device. Nevertheless, a strong correlation was observed between the two experiment methodologies.

Despite all these efforts using multiple settings, no subjective experiments have been conducted so far on autostereoscopic displays using point clouds. Although previous works already studied different methods used to produce depth perception without external glasses [23], this is the first study to evaluate compressed 3D models using spatial rendering.

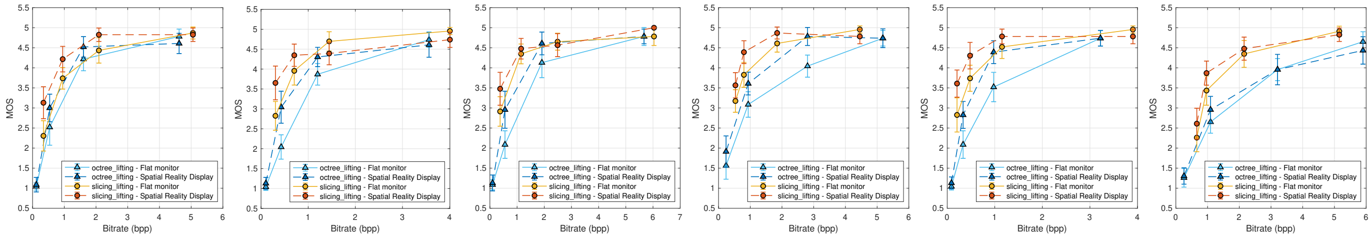
III. Subjective Experiment

A. Experiment setup

The subjective visual scores were collected in a controlled environment setup, where the subjects were asked to assess the visual impairment between pairs of point clouds on both a flat monitor and on an ELFD. In order to prevent excessive tiredness of the subjects, the experiment was organized over two consequent days, where one monitor was randomly assigned to each subject during the first session and, upon completion of the experiment, they were asked to return the consecutive day to complete the experiment on the second monitor. At the end of the second evaluation session, a short survey, designed using [24], [25] as baseline and adapted for the specific target of the paper, was conducted to assess the quality of experience. Specifically, the beginning of the survey contained questions on the users' previous experience with 3D data, including the estimated number of times they experienced immersive contents and the utilized type of 3D devices. In the subsequent questions, the subjects rated the overall immersion level, involvement in the experiment, and quality of experience on a 5-level discrete rating scale. Successively, a number of questions on the naturalness of the displayed contents were presented. The subjects then evaluated the discomfort experienced during the experiment on both display devices, including the indicative time when the discomfort appeared and the kind and strength of the sensation. The final question inquired the preferred display type for future experiments.

The simultaneous Double-Stimulus Impairment Scale (DSIS) experiment with a 5-scale rating and hidden reference was chosen for the experiment. The subjects were asked to rate the impairment between the original and the distorted point clouds on the following scale: "5 - Imperceptible", "4 - Perceptible, but not annoying", "3 - Slightly annoying", "2 - Annoying" and "1 - Very annoying". The two models were displayed side by side, with the position of the original being randomly set to the left or right side of the screen

[1]https://www.epfl.ch/labs/mmspg/downloads/sr-pcd/



(a) *CITIUSP* (b) *longdress* (c) *ipanemaCut* (d) *fruits* (e) *mitch* (f) *wooden_dragon*

Fig. 2: Content-based rate distortion plots.

and remaining the same during the entire session. Before the experiment, instructions were given in written and oral form, and a short training session was conducted to get the subjects acquainted with the grading scale. All subjects rated the entire compressed dataset in random order, and care was taken to not display the same content consecutively. Four dummy stimuli were included at the beginning of the experiment and their scores were excluded from the analysis. At the end of the experiment, the original stimuli were displayed once more in a single stimulus fashion, where the subjects were asked to assess the naturalness of the displayed content on a scale from 1 (*very artificial*) to 5 (*very natural*).

Two different viewing conditions were arranged for the different displays, specifically: (1) as a *flat monitor*, a DELL UltraSharp U3219Q with a native resolution of 3840 x 2160 pixels and 31.5 inches diagonal size was used. The viewing conditions were selected according to ITU-R Rec. BT.500 [26], i.e. the brightness of the monitor was set to 120 cd/m2 with a D65 white point profile, the viewing distance was initially set to 35 cm proportionally to the height of the rendered point clouds, and the ambient light measured in the proximity of the monitor was set to 15 lux; (2) as an *ELFD*, the Sony's Spatial Reality Display with a resolution of 3840 x 2160 pixels and of 15.6 inches diagonal size was used. The viewing conditions were selected according to the recommendations from the manufacturer [8], i.e. the viewing distance was initially set to 30 cm, while the ambient light measured in the proximity of the monitor was set to 100 lux.

The subjects were instructed to explore the viewing distance that was most comfortable for them, while staying in the proximity of the distance they were instructed to keep. Both monitors and lights were switched on at least 30 minutes prior to the experiment to avoid excessive fluctuations of the reported values. The brightness measurements were collected through an X-Rite i1 Display Pro calibration device.

Prior to the first evaluation session, the subjects were asked to take multiple vision tests, i.e. Snellen visual acuity test, Ishihara color vision test, VT-04 test for fine stereopsis, and VT-07 test for dynamic stereopsis, as recommended in ITU-R Rec. BT.500 [26]. Three subjects presented vision capabilities below the average in one or more tests, and therefore their scores were excluded from the analysis. Accordingly, the scores from a total of 23 subjects were collected and used for the experiment, being 11 females and 12 males. The average and median age of the subjects are respectively 21.35 and 21 years, while the minimum and maximum age are respectively 18 and 25 years.

B. Test Material

The six point clouds presented to the subjects during the experiment were selected to be representative of content from relevant use cases. The test models *longdress* [27] and *mitch* were recruited from the MPEG repository and depict full-body human figures. *CITISUP* and *ipanemaCut* were chosen from the University of São Paulo dataset [28] to represent large-scale outdoor scenes. Finally, *wooden_dragon* and *fruits* were chosen as two models of small objects. These last two point clouds were generated using photogrammetry algorithms taking as input multiple pictures of these objects from different angles captured with a mobile device. The entire dataset can be observed in Figure 1.

The original test models were distorted using two compression strategies at four quality levels. In order to assess a range of distortions representative of current compression techniques, both a conventional standard and a recent learning-based method were selected. The first employed codec was the MPEG coding standard G-PCC [12] (Geometry-based Point Cloud Compression). The geometry was compressed using the octree module while the lifting algorithm was used for color coding, using version 13 of the open source implementation. The second compression method was selected from [16], and the implementation released with the paper was employed. Since this learning-based algorithm is only capable of compressing geometric data, it was combined with the lifting module of G-PCC in order to encode color attributes as well. Both compression strategies are henceforth referred to as *octree_lifting* and *slicing_lifting*.

Since both geometry coding modules can only encode models with quantized geometric coordinates, all point clouds were voxelized with bit depth 10 prior to compression. The original and distorted point clouds were also translated to have their centroid in line with the vertical axis that goes through the origin. *CITISUP* and *ipanemaCut* were additionally rotated to have the ground aligned with the horizontal plane. The *loot* [27] model was additionally selected for the training phase and compressed with three different levels using *octree_lifting*.

TABLE I: Performance indexes computed over the entire dataset.

Fitting	PLCC	SROCC	RMSE	OR	CE	UE	OE	CD	FR	FD	FT
Flat monitor experiment as ground truth											
No fitting	0.963	0.909	0.424	0.574	90.7%	0%	9.3%	86.2%	0%	1.3%	12.5%
Linear	0.963	0.909	0.343	0.593	100%	0%	0%	87.0%	0%	2.4%	10.6%
Cubic	0.978	0.909	0.265	0.370	100%	0%	0%	88.1%	0%	2.9%	8.9%
Eye-Sensing Light Field experiment as ground truth											
No fitting	0.963	0.909	0.424	0.5	90.7%	9.3%	0%	86.2%	0%	12.5%	1.3%
Linear	0.963	0.909	0.308	0.463	100%	0%	0%	87.0%	0%	10.6%	2.4%
Cubic	0.988	0.909	0.176	0.167	100%	0%	0%	93.4%	0%	3.1%	3.5%

C. Visualization Framework

The framework for the visualization and rating of the point cloud models was created using Unity. A virtual scene was created as an initially empty space where a large ground plane with dark gray tonality was positioned and a camera was placed above the plane, facing down with an angle of 20 degrees. Both the distorted and reference point clouds were scaled by a 0.001 factor and placed with their bases touching the ground plane, at the same distance from the observer and with an equal gap from the center of the screen. As an exception, *CITIUSP* and *ipanemaCut* were scaled by a factor of 0.0013, and were placed higher above the ground, in order to ensure better inspection. A floating text box would indicate at all times the distorted and reference models. The point clouds were rendered using the Pcx package [2], and each point was rendered as a disk with constant size determined individually for each reference and distorted point cloud in order to result in a visually watertight object.

During the subjective inspection, the models were rotated through 360° for 12 seconds around their vertical axis to reveal the most relevant viewpoints to the subjects, similarly to methods adopted in previous passive subjective studies for point clouds [6]. Afterward, following the recommendation in [25], the framework would move to the voting interface where the subjects would rate the stimulus without the possibility of further inspection. Thereby, all the subjects are exposed to all stimuli for the same amount of time, having consequently an identical experience. A similar protocol was adopted for the naturalness evaluation section, with the single model being displayed in the center of the screen.

IV. Statistical Analysis

The collected subjective scores were grouped into two different experiments according to the visualization device. The outlier detection algorithm from ITU-R recommendation BT.500 [26] was separately applied to each experiment, in order to identify and exclude scores from subjects with deviating behavior. The mean opinion scores (MOS) and 95% confidence intervals (CI) were subsequently computed, assuming a t-Student distribution. The stimuli used for the naturalness were also considered in the computation of the MOS and CI. These values were then examined through two different perspectives: first, the subjective opinions on each compression method were plotted against the bitrate in order to assess the performance of both codecs; moreover, the scores from different experiments were compared to each other in order to evaluate the impact of the visualization device on the subjective opinions.

[2]https://github.com/keijiro/Pcx

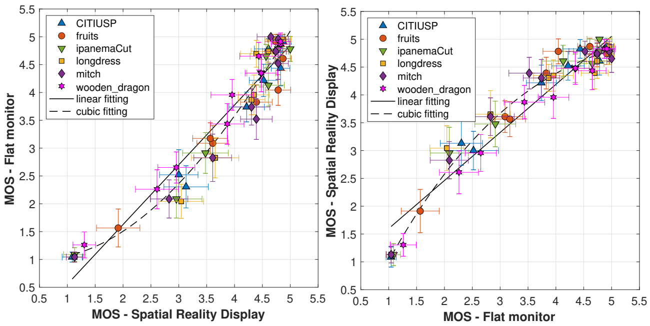


Fig. 3: Scatter plots of MOS collected on the two visualization devices and corresponding CI.

In order to evaluate linearity, monotonicity, accuracy and consistency, four performance indexes were computed between the MOS from both experiments: the Pearson linear correlation coefficient (PLCC), the Spearman rank order correlation coefficient (SROCC), the root-mean-square error (RMSE) and the outlier ratio (OR). The correct estimation (CE), over-estimation (OE) and under-estimation (UE) percentages were also obtained according to [29] to evaluate the statistical equivalence between MOS results. Moreover, the percentage of correct decision (CD), false ranking (FR), false differentiation (FD), and false tie (FT) was obtained for all pairs of distorted stimuli following recommendation ITU-T J.149 [30]. Linear and cubic fitting were employed, following Recommendation ITU-T P.1401 [29], and scores from both experiments were used as ground truth for the computation of all above metrics.

Finally, a multi-way ANOVA with repeated measures is performed after excluding scores given to the reference point clouds. For each stimulus and subject, the scores obtained with both rendering devices are regarded as repeated measures. The content, compression method and rate level are regarded as independent categorical variables. The p-value for the rendering device as an influencing factor is computed, as well as the p-values corresponding to the interactions between the rendering device and each of the independent variables.

V. Results and Discussion

The MOS and CI values of each stimulus against corresponding bitrates are portrayed in Figure 2. It is clear that, regardless of the display, *slicing_lifting* has in general better performance than *octree_lifting*, which is mainly observed at low and mid-range bitrates. This difference is smaller for MOS values at bitrates corresponding to transparent quality, or for the *CITIUSP* point cloud model, where there is no clear distinction between the performance of both codecs. Overall, we can safely state that the learning-based codec outperforms the G-PCC standard, endorsing the recent research trends toward such techniques for compression applications.

Another observation regards the MOS values for the ELFD, which are on average higher than for the flat monitor. For some stimuli, the distance between the MOS values is so high that there is no overlap between the confidence intervals. This trend can also be observed in Figure 3, where scatter plots are presented for the MOS and CI values using both the flat monitor and the eye-sensing light field display experiments as the ground truth, along with the corresponding linear and cubic fitting curves. The plots underline the tendency of higher scores for the ELFD experiment, mainly visible at mid-range quality levels. One likely reason for this observed difference is the size of the rendered content, which is much smaller in the ELFD. The compression artifacts are thus harder to discern, and the overall scores are higher. Another potential factor influencing the overall scores might be the pixel density expressed in Pixels-Per-Degree (PPD). However, since the PPD value depends on the viewing distance and since the subjects were free to adjust to their preferred position, the influence of this factor cannot be accurately estimated and further studies are required to bring insights on this regard.

The confidence interval is also found 6.5% higher for the ELFD, indicating that the subjects were less in agreement when rating stimuli on this device. This behavior can be explained by the higher level of discomfort generated during this experiment, as demonstrated by the results of the final survey presented at the end of this section. Additionally, most subjects had never previously experienced displays with spatial rendering, and were therefore less familiar with this form of visualization.

The performance indexes reported in Table I suggest that the scores from the two experiments were highly correlated. Even without fitting, the MOS values for around 90% of the stimuli could be correctly estimated from either experiment, while the remaining consisted in under-estimations by the flat monitor experiment, confirming that the scores for these stimuli were indeed higher with the ELFD. The increase in the CD percentage with cubic fitting also suggests that the relationship between the MOS from both experiments is not linear, which is also supported by the scatter plots from Figure 3.

The results for the ANOVA with repeated measures are presented in Table III. The first p-value confirms that the rendering device is a significant influencing factor on the scores, which was already suggested in Figures 2 and 3. A significant interaction is also observed between rendering device and rate level with low p-value, revealing differences in the influence of the display on the scores between levels. Although scores are higher for the ELFD at lower levels, and a crossing between curves is often present between the third and the fourth level, where the MOS is higher with the flat monitor. Moreover, the average of the scores given to the reference was higher in the flat monitor (4.93) than in the ELFD (4.79), indicating that subjects could discriminate more clearly in the flat monitor between point clouds where perceptible distortion was present versus point clouds with transparent quality. Finally, the ANOVA also reveals that the difference between scores of both experiments varies according to the content. The average MOS difference between experiments is lowest for *wooden_dragon* (0.12), this value being higher than 0.26 for all the remaining models. A possible explanation is that the color values in this point cloud are more uniform, but its geometric shape has more fine details and large interesting features. In such case, the advanced depth perception made possible in the ELFD might have compensated for the smaller screen size to help subjects distinguish the added distortion. However, further studies are needed to evaluate which types of content are able to better take advantage of spatial rendering.

TABLE II: Average naturalness scores per content.

Display	*CITIUSP*	*fruits*	*ipanemaCut*	*longdress*	*mitch*	*wooden_dragon*
Flat	3.61 ±0.39	4.35 ±0.31	3.65 ±0.56	4.04 ±0.40	4.39 ±0.34	4.52 ±0.22
ELFD	4.13 ±0.40	4.61 ±0.25	4.13 ±0.35	4.17 ±0.38	4.13 ±0.38	4.65 ±0.34

TABLE III: Results of the ANOVA with repeated measures.

Source	DF	F	p
Display	1	140.48	<0.001
Content:Display	5	2.495	0.029
Compression method:Display	1	0.121	0.728
Rate level:Display	3	31.338	<0.001
Error	1094		

In order to evaluate the naturalness of the rendering for each uncompressed point cloud, the average scores and confidence intervals for the final rating questions are presented in Table II. Due to the increased immersion level, the ELFD achieves higher naturalness scores across the majority of the contents. However, the confidence intervals for both rendering devices overlap, indicating that this difference may not be considered as significant in this context. Also, the point cloud contents representing small objects are considered the most natural in both displays, with *wooden_dragon* reaching the highest scores. Point cloud models representing humans received intermediate naturalness scores, which can be explained by the fact that subjects are more sensitive to rendering imperfections on human features. The large-scale scenes are considered the least natural, likely due to the missing points as a consequence of occlusions during acquisition.

The final survey on the quality of experience underlined that most subjects (60.9%) had a previous experience with 3D contents with rendered depth, mainly on passive 3D monitors with glasses (92.9%) or VR devices (71.4%). The survey also revealed that, while the general quality of the experience was higher on the flat monitor, the immersion level and the involvement in the experiment were higher on the Sony's Spatial Reality Display. Additionally, most of the subjects felt encouraged in exploring different viewpoints on the ELFD. Almost half of the subjects (43.5%) reported mild discomfort on the Spatial Reality Display, being headaches, eye strain, and blurred vision the most common symptoms. This might explain the lower overall quality of experience on the ELFD. Lastly, subjects did not express a clear preference for a visualization device, as 30.4% preferred the flat monitor, 34.8% the Spatial Reality Display and another 34.8% declared to prefer both. The complete results of the survey are made available on the project's repository.

VI. Conclusions

In this paper, we describe a point cloud subjective experiment designed to investigate the impact of spatial rendering on human perception when compared to a flat monitor, using an eye-sensing light field display. The results indicate that this novel visualization device allows for enhanced immersion and naturalness, but decreases the discriminatory power of the evaluation regarding compression artifacts, mainly due to its reduced size. Although the scores from both experiments were globally highly correlated, detailed analysis shows that they are statistically different. This study suggests that there are benefits to using spatial rendering in immersive applications, although there are still drawbacks related to experienced discomfort and reduced display size.

References

[1] E. Alexiou, N. Yang, and T. Ebrahimi, "Pointxr: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020, pp. 1–6.

[2] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6dof head-mounted display," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4630–4644, 2021.

[3] I. Viola, S. Subramanyam, J. Li, and P. Cesar, "On the impact of vr assessment on the quality of experience of highly realistic digital humans," *arXiv preprint arXiv:2201.07701*, 2022.

[4] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2017, pp. 1–6.

[5] E. Alexiou and T. Ebrahimi, "Impact of visualisation strategy for subjective quality assessment of point clouds," in *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2018, pp. 1–6.

[6] S. Perry, H. P. Cong, L. A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, and T. Ebrahimi, "Quality evaluation of static point clouds encoded using mpeg codecs," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3428–3432.

[7] K. Aoyama, K. Yokoyama, T. Yano, and Y. Nakahata, "Eye-sensing light field display for spatial reality reproduction," in *SID Symposium Digest of Technical Papers*, vol. 52, no. 1. Wiley Online Library, 2021, pp. 669–672.

[8] "Spatial Reality Display ELF-SR1 Technical Background," https://www.sony.com/en/SonyInfo/sony_ai/siggraph2021/assets/img/event_session04-01.pdf, accessed: 2022-03-12.

[9] T. Itamiya, M. To, T. Oguchi, S. Fuchida, M. Matsuo, I. Hasegawa, H. Kawana, and K. Kimoto, "A novel anatomy education method using a spatial reality display capable of stereoscopic imaging with the naked eye," *Applied Sciences*, vol. 11, no. 16, p. 7323, 2021.

[10] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, 2016.

[11] K. Mammou, A. M. Tourapis, D. Singer, and Y. Su, "Video-based and hierarchical approaches point cloud compression," ISO/IEC JTC1/SC29/WG11 Doc. M41649, Macau, China, Oct. 2017.

[12] MPEG Systems, "Text of ISO/IEC DIS 23090-18 Carriage of Geometry-based Point Cloud Compression Data," ISO/IEC JTC1/SC29/WG03 Doc. N0075, Nov. 2020.

[13] MPEG 3D Graphics Coding, "Text of ISO/IEC CD 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition," ISO/IEC JTC1/SC29/WG07 Doc. N0003, Nov. 2020.

[14] C. Zhang, D. Florêncio, and C. Loop, "Point cloud attribute compression with graph transform," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2066–2070.

[15] R. L. de Queiroz and P. A. Chou, "Compression of 3d point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.

[16] N. Frank, D. Lazzarotto, and T. Ebrahimi, "Latent space slicing for enhanced entropy modeling in learning-based point cloud geometry compression," in *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[17] L. Wiesmann, A. Milioto, X. Chen, C. Stachniss, and J. Behley, "Deep compression for dense point cloud maps," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2060–2067, 2021.

[18] E. Alexiou, K. Tung, and T. Ebrahimi, "Towards neural network approaches for point cloud compression," in *Applications of Digital Image Processing XLIII*, 08 2020, p. 4.

[19] M. Quach, G. Valenzise, and F. Dufaux, "Folding-based compression of point cloud attributes," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3309–3313.

[20] B. Isik, P. A. Chou, S. J. Hwang, N. Johnston, and G. Toderici, "Lvac: Learned volumetric attribute compression for point clouds using coordinate based networks," *arXiv preprint arXiv:2111.08988*, 2021.

[21] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. De Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.

[22] D. Lazzarotto, E. Alexiou, and T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression," in *IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, no. CONF, 2021.

[23] M. Řeřábek and T. Ebrahimi, "Comparison of 3D portable display restitution techniques based on stereo and motion parallax," in *2012 Fourth International Workshop on Quality of Multimedia Experience*. IEEE, 2012, pp. 80–85.

[24] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[25] ITU-R Rec. P.919, "Subjective test methodologies for 360º video on head-mounted displays," 2020.

[26] ITU-R Rec. BT.500, "Methodologies for the subjective assessment of the quality of television images," 2012.

[27] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies - a voxelized point cloud dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva, Switzerland, Jan. 2017.

[28] "University of São Paulo Point Cloud Dataset [Online], http://uspaulopc.di.ubi.pt/."

[29] ITU-T Rec. P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," 2020.

[30] ITU-T J.149, "Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)," International Telecommunication Union, Mar 2004.