# Will Neural 3D Object Representations be the Silver Bullets for Improving VR Experience in HMDs?

Charlie Hsu
International Bilingual School
National Experimental High School
Hsin-Chu, Taiwan

Yuan-Chun Sun
Institute of Information Systems and Applications
National Tsing Hua University
Hsin-Chu, Taiwan

Kuan-Yu Lee
Department of Computer Science
National Tsing Hua University
Hsin-Chu, Taiwan

Chun-Ying Huang
Department of Computer Science
National Yang Ming Chiao Tung University
Hsin-Chu, Taiwan

***Abstract*—We evaluate the presence and cybersickness scores of Virtual Reality (VR) applications using a Head-Mounted Display (HMD) with different: (i) 3D objects in non-neural and neural representations and (ii) interaction modes in 0-, 3-, and 6-DoF (Degrees-of-Freedom). The presence score indicates how much an HMD user feels being there, while the cybersickness score represents how much an HMD user reports discomfort. These two scores impose crucial impacts on VR experience in HMDs. To the best of our knowledge, this paper is the first work investigating the VR Quality-of-Experience (QoE) of the latest neural 3D object representations, including 3D Gaussian Splatting (3DGS). The results from our user study reveal the benefits of 3DGS, it: (i) delivers the highest presence scores, (ii) achieves the lowest cybersickness scores, (iii) reaches the full frame rates, and (iv) consumes the least storage space. We also find that the 6-DoF interaction mode increases the presence scores and reduces the cybersickness scores, compared to 0- and 3-DoF interaction modes in most cases. The findings from our user study shed some light on future investigations into many other factors of VR QoE in HMDs.**

***Index Terms*—Quality of experience, 3D representation, realism, presence, interaction mode, cybersickness**

## I. Introduction

A recent report indicates that the Virtual Reality (VR) and Augmented Reality (AR) display market is expected to expand from 1.8 in 2023 to 8.2 billion USD in 2028, in which Head-Mounted Display (HMD) hardware and relevant applications represent a significant portion [1]. We collectively refer to these HMDs as VR HMDs, which can be used in entertainment, education, healthcare, manufacturing, aerospace, tourism, heritage, etc. The success of VR (and AR[1]) applications highly relies on good Quality of Experience (QoE), defined as "the degree of delight or annoyance of the user of an application or service" [2], to attract and retain many more VR HMD users.

From the "delight" and "annoyance" aspects, VR HMD users ask for high visual *realism* and to low *cybersickness*, respectively [2], [3]. Here, cybersickness refers to uncomfortable feelings due to motion-sickness in VR environments [3], while realism depends on two key factors [2]: *presence*, referring to the sense of "being there[2]", and *interaction mode*, defining how HMD users interact with the environments. In particular, interaction modes can be described by the number of Degrees-of-Freedom (DoF). 0-DoF is offered by non-interactive VR applications, where an HMD user's head rotation and movements do not affect the rendered viewports. 3- and 6-DoF refer to dynamically rendered viewports based on head rotation and head rotation/movement, respectively, which enable interactive VR applications.

(a) (b) (c) (d)

Fig. 1. Ficus in diverse 3D representations: (i) meshes, (ii) point clouds, (iii) NeRF, and (iv) 3DGS.

Although higher DoFs contribute to higher realism, 3D representations of virtual objects may also affect presence, and therefore realism in VR HMDs. Fig. 1 shows a sample object in four popular 3D representations: (i) 3D meshes, where an object (or scene[3]) is represented by a set of textured triangles, (ii) 3D point clouds, where an object is represented by a set of colored, unconnected points, (iii) Neural Radiance Field (NeRF) [5], where an object is represented by a Multi-Layer

[1]The boundary between modern VR and AR HMDs is getting increasingly blurrier, and we use VR to refer to VR, AR, and Extended Reality (XR) technologies throughout the paper.

[2]Different disciplines may have diverse definitions of presence and even introduce similar yet alternative concepts, which are out of the scope of this paper; interested readers are referred to Goncalves et al. [4] for a complete treatment.

[3]In this paper, we consider each scene consisting of a main 3D object without background. Hence, we use *object* and *scene* interchangeably. QoE evaluations of a complex 3D scene with multiple objects are our future works.

Perception (MLP) neural network that maps 3D position and 2D orientation (of an HMD) to RGB color and volume density, and (iv) 3D Gaussian Splatting (3DGS) [6], where an object is represented by a set of 3D Gaussians that are composed of 3D positions in coordinates, rotation in covariance matrices, color in Spherical Harmonic (SH) coefficients, and opacity. Among these representations, 3D meshes and 3D point clouds are traditional, non-neural representations, while NeRF and 3DGS are neural representations.

In this paper, we carry out a user study to quantify the presence and cybersickness scores of VR HMD users with different 3D representations and interaction modes, which has never been done in the literature. Our user study adopts well-established QoE questionnaires [7], [8], [9], including both coarse- and fine-grained questions. Particularly, we consider two coarse-grained questions on presence and cybersickness from Tran et al. [7]. We also consider 6 fine-grained questions on presence from Usoh et al. [8] and 9 fine-grained questions on cybersickness from Singla et al. [9]. All these 17 questions are given in Sec. III-B. Our user study revealed that: (i) 3DGS delivers the highest presence scores, (ii) 3DGS achieves the lowest cybersickness scores, (iii) 6-DoF increases the presence scores and reduces the cybersickness scores in most cases, (iv) NeRF and 3DGS both achieve the highest frame rate, and (v) 3DGS consumes the least storage space.

## II. Related Work

In this section, we survey representative QoE evaluations of different 3D object representations. Due to the space limit, we could not provide an exhaustive survey; interested readers are referred to Alexiou et al. [3].

**QoE evaluations of homogeneous non-neural representations.** Most user studies were conducted with homogeneous non-neural representations, which can be classified into non-interactive [10], [11], [12], [13], [14] and interactive [15], [16], [17], [18], [19], [20] ones. Both 3D meshes [10], [11] and 3D point clouds [12], [13], [14] have been considered in non-interactive QoE evaluations, where user head movements (or other inputs) do not affect the viewport. In an earlier work, Guo et al. [10] produced 2D videos by rotating each 3D object in textured meshes along the vertical axis. Nehme et al. [11] carried out a similar study with 3D meshes in slowly rotated and zoomed-in viewports shown in an HTC HMD. Cruz et al. [12] evaluated the QoE of compressed 3D point clouds to render 2D videos using a fixed camera trajectory, while Alexiou et al. [13] considered colorless point clouds instead. Weil et al. [14] employed crowdsourcing for a user study to assess the subjective quality of videos non-interactively rendered from 3D point clouds. These studies [10], [11], [12], [13], [14] were done with rendered viewpoints in non-interactive mode, making them quite different from our work.

2D monitors have been used for interactive QoE evaluations with 3D meshes [21], [22] and 3D point clouds [19], [20]. Corsini et al. [21] and Lavoue [22] allowed users to control the viewport of a 2D monitor when evaluating the geometry quality degradation of 3D meshes due to compression, downsampling, and watermarking. Alexiou and Ebrahimi [19] studied the negative implications of Gaussian noise and octree pruning on the geometry degradation, while Yang et al. [20] quantified both the geometry and color quality degradation due to Gaussian noise, octree pruning, and downsampling of 3D point clouds. These studies [21], [22], [19], [20] were performed on 2D monitors, which were fundamentally different from our work.

More recently, HMDs have also been used for QoE evaluations with interactive VR applications using 3D meshes [23], [24], [25] and 3D point clouds [15], [16], [17], [18]. For example, Christaki et al. [23] conducted QoE evaluations on texture-less 3D meshes using an HTC HMD, while Gutierrez et al. [24] also carried out similar evaluations on textured 3D meshes using a Microsoft HMD. Damme et al. [25] employed a Meta HMD to study how the camera distance, mean/variance of quality, and content type affect the QoE of 3D meshes built from 3D point clouds. Alexiou et al. [15] and Wu et al. [16] derived the QoE levels of 3D point clouds using HTC HMDs. Furthermore, Subramanyam et al. evaluated point cloud compression algorithms [17] and adaptive streaming algorithms [18] in 6-DoF interaction mode. Different from our work, these papers [23], [24], [25], [15], [16], [17], [18] only considered homogeneous, and non-neural 3D object representations.

**Comparison among heterogeneous non-neural representations.** A few prior arts [26], [27], [28] compared heterogeneous 3D object representations. For example, Zerman et al. [26] compared QoE levels of static 3D meshes compressed by Draco and JPEG against static 3D point clouds compressed by V-PCC and G-PCC. Using rendered videos shown on 2D monitors, their paper revealed that 3D point clouds lead to better QoE at lower bitrates, while 3D meshes result in better QoE at higher bitrates. Cao et al. [27] evaluated dynamic 3D meshes and point clouds using a 2D monitor. Javaheri et al. [28] qualified the QoE levels of static: (i) colored 3D point clouds, (ii) colorless 3D point clouds, and (iii) colorless 3D meshes using 2D monitors. Different from our current paper, these previous works [26], [27], [28] considered non-neural 3D object representations and used 2D monitors in their experiments.

**QoE evaluations of neural 3D object representations.** Neural representations, such as NeRF [5] and 3DGS [6], have not been thoroughly evaluated in terms of QoE. To the best of our knowledge, there were only a couple of studies [29], [30] that evaluated the QoE of NeRF variants. In particular, Liang et al. [29] used natural NeRF scenes with uniformly sampled viewports, where rendered videos are shown on a 2D monitor. While also playing rendered videos on a 2D monitor. Martin et al. [30] considered arbitrary viewports with both natural and synthetic NeRF scenes. Compared to our work, these QoE evaluations [29], [30] did not: (i) employ HMDs, (ii) consider 3- and 6-DoF interactive modes, nor (iii) include the latest 3DGS representation.

**Comparison among heterogeneous interaction modes.** A couple of QoE evaluations [31], [32] compared the implica-

TABLE I
OUR ADOPTED QOE QUESTIONS

| Q. # | Description |
|---|---|
| | **Presence Questions [7], [8]**: 1–5, Higher is More Delightful |
| **PQ0** | How is your assessment about the sense of presence in VR environment? |
| **PQ1** | Please rate your sense of being there. |
| **PQ2** | To what extent where there times during the experience when the scene was the reality for you? |
| **PQ3** | When you think back about your experience, do you think of the scene as images that you saw, or more as somewhere that you visited? |
| **PQ4** | During the time of the experience, which was strongest on the whole, your sense of being in front of the object or being elsewhere? |
| **PQ5** | How similar in terms of the structure of the object is this to the structure of the memory of other objects you have interacted with today? |
| **PQ6** | During the time of the experience, did you often think to yourself that the object was real? |
| | **Cybersickness Questions [7], [9]**: 1–5, Higher is More Annoying |
| **CQ0** | How is the level of dizziness or nausea during in VR experiment? |
| **CQ1** | I felt dizzy. |
| **CQ2** | I had eyestrain. |
| **CQ3** | I had a headache. |
| **CQ4** | I felt fatigued during the experience. |
| **CQ5** | I had difficulty focusing. |
| **CQ6** | I felt general discomfort during the experience. |
| **CQ7** | I had difficulty concentrating. |
| **CQ8** | I experienced blurred vision during the experience. |
| **CQ9** | I felt an fullness of head during the experience. |

tions of the interaction modes on 3D object representations. In particular, Torkhani et al. [31] employed a 2D monitor to evaluate the impacts of 0- versus 6-DoF interaction modes on QoE levels of dynamic 3D meshes. Viola et al. [32] compared the impacts of both: (i) 3- versus 6-DoF and (ii) 2D monitors versus HMDs on human avatars in dynamic 3D point clouds. In contrast to non-neural 3D meshes and point clouds used in these papers [31], [32], our work also considers the implications of 0-, 3-, and 6-DoF interaction modes on QoE with neural representations.

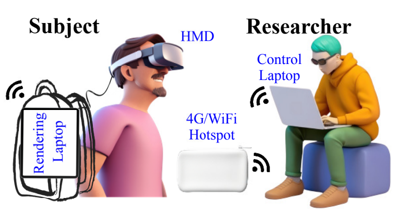



Fig. 2. The testbed used in our QoE evaluations.

## III. USER STUDY DESIGN

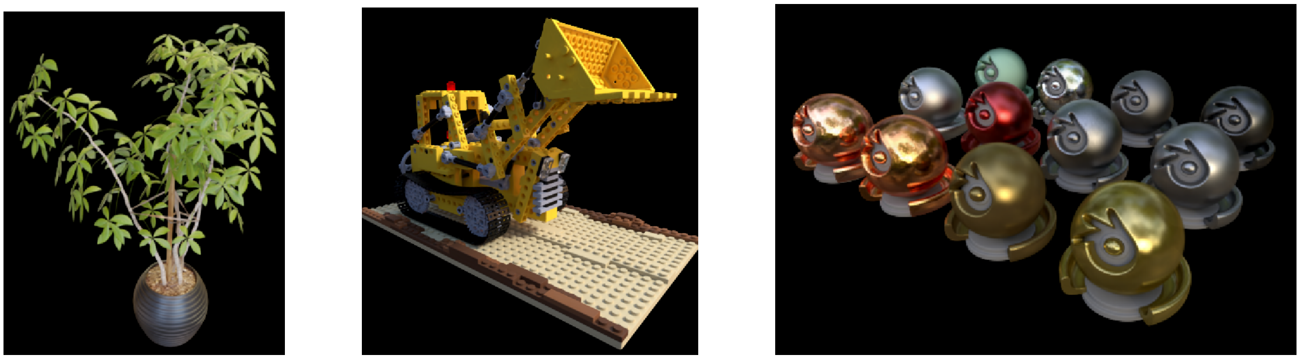
### A. Testbed and Implementations

Fig. 2 shows the testbed for our QoE study, where each subject wears an HMD that is tethered to a rendering laptop inside a backpack. By doing so, the subject could move freely when viewing each 3D object. A researcher supervises the QoE study using a control laptop, which allows him to: (i) remotely access the rendering laptop and (ii) record the subject's verbal feedback to QoE questions. We adopt a 4G WiFi hotspot to interconnect the two laptops and the Internet. Note that only commands to switch among 3D objects and answers to questionnaires are sent over the network, which incur negligible network traffic. In this paper, we use a Meta Quest as our HMD, which has a Qualcomm Snapdragon 835 CPU at 2.45 GHz and 4 GB RAM. We use an ASUS ROG G15 as the rendering laptop, which has an AMD Ryzen 9 5900HS CPU with 32 GB of RAM, and an Nvidia GeForce RTX 3080 GPU with 24 GB of VRAM. They are tethered via a USB-C cable. We use a Dell XPS 13 7390 as the control laptop, which has an Intel i7-10710U CPU at 1.10 GHz and 16 GB of RAM.

We have implemented a virtual scene in Unity version 2022.3.7f1, which could render 3D objects represented in 3D meshes, 3D point clouds, NeRF, and 3DGS. Among too many NeRF and 3DGS variants, we chose to use the MobileNeRF [33] and the original 3DGS [6] for two reasons; they: (i) were proposed recently, (ii) can be rendered in real-time, and (iii) have been ported to Unity. More specifically, we have imported the Unity assets from two GitHub projects [34], [35] for MobileNeRF and 3DGS Unity viewers, respectively. Our virtual scene supports 0-, 3-, and 6-DoF interaction modes, which can be switched by the researcher using hotkeys. In 0-DoF, the viewport is randomly placed within three meters of the object center, and the orientation is fixed toward the object center. We limit the HMD position between the Tropics of Cancer and Capricorn of a sphere centered at the object to avoid rarely observed angles. In 3-DoF interaction mode, the HMD orientation can be changed when a subject rotates his/her head, while in 6-DoF, the HMD position can also be changed when a subject moves his/her head or walks.

### B. Procedure

As subjects feel fatigued in HMDs quickly, we opted for Absolute Category Rating (ACR) ratings between 1 and 5 in our evaluations. In addition to pre-examination questions for demographics, Table I gives our questionnaire, composed of coarse-grained presence and cybersickness questions from Tran et al. [7], which are PQ0 and CQ0. Fine-grained presence questions PQ1–PQ6 come from Usoh et al. [8], where higher numbers represent a more delightful experience. For example, PQ1 asks the subject to rate their sense of being in the scene.



(a) (b) (c)

Fig. 3. Three diverse 3D objects in 3DGS representations: (i) ficus, (ii) lego, and (iii) materials.

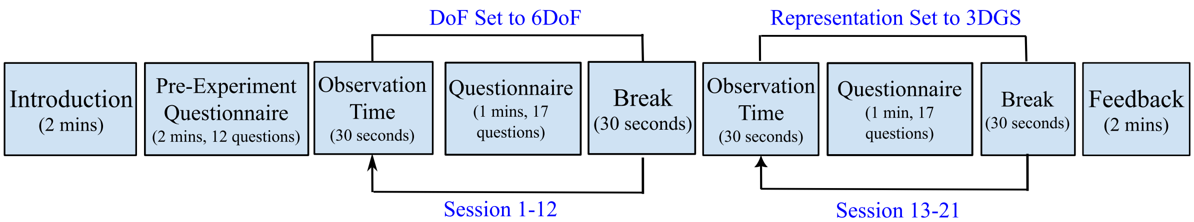



Fig. 4. The procedure of our QoE evaluations.

TABLE II
CONSIDERED 3D OBJECT REPRESENTATIONS

| **Representation** | | **Ficus** | **Lego** | **Materials** |
|---|---|---|---|---|
| **Meshes** | **File Size (MB)** | 680 | 6950 | 4730 |
| | **No. Triangles (k)** | 6426 | 68080 | 45835 |
| **Point Clouds** | **File Size (MB)** | 86 | 953 | 627 |
| | **No. Points (k)** | 6023 | 66626 | 43869 |
| **NeRF** | **File Size (MB)** | 76 | 188 | 180 |
| | **No. Objects** | 16 | 32 | 32 |
| | **No. Texture Images** | 4 | 8 | 8 |
| **3DGS** | **File Size (MB)** | 69 | 77 | 43 |
| | **No. Gaussians (k)** | 303 | 342 | 191 |

A 5 would demonstrate that they feel immersed in the scene, and a 1 would demonstrate that they feel completely alienated from the scene. Fine-grained cybersickness questions CQ1–CQ9 correspond to nine symptoms, largely inspired by Singla et al. [9], where higher numbers demonstrate a more annoying experience. For example, in CQ1, 5 would demonstrate that a subject is extremely dizzy, and a 1 would demonstrate that he/she is not dizzy at all. Notice that we align the coarse-grained cybersickness question CQ0 with CQ1–CQ9 so that 5 indicates very dizzy and 1 indicates absolutely not dizzy, which is the opposite of its original definition [7].

From the 3D objects widely used in recent neural representation papers [5], [33], we chose three challenging ones with diverse characteristics: (i) *ficus*, which contains detailed leaves; (ii) *lego*, which contains many small parts; and (iii) *materials*, which contains metallic balls with extremely diverse textures. Fig. 3 shows these 3D objects in 3DGS representations. Table II gives the basic statistics of the considered objects. To get the 3D mesh representation, we first export the 3D geometry of the objects from a dataset [5], created in Blender [36]. We then bake the texture of each object in the dataset into the texture images to better replicate the colors, materials, and lighting of the dataset. Last, we merge the 3D meshes with these texture images in Unity using UV mapping for 3D meshes. We generate the 3D point clouds from 3D meshes. To achieve sufficient point density, we connect the midpoints of the edges to divide each mesh face into four smaller faces. We repeat this three times to upsample the point cloud before exporting and baking each object. Last, we map the texture images with the 3D meshes and extract the vertices to get 3D point clouds. For NeRF representations, we download the pre-trained models from the official site of MobileNeRF [33]. Finally, we use NeRF's training images to train the 3DGS models ourselves with the default training setting [33] for 30k iterations.

Fig. 4 outlines the procedure of our QoE evaluations on a single subject. During the introduction, the subjects are given an overview of what they will be experiencing during the entire experiment. They are also taught how to use the VR HMD for a comfortable experience. They are then given a quick pre-experiment questionnaire to gather subject demographics, which includes inquiries on age, gender, education, and experience with VR. They also self-report their Snellen and Ishihara test results. All subjects have 20/20 corrected and normal color vision. We then proceed to the first 12 sessions of the QoE, where users are given 30 seconds to observe the

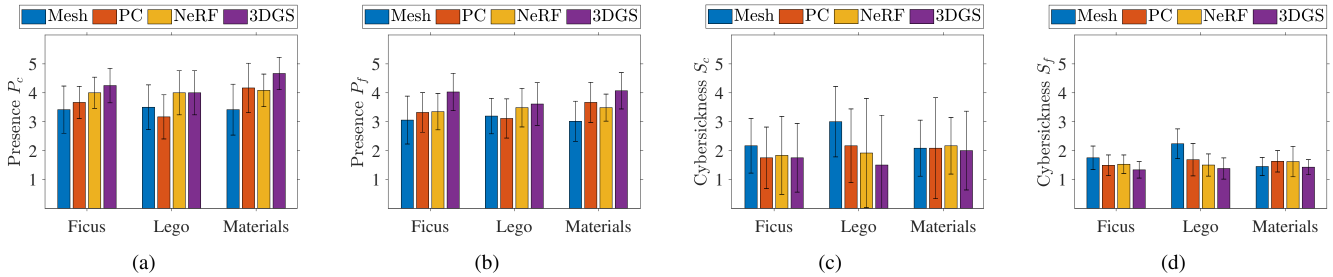



Fig. 5. QoE results from different representations: (a) coarse-grained presence, (b) fine-grained presence, (c) coarse-grained cybersickness, and (d) fine-grained cybersickness.

object, ~1 minute to answer the questions verbally, and up to 30 seconds to opt for a break. The first 12 sessions focus on observing differences across different representations and objects. We follow with the next 9 sessions to observe the differences regarding the changes in interaction modes with different objects. Here, we only consider objects in the 3DGS representations[4] to avoid subject fatigue. Last, we ask subjects for general feedback.

## IV. User Study Analysis

### A. Methodology

We define four QoE scores based on the subjects' answers to the questions in Table I:

1) *Coarse-grained presence score* $P_c$, which is the mean opinion score of PQ0.
2) *Coarse-grained cybersickness score* $S_c$, which is the mean score of CQ0.
3) *Fine-grained presence score* $P_f$, which is computed by $P_f = \sum_{i=1}^{6} \mathrm{PQ}_i/6$ following the SUS (Slater, Usoh, and Steed) mean described in Usoh et al. [8].
4) *Fine-grained cybersickness score* $S_f$, which is inspired by Singla et al. [9], written as $S_f = \mathrm{CF1} + \mathrm{CF1} + \mathrm{CF3}$, as detailed in the following.

Here, CF1, CF2, and CF3 represent *uneasiness*, *visual discomfort*, and *loss of balance*, respectively. They defined as:

$$\mathrm{CF1} = (\mathrm{CQ2} + \mathrm{CQ3} + \mathrm{CQ4} + \mathrm{CQ6} + \mathrm{CQ7} + \mathrm{CQ9})/18; \quad (1)$$

$$\mathrm{CF2} = (\mathrm{CQ5} + \mathrm{CQ8})/6; \quad (2)$$

$$\mathrm{CF3} = \mathrm{CQ1}/3. \quad (3)$$

It is not hard to see that $P_c, S_c, P_f, S_f \in [1, 5]$, where higher $P_c, P_f$ and lower $S_c, S_f$ lead to better VR experience.

We recruited twelve subjects throughout our experimentation, who are between 16 and 54 years old (39.25 years old on average with a standard deviation of 16.73). One-third of the subjects identified themselves as male, while the remaining two-thirds identified themselves as female. Half of the users reported sporadic usage of VR, while the other half have never interacted with VR. We believe the distribution is similar to reality, as VR HMDs have yet to be widely adopted. It took each subject ~40–60 minutes to complete the experiment, including the system setup, introduction, and feedback interview. Each subject receives a gift voucher of about 3.20 USD as compensation for their time.

[4]Later in Sec. IV, our analysis reveals that 3DGS results in the best presence and cybersickness scores, justifying our design decision.

### B. Results

This section reports the average QoE in presence and cybersickness scores with 95% confidence intervals. We first compare the QoE achieved by different 3D object representations, including 3D Meshes (Mesh), 3D Point Clouds (PC), MobileNeRF (NeRF), and 3DGS. We then compare the QoE achieved by diverse interaction modes, including 0-, 3-, and 6-DoF, using the 3DGS representation.

**Neural representations result in higher presence scores.** Figs. 5(a) and 5(b) present the coarse- and fine-grained presence scores of different 3D object representations, respectively. We make the following observations. First, 3DGS always leads to the highest average presence scores compared to NeRF and non-neural representations. Second, NeRF outperforms Mesh and PC with ficus and lego in terms of presence scores; but PC slightly outperforms NeRF with materials. We note that 3DGS does not trade larger file sizes for higher presence scores. Table II indicates that the size of 3DGS is as small as: (i) 23.89% of that of NeRF, (ii) 6.86% of that of PC, and (iii) 0.91% of that of Mesh. We conclude that 3DGS delivers the highest presence scores and yet takes much smaller spaces than NeRF and non-neural representations.

**Neural representations result in lower cybersickness scores.** Figs. 5(c) and 5(d) give the coarse- and fine-grained cybersickness scores of different 3D object representations, respectively. We observe that 3DGS generally leads to the lowest average cybersickness scores between 1.50–2.00 in $S_f$ and 1.33–1.42 in $S_c$, which is quite close to no cybersickness (1). Moreover, as another neural representation, NeRF also leads to slightly slower cybersickness scores, except with materials. We believe the more complex reflections in materials cause cybersickness, which may indirectly cause lower presence scores of NeRF as reported above. In summary, 3DGS achieves the lowest cybersickness scores; NeRF also improves cybersickness scores, compared to non-neural representations.

**Neural 3D object representations achieve the full frame rate.** The hardware of our HMD caps its refresh rate at 72

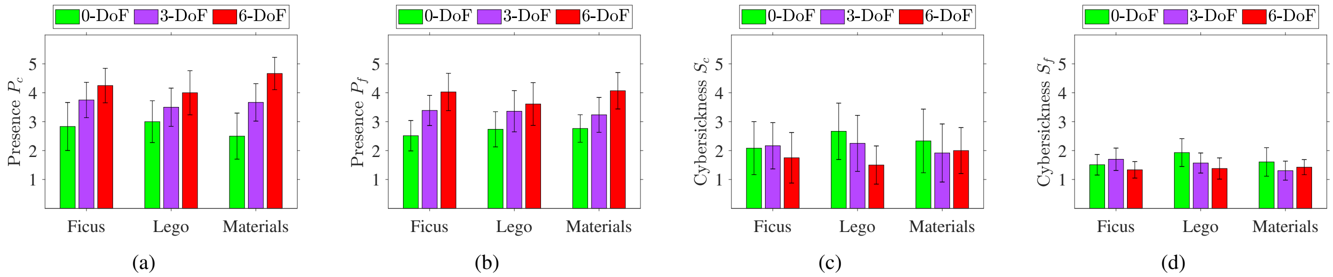



Fig. 6. QoE results from different interaction modes: (a) coarse-grained presence, (b) fine-grained presence, (c) coarse-grained cybersickness, and (d) fine-grained cybersickness.

TABLE III
AVERAGE FRAME RATE IN FRAME PER SECOND (FPS)

| Object | Mesh | PC | NeRF | 3DGS |
|---|---|---|---|---|
| **Ficus** | 70 | 72 | 71 | 72 |
| **Lego** | 1 | 36 | 72 | 72 |
| **Materials** | 9 | 58 | 72 | 72 |

frame-per-second (fps). Table III reports the average frame rates from different representations of individual objects measured from 30-sec sessions. This table shows that both neural representations, NeRF and 3DGS, can achieve full frame rate with all objects, while non-neural representations, Mesh and PC, can only reach full frame rate with the simpler ficus[5].

**Interaction modes with higher DoFs lead to higher presence scores.** Figs. 6(a) and 6(b) depict the coarse- and fine-grained presence scores of diverse interaction modes, respectively. We note that higher DoFs lead to higher presence scores, e.g., 6-DoF results in 4.00–4.66 average coarse-grained presence score across three objects, while 0-DoF only leads to 2.50–3.00. Such difference is not surprising because, in our user study, all subjects realized that: (i) they essentially evaluated images in 0-DoF mode and (ii) their distances to 3D objects were fixed in 3-DoF mode. To summarize, higher DoFs improve the QoE in presence scores.

**Interaction modes with higher DoFs lead to lower cybersickness scores.** Figs. 6(c) and 6(d) show the coarse- and fine-grained cybersickness scores of diverse interaction modes, respectively. We observe that 6-DoF causes lower cybersickness scores in general, e.g., it achieves 1.33–1.42 average fine-grained cybersickness scores across three objects, which is very close to no cybersickness (1). During our user study, we noticed that subjects who rotated their heads more often in 0-DoF mode or walked forward/backward more in 3-DoF mode tended to report higher cybersickness scores. In addition, younger subjects are more resilient to cybersickness, which is in line with earlier work [37]. In our prior work [38], we also observed that cybersickness depends on human factors a lot, which is an interesting future research direction. Overall, although the difference among subjects is clear, 6-DoF does not worsen cybersickness and could mitigate cybersickness in most cases.

[5]For the complexity levels of different 3D objects, please see Table II.

**Subject feedback.** Multiple subjects provided their opinions in feedback interviews. Upon checking the actual representations (unknown to subjects), we have found that:

- Mesh objects often cause lagging, while 3DGS ones lead to the smoothest experience.
- 3DGS objects have the most natural reflections compared to other representations.
- PC objects become visually rough when they are closely inspected.
- Once subjects walk or rotate their heads in 0-DoF interaction mode, they feel dizzy quickly.

These findings are consistent with our analysis of presence and cybersickness scores reported above.

## V. CONCLUSION

In this paper, to better understand the VR experience in HMDs, we designed and conducted a user study to quantify the presence and cybersickness scores with: (i) diverse 3D object representations, including non-neural and neural ones and (ii) different interaction modes, from 0- to 6-DoF. To the best of our knowledge, we are the first to study the VR QoE with cutting-edge NeRF and 3DGS representations using an HMD. Through our analysis of subjects' feedback to QoE questions, we have witnessed the merits of 3DGS: compared to non-neural 3D meshes and point clouds, 3DGS leads to higher frame rates, much higher presence scores, and relatively lower cybersickness scores. NeRF, on the other hand, achieves performance somewhere between the non-neural representations and 3DGS. In addition, we have also validated that the 6-DoF interaction mode is critical for higher presence scores, although its implications on cybersickness are not clear. We believe this is partially because most subjects were not bothered too much by cybersickness throughout our user study.

Our work took the first step in demonstrating that neural 3D representations are promising for improving the QoE of the current upcoming VR applications in HMDs. We hope our findings could stimulate many more follow-up works on the VR QoE of 3D representations in HMDs, as overall QoE is affected by too many system, human, content, and context factors, which cannotlbe thoroughly investigated in a single paper.

## References


[1] Markets and Markets, "AR and VR display market–global forecast to 2028," 2023, https://www.marketsandmarkets.com/Market-Reports/ar-and-vr-display-market-180787629.html.

[2] M. Alain, E. Zerman, C. Ozcinar, and G. Valenzise, "Introduction to immersive video technologies," in *Immersive Video Technologies*, G. Valenzise, M. Alain, E. Zerman, and C. Ozcinar, Eds. Academic Press, 2023, ch. 1, p. 3–24.

[3] E. Alexiou, Y. Nehme, E. Zerman, I. Viola, G. Lavoue, A. Ak, A. Smolic, P. L. Callet, and P. Cesar, "Subjective and objective quality assessment for volumetric video," in *Immersive Video Technologies*, G. Valenzise, M. Alain, E. Zerman, and C. Ozcinar, Eds. Academic Press, 2023, ch. 18, p. 501–552.

[4] G. Goncalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa, "Systematic review of comparative studies of the impact of realism in immersive virtual experiences," *ACM Computing Surveys*, vol. 55, no. 6, p. 115:1–115:36, June 2023.

[5] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, p. 99–106, January 2022.

[6] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, "3d Gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, p. 139:1–139:14, August 2023.

[7] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *Proc. of IEEE International Workshop on Multimedia Signal Processing (MMSP'17)*, Luton, UK, October 2017, pp. 1–6.

[8] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence*, vol. 9, no. 5, pp. 497–503, 2000.

[9] A. Singla, S. Göring, D. Keller, R. R. Ramachandra Rao, S. Fremerey, and A. Raake, "Assessment of the simulator sickness questionnaire for omnidirectional videos," in *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR'21)*, Virtual, March 2017, pp. 198–206.

[10] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoue, "Subjective and objective visual quality assessment of textured 3D meshes," *ACM Transactions on Applied Perception*, vol. 14, no. 2, p. 11:1–11:20, October 2016.

[11] Y. Nehme, F. Dupont, J.-P. Farrugia, P. L. Callet, and G. Lavoue, "Visual quality of 3D meshes with diffuse colors in virtual reality: Subjective and objective evaluation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2202–2219, March 2021.

[12] L. A. da Silva Cruz, E. Dumic, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'19)*, Berlin, Germany, June 2019, pp. 1–6.

[13] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. D. S. Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics, and A. Skodras, "Point cloud subjective evaluation methodology based on 2D rendering," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'18)*, Sardinia, Italy, May 2018, pp. 1–6.

[14] J. Weil, Y. Alkhalili, A. Tahir, T. Gruczyk, T. Meuser, M. Mu, H. Koeppl, and A. Mauthe, "Modeling quality of experience for compressed point cloud sequences based on a subjective study," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'23)*, Karlshamn, Sweden, June 2023, pp. 135–140.

[15] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'20)*, Athlone, Ireland, May 2020, pp. 1–6.

[16] 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, December 2021.

[17] 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 *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*, Atlanta, USA, March 2020, pp. 127–136.

[18] S. Subramanyam, I. Viola, J. Jansen, E. Alexiou, A. Hanjalic, and P. Cesar, "Subjective QoE evaluation of user-centered adaptive streaming of dynamic point clouds," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'22)*, Lippstadt, Germany, September 2022, pp. 1–6.

[19] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'17)*, Erfurt, Germany, May 2017, pp. 1–3.

[20] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration," *IEEE Transactions on Multimedia*, vol. 23, pp. 3877–3891, October 2021.

[21] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3-D mesh quality assessment," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 247–256, February 2007.

[22] G. Lavoue, "A local roughness measure for 3D meshes and its application to visual masking," *ACM Transactions on Applied Perception*, vol. 5, no. 4, pp. 21:1–21:23, January 2009.

[23] K. Christaki, E. Christakis, P. Drakoulis, A. Doumanoglou, N. Zioulis, D. Zarpalas, and P. Daras, "Subjective visual quality assessment of immersive 3D media compressed by open-source static 3D mesh codecs," in *Proc. of International Conference on MultiMedia Modeling (MMM'18)*, Thessaloniki, Greece, January 2018, pp. 80–91.

[24] J. Gutierrez, T. Vigier, and P. L. Callet, "Quality evaluation of 3D objects in mixed reality for different lighting conditions," *Electronic Imaging*, vol. 32, no. 11, pp. 128:1–128:7, January 2020.

[25] S. V. Damme, I. Mahdi, H. K. Ravuri, J. van der Hooft, F. D. Turck, and M. T. Vega, "Immersive and interactive subjective quality assessment of dynamic volumetric meshes," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'23)*, Karlshamn, Sweden, June 2023, pp. 141–146.

[26] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, "Textured mesh vs coloured point cloud: A subjective study for volumetric video compression," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'20)*, Athlone, Ireland, May 2020, pp. 1–6.

[27] 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, September 2020.

[28] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point cloud rendering after coding: Impacts on subjective and objective quality," *IEEE Transactions on Multimedia*, vol. 23, pp. 4049–4064, November 2021.

[29] H. Liang, T. Wu, P. Hanji, F. Banterle, H. Gao, R. Mantiuk, and C. Oztireli, "Perceptual quality assessment of NeRF and neural view synthesis methods for front-facing views," 2023.

[30] P. Martin, A. Rodrigues, J. Ascenso, and M. P. Queluz, "NeRF-QA: Neural radiance fields quality assessment database," in *Proc. of IEEE International Workshop on Quality of Multimedia Experience (QoMEX'23)*, Karlshamn, Sweden, June 2023, pp. 107–110.

[31] F. Torkhani, K. Wang, and J.-M. Chassery, "Perceptual quality assessment of 3D dynamic meshes: Subjective and objective studies," *Signal Processing: Image Communication*, vol. 31, pp. 185–204, February 2015.

[32] 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," *Springer Quality and User Experience*, vol. 7, pp. 3:1–3:32, May 2022.

[33] Z. Chen, T. Funkhouser, P. Hedman, and A. Tagliasacchi, "c: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures," in *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22)*, Vancouver, Canada, June 2023, pp. 16 569–16 578.

[34] J. Kipp, "MobileNeRF Unity viewer," 2023, https://github.com/julienkay/MobileNeRF-Unity-Viewer.

[35] A. Pranckevicius, "Gaussian splatting playground in Unity," 2023, https://github.com/aras-p/UnityGaussianSplatting.

[36] B. O. Community. (2018) Blender–a 3d modelling and rendering package. Http://www.blender.org.

[37] L. Arns and M. Cerney, "The relationship between age and incidence of cybersickness among immersive environment users," in *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR'05)*, Bonn, Germany, March 2005, pp. 267–268.

[38] K.-Y. Lee, J.-W. Fang, Y.-C. Sun, and C.-H. Hsu, "Modeling gamer Quality-of-Experience using a real cloud VR gaming testbed," in *Proc. of ACM International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE'23)*, Vancouver, BC, Canada, June 2023, p. 12–17.