Journal on Interactive Systems, 2025, 16:1 doi: 10.5753/jis.2025.5593

Comparing Perceptual Visual Quality of Hybrid Neural and Ray Tracing in Foveated Rendering

Horácio Henriques [**Fluminense Federal University** | ***horaciomacedo@id.uff.br***]
Alan de Oliveira [**Federal Institute of Education, Science and Technology of Rondonia** | ***alan.oliveira@ifro.edu.br***]
Eder Oliveira [**Fluminense Federal University** | ***eder_oliveira@id.uff.br***]
Esteban Clua [**Fluminense Federal University** | ***esteban@ic.uff.br***]
Daniela Trevisan [**Fluminense Federal University** | ***daniela@ic.uff.br***]

Institute of Computing, Fluminense Federal University, Av. Gal. Milton Tavares de Souza, s/n, São Domingos, Niterói, RJ, 24210-590, Brazil.

Received: 20 February 2025 • **Accepted:** 08 May 2025 • **Published:** 15 May 2025

Abstract: Enabling Foveated Rendering for VR devices displays is fundamental when dealing with real-time ray tracing. Combining traditional methods with Neural based strategies, such as NeRFs and 3D Gaussian Splatting, may impact on leveraging performance even more. In this work we enhance and validate how well our traditional Instant-NeRF reconstructs common ray-traced effects through user metrics and quality metrics. We also show that 3D Gaussian Splatting used in the periphery vision area presents better results than the perceptual quality achieved through NeRFs. We present a deep human perception experiment through different global illumination light effects.

Keywords: Virtual Reality, Foveated Rendering, Perception-based Rendering, Ray Tracing, NeRF, 3D Gaussian Splatting

1 Introduction

Recent technological advancements have allowed consumers to experience real-time ray tracing in interactive environments like video games. This is made possible by the latest generation of graphics processing units (GPUs), which are specifically designed to handle ray tracing and AI-driven graphics enhancements, including denoising and DLSS [Kilgariff *et al*., 2018], [Einhorn and Mawdsley, 2023]. Even so, these effects may be unfeasible in certain conditions, such as unthetered Head Mounted Displays, with lower hardware resources and high resolution screens [Ubrani *et al*., 2024]. AIn this sense, different optimizations are required in order to guarantee stable and acceptable frame rate [Swafford *et al*., 2016; Albert *et al*., 2017], given that high latencies in VR are one of the reasons for *cybersickness* [Porcino *et al*., 2020]. The need for optimizing rendering also increases with the advent of untethered HMDs, which handles with constraints related to energy consumtion [Ujjainkar *et al*., 2024]. One such optimization is Foveated Rendering (FR), a set of techniques that take advantage of the limitations of the human visual system to improve performance without sacrificing the user's visual experience [Mohanto *et al*., 2022].

AI has fostered the development of a new field in computer graphics called neural rendering, with radiance fields being one of its key practical approaches [Yan *et al*., 2024]. In its first developments, radiance fields were achieved solely by training neural networks over images and coordinates through a technique named NeRF, as an alternative for voxel grids. NeRFs learn to infer the scene's volume, whether it's simple objects or vast landscapes [Mildenhall *et al*., 2021].More recently, 3D Gaussian Splatting (3DGS) achieved comparable results to NeRF, as this solution represents volumes as a collection of Gaussian surfaces that can be rastered by a regular rendering pipeline [Kerbl *et al*., 2023].

Regarding real-time rendering, 3DGS is a necessary alternative to NeRFs due to its rastering pipeline, allowing for better high-resolution performance. Compared to other NeRF-like methods that precompute their reconstruction for rendering without relying directly on a neural network, such as Plenoxels, 3DGS models achieve faster frame rates and superior visual quality. They do not require the tradeoffs that real-time NeRF solutions have, such as operating in lower qualities for a fraction of the performance or compromising reconstruction quality.

This paper aims to expand the findings of our previous work [Henriques *et al*., 2024b], which presented which NeRFs light effects can accurately represent the real global illumination effects achieved by real-time ray tracing as a surrogate for peripheral vision in VR. For so, we conducted a more detailed user study to assess user perception of the reconstruction of specific ray-traced effects in the peripheral vision. The study also examines how effectively these reconstructions can simulate such effects and the significance of an artificial reticle in keeping the screen's foveal region aligned with the user's fovea on an HMD without eye-tracking capabilities.

Our previous work showed that NeRF is capable of reconstructing some ray tracing effects in a near imperceptible manner, such as Subsurface Scattering and translucent objects with colored shadows. We also confirmed that the reticule in a testing setup without eye tracking is useful, given that users who were exposed to the experiment without the reticule, in general, perceived the effects as worse when compared to users who experimented with the reticule.

This work is hereby expanded by conducting a quality

analysis of the foveated rendering through various quality metrics, such as Peak Signal-to-Noise Ratio (PSNR) [Hore and Ziou, 2010], Structural Similarity Index Measure (SSIM) [Wang *et al*., 2004], Learned Perceptual Image Patch Similarity (LPIPS) [Zhang *et al*., 2018] and FovVideoVDP [Mantiuk *et al*., 2021]. We correlate these scores to the perceived quality of each effect, trying to figure out a link between the user feedback and these metrics. We then conduct the same quality analysis with 3D Gaussian Splatting enhanced frames of the same 3D scenes, as a way to approximate the possible user feedback under these circumstances. This is relevant since that, as previously stated, 3D Gaussian Splatting is more efficient for real-time renders.

This work is organized as follows: The Related Subjects section discusses important subjects to the comprehension of this work. Following, in the Hybrid rendering with NeRF and Ray tracing section we provide a more detailed explanation of the technique employed; then, the Perception Experiment section describes our user experiment and profile, followed by a Results Assessment, where we analyze and discuss the results of our user study; lastly, a Final Remarks section summarizes on what can be improved and expanded upon our results and recapitulate our hypothesis, our experiment, our results and further contributions.

1.1 Ethical issues

Before experimenting on users, we obtained approval from an ethics committee (CAAE 75192823.2.0000.5243) to run the procedure on 60 (sixty) consenting adults from Universidade Federal Fluminense. The users had the right to stop the experiment at any time for any reason and request that their data be deleted from the sample. They were given constant check-ups throughout the experiment to see if anyone exhibited *cybersickness* symptoms. In case of any *cybersickess* symptoms, the procedure would be stopped, the data generated up to that point would be erased and the experiment would continue with another user only after the sick user recovered. Researchers were also obligated to pay compensation if user participation in the experiment caused any expenses. No user requested compensation or reported any cybersickness symptoms during or after the experiment.

Every single user was informed of the rights stated above, signed an informed consent form and voluntarily gave non-identifying data pertinent to our experiment, such as age, biological sex, any vision impairments and prior experience with VR, which was given on a scale of:

- never had any prior experience with VR;
- had very few prior experiences with VR;
- had some experiences with VR;
- was an expert on VR.

2 Related Subjects

In this section we will present related topics required for understanding our proposal. First, we present basic concepts and challenges of real time Ray Tracing methods. Following, we discuss Foveated Rendering and how it became relevant in recent years. Lastly we discuss Radiance Fields, its most popular implementations and why it deserves our attention.

2.1 Real-time Ray Tracing

Ray tracing is an illumination model used to accurately simulate light transfer on scene, launching rays from light sources over to geometry [Whitted, 1979]. This model was modified in later works in order to launch one ray per ray bounce per pixel, not wasting resources by tracing rays that may never reach the user's camera [Kajiya, 1986]. This latter approach is named path tracing.

For years, path tracing was limited to offline rendering. This limitation was overcome in 2019 with the advent of specialized hardware that enables the real-time display of path tracing effects, such as reflections, refraction and shadows [Caulfield, 2022]. Nevertheless, the usage of path tracing in real time is contingent upon the availability of an effective denoising solution. This is because path tracing, which is a light effect, can result in a noisy outcome due to the statistical nature of the approximation [Koskela *et al*., 2019]. To this end, AI tools have been developed to enable real-time denoising, thereby reducing the effort required to create convincing graphics [Einhorn and Mawdsley, 2023].

Despite the construction of new graphic processing units designed with the specific goal of displaying ray tracing in real-time, and the utilization of IA to denoise the scene, these effects are still expensive to render on conventional displays. That cost rises up significantly when discussing their probable application in VR, given the higher refresh rates, higher resolutions and the need for keeping those high standards consistently in HMDs [Weier *et al*., 2016; Albert *et al*., 2017]. In our approach we intend to bake Ray tracing through neural rendering approaches and spend Ray tracing efforts only at small fovea regions.

2.2 Foveated Rendereing

One solution for outputting realistic graphics in VR comes from foveated rendering, which is a technique that leverages the human visual system's limitations in order to save computational resources and maintain latency without compromising immersion. This technique was first introduced in 1990 [Levoy and Whitaker, 1990], but the popularization of VR devices has reignited interest in this field of study.

The main idea underlying Foveated Rendering is the understanding that the human eye does not uniformly process visual information. The retina captures images mainly through two distinct types of photoreceptor cells: the rods, which are located in the periphery of the retina, and the cones, which are primarily concentrated in a narrow area at the center of the retina. Cones are more adept at capturing fine details due to their activation by color stimuli, while rods capture less information due to their activation by motion [Kim *et al*., 2024]. Studies have indicated that although cones are concentrated in a 5-degree region surrounding the center of the retina, humans experience relatively high optical quality up until 30 degrees of the fovea [Ogboso and Bedell, 1987; Banks *et al*., 1991]. The photoreceptor distribution on the retina is illustrated in Figure 1.

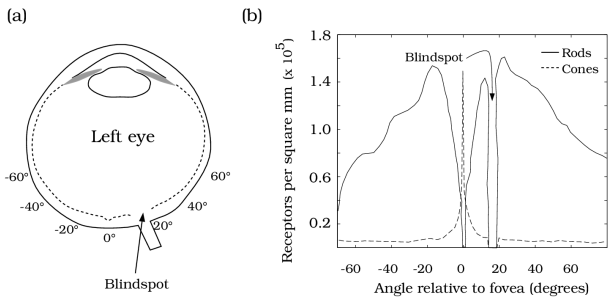

Figure 1. Photoreceptor distribution on the retina according to the distance from the fovea, as shown in Wandell [1995].

Photoreceptor distribution on the retina according to the distance from the fovea, as shown in Wandell [1995].

In HMDs, in which the screens are positioned in close proximity to the human eye, this effect is particularly significant due to the limited area of the screen that falls within the center of the retina, named the foveal region. That means that a huge portion of the screen is captured by photoreceptors that are not capable of capturing fine detail, which means that rendering those regions at full capacity is a waste of computational power [Guenter *et al*., 2012].

Given that the fundamental concept is straightforward, there are numerous approaches to achieving this foveation effect [Mohanto *et al*., 2022; Wang *et al*., 2023]. The most common approach is to vary the screen resolution according to the distance to the foveal region of the screen, which is determined by eye tracking or some other parameter, displaying the foveal portion of the screen in full resolution and decaying the further it goes [Levoy and Whitaker, 1990; Koskela, 2020]. Other techniques revolve around varying the refresh rate of the screen according to the distance to the fovea [Yee *et al*., 2001; Patney *et al*., 2016], manipulating the displayed colors or the shading complexity of a scene [Duchowski *et al*., 2009; Tursun *et al*., 2019], or using adjustable data to display cheaper information in the periphery of the screen [Schütz *et al*., 2019].

Foveated rendering can be applied to a variety of different rendering pipelines. From simple rasterization techniques to advanced path tracing in real-time, the more expensive the render process is, the more important it is to use an appropriate technique. For instance, in the context of real-time path tracing, Koskela *et al*. [2019] employs a sampling strategy on the screen according to human acuity to achieve real-time path tracing even without dedicated hardware [Koskela *et al*., 2019].

To the best of our knowledge, there is a lack of works that explore the usage of radiance fields on Foveated Rendering. Although there is at least one highly interesting solution that trains NeRFs in such a way that optimizes rendering due to inferring a lighter model on the periphery of the screen [Deng *et al*., 2022], we did not found works that attempt to combine rendering pipelines. We believe that as radiance fields become more powerful and efficient to render, there will be more proposals on how to mix radiance field solutions with more expensive rendering methods.

2.3 Radiance Fields

Computer graphics has been looking into rendering novel views given a set of observable images ever since the nineties [Avidan and Shashua, 1997], and since then, many relevant works have been proposed, such as light field sample interpolation [Gortler *et al*., 1996; Davis *et al*., 2012], mesh-based representation of scenes optimized by differentiable renderers [Buehler *et al*., 2001; Debevec *et al*., 2023; Genova *et al*., 2018; Li *et al*., 2018] and voxel grids. Machine Learning has been especially present in works that optimize voxel grids based on input images.

Neural Radiance Fields, or NeRFs, introduced by Mildenhall *et al*. [2021], are a novel approach to View Synthesis. The main idea involves training a multilayer perceptron network with images and coordinates to generate novel views from a given scene. The network receives as an input a three-dimensional vector denoting space position and a two-dimensional vector denoting gaze direction, and from that the network infers color and volume information [Mildenhall *et al*., 2021].

This foundational work has served as the base for hundreds of subsequent works in the field, with researchers adapting NeRFs, either to achieve specific objectives such as large-scale reconstructions [Turki *et al*., 2022; Lu *et al*., 2023], real-time rendering of NeRF reconstructions Müller *et al*. [2022]; Yu *et al*. [2021]; Chen *et al*. [2023b] and even foveated rendering [Deng *et al*., 2022] or to enhance the original work [Reiser *et al*., 2021; Barron *et al*., 2022, 2023].

NeRF has also inspired other works that aim to reconstruct scenes and allow for novel views without the neural render. There are works, such as Yu *et al*. [2021], that aims at distilling NeRF into easily navigable data structures to allow for faster rendering. Other works, however, aim at optimizing alternative data structures [Liu *et al*., 2020; Fridovich-Keil *et al*., 2022; Yang *et al*., 2024].

In 2023, Kerbl et al. proposed optimizing three-dimensional Gaussian surfaces (referred to as "*splats*") to achieve novel views by utilizing the same input data as NeRFs. This technique, designated 3D Gaussian Splatting, also inspired hundreds of other works, which expanded upon the original contribution or optimized its offerings. The original work, however, describes the process of optimizing a scene according to the number of colors on surfaces and sorting those same *splats* according to their prominence on screen [Kerbl *et al*., 2023]. This allows for 3DGS to be efficiently rendered by regular rasterization methods. This implies that 3DGS can function on less powerful computers, a capability most NeRF-based solutions struggle with.

3 Hybrid rendering with NeRF and Ray tracing in Foveated Rendering

We propose the idea of substituting Ray tracing rendering at the peripheral regions of the HMD display using alternative neural rendering methods. While there are solutions for hybrid rendering [Henriques *et al*., 2024a], there is a lack in the literature related to the perception of illumination features when combining both. This is especially relevant when conspiring that the Neural rendering approaches are related to pre-processed scenes.

This proposal stems from the notion that neural radiance fields are more capable of representing light-based effects

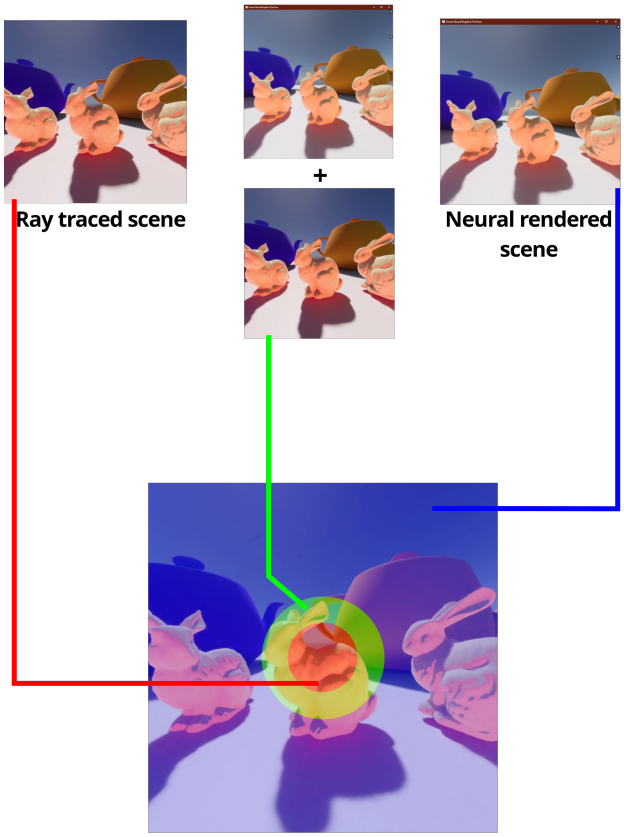

Figure 2. Composition of NeRF and Ray traced images as described by Henriques *et al*. [2024a]. The peripheral region is painted in blue, the parafoveal region that covers about 35% of the screen is painted green and the foveal region that covers about 20% of the screen is painted red.

than other light baking methods. Since color information is dependent on world position and gaze direction [Mildenhall *et al*., 2021], one could bake reflective properties into those radiance fields given a single reflective surface is stationary on a scene.

The output radiance formula for each pixel is shown in the equation below [Mildenhall *et al*., 2021]. The expected color of a single camera ray, where each ray r, constrained between Tn and Tf, outputs a color $C(r)$ given by integrating the accumulated transmittance $T(t)$, the volume density $\sigma(t)$ and the albedo color $c(r(t), d)$. If a single scene is trained upon reflective images, it will learn to reconstruct this reflection upon the training viewpoints.

$$C(r) = \int_{Tn}^{Tf} T(t)\sigma(r(t))c(r(t), d)\, dt \qquad (1)$$

This approach to image generation has the potential to generate more realistic effects without resorting to expensive ray tracing techniques or more traditional environment mapping techniques. This is true, however, when the training data used by NeRF is of high quality.

Given a virtual scene, each scene would have two representations: a polygonal mesh, rendered in real time with path tracing, and an equivalent reconstruction achieved by any NeRF-like method. The final image would be composed by rendering the path traced representation only in the foveal region of the screen, while rendering the periphery solely by the radiance field. The parafoveal region, which encompasses the middle region between fovea and periphery, has an alpha blending of both regions, transitioning from fovea to periphery [Henriques *et al*., 2024a]. The technique is illustrated in Figure 2.

The surrogate scene used in the periphery of the screen is trained based on images captured from the polygonal scene rendered in the foveal region of the screen. We recommend to setup cameras in a spherical disposition around the center of the scene, increasing the density of captures to account for scene details, such as low visibility regions. The pairs of images used in the experiment are represented on Figure 3.

In this experiment, we used Instant-NeRF as the NeRF solution for the peripheral image [Müller *et al*., 2022] due to its training time and rendering performance. However, there are other NeRF solutions focused on displaying ray tracing effects, focused on delivering better reflections than some of the best known NeRFs in the literature [Verbin *et al*., 2022; Chen *et al*., 2023a].

4 Human Perception Experiment

The experimental study has two main goals: to understand the impact of replacing the peripheral vision of specific ray tracing effects with neural rendered reconstructions, and to understand the impact of an artificial reticule while conducting such an experiment on devices without eye tracking capabilities. Our hypothesis is that neural rendering will reconstruct some effects better than others, and that the reticule is important to assure good results in this scenario.

In this section, we describe the user profile and the experiment protocol we conducted with users.

4.1 User Profiling

To validate our hypothesis, we conducted the test with forty-two users. From those forty-two, twenty-one conducted the test without the reticule on the screen to fixate their gaze, while the other twenty-one conducted the test with the reticule. In the end, we had 630 (six hundred and thirty) user inputs, half of them being without the reticule and the other half being with the reticule.

Most of the users (90,5%) were male, 83,4% of them had little or no previous experiences with VR, and 61,3% had no diagnosed vision impairments. Among the 38,7% users with some kind of vision impairment, all of them had their vision corrected by the usage of glasses, which they wore during the experiment. Most of them reported having nearsightedness and astigmatism. The users' age ranged from 18 to 44 years old, with average age being 22,02 and a standard deviation of 5,02.

4.2 Experiment Protocol

Although the reliability for reconstructing global illumination in closed scenes was assessed in Henriques *et al*. [2024a], we have yet to assess a larger number of users while testing other ray tracing based effects during the interactive applications. The aim of this user study is to assess how convincingly can these reconstructions be to the human eye. Our

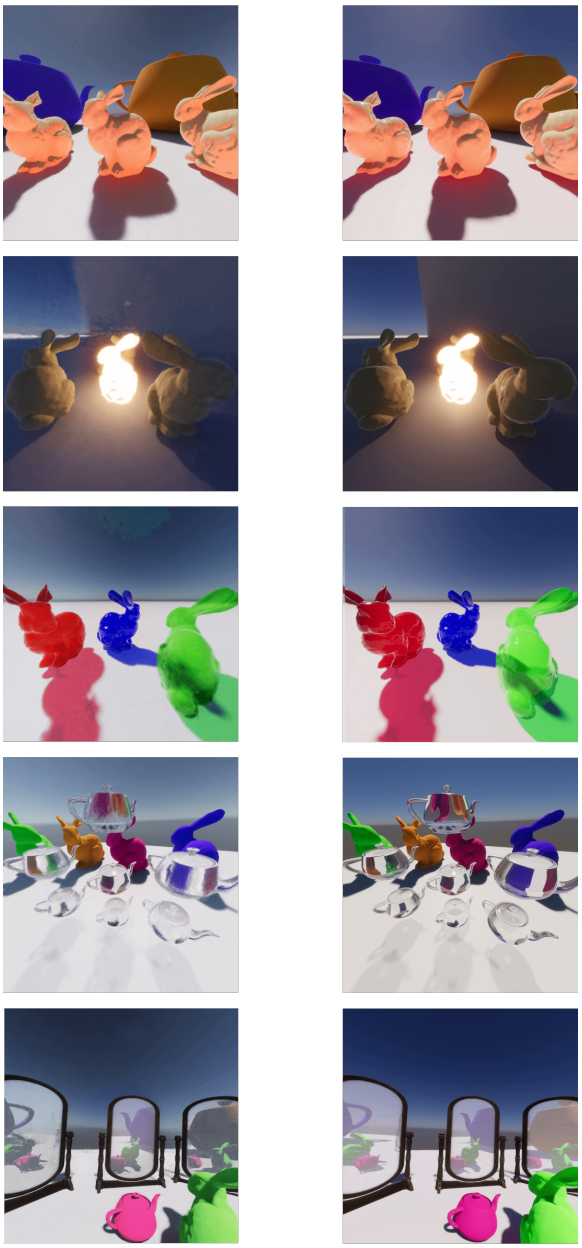

Figure 3. Pairs of images used for the user experiment. Each pair of images represents a different ray tracing effect: Subsurface Scattering (the rose coloring on the rabbit's body), Emission (the rabbit at the center emitting light), Colored Shadows (the three translucent rabbits), Refraction (the floating teapots refracting the rabbits) and Reflection (the mirrors reflecting the rabbit and teapot.). The left image contains the foveated image, composed using the technique; the right image contains the fully raytraced image.

primary hypothesis is that NeRF will be sufficient to approximate ray tracing effects convincingly enough for peripheral view, with some effects being less perceived than others.

To achieve this goal, we identified five main effects often achieved through ray tracing pipeline: Subsurface Scattering (SS), Light Emission (E), Colored Shadows (CS), Reflections (RL) and Refraction (RR), and conducted a Double-Stimulus Impairment Scale (DSIS) test in order to assess how well executed were such effects.

The DSIS test [Series, 2012] is a traditional testing routine used to measure the degradation of screens according to user feedback given on a DCR (Degradation Category Rating) scale from 1 (very damaging) to 5 (not perceptible). On DSIS, a round consists of the user being first exposed to a reference, non-modified image for a small amount of time. Following, the user is exposed to a rest period, without visual stimuli, for a significantly smaller amount of time, and then to a modified image for the same amount of time they had seen the reference image. The round ends with the user giving their input through the DCR scale. The method is illustrated by Figure 4 [Series, 2012]. In our experiment, we conducted a script of questions that allowed the user to give an assessment that fits the DCR without taking off the HMD, reducing the length of the test and minimizing exhaustion.

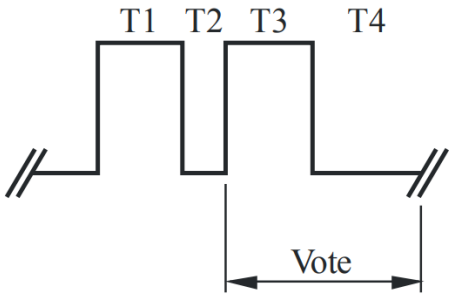

Figure 4. A single round in a DSIS experiment, adapting the length of each part to our needs [Series, 2012]: T1 = 5s (reference image), T2 = 2s (resting black screen), T3 = 5s (foveated image), T4 = from 5s to 11s (resting black screen).

The aforementioned script consists of three simple questions:

1. **"Have you noticed any difference between the images?"**, to which the user could answer "yes" or "no". If the answer is "yes", then we move to the second question. Otherwise, if the user answers with a "no", it is equivalent to a score of 5; there are no perceptible differences between images.
2. **"Would you say you noticed the image getting worse?"**, to which the user also can answer with "yes" or "no". If the answer is "yes", then we jump to the third question. Otherwise, we assume that the score of the answer is 4 on the DCR scale, which means that there is a perceptible but non-damaing change to the screen.
3. **"How would you say the worsening damaged your perception of the scene?"**, to which the user can answer with "a little", "mildly" or "a lot". Each one of these answers has different scores on the DCR; respectively, scores of 3 (perceived little damage), 2 (perceived some damage) or 1 (perceived a lot of damage).

Each user underwent fifteen rounds of DSIS: three repetitions of each one of the five effects being tested. This repetition is a strategy to ensure better feedback from the user, accounting for any change of perception that repeated exposures to the same effect may have [Sorokin and Forsyth, 2008]. The effects were picked at random every single time to mitigate any bias perception one effect may have over another effect.

To better control the focus of the users, we baked examples into pairs of static images. We intend to guarantee that the user would see each scene in the most interesting way possible to test their perception of the effects, in such a way that both the fovea and periphery would have at least an example of an effect on display. This setup is solely for testing user perception, negating any problem that might arise from rendering the images in real-time (misalignment of both reconstructions, stuttering caused by an unpredictable performance dip) and diminishing the risk of breaking user's immersion. We conducted the experiment using two Meta Quest 2 HMDs plugged in different computers with non-relevant specifications.

Since we were testing with pairs of static images, a question arose during the pilot tests: does the user need a specific reticule to fixate the gaze on the appropriate region and avoid looking at the peripheral area with the fovea, since the hardware used does not have an eye-tracking system? To explore this possibility, we divided the users into two separate groups: one group performed the test with a green reticule in the center of the screen and received explicit instructions to fixate on it. The other group experimented without such a crosshair and was only instructed to look forward. Both groups were periodically reminded during the test to keep looking forward, or in the case of the reticule group, to keep looking at the reticule. Figure 5 features a comparison between an image with reticule and the same image without it.

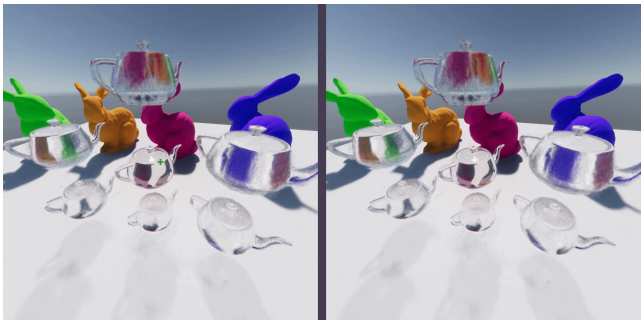

Figure 5. The foveated refraction testing view. On the left, the refraction testing view has a reticule upon which the users could fix their gaze during experiment. On the right, the image features no reticule.

5 Human Results Assessments

In this section, we assess the results collected from the experiments listed in the previous section. In order, we first describe the statistical approach to the collected data, then analyze the perceived quality reported by users and discuss the role that the reticule played on the data we gathered.

5.1 Statistical Analysis

Research in human-computer interaction (HCI) frequently employs quantitative methods and parametric analysis to evaluate human performance metrics, such as task completion metrics [Dix *et al.*, 2003]. However, the Shapiro-Wilk normality test indicated that the dependent variable $dScore$ does not follow a normal distribution, rendering parametric methods unsuitable for this analysis. Consequently, we applied the aligned rank transform (ART) method [Wobbrock *et al.*, 2011; Bates *et al.*, 2015] and aligned rank transform contrasts (ART-C) [Elkin *et al.*, 2021] to examine the impact of the independent variables $iReticule$ and $iEffect$ (Table 1) on the dependent variable $dScore$.

Table 1. Values of the independent variables within the experimental context.

Variables & Levels	
$iReticule$	No, Yes
$iEffect$	CS, E, RL, RR, SS

Following the analysis of variance and the identification of a significant overall difference between groups, it is necessary to perform post-hoc tests to determine which specific groups differ from each other. The use of Holm's Method (Holm-Bonferroni) [Holm, 1979] is justified to control the risk of Type I errors (false positives) in multiple comparisons. The choice of Holm's Method is appropriate for this study due to the large number of post-hoc comparisons required. With multiple combinations of $iReticule$ and $iEffect$, Holm's more balanced approach allows for more effective detection of true differences between groups without sacrificing control over false positives.

We performed the statistical analysis using R (4.4.0)[1] and RStudio (2024.4.1.748)[2]. As statistical significance parameters, we explored using three p-values: $p < .0001$, $p < .001$, and $p < .05$.

5.2 Perceived Effect Quality

Table 2. Mean user scores per tested lighting effect

Effect & Reticule & Mean Score		
Colored Shadows	No	3.89
	Yes	4.17
Emission	No	2.51
	Yes	3.13
Reflections	No	3.62
	Yes	4.24
Refraction	No	2.65
	Yes	3.33
Subsurface Scattering	No	4.48
	Yes	4.65

The values in Figure 6 reveal that the peripheral vision was better received by users when displaying certain effects. From the five effects we tested (Colored Shadows, Emission, Reflection, Refraction and Subsurface Scattering), Colored

[1] https://cran.r-project.org/
[2] https://posit.co/

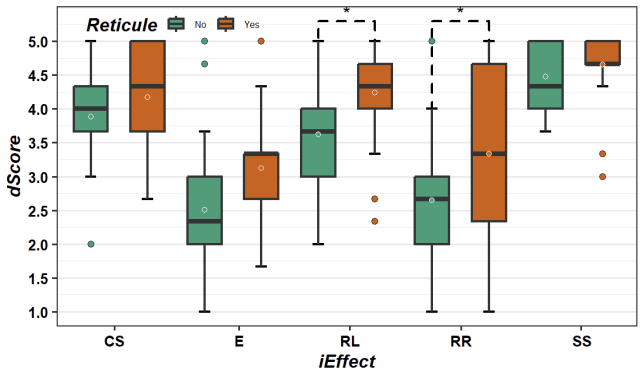

Figure 6. Boxplot featuring scores of each effect according to user assessments. To better assess users' scores, we divided these scores by whether the user participated in the experiment with reticule on screen (in orange) or not (in green). In order, the featured effects are: Colored Shadows (**CS**), Emission (**E**), Reflections (**RF**), Refraction (**RR**) and Subsurface Scattering (**SS**).

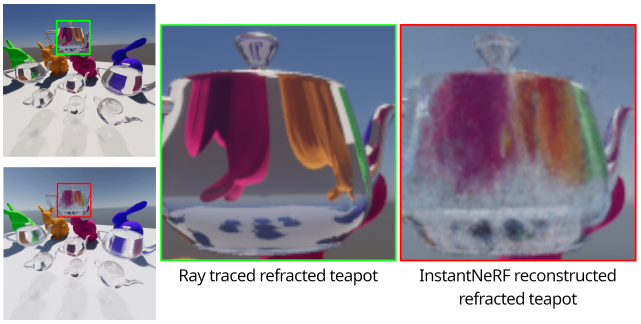

Figure 7. View on the refraction pair of images. The refracted image through the teapots suffers from the lack of detail present on the ray-traced images, which hinders NeRF's capacity to properly reconstruct such an effect. The refractive teapot's specularity also hinders its reconstruction.

Shadows and Subsurface Scattering performed better than their peers. The mean scores given by users are featured in Table 2.

Users perceived Subsurface Scattering as the better reconstructed effect on the peripheral vision, given its higher score. This is so due to NeRF's capability of reconstructing well-lit scenes in convincing ways. Users also perceived Colored Shadows as the second better reconstructed effect. Reflection reconstrucions were also largely well received by users, even though more users would perceive them as worse quality than their full ray traced counterpart, with the presence of reticule playing a part in this perception. This feature reinforces our original hypothesis that NeRF could be a good surrogate for peripheral vision while being insufficient for foveal vision.

Users perceived the other effects as damaged in some way. Refraction and Emission were worse received due to NeRF's limitation on displaying refracted coloring and reconstructing emissive scenes. This is a known limitation of NeRF to some extent, with literature around possible solutions for more convincing reconstruction of these scenarios [Cui *et al.*, 2024; Deng *et al.*, 2024]. Examples of these limitations can be seen in more detail on Figures 7 and 8.

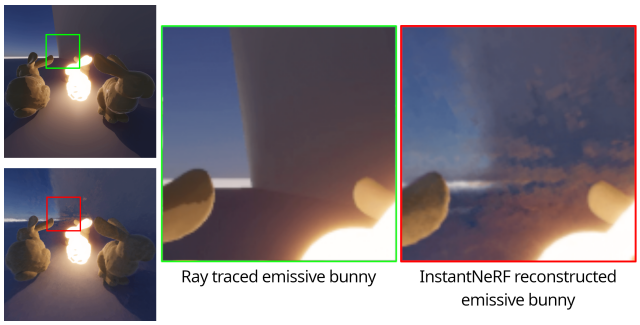

Figure 8. View on the emissive pair of images. We remark that the emissive effect shows artifacts from the volume accumulation, which leaves a smoke-like visual artifact in the reconstruction. The low light in the scene also represents another challenge for the reconstruction, since the lack of light represents fewer features to reconstruct the scene.

5.3 Reticule's Influence

As previously stated, we conducted the experiment without access to eye tracking, which means that we were unable to adjust the foveal region onto the images in real-time. That limitation led us to question if users could conduct such an experiment without a visual aid to fixate their gaze at the center of the foveal region.

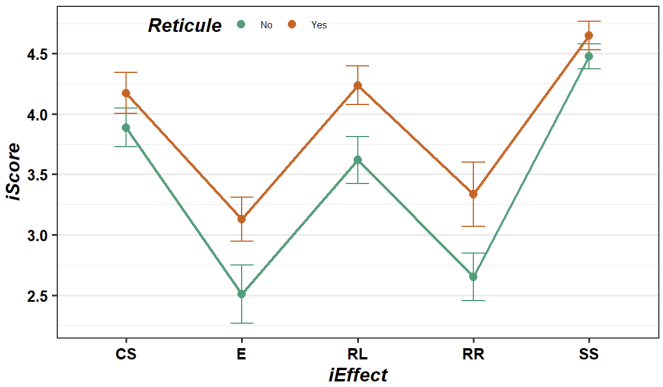

Figure 9. Difference between user scores that participated in the experiment with the reticule, in orange, and user scores that participated in the experiment without the reticule, in green. In order, the featured effects are: Colored Shadows (**CS**), Emission (**E**), Reflections (**RF**), Refraction (**RR**) and Subsurface Scattering (**SS**).

Given the values presented in Figure 6, we can perceive the reticule playing a role in improving the experiment results. With a reticule to focus their attention, users would end up fixating their gaze on the best quality, minimizing the risk of drifting their foveal vision over regions that were suitable only for their peripheral vision.

We observed in Figure 9 that the group of participants who used the Reticule (iReticule = Yes) during the experiment exhibited higher mean scores across all effects compared to the group of participants who did not use the Reticule (iReticule = No). A significant interaction was noted between $Reticule \times iEffect$ ($F_{4,580} = 2.6860, p < .05$) on $dScore$. However, posthoc tests revealed that only the RL and RR effects showed significant p-values, as illustrated in Figure 6.

6 Image Quality Assessment

In this section, we will compare the user feedback provided in Section 5 with well-established quality metrics present in the literature. We want to observe the relation between the perceived quality and the aforementioned metrics to predict if the usage of 3DGS could be considered appropriate for substituting the Instant-NeRF rendering in the periphery of the vision.

Even if we conduct this analysis to try to figure out a relationship between quality measurements and user perception, it is important to point out that no testing image scored poorly in either scenario. Even the worst perceived effects had decent quality scores, with Emission being the worst offender, as the next subsections will expand upon.

The subsection 6.1 introduces the quality metrics chosen for this comparison and their limitations. The subsection 6.2 will compare the user feedback with the Learned Perceptual Image Patch Similarity (LPIPS), Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) quality metrics. The subsection 6.3 will compare the scores from the 3DGS-generated frames to the NeRF ones.

6.1 Quality Metrics

This work uses four different quality metrics to assess the quality of our NeRF reconstructions for peripheral vision and compare them to user feedback. These quality metrics also compare equivalent compositions using a Gaussian Splatting reconstruction instead of the NeRF for peripheral view.

This replacement is interesting due to performance reasons. Since 3DGS renders its reconstructions through rastering the splats over the screen, they can easily render high-resolution reconstructions at low latencies. As stated in Section 3, achieving low latencies without disturbing user experience is the point of this study.

This work will compare the images under the following metrics: PSNR, SSIM, LPIPS and FovVideoVDP. The choice behind those metrics is motivated by the literature surrounding them: PSNR and SSIM are classic metrics that are featured in several other works that explore image quality, but more importantly, imperceptibility [Setiadi, 2021]. PSNR works by measuring the logarithm of the mean square error of an image, as SSIM measures the difference between two images by comparing luminance, contrast and structure between one image as a reference and the other image as a test case.

But PSNR is known not to be conforming with human visual perception, and SSIM is also known not to be perfectly representative of human visual perception as stated by Nilsson and Akenine-Möller [2020]. These metrics are better suited to objective image quality assessments, which may not be entirely compliant to visual quality assessment, but still are important in themselves as quality metrics of these images specially when comparing the reconstruction quality of the scenes.

These limitations pushed us to pursue alternative quality metrics better aligned with human perception for a more complete and valuable quality assessment. For that kind of assessment, we chose LPIPS and FovVideoVDP due to their goal of approximating human perception and previous works attesting to the success of these techniques.

LPIPS is a quality metric based upon human perception learned by a deep neural network, which has proven itself more capable of approximating human perception instead of more traditional metrics. This is due to its training dataset, composed of images distorted in many ways and human assessments of said images [Zhang *et al*., 2018]. While executing this work, we used the three networks available for LPIPS testing: AlexNet, VGG and SqueezeNet, trained with the entire LPIPS dataset.

FovVideoVDP, on the other hand, is a quality metric aimed at measuring quality of foveated videos. It works based on a decomposition of spatial and temporal factors over foveated videos pooled over spatial frequency bands, temporal channels, and all the frames. Ultimately, this entire process composes just-objectionable-difference (JOD) units that represent differences that may present themselves objectionable in foveated scenarios since it considers foveal eccentricity to evaluate the damage. This metric has been validated by psychometric studies and comparisons with other quality metrics, and we tend to believe that FovVideoVDP is the most relevant in this scenario due to its specific usage on assessing quality on Foveated Rendering images, an explicit use case of this metric [Mantiuk *et al*., 2021].

6.2 NeRF Scores

In this section, we pay attention to the images used during the user tests. Those were submitted over the four quality metrics, with their default parameters. We ran FovVideoVDP after setting the parameter "Display Name" as "standard_hmd" for running the test considering the resolution of 1440 by 1600 with a diagonal FOV of 110. We ran LPIPS tests three times, each with a different network mentioned at Zhang *et al*. [2018].

Figures 10, 11, 12 and 13 display the comparison between the user assessment to each image and the aforementioned quality metrics. The user assessment is represented in a scale from 1 to 5 shown in subsection 5.2 that measures how damaging the effect of NeRF was on the user perception.

Figure 10 points to a more fragile relation between the quality and user scores. This is obvious when the worst perceived effect of the bunch, Emission, is the effect with the greater PSNR score from the testing group. Another curious result is the Reflection score, which performed fairly well with users and had the second-lower PSNR score.

Figure 11 shows a similar trend, under which Emission is practically tied for the second-best-looking effect of the bunch with Refraction, surpassed only by Colored Shadows. SSIM values also point to Subsurface Scattering as the worst reconstruction, and its composition the most damaged. However, this does not match the user assessment, which gave it

Figure 10. Difference between user scores that participated in the experiment with reticule (in green), user scores that participated in the experiment without the reticule (in orange) and the PSNR value for each effect (on the bottom of the horizontal axis). In order, the featured effects are: Colored Shadows (**CS**) with PSNR of 30.24, Emission (**E**) with PSNR of 32.29, Reflections (**RL**) with PSNR of 29.003, Refraction (**RR**) with PSNR of 28.85 and Subsurface Scattering (**SS**) with PSNR of 30.24.

Figure 11. Difference between user scores that participated in the experiment with reticule (in green), user scores that participated in the experiment without the reticule (in orange) and the SSIM value for each effect (on the bottom of the horizontal axis). In order, the featured effects are: Colored Shadows (**CS**) with SSIM of 0.971, Emission (**E**) with SSIM of 0.965, Reflections (**RL**) with SSIM of 0.9598, Refraction (**RR**) with SSIM of 0.9651 and Subsurface Scattering (**SS**) with SSIM of 0.9367.

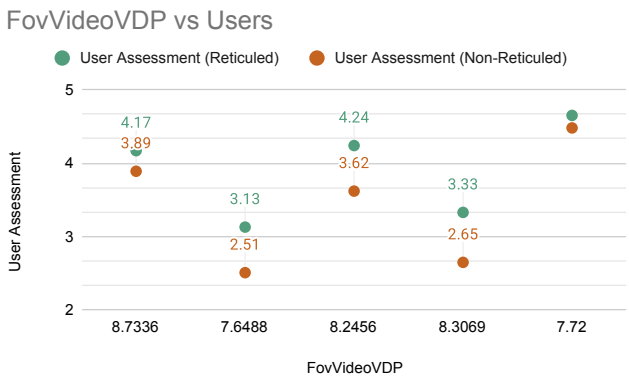

Figure 12. Difference between user scores that participated on the experiment with reticule (in green), user scores that participated on the experiment without the reticule (in orange) and the FovVideoVDP value for each effect (on the bottom of the horizontal axis). In order, the featured effects are: Colored Shadows (**CS**) with FovVideoVDP of 8.7336, Emission (**E**) with FovVideoVDP of 7.6488, Reflections (**RL**) with FovVideoVDP of 8.2456, Refraction (**RR**) with FovVideoVDP of 8.3069 and Subsurface Scattering (**SS**) with FovVideoVDP of 7.72.

the higher score during testing.

The FovVideoVDP metric, shown in Figure 12, is closer to the user assessment than the aforementioned metrics. As can be seen, the lowest FovVideoVDP value matches the lowest-rated effect by users. These results don't match when we look at Refraction's quality, which is scored as the second-best reconstruction according to FovVideoVDP, and yet is the second-worst NeRF reconstruction according to users. Another perplexing result is the Subsurface Scattering one, rated by users as the best effect but with the second-lowest FovVideoVDP value.

Figure 13 brings different information, and more closely matches user assessment. In this metric, the closer the score is to zero, the best is the image, since it measures difference from a reference value.

Even though LPIPS' default model is the AlexNet, the SqueezeNet-based one scored closer to the user assessments by ranking Emission as the worst effect (a common trend between the three LPIPS tests) and Refraction as the second-worst (such as the VGG-net graph). SqueezeNet also performed better ranking the three best effects than its counterparts, besides practically tying up Subsurface Scattering and Colored Shadows for first place (Subsurface Scattering was lower than Colored Shadows for a difference of 0.001) and keeping Reflection in the third place. VGG also kept the same order but scored Subsurface Scattering closer to Reflection. AlexNet also performed well, but not as well as its counterparts. Refraction was the second-best-rated effect, tailing Colored Shadows, while Subsurface Scattering and Reflections scored close to one another.

Inspection of these metrics shows a closer correspondence between user feedback and the LPIPS, especially with SqueezeNet, even though it does not perfectly match user feedback. PSNR and SSIM showed similar patterns in their distributions, which points to the fact that they are both unsuitable to match the user's perception and is to be expected [Nilsson and Akenine-Möller, 2020]. FovVideoVDP, created to approximate user perception, does not perfectly match the user feedback but is closer to the user scoring.

LPIPS' scores are generally distributed just like the user feedback except for the Reflections, which scored worse than every single one of the effects, and FovVideoVDP ranked Subsurface Scattering below many other effects while scoring Refraction as the second-best reconstruction. This assessment leads us to conclude that, despite the discrepancies, LPIPS and FovVideoVDP are more suited to infer the quality of reconstructions produced by other means.

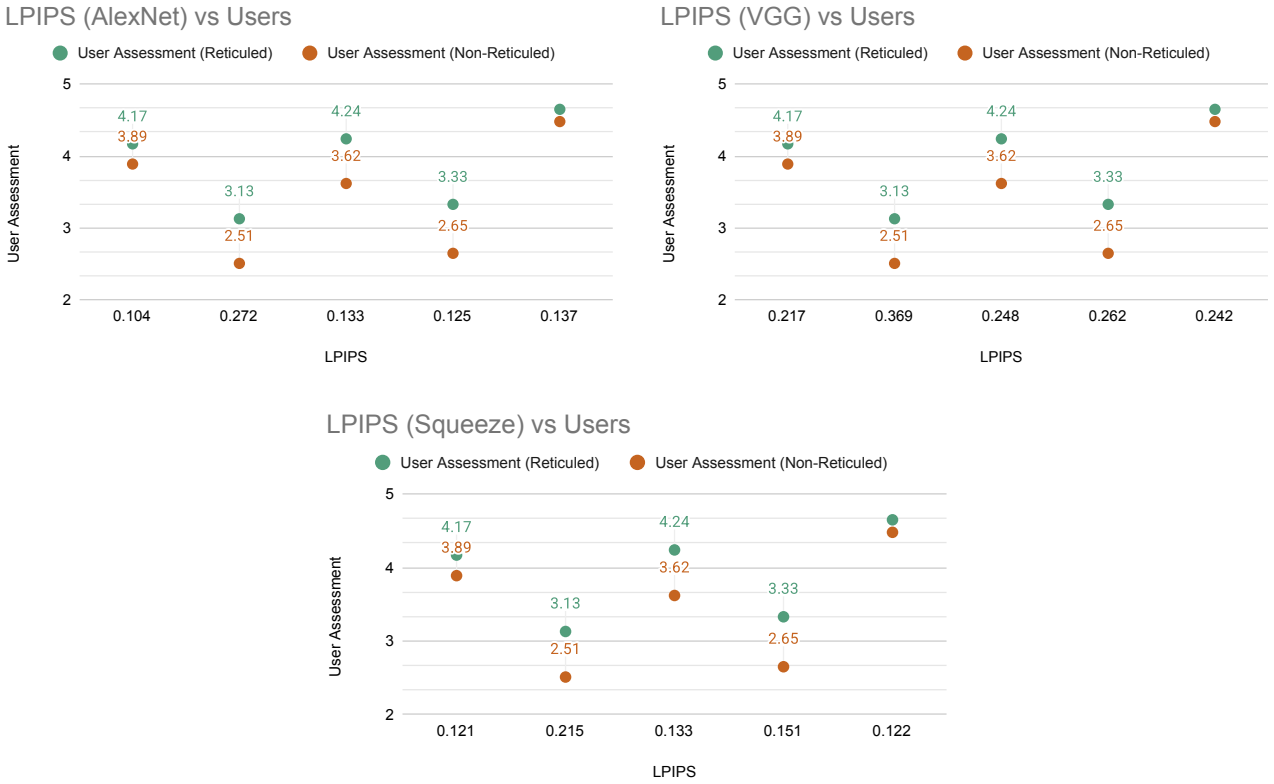

Figure 13. Difference between user scores that participated in the experiment with reticule (in green), user scores that participated in the experiment without the reticule (in orange) and the LPIPS value for each effect (on the bottom of the horizontal axis). In order, the featured effects are: Colored Shadows (**CS**) with LPIPS of 0.104 (AlexNet)/0.217 (VGG)/0.121 (SqueezeNet), Emission (**E**) with LPIPS of 0.272 (AlexNet)/0.369 (VGG)/0.215 (SqueezeNet), Reflections (**RL**) with LPIPS of 0.133 (AlexNet)/0.248 (VGG)/0.133 (SqueezeNet), Refraction (**RR**) with LPIPS of 0.125 (AlexNet)/0.262 (VGG)/0.151 (SqueezeNet and Subsurface Scattering (**SS**) with LPIPS of 0.137 (AlexNet)/0.242 (VGG)/0.122 (SqueezeNet).

6.3 3D Gaussian Splatting Scores

After comparing the user feedback with quality scores from the NeRF reconstructions, we constructed 3DGS-based reconstructions of the same scenes. The peripheral view of each ray-traced scene was replaced by a standard 3D Gaussian Splatting reconstruction after 30000 (thirty thousand) iterations of training. These reconstructions were trained upon datasets of 400 images surrounding the artifacts of each image with known extrinsic parameters.

Table 3 contains scores from foveated NeRF reconstructions and foveated 3DGS reconstructions from each frame used in the experiment lined out at Section 5. It points to a clear improvement all over the reconstructions of the effects, having the least performing effect to be Reflection over any observable metric.

SSIM was the effect that accused the least amount of improvement, having Reflections as good with NeRF and 3DGS, and pointing to a modest downgrade with Refraction. This is also observed with FovVideoVDP, which pointed to a very modest decrease in quality when comparing Reflections and Refraction. This result is a surprise on some level, given that FovVideoVDP considers screen eccentricity to measure the objectionable differences metric necessary for its scoring. On the other hand, we must mention the huge increase of quality when using 3DGS to reconstruct Emission in the periphery, which was the worst perceived effect by the public in its NeRF reconstruction.

PSNR, on the other hand, only pointed to a 1.325-point drop in quality in Reflections, pointing to enhancements over every other effect. According to PSNR alone, the best-increasing effect was Colored Shadows, Emission being the second-best enhanced effect of the bunch.

LPIPS trained in an AlexNet showed improvement in four different effects (Colored Shadows, Emission, Reflections and Refraction). The effect that most benefitted from 3DGS was Emission, with an LPIPS value decrease of 0.167. In the case where we observed a decrease in visual quality through AlexNet's LPIPS, the value increase is very modest. VGG and SqueezeNet's LPIPS performed better, universally pointing to enhancements in every effect. VGG pointed at the biggest enhancements, showing the biggest drops in scoring over every effect when compared to other LPIPS test.

Reflections probably fared worse in these tests due to the reconstruction conditions. The NeRF reconstruction of this effect was troublesome already, needing manual tuning of training parameters to learn these reflections since the mirrors were relatively small amidst the scene, causing the mirror image to become less prioritized. The 3DGS effect suffered even more, optimizing fewer details on the reflective surface. A more detailed comparison can be seen at Figure 14. This problem may be circumvented by optimizing such reconstructions with more images focusing on the mirrored surfaces, but that solution may lead to overfitting problems and generate undesirable artifacts. Other alternative would be exploring a different 3DGS-based method, one more suited to reconstructing scenes with less amount of

Table 3. NeRF quality scores versus 3DGS quality scores. 3DGS compositions that had better scoring than their NeRF counterparts are marked in bold. The different LPIPS models were shortened for brevity; LPIPS with AlexNet was shortened to "LPIPSA.", LPIPS with VGG was shortened to "LPIPSV." and LPIPS with SqueezeNet was shortened to "LPIPSS.". FovVideoVDP was shortened to "VDP" for brevity.

Effect	Metric	NeRF	3DGS	Difference
Colored Shadows	$PSNR$	31.99	34.414	↑ **2.424**
	$SSIM$	0.971	0.98	↑ **0.009**
	$LPIPSA.$	0.104	0.074	↓ **0.03**
	$LPIPSV.$	0.217	0.1	↓ **0.117**
	$LPIPSS.$	0.121	0.058	↓ **0.063**
	VDP	8.733	8.887	↑ **0.154**
Emission	$PSNR$	30.351	32.404	↑ **2.053**
	$SSIM$	0.965	0.984	↑ **0.019**
	$LPIPSA.$	0.272	0.105	↓ **0.167**
	$LPIPSV.$	0.369	0.11	↓ **0.259**
	$LPIPSS.$	0.215	0.079	↓ **0.136**
	VDP	7.648	8.611	↑ **0.963**
Reflections	$PSNR$	30.105	28.78	↓ 1.325
	$SSIM$	0.959	0.959	0.000
	$LPIPSA.$	0.133	0.136	↑ 0.003
	$LPIPSV.$	0.248	0.165	↓ **0.083**
	$LPIPSS.$	0.133	0.109	↓ **0.024**
	VDP	8.245	8.237	↓ 0.008
Refraction	$PSNR$	32.693	34.272	↑ **1.579**
	$SSIM$	0.965	0.962	↓ 0.003
	$LPIPSA.$	0.125	0.091	↓ **0.034**
	$LPIPSV.$	0.262	0.144	↓ **0.118**
	$LPIPSS.$	0.151	0.078	↓ **0.073**
	VDP	8.306	8.267	↓ 0.039
Subsurface Scattering	$PSNR$	30.865	31.618	↑ **0.753**
	$SSIM$	0.936	0.972	↑ **0.036**
	$LPIPSA.$	0.137	0.083	↓ **0.054**
	$LPIPSV.$	0.242	0.125	↓ **0.117**
	$LPIPSS.$	0.122	0.069	↓ **0.053**
	VDP	7.72	8.476	↑ **0.756**

data.

Considering the comparisons made in Section 6.1, the 3DGS reconstructions seem to score generally better than their NeRF counterparts, with a specific effect. These results are exciting since 3DGS reconstructions are bound to have higher resolutions at interactive frame rates according to Henriques *et al*. [2024a].

7 Conclusion

In this work, we validated a novel approach for the field of VR and Foveated Rendering, enabling ray-traced effects in virtual reality setups by replacing peripheral view by neural rendering approximations. We believe that, since NeRFs' color and depth information are world position and gaze direction dependent, they pose a fine candidate for replacing ray-traced effects in the peripheral view while preserving the same effects.

To achieve this goal, we used Nvidia's Instant-NeRF to generate scene reconstructions based on ray-traced effects and conducted a user study to understand better how people

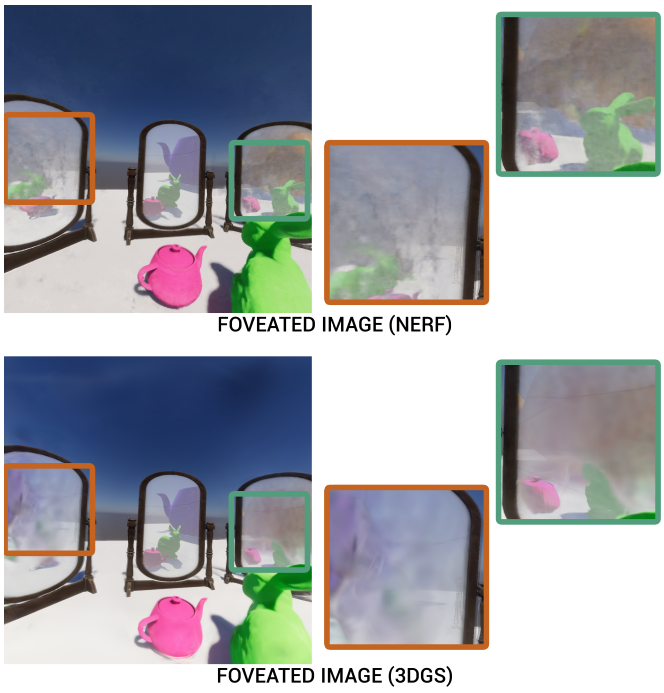

Figure 14. Comparison of the same frame generated with NeRF in the periphery, after manual tuning of training parameters, against the same view of the same scene with 3DGS in the periphery of the screen. We can see that the 3DGS reconstruction got the mirror blurrier than NeRF due to a combination of factors: number of epochs for optimizing the splats, input data focusing on the surface of the mirror, training parameters.

perceive such approximations. At the same time, we tested how necessary a visual aid is to fixate the user's gaze upon a foveal region in setups where there is no user gaze tracking.

According to our findings, NeRF is capable of better reproducing some ray-traced effects better than others. Effects that are clearly reproduced in well-lit scenes and mostly display other colors can be reproduced quite seamlessly, such as colored shadows or subsurface scattering. At the same time, effects that are only clearly visible in darkened scenes, or that depend on more complicated ray behavior, have more perceptible artifacts that users perceive as quality loss. Meanwhile, our observations pointed to the relevance of using a visual aid, such as a reticule, at the center of the foveal region to experiment in scenarios where eye tracking is impossible.

This work also explores 3DGS' capabilities as a surrogate for peripheral vision in VR, focusing on enhancing ray-traced effects that NeRF may have struggled to reconstruct. These new comparisons were made using image quality metrics (PSNR and SSIM) and visual quality metrics (FovVideoVDP and LPIPS). But first, we compared user assessment with these metrics over NeRF reconstructed images to see if one metric would better approximate the user feedback. These tests, although valuable, do not fully replace human assessment in a study such as this.

FovVideoVDP and LPIPS better captured the relation between user feedback and visual quality, but not perfectly. Even that being the case, every single metrics points to better visuals when using 3DGS over NeRF for rendering ray-traced effects, which is very exciting considering its better latency at higher resolutions.

This work was a deeper exploration of the limitations of radiance fields as a peripheral view candidate in ray-traced

environments, while at the same time providing insight on the role of having the foveal region under the user's fovea. As we deepen our understanding of these features, we can better direct future endeavors on the field as we explore new, inventive ways to incorporate these new technologies into foveated rendering pipelines.

7.1 Future Works

Although this study explored visual perception using one specific NeRF implementation for reconstructing the peripheral vision, we haven't assessed the cost of displaying such effects. However, the computational cost for real-time rendering, which is the main interest of such an application, will largely depend on the chosen technique to render the peripheral image.

Some other future works would include comparing these results against other radiance field-based techniques. Given the observable upsides of using 3DGS instead of NeRF for rendering peripheral vision of ray-traced scenes shown by Henriques *et al*. [2024b] and by this work, we believe it would be wise to experiment with recent solutions that explore light interaction with 3DGS. Some notable works are 3DGUT [Wu *et al*., 2024], aimed at reconstructing real-time mirrors in 3DGS, and GaussianShader [Jiang *et al*., 2023], aimed at applying some simplified shading to 3DGS reconstructions.

It must be noted that the technique described in this paper was taught for static scenes, which greatly limits its usage in interactive applications. New works such as GaussianFlow [Gao *et al*., 2024] and Motion-Aware 3DGS [Guo *et al*., 2025] present promising results for representing dynamic scenes, and extending such dynamism to be interactive would greatly expand the possibilities. On the subject of interactive scenes, works such as Gaussian Splashing [Feng *et al*., 2024] and GaussianAvatars [Qian *et al*., 2023] seem to be strong candidates for further investigation as well.

At last, it would be valuable to repeat the user experiments using the 3DGS rendered scenes instead of the NeRF ones to double-check if the metrics comparison from Section 6.3. As stated previously, although LPIPS and FovVideoVDP approximate user feedback, they do not perfectly correspond to user feedback.

Declarations

Acknowledgements

The authors would like to thank all volunteers who have participated in the experiment and to the entirety of the Medialab-UFF crew for their incredible support. The authors would like to specially thank José Machado from the Medialab-UFF for his contributions.

To support this study, the following software tools were used: Unity Engine, Instant-NGP and 3D Gaussian Splatting tools were constructed to compose the testing images; R, Excel and Python scripts were used to run statistical tests on the data collected via human testing and bench testing. The final version of this paper benefited from the use of Grammarly for improving spelling, grammar and vocabulary.

Funding

This research was funded by CAPES - Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and by FAPERJ - Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro.

Authors' Contributions

HM, EC and DT contributed to the conception of this study. EO and HM analyzed the results, with EO orienting on the most suitable statistical methods to employ. AO oriented while generating the artifacts for the experiments reported in Section 6. HM is this manuscript's main contributor and writer, as EC reviewed the entire paper, issuing corrections and clarifications, AO contributed to Sections 2 and 3, and EO contributed to Section 6. HM conducted the experiments. All authors read, reviewed and approved the final manuscript.

Competing interests

The authors declare they have no following competing interests.

Availability of data and materials

The application used on the user experiment (an Unity based VR-adapted slideshow for random pairs of images), and its images, can be found at `https://github.com/pusacaspica/HMD-Image-Viewer`. Images used in Section 6's experiment will be att the Misc folder on the same repository.

Both Instant-NeRF and 3D Gaussian Splatting are widely available to the public. The quality metrics' codes are also available to the public, be them implemented in traditional libraries .

References

Albert, R., Patney, A., Luebke, D., and Kim, J. (2017). Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 14(4):1–13. DOI: https://doi.org/10.1145/3127589.

Avidan, S. and Shashua, A. (1997). Novel view synthesis in tensor space. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pages 1034–1040. IEEE.

Banks, M. S., Sekuler, A. B., and Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *JOSA a*, 8(11):1775–1787. DOI: https://doi.org/10.1364/josaa.8.001775.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2022). Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479. DOI: https://doi.org/10.1109/CVPR52688.2022.00539.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. (2023). Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19697–19705.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). Cited by: 37248;

All Open Access, Gold Open Access, Green Open Access. DOI: https://doi.org/10.18637/jss.v067.i01.

Buehler, C., Bosse, M., McMillan, L., Gortler, S., and Cohen, M. (2001). Unstructured lumigraph rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 425–432.

Caulfield, B. (2022). "What Is Path Tracing?". NVIDIA Blog. https://blogs.nvidia.com/blog/2022/03/23/what-is-path-tracing/ (accessed May 08, 2025.).

Chen, X., Liu, J., Zhao, H., Zhou, G., and Zhang, Y.-Q. (2023a). Nerrf: 3d reconstruction and view synthesis for transparent and specular objects with neural refractive-reflective fields. *arXiv preprint arXiv:2309.13039*. DOI: https://doi.org/10.48550/arXiv.2309.13039.

Chen, Z., Funkhouser, T., Hedman, P., and Tagliasacchi, A. (2023b). Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16569–16578. DOI: https://doi.org/10.1109/cvpr52729.2023.01590.

Cui, Z., Gu, L., Sun, X., Ma, X., Qiao, Y., and Harada, T. (2024). Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 1435–1444. DOI: https://doi.org/10.1609/aaai.v38i2.27908.

Davis, A., Levoy, M., and Durand, F. (2012). Unstructured light fields. In *Computer Graphics Forum*, volume 2, pages 305–314.

Debevec, P. E., Taylor, C. J., and Malik, J. (2023). Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 465–474.

Deng, N., He, Z., Ye, J., Duinkharjav, B., Chakravarthula, P., Yang, X., and Sun, Q. (2022). Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3854–3864. DOI: https://doi.org/10.1109/TVCG.2022.3203102.

Deng, W., Campbell, D., Sun, C., Kanitkar, S., Shaffer, M., and Gould, S. (2024). Ray deformation networks for novel view synthesis of refractive objects. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3118–3128. DOI: https://doi.org/10.1109/wacv57701.2024.00309.

Dix, A., Finlay, J., Abowd, G. D., and Beale, R. (2003). *Human-computer interaction*. Pearson Education.

Duchowski, A. T., Bate, D., Stringfellow, P., Thakur, K., Melloy, B. J., and Gramopadhye, A. K. (2009). On spatiochromatic visual sensitivity and peripheral color lod management. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–18. DOI: https://doi.org/10.1145/1498700.1498703.

Einhorn, E. and Mawdsley, J. (2023). Generate Groundbreaking Ray-Traced Images with Next-Generation NVIDIA DLSS. https://developer.nvidia.com/blog/generate-groundbreaking-ray-traced-images-with-next-generation-nvidia-dlss/ (accessed May 13, 2025).

Elkin, L., Kay, M., Higgins, J., and Wobbrock, J. (2021). An aligned rank transform procedure for multifactor contrast tests. In *UIST 2021 - Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*, pages 754–768. Association for Computing Machinery, Inc. cited By 51. DOI: https://doi.org/10.1145/3472749.3474784.

Feng, Y., Feng, X., Shang, Y., Jiang, Y., Yu, C., Zong, Z., Shao, T., Wu, H., Zhou, K., Jiang, C., and Yang, Y. (2024). Gaussian splashing: Dynamic fluid synthesis with gaussian splatting. *CoRR*, abs/2401.15318.

Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., and Kanazawa, A. (2022). Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510. DOI: https://doi.org/10.1109/CVPR52688.2022.00542.

Gao, Q., Xu, Q., Cao, Z., Mildenhall, B., Ma, W., Chen, L., Tang, D., and Neumann, U. (2024). Gaussianflow: Splatting gaussian dynamics for 4d content creation. *arXiv preprint arXiv:2403.12365*. DOI: https://doi.org/10.48550/arXiv.2403.12365.

Genova, K., Cole, F., Maschinot, A., Sarna, A., Vlasic, D., and Freeman, W. T. (2018). Unsupervised training for 3d morphable model regression. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8377–8386.

Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F. (1996). The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, page 43–54, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/237170.237200.

Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. (2012). Foveated 3d graphics. *ACM Transactions on Graphics (TOG)*, 31(6):1–10. DOI: https://doi.org/10.1145/2366145.2366183.

Guo, Z., Zhou, W., Li, L., Wang, M., and Li, H. (2025). Motion-aware 3d gaussian splatting for efficient dynamic scene reconstruction. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(4):3119–3133. DOI: https://doi.org/10.1109/TCSVT.2024.3502257.

Henriques, H., de Oliveira, A., Oliveira, E., Trevisan, D., and Clua, E. (2024a). Foveated path culling: A mixed path tracing and radiance field approach for optimizing rendering in xr displays. *Journal on Interactive Systems*, 15(1):576–590. DOI: https://doi.org/10.5753/jis.2024.4352.

Henriques, H., Oliveira, E., Clua, E., and Trevisan, D. (2024b). Analysing hybrid neural and ray tracing perception for foveated rendering. In *Proceedings of the 26th Symposium on Virtual and Augmented Reality*, pages 21–30. DOI: https://doi.org/10.1145/3691573.3691580.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE. DOI:

https://doi.org/10.1109/ICPR.2010.579.

Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., and Ma, Y. (2023). Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. DOI: https://doi.org/10.48550/arXiv.2311.17977.

Kajiya, J. T. (1986). The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 143–150. DOI: https://doi.org/10.1145/15922.15902.

Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14. DOI: https://doi.org/10.1145/3592433.

Kilgariff, E., Moreton, H., Stam, N., and Bell, B. (2018). NVIDIA Turing Architecture In-Depth | NVIDIA Technical Blog. https://developer.nvidia.com/blog/nvidia-turing-architecture-in-depth/ (accessed May 13, 2025).

Kim, J., Kim, J., Jung, M., Kwon, T., and Kim, K. K. (2024). Individualized foveated rendering with eye-tracking head-mounted display. *Virtual Reality*, 28(1):25. DOI: https://doi.org/10.1007/s10055-023-00931-8.

Koskela, M. (2020). Foveated path tracing with fast reconstruction and efficient sample distribution.

Koskela, M., Lotvonen, A., Mäkitalo, M., Kivi, P., Viitanen, T., and Jääskeläinen, P. (2019). Foveated Real-Time Path Tracing in Visual-Polar Space. In Boubekeur, T. and Sen, P., editors, *Eurographics Symposium on Rendering - DL-only and Industry Track*, pages 1–12. The Eurographics Association. DOI: https://doi.org/10.2312/sr.20191219.

Levoy, M. and Whitaker, R. (1990). Gaze-directed volume rendering. In *Proceedings of the 1990 symposium on interactive 3d graphics*, pages 217–223. DOI: https://doi.org/10.1145/91385.91449.

Li, T.-M., Aittala, M., Durand, F., and Lehtinen, J. (2018). Differentiable monte carlo ray tracing through edge sampling. *ACM Transactions on Graphics (TOG)*, 37(6):1–11. DOI: https://doi.org/10.1145/3272127.3275109.

Liu, L., Gu, J., Lin, K. Z., Chua, T., and Theobalt, C. (2020). Neural sparse voxel fields. *CoRR*, abs/2007.11571.

Lu, C., Yin, F., Chen, X., Liu, W., Chen, T., Yu, G., and Fan, J. (2023). A large-scale outdoor multi-modal dataset and benchmark for novel view synthesis and implicit scene reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7557–7567. DOI: https://doi.org/10.48550/arXiv.2301.06782.

Mantiuk, R. K., Denes, G., Chapiro, A., Kaplanyan, A., Rufo, G., Bachy, R., Lian, T., and Patney, A. (2021). Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)*, 40(4):1–19. DOI: https://doi.org/10.1145/3450626.3459831.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106. DOI: https://doi.org/10.1145/3503250.

Mohanto, B., Islam, A. T., Gobbetti, E., and Staadt, O. (2022). An integrative view of foveated rendering. *Computers & Graphics*, 102:474–501. DOI: https://doi.org/10.1016/j.cag.2021.10.010.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):1–15. DOI: https://doi.org/10.1145/3528223.3530127.

Nilsson, J. and Akenine-Möller, T. (2020). Understanding ssim. *arXiv preprint arXiv:2006.13846*. DOI: https://doi.org/10.48550/arXiv.2006.13846.

Ogboso, Y. U. and Bedell, H. E. (1987). Magnitude of lateral chromatic aberration across the retina of the human eye. *JOSA A*, 4(8):1666–1672.

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, 35(6):1–12. DOI: https://doi.org/10.1145/2980179.2980246.

Porcino, T., Trevisan, D., and Clua, E. (2020). Minimizing cybersickness in head-mounted display systems: causes and strategies review. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*, pages 154–163. IEEE. DOI: https://doi.org/10.1109/SVR51698.2020.00035.

Qian, S., Kirschstein, T., Schoneveld, L., Davoli, D., Giebenhain, S., and Nießner, M. (2023). Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. *arXiv preprint arXiv:2312.02069*. DOI: https://doi.org/10.48550/arXiv.2312.02069.

Reiser, C., Peng, S., Liao, Y., and Geiger, A. (2021). Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. *arXiv e-prints*, pages 1–11. DOI: https://doi.org/10.48550/arXiv.2103.13744.

Schütz, M., Krösl, K., and Wimmer, M. (2019). Real-time continuous level of detail rendering of point clouds. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 103–110. DOI: https://doi.org/10.1109/VR.2019.8798284.

Series, B. (2012). Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT*, 500:1–46.

Setiadi, D. R. I. M. (2021). Psnr vs ssim: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6):8423–8444. DOI: https://doi.org/10.1007/s11042-020-10035-z.

Sorokin, A. and Forsyth, D. (2008). Utility data annotation with amazon mechanical turk. In *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 1–8. IEEE. DOI: https://doi.org/10.1109/CVPRW.2008.4562953.

Swafford, N. T., Iglesias-Guitian, J. A., Koniaris, C., Moon, B., Cosker, D., and Mitchell, K. (2016). User, metric, and computational evaluation of foveated rendering methods. In *Proceedings of the ACM Symposium on Applied Perception*, pages 7–14, Anaheim California. ACM. DOI: https://doi.org/10.1145/2931002.2931011.

Turki, H., Ramanan, D., and Satyanarayanan, M. (2022). Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12922–12931. DOI: https://doi.org/10.1109/CVPR52688.2022.01258.

Tursun, O. T., Arabadzhiyska-Koleva, E., Wernikowski, M., Mantiuk, R., Seidel, H.-P., Myszkowski, K., and Didyk, P. (2019). Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)*, 38(4):1–14. DOI: https://doi.org/10.1145/3306346.332298.

Ubrani, J., Llamas, R., and Reith, R. (2024). Ar vr headsets market insights. https://www.idc.com/promo/arvr (accessed Jan. 20, 2025).

Ujjainkar, N., Shahan, E., Chen, K., Duinkharjav, B., Sun, Q., and Zhu, Y. (2024). Exploiting Human Color Discrimination for Memory- and Energy-Efficient Image Encoding in Virtual Reality. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1*, pages 166–180, La Jolla CA USA. ACM. DOI: https://doi.org/10.1145/3617232.3624860.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J. T., and Srinivasan, P. P. (2022). Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE. DOI: https://doi.org/10.1109/CVPR52688.2022.00541.

Wandell, B. A. (1995). *Foundations of vision.* Sinauer Associates.

Wang, L., Shi, X., and Liu, Y. (2023). Foveated rendering: A state-of-the-art survey. *Computational Visual Media*, 9(2):195–228. DOI: https://doi.org/10.1007/s41095-022-0306-4.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612. DOI: https://doi.org/10.1109/TIP.2003.819861.

Weier, M., Roth, T., Kruijff, E., Hinkenjann, A., Pérard-Gayot, A., Slusallek, P., and Li, Y. (2016). Foveated Real-Time Ray Tracing for Head-Mounted Displays. pages 289–298. The Eurographics Association and John Wiley & Sons Ltd.. DOI: https://doi.org/10.1111/cgf.13026.

Whitted, T. (1979). An improved illumination model for shaded display. In *Proceedings of the 6th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '79, page 14, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/800249.807419.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. page 143 – 146. Association for Computing Machinery. Cited by: 1112; All Open Access, Green Open Access. DOI: https://doi.org/10.1145/1978942.1978963.

Wu, Q., Martinez Esturo, J., Mirzaei, A., Moenne-Loccoz, N., and Gojcic, Z. (2024). 3dgut: Enabling distorted cameras and secondary rays in gaussian splatting. *arXiv e-prints*, pages arXiv–2412. DOI: https://doi.org/10.48550/arXiv.2412.12507.

Yan, X., Xu, J., Huo, Y., and Bao, H. (2024). Neural rendering and its hardware acceleration: A review. *arXiv preprint arXiv:2402.00028*. DOI: https://doi.org/10.48550/arXiv.2402.00028.

Yang, W., Shi, J., Bai, S., Qian, Q., Ou, Z., Xu, D., Shu, X., and Sun, Y. (2024). Clear-plenoxels: Floaters free radiance fields without neural networks. *Knowledge-Based Systems*, 299:112096. DOI: https://doi.org/10.1016/j.knosys.2024.112096.

Yee, H., Pattanaik, S., and Greenberg, D. P. (2001). Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics (TOG)*, 20(1):39–65. DOI: https://doi.org/10.1145/383745.383748.

Yu, A., Li, R., Tancik, M., Li, H., Ng, R., and Kanazawa, A. (2021). Plenoctrees for real-time rendering of neural radiance fields. volume abs/2103.14024. DOI: https://doi.org/10.48550/arXiv.2103.14024.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. volume abs/1801.03924. DOI: https://doi.org/10.48550/arXiv.1801.03924.