Foveated RTX Ray Tracing in Virtual Reality

Uroš Šmajdek*

Supervised by: Ciril Bohak†

Faculty of Computer and Information Science
University of Ljubljana
Ljubljana / Slovenia

Abstract

In this paper, we present a foveated ray tracing method for enhancing the viewing experience of virtual reality environments featuring dense molecular structures. Achieving immersion in virtual reality requires both high-resolution and high frame-rate rendering combined with the ability to generate realistic images. Our approach combines the power of the Nvidia RTX framework with a dynamic rendering rate that is based on the direction of the user’s gaze. Foveated rendering is applied in the ray-generation stage of the ray-tracing pipeline, in order to reduce the ray-tracing rate depending on the distance from the center of the user’s gaze on the image. This takes advantage of the way the human eye perceives visual detail, allowing us to spare processing power on areas that are largely imperceptible. The nature of the ray-tracing pipeline also allows us to easily apply any post-processing effects, such as screen-space ambient occlusion and temporal anti-aliasing, in a matter similar to the standard deferred rendering pipeline. We demonstrate our technique by rendering a model of SARS-CoV-2 on Meta Quest Pro in 3776 × 3904 × 2 stereo resolution, where we observed a higher performance in comparison to the non-foveated counterpart of our method.

Keywords: Dense environments, ray tracing, foveated rendering, virtual reality

1 Introduction

Virtual reality (VR) is one of the key factors driving innovation in modern computer graphics. Development of immersive VR technologies has begun to encompass various fields from medicine [5, 10] to entertainment [30] and education [29], and as such, there is an increasing need for interactivity and real-time rendering. Additionally, over the last two decades, the pixel densities of high-quality head-mounted displays, also known as VR headsets, have dramatically increased from 263 × 480 × 2 in 1997 (Forte VFX 3D) to 2448 × 2448 × 2 in 2021 (HTC Vive Pro 2). To achieve the visual quality of a human eye, however, a display would require a resolution of about 32k × 24k [12], making such achievement far beyond the reach of the current hardware and software, but nonetheless ensuring it will continue for decades to come. Not only that but to achieve immersion and limit the potential motion sickness, such rendering would have to exceed 60 updates per second.

One of the recent applications for such rendering is virtual visualization of the molecular structures of biological specimens [21, 1], which aims to immerse the user in the molecular environment, using highly detailed scenes coupled with interactivity. Such software can then be used for science dissemination to the broader public, an increasingly relevant topic since the Coronavirus pandemic. It can also aid in scientific presentations, or be used as a tool to gather additional information from such structures. On the other hand, dense molecular environments provide additional challenges, as scenes contain hundreds of thousands and up to billions of objects, that form closely packed structures, representing proteins, RNA, membranes, etc.

Foveated rendering presents one of the approaches to solving this problem. It describes the dynamic adaptation of rendering based on the user’s gaze, specifically by limiting the details of the scene in the peripheral visual area that are largely imperceptible. The concept of foveated displays is not new, and gaze-reactive displays were already used in building flight simulators back in 2001 [25]. Until the recent generation of VR headsets, real-time eye tracking, which is required for its interactive use, was not commercially available, thus the approach was of lesser interest to a wider computer graphics community.

Hardware-accelerated RTX Ray Tracing is a recent breakthrough that promises to revolutionize the rendering scene, traditionally dominated by rasterization techniques. With the help of other Nvidia technologies, such as DLSS, it enables higher visual fidelity without compromising performance. Past research also indicates that the computational complexity of interactive ray tracing increases logarithmically with scene complexity [33], making it a compelling choice for addressing the above-mentioned interactive molecular visualization.

The main contributions of our work are:
• combining hardware-accelerate RTX Ray Tracing with Multi-spatial resolution-based foveated render-
Figure 1: A visualization of dense molecular environment depicting SARS-COV-2 using our rendering technique (left), with individual structure labels and hierarchy (right).

In section 2, we present the previous work on molecular visualization and resolution-based foveated ray tracing and differentiate our contributions from it. In section 3, we give a brief overview of the molecular environment we use as a basis. In section 4, we present our approach. The results and evaluation of our method are presented in sections 5 and 6. In section 7, we present the conclusions and give possible extensions as part of further work.

2 Related Work

Molecular Visualization: There are various tools for smaller-scale molecular visualization, such as VMD [11], Mol* [32], or PyMOL [31] which are unsuited for larger data sizes exceeding tens of millions of individual atoms. Over the years, numerous solutions to interactive molecular visualization have been proposed; glyph-based rendering [9], mapping structures onto a mesh geometry [36], using meshlets with probabilistic occlusion culling [13], instancing to repeat the recurring proteins in the scene [22, 6]. In 2015 Le Muzic et al. introduced cellVIEW [26], which depends on the LoD scheme and dynamic sphere primitives injection into primitives to be able to render 250 copies of the HIV virus model in blood plasma (16 billion atoms) at 60 FPS on NVIDIA GTX Titan. All of the above-mentioned techniques use procedural impostors to simplify the geometry and accelerate the rendering. Recently Alharbi et. al. [2] devised a solution that relies on hardware ray tracing instead and is able to render trillions of atoms using on-the-fly data construction and parallel rendering. In this paper, we expand on this paradigm but move the focus from scalability to higher performance, suitable for interactive VR visualization, which requires stereo rendering with a high and steady framerate. To this end, we supplement the hardware ray tracing with DLSS-enhanced foveated rendering but limit ourselves to data scales that directly fit into the GPU memory. To the best of our knowledge, there are no alternative foveated rendering techniques of dense molecular environments.

Multi-spatial resolution-based foveated ray tracing: Koskela et al. [20] proposed a theoretical estimate in 2016 that by integrating foveated rendering with ray tracing, 94% of the rays could potentially be omitted. This is based on the fact that ray tracing inherently supports spatial multi-resolution rendering, as it has the capacity to effortlessly adjust the number of rays emitted from a single pixel. Fujita and Harada [8] first implemented the foveated rendering system based on ray tracing, which relied on sparse sampling and kNN for image reconstruction. The system did not consider the eye sensitivity to contrast and lacked pertinent input from relevant user studies. To address these challenges, Weier et al. [38] combined ray tracing-based foveated rendering with reprojection rendering, using information from the previous frame to reduce the sampling rays for new frames. From there on out, different sampling models were proposed; Molenaar [24] traced rays based on the visual acuity fall-off model, Kim et al. [17] proposed a perceptually efficient pixel sampling method suitable for HMD ray tracing, which combined the Jin et al. [16] selective oversampling technique with the foveated rendering scheme and Koskela et al. [19] traced rays and denoised in Visual-Polar space and subsequently mapped the results to the screen space. Hybrid approaches also emerged with Blackmon et al. [4] using ray tracing to render the foveal region and rasterization to render the peripheral region. To address the lack of consideration for the displayed content, as opposed to the eccentricity, when
determining the number of rays emitted, Tursun et al. [35] proposed a new luminance-contrast-aware foveated ray tracing technique. The technique significantly reduced the number of traced rays but required a low-quality image to be generated for each frame, indicating areas with different luminances. Similarly, Yang et al. [39] introduced a trio of simple policies to achieve a variable ray tracing rate. The method integrates the foveate perspective, material, and depth information in order to dynamically reduce the number of tracing rays in certain steps in the ray tracing pipeline. Recently Wang et al. [37] compiled an in-depth overview of state-of-the-art foveated methods for rasterization, ray tracing, and volumetric rendering techniques. In our approach, we implement a low-overhead spatial policy to determine the local number of rays, similar to the one introduced in [39], but we further supplement it using a hardware-accelerated Nvidia Ray Tracing framework [18]. Additionally, we use traditional spatial and temporal anti-aliasing techniques in favor of Nvidia DLSS [23]. While there is no substantial contribution from individual methods, their combination, however, is a novelty. Moreover, some components of the pipeline utilize methods that we were unable to find elsewhere, including a single multi-resolution ray tracing pass and blending between foveal regions.

3 Molecular Environment

As input, we used the atom-based hierarchical 3D models of biological mesoscale organisms, such as viruses. The model may consist of an arbitrary number of instances, each composed of densely packed atoms, with the rest of the space implicitly filled with water molecules. We choose not to render those molecules as they are of less interest to the viewer and would render the scene completely packed. Those spaces are instead treated as empty, allowing the user to focus and travel between important structures found inside a virus. An example of a model's hierarchy can be seen in Figure 1. Additionally, employing such a hierarchical structure allows us to dynamically highlight different levels of structures based on distance and type, further enhancing the user experience.

4 Foveated Ray Tracing

In this section, we describe our proposed foveated rendering system. An overview of the rendering pipeline can be seen in Figure 2, and we describe the individual stages in the following subsections. The pipeline is modular, the only static part being the visual acuity test and the hardware-accelerated ray tracing. Other parts can be easily swapped out or new ones added, which is made even easier by its strong resemblance to a classic deferred rendering pipeline.

4.1 Visual Acuity Test and Ray Generation

Visual acuity generally refers to our ability to distinguish the shapes and details of the object we are observing. One of its key properties is the foveal/peripheral vision, which recognizes that human visual acuity is not uniform over the entire visual field. Objects in the periphery are significantly harder to recognize than those in the center of the fovea [34], as illustrated in Figure 3.

We use this foveal/peripheral property of the visual acuity to determine the resolution at which we render different parts of the final image. To achieve this, we perform a vi-
Figure 4: Visual representation of the multi-resolution rendering based on the eccentric angle from the user’s gaze. The red squares in the center each cover one pixel, and larger squares denote rendering in lower resolution.

4.2 Ray Tracing Pipeline

Our ray tracing pipeline heavily relies upon the Nvidia RTX ray tracing extension for Vulkan [18], which is the continuation of Nvidia OptiX [28], a general-purpose SDK for accelerating ray casting applications. Using the framework, we pack the atoms that build our model into Accelerated Structures, a specialized high-performance spatial structure built for ray tracing. The framework then takes care of optimizing the ray collision detection using bounding volume hierarchy traversal while we still retain full control of the ray generation and after-collision events. To be able to use deferred rendering techniques later, we store not only the colors of hit objects but also the hit location, the normal of the hit surface, and the hit object’s numerical identifier.

To further reduce the overhead in the rendering pipeline, we only perform the ray tracing procedure once instead of separately for each resolution we render. We achieve this by inherently rendering at full resolution and then discarding the unneeded rays in the periphery.

4.3 Composition

During composition, we combine the 2D multi-resolution output of the ray tracing procedure. Since rendering at different resolutions results in spatial discontinuities between the rendering regions, we blend different resolutions at their border, as opposed to simply taking the highest resolution present at that location, which results in smoother transitions. To that end, we employ an alpha-blending technique to reduce the computational overhead of the procedure.

4.4 Shading

As our model does not depict an environment in which realistic lighting plays a role, we can simplify the illumination in ray tracing algorithm to focus solely on local illumination with Phong shading. This optimization not only improves performance by reducing the number of traced rays but also eliminates the probabilistic computations typically found in ray tracing algorithms, resulting in immediate convergence of the final result.

Another challenge that the molecular environment presents is the discrepancy between the macro-level and micro-level perspectives. At the macro level, we are interested in the various structures that make up our model, such as membranes or RNA, while at the micro level, we focus on individual atoms that comprise those structures. To seamlessly bridge these two levels, we modulate the color of each atom by interpolating between its inherent color, $c_a$, and the color of its corresponding structure, $c_s$, based on the atom’s distance from the camera, $d$:

$$c = \begin{cases} (1 - \frac{d}{D_{\text{max}}}) \cdot c_a + \frac{d}{D_{\text{max}}} \cdot c_s & : d < D_{\text{max}} \\ c_s & : d \geq D_{\text{max}} \end{cases}$$

Any atom located beyond a pre-determined distance, $D_{\text{max}}$, is assigned the color of its structure, making it easier to differentiate between the structures from a distance. The colors are user-defined.

4.5 Screen-Space Ambient Occlusion

As we are dealing with primarily mono-colored, locally illuminated base shapes of the same size, an ambient occlusion technique is needed to help the user distinguish between them and also improve depth perception. Screen Space Ambient Occlusion (SSAO) is well suited for this...
task, as its complexity does not scale with the number of objects in the scene and is generally better in dense environments due to the improved locality of GPU memory accesses. Since our model only contains spheres, we use the hemisphere sampling variant of the technique, first described by Fillion and McNaughton [7].

4.6 Post processing

In the post-processing stage, we draw atom contours in order to ease the distinction between neighboring objects. This is especially important when working in dense molecular environments, as the high number of small objects is inherently difficult to distinguish. To determine whether to draw a contour at a pixel or not, we use previously computed object identifiers to check if all neighboring pixels belong to the same object. If they do not we draw a semi-transparent contour and blend it with the pixel's color.

5 Results

We evaluated the proposed technique on the SARS-CoV-2 model [27], which consists of 74,724,996 atoms. Experiments were rendered to a Desktop application window with a rendering resolution of 3840 × 2160, and the Oculus Pro VR headset with a stereo rendering resolution of 3776 × 3904 × 2. All rendering was done using a Nvidia GeForce RTX 4090 graphics card with 24 GB of VRAM, an AMD Ryzen Threadripper PRO 3995WX 64-core processor, and 256 GB of RAM.

To discern the impact of the inclusion of foveated rendering, we measured the number of rendered frames per second (FPS). We performed 4 different tests for different combinations of the inclusion of foveated rendering and SSAO. During each experiment, we measured the average framerate over a duration of 30 seconds. To better represent a realistic usage scenario we slowly moved the camera around for the entire duration. As the framerate is restricted to 72 FPS on Meta Quest Pro, we chose the scene that closely matches this framerate to minimize the impact on final results. The results can be seen in Table 1.

The addition of foveated rendering almost doubled the performance both on Desktop and in VR as long as SSAO is not used. Enabling the SSAO resulted in a massive performance drop, especially when foveated rendering was also enabled. For qualitative evaluation, we present renderings of all four states in Figure 5. The two left segments show the rendering without foveation and the two right segments with foveation. The top two segments show the result with SSAO disabled, and the bottom two segments show the result with SSAO enabled.

<table>
<thead>
<tr>
<th>Display</th>
<th>SSAO</th>
<th>Foveated</th>
<th>FPS</th>
<th>rFPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop</td>
<td>No</td>
<td>No</td>
<td>288.3</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>553.0</td>
<td>92%</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>126.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>159.0</td>
<td>26%</td>
</tr>
<tr>
<td>Oculus Pro</td>
<td>No</td>
<td>No</td>
<td>39.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>71.8</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>24.6</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>36.1</td>
<td>47%</td>
</tr>
</tbody>
</table>

Table 1: Performance evaluation of the proposed foveated rendering pipeline for dense molecular data. For reference, we rendered the same model without foveated rendering and also repeated both experiments without the inclusion of SSAO. The maximum refresh rate on Oculus Pro is 72 Hz, which represents the maximum measured framerate on the system. rFPS is defined as relative improvement using foveation on the specific display and SSAO setting.

6 Discussion

Our tests have revealed that foveated rendering by itself massively improves the performance (see Table 1). The results are less significant when the technique is used simultaneously with SSAO. We believe this is due to the lower resolution producing higher spatial discontinuities between neighboring pixels in low-resolution regions, resulting in a higher number of GPU cache misses. By itself, SSAO also massively reduces performance, as observed by the 38%-56% framerate drop when applied without foveation. This likely signifies that SSAO is not a suitable candidate for rendering dense molecular data in high resolutions.

From Figure 5, we conclude that the visual differences between the foveated and non-foveated results are rather minimal, especially close to the gaze direction. Internal user testing has shown that during movement, the aliasing in the distant object in the peripheral regions can be distracting. To address this issue, a performance-friendly solution would be to apply a blur filter to both peripheral regions in the post-processing stage. On the other hand, the figure also showcases the importance of ambient occlusion, without which it becomes difficult to distinguish between the objects, but we do not believe it justifies the performance loss in terms of user experience.

7 Conclusions

In this paper, we present a multi-spatial resolution-based foveated ray tracing method for the visualization of dense molecular environments in virtual reality. We evaluate the technique, both with and without the addition of ambient occlusion, showing a moderate to large increase in performance, depending on whether SSAO was used or not. We
also showcase the need for ambient occlusion in terms of being able to tell the objects apart in a dense molecular environment.

As a future extension of this work more performant ambient occlusion techniques could be employed [3, 15], as the usage of SSAO resulted in a massive performance loss. Additionally, the method could be expanded by also considering the context of the scene during the visual acuity test, and thus being able to modulate the rendering resolution of different structures depending on other criteria, such as their depth and importance to the user.

References


