Acceptable System Latency for Gaze-Dependent Level of Detail Rendering

Michał Chwesiuk*
Krzysztof Wolski†
Supervised by: Radosław Mantiuk ‡

West Pomeranian University of Technology
Szczecin / Poland

Abstract

The human visual system is unable to perceive all details in the entire field of view. High frequency features are noticeable only at a small angle of 1-2 degrees around the viewing direction. Therefore, it is a reasonable idea to render a coarser object representations for the parafoveal and peripheral visions. A core problem of this gaze-dependent level-of-detail rendering is the minimisation of the system latency. In this work we measure how fast the whole process of rendering and visualisation should be to prevent that a level-of-detail change will be visible for human observers. We noticed that even for distant periphery, the change from coarser to fine object representation should take less than 24 ms. It can be obtained only in systems equipped with the high-end eye tracker and a display with a refresh rate of 120 Hz or faster.

Keywords: system latency, gaze-contingent display, level-of-detail, LOD, eye tracking, real time computer graphics

1 Introduction

One goal of the *level-of-detail* (LOD) technique is to quickly change between coarse and fine representations of the object geometry [3]. Objects with smaller number of polygons can be rendered faster than their fine representation. Therefore, if an object occupies limited number of pixels in the final rendering or it is merely visible, it is more efficient to use its coarser representation. The goal is to find a proper level of detail for an object, taking into account its visibility on the screen.

The human visual system (HVS) is unable to perceive all details in the entire field of view. High frequency features are noticeable only at a small angle of 1-2 degrees around the viewing direction, otherwise details are imperceptible. Areas outside foveal are called parafoveal and peripheral regions. This nonlinear sensitivity of the eye is defined by the gaze-dependent contrast sensitivity function (CSF) [15], which models the sensitivity to contrast as function of eccentricity (i.e. distance from the gaze direction).

In order to increase performance of rendering, it is a reasonable idea to render the coarser object representations for the parafoveal and peripheral visions. In the gaze-contingent graphics systems, information about the gaze direction must be delivered to the rendering engine. The angular distance between momentary gaze location and position of the object in the screen space, will be a determinant of the model simplification.

A core problem with the implementation of such systems is the minimisation of the *system latency*. The gaze direction must be captured by the eye tracker, the image must be rendered, and finally the display device needs some time to present the image on the screen. If the total processing time would be too long, the observer could see the object changing between the coarse and fine representation. In this work we investigate how short the system latency should be to make LOD modifications imperceptible to a human observers.

We perform a perceptual experiment, in which two geometric objects are rendered on one side of the screen. The first object consists of a large number of polygons and acts as the reference (or fine) representation. The second object is its simplified (or coarse) version with a reduced number of polygons. We asked observers to look at the marker located on the opposite side of the screen. The eye tracker is used to detect the moment, in which observer turns his/her eyes to look at the objects. In this moment, the image is redrawn with both objects using the fine representation. The task of the observer is to identify which of the objects were rendered with the reduced number of polygons. During the experiment we changed the display refresh rate to differentiate the system latencies.

In Section 2 of this paper we provide basic information regarding the human peripheral vision, eye tracking, and latencies of the gaze-dependent rendering systems. We present our gaze-dependent LOD rendering environment and review the previous work related to similar systems in Section 3. Section 4 presents details of the conducted ex-

*mchwesiuk@wi.zut.edu.pl
†krwolski@wi.zut.edu.pl
‡rmantiuk@wi.zut.edu.pl

periment and we discuss acceptable latencies of the gaze-dependent LOD rendering in Section 4.4. The paper ends with conclusions and propositions for future work in Section 5.

2 Background

2.1 Visual resolution

The *visual resolution* of the human eye is measured in terms of contrast sensitivity [8]. A stimulus consisting of the alternating bars of a grating (e.g. Gabor pattern) is presented to observers. They decide what contrast is needed to see the bars at each frequency, while the contrast is defined as the difference in brightness between light gray and dark gray bars. The threshold contrast values as a function of spatial frequencies form the *contrast sensitivity function* (CSF, [2]).

However, people can see details with the frequency defined by the CSF only in a small viewing angle, which subtends 1-2 degrees around the gaze direction. The loss of visual resolution increasing of viewing angle is caused by decreasing number of cones (light-sensitive cells) in the parafoveal and peripheral regions of the retina. This trait is described by a *gaze-dependent contrast sensitivity function*, which shows how contrast sensitivity varies as a function of distance from the fovea [15, 8]. Fig. 1 presents a plot of the perceptible signal frequency as a function of eccentricity . This frequency defines the highest frequency of the Gabor pattern, which is still recognized by human observer. An important observation is that the visual resolution decreases rapidly for higher spatial frequencies and e.g. for a eccentricity of 20 degrees it becomes one-tenth of the maximum resolution. A typical 22-inch LCD display is seen in a viewing angle of 40 degrees, therefore, the geometry of the object located at the screen corner can be significantly reduced if the observer does not look directly at it.

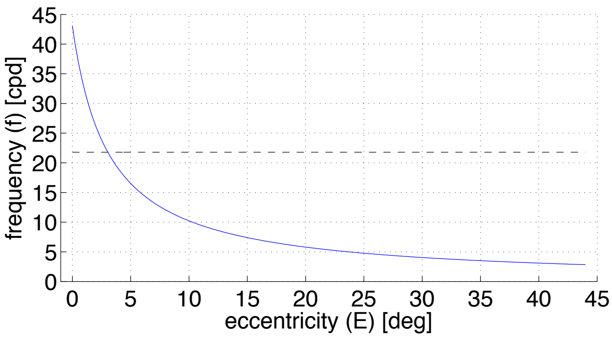

Figure 1: Visible spatial frequencies as a function of viewing angle (the plot is based on the formulas delivered in [8]). The dashed line shows the maximum frequency of a typical LCD display.

2.2 Eye tracking

The principle of eye tracking is based on the observation that the pupil follows the gaze direction during eye movement [4]. Therefore, the location of the pupil centre can be used to estimate the gaze direction. A popular technique employed to localized the pupil center is the modeling of the iris shape (for an excellent review of models for the eye detection we refer to Hansen and Ji work [5]). The eye tracker camera captures an image of the eye. The location of the pupil centre is detected in this image. This location must be transformed from the camera space to the screen space to estimate the gaze position on the screen. This is done using a polynomial transformation as a mapping, which parameters are determined during eye tracker calibration. During calibration, people are asked to look at the target points displayed on the screen. Then, known locations of the target points and data captured by the eye tracker are used to compute the polynomial coefficients. Finally, this polynomial is applied to transform the pupil centre from the camera to screen space.

The human visual system scans the surrounding withits eyes to build a complete view of the environment. The rapid repositioning of the pupil (called saccadic movement) can reach up to 900°/sec. To capture this movement, the eye tracker should work with a latency less than 5 ms [12], which is equivalent to a frequency of 200 Hz. In practical systems, this frequency needs to be even higher, because of the additional time needed to render and display the image.

2.3 System latency

The gaze-dependent rendering system uses the gaze direction captured by the eye tracker to control the image rendering process. For example, an object's geometry can be simplified if the object is positioned far away from the gaze location. Gaze-dependent systems work in real time, i.e. the image redrawing (including its visualisation) must be imperceptible to the human. According to Loschky and McConkie [12, 9], the *latency* of such systems should be less than 22 ms (5 ms for gaze capture and rendering, and additional 17 ms for visualisation on a 60 Hz display).

As shown by Saunders and Woods [17], the latency of the gaze-dependent rendering system ranges from 12 ms for CRT display, 18 ms for DLP projectors, to over 30 ms for low quality LCD displays. However, high-end LCDs with a short display lag can speed-up this process to about 18 ms, which is enough for the gaze-dependent LOD simplification. Loschky et al. [10] measured the system latency using a technique proposed in [1]. They report mean latency of 20 ms for the 1000 Hz EyeLink eye tracker working with a 85 Hz CRT display .

In this work, we use a LCD display, which has a maximum display frequency of 144 Hz (or a display latency of less than 7 ms).

2.4 Previous work

An early work on the gaze-dependent level-of-detail was presented by Mark Levoy [6]. The complexity of the volumetric data was reduced to speed-up the volume rendering method. The author used a precomputed pyramid of 3D texture volumes to skip some complex data structure that were far away from the viewing direction.

In Ohshima et al. [14] a concept for the visual acuity was proposed. This model examines the central/peripheral vision, kinetic vision, and fusional vision to cluster objects of low acuity and render them using simplified versions. The model was tested using a head tracker.

Reddy [16] investigates the perceptual content of a computer-generated image in terms of spatial frequency. The level-of-detail of each object is based on a screen-based measure of the degree of spatial detail which the user can perceive at different distance, angular velocity, and the degree to which it exists in the peripheral field. The author reports a factor 4.5 improvement in frame rate. Reddy also proposes a polygon simplification framework to complement the use of perceptually modulated LOD. However, it is not clear how this framework was used and whether a eye tracker was applied during experiments.

Watson et al. [18] studied the effect of peripheral LOD degradation on the visual search performance. He used a head mounted display to show a high resolution inset within a low resolution display field. The obtained results indicate that the area of the high detail central inset is a significant factor in search performance. However, Watson suggests that visual spatial and chrominance complexity can be reduced by almost half without degrading perceived quality.

A remote monocular eye tracker was used in [11] to measure the viewers real- time gaze location. The authors developed a classic LOD technique, in which objects geometry is simplified according to eccentricity.

In Murphy and Duchowski [13] objects degradation is applied nonisotropically, i.e. only a parts of large object are smoothly reduced. A three-dimensional spatial degradation function is obtained from human subject experiments and applied directly to object geometries prior to rendering. The technique was implemented in the rendering system integrated with a binocular eye tracker. The results indicate a frame rate improvement ranging from a factor of at least 2, up to a 15-fold gain in performance over full resolution display.

Players perception to level of detail (LOD) changes while playing a computer game is investigated in Lopez et al. [7]. The simplified models were unrelated to the task assigned to the player and located away from the area in which the task was being accomplished. Thus, a perception of LOD modification was tested under the inattentional blindness. The results show that players were able to detect only about 15% of LOD changes during the game.

In this work we implemented the LOD simplification approach similar to Luebke [11] technique. However, our main goal is to investigate a perception of the LOD change in a real time rendering application. Therefore, we used apparatuses and techniques that enable the fastest rendering and visualisation of an image possible.

3 Gaze-dependent LOD

3.1 Implementation

Fig. 2 illustrates a gaze-dependent rendering and visualization system. The observer looks on the display. Her/his gaze direction is captured by the eye tracker, which computes the gaze point location on the screen. The graphics engine uses this gaze location to render the scene. The scene contains objects whose complexity depends on the eccentricity. The object close to the gaze point consists of a larger number of triangles than its simplified version seen from a high angle.

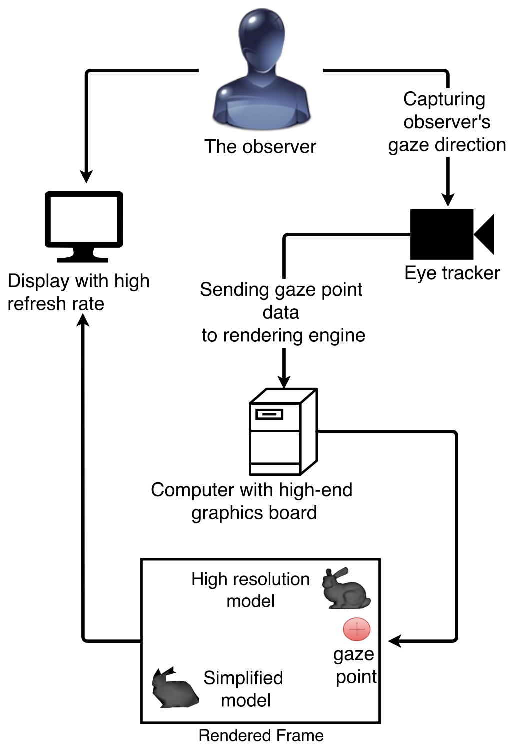

Figure 2: Diagram of the gaze-dependent rendering system.

3.1.1 Eye tracker

In our rendering system we use the Mirametrix S2 eye tracker equipped with a 60 Hz camera. The S2 is a portable device, which should be placed under the display in front of the observer. Before each session the eye tracker must be calibrated. After successful calibration the software sends gaze locations to our eye tracker communication server using the TCP/IP protocol (see Fig. 3), which collects the gaze data and send them to the rendering engine using the shared memory. We obtained an accuracy of Mirametrix S2 close to 1 degree of visual angle, which is sufficient for our experimental application.

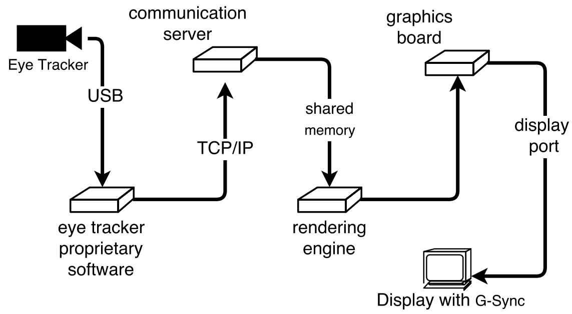

Figure 3: Hardware and software architecture.

3.1.2 Rendering framework

We developed a test framework for fast visualisation of the complex objects. This application is able to render objects consisting of more than 4k triangles in less than 1 ms. It was implemented in C++ and is based on the OpenGL library supported by GLEW and GLFW extensions. The application applies Phong shading and 16-samples multisample antialiasing.

Figure 4: Example screenshot from our application. The green cross on the left side depicts location of the gaze point captured by the eye tracker.

In Fig. 4 an example screenshot from our application is presented. The Stanford Bunny models are rendered in 0.89 ms. The object at the top was reduced to 100 triangles, while the bottom one consists of 4k triangles.

3.1.3 Visualisation

The rendered images are displayed on the fast LCD with a display lag of 1 ms and the maximum screen refresh rate of 144 Hz. This display is equipped with the G-Sync electronic, which has a positive impact on the gaze-dependent LOD systems that works on slower hardware. However, we do not use this feature because our rendering application is fast enough to finish the calculations in the required time interval. The main advantage is that G-Sync-supported displays were the fastest commercially available LCD display at the time when we performed our experiments (recently, 165 Hz displays were issued).

Figure 5: Hardware setup used in the experiment.

4 Experiment

The main goal of the experiment was to find the acceptable system latency, i.e. how fast the object should be redrawn on the display after changing the LOD level to avoid that a human observer would notice this change.

4.1 Procedure

The observer sat in front of the display and used the chin-rest adopted from an ophthalmic slit lamp to stabilise her/his eyes in 75 cm distance from the screen (see Fig. 5). The experiment started with a 9-point calibration of the eye tracker. This procedure took about 20 seconds and involved observation of the markers displayed in different areas of the screen. The data processing related to the calibration and further gaze location computation was performed by the proprietary eye tracker software.

During the actual experiment the observer was asked to look at a red cross presented on a 18% grey background (see Fig. 6, top row). After half a second two objects were shown on the left side of the screen in 10°, 20°, or 35° angular distances from the red cross (see Fig. 6, middle row). One of the objects was composed of a large number of polygons and considered as a reference. We reduced the mesh complexity of the second object to a number of polygons that prevent distinguishing this object from the reference. This simplification depends on the angular distance between the observer's gaze point and the object (based on the lower resolution of the human eyes in the periphery). The objects were displayed above each other. Each time it was randomly chosen whether the high or low resolution mesh would be displayed at the top.

Then, the observer's task was to look at the objects and decide which one was a simplified version. She/he pressed the up/down cursor buttons to indicate the choice. As the observer's gaze were captured by eye tracker, we could replace the simplified version of the object with the reference consisting of 4000 polygons as soon as the gaze moved away from the initial position (see Fig. 6, bottom row). More precisely, we switched the level-of-detail when the gaze location moved by 4 degrees from the initial position.

Our eye tracker operates at a 60 Hz frequency, i.e. we were able to replace objects not earlier than after a 17 ms delay. Additional latency derived from the display. We tested the display working at 30 Hz, 60 Hz, 120 Hz, and 144 Hz, which corresponds to the delays of 33 ms, 16 ms, 8 ms, and 7 ms, respectively.

The experiment was repeated for 3 angular distances and 4 display frequencies resulting in 12 trials per observer. Additionally, we repeated each trial 30 times in random order to obtain averaged results. The experiment was performed in a darkened room. We used 22 ASUS ROG SWIFT PG278Q LCD display with native resolution of 2560 x 1440 pixels. The rendering was performed on a PC equipped with NVIDIA 780 GTX graphic card.

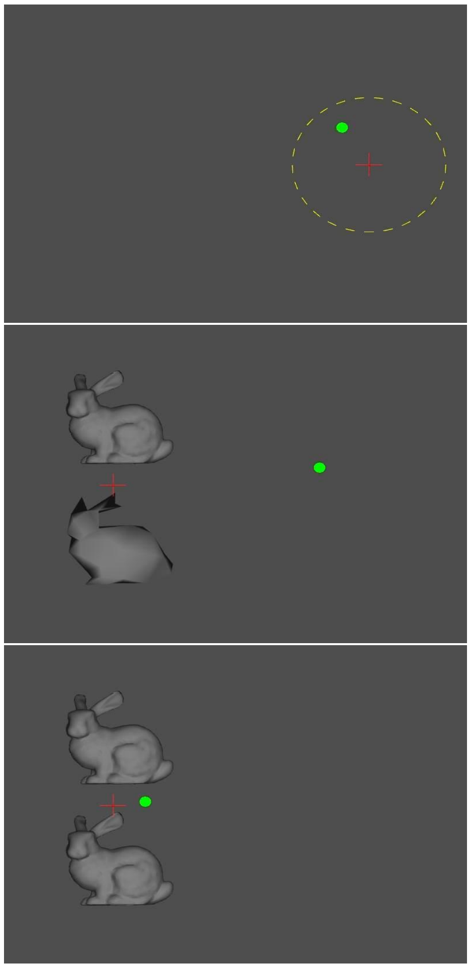

Figure 6: Succeeding phases of the experiment. The green spot depicts the observer's gaze location.

4.2 Stimuli

We generated simplified versions of the Stanford Bunny geometric model using the Quadric Edge Collapse Decimation algorithm in MeshLab [1]. The degree of simplification has been chosen in a separate pilot experiment. We searched for a minimum number of polygons, which do not cause the perceptual difference in comparison with the reference model consisting of 4,000 polygons. The experiment was repeated for 3 angular distances because, as we noticed, smaller distances require more precise models. The results of this pilot experiment show that object consisting of 2000, 1600, and 1000 polygons are suitable for 10°, 20°, and 35°, respectively (see Fig. 7).

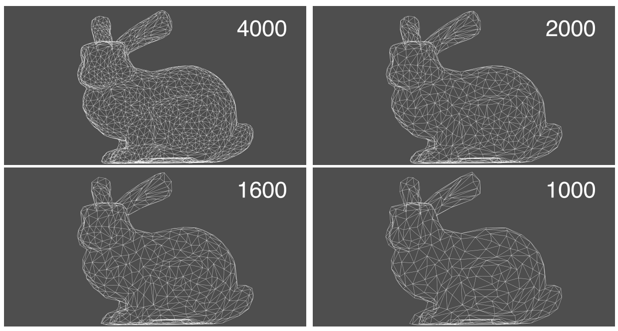

Figure 7: Stanford Bunny reference object (4000 polygons) and its simplified versions with 2000, 1600, and 1000 polygons.

4.3 Participants

We performed the experiment for a group of 10 volunteer observers (age between 20 and 23 years, 2 females and 8 males). They declared normal or corrected to normal vision and correct color vision. The participants were aware what they should do, but they were naïve about the purpose of the experiment. An average experimental session lasted approximately 12 minutes.

4.4 Results

The results of the experiment are presented in Fig. 8. The plot shows the normalised ratio of correct answers (correct

[1] `http://meshlab.sourceforge.net/`

indications on simplified objects) as a function of the display frame rate. The ratio of 0.5 (horizontal dashed line in Fig. 8) is equivalent to the random choice, i.e. indicates inability to distinguish between reference and simplified models. In our study only for the display refresh rate of 144 Hz and the angular distance of 35° the results are close to this line. In all other cases the system latency was to long to ensure imperceptible change of the level-of-detail. Especially, for smaller viewing angles the redrawing is clearly visible.

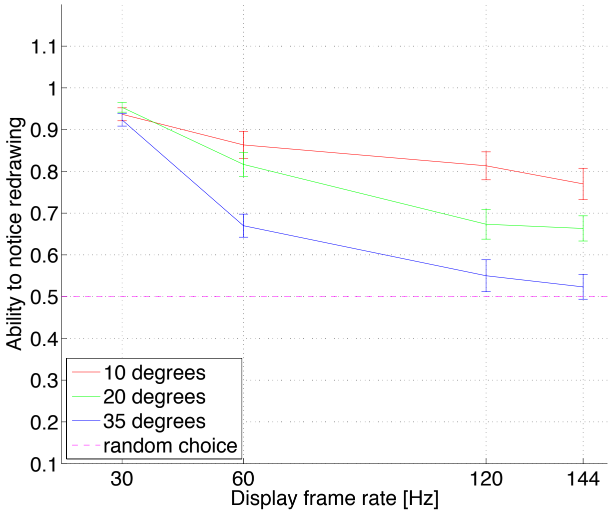

Figure 8: Results of the perceptual experiment. The error bars show the standard error of the mean.

5 Conclusions and Future Work

In this work we conducted a comprehensive evaluation of the acceptable system latency for the gaze-dependent LOD rendering. Our study included a psychophysical experiment which allowed us to evaluate perception of the LOD change for various display refresh rates, ranging from 33 to 7ms, and for different viewing angles. In the experiment we used a fast 144 Hz display but also slow a 60 Hz eye tracker, which introduced additional 17 ms delay. The results of the experiment show that the total system latency in our gaze-contingent system is too long for the imperceptible LOD change. Only for the angular distance of 35° and the latency close to 24 ms (17 ms for eye tracker and 7 for display), the LOD redrawing was unnoticeable for observers.

In future work we plan to test faster eye trackers, which captures the gaze location in less than one refresh cycle of the display (less than 7 ms in our case). We also plan to develop a technique of the LOD blending instead of immediate switching object's geometry from coarse to fine. This solution would increase the acceptable system latency. Finally, we would like to perform a user study of the gaze-dependent level-of-detail rendering in a complex computer game environment.

References

[1] Jean-Baptiste Bernard, Scherlen Anne-Catherine, and Castet Eric. Page mode reading with simulated scotomas: A modest effect of interline spacing on reading speed. *Vision research*, 47(28):3447–3459, 2007.

[2] Fergus W Campbell and JG Robson. Application of fourier analysis to the visibility of gratings. *The Journal of physiology*, 197(3):551–566, 1968.

[3] James H Clark. Hierarchical geometric models for visible surface algorithms. *Communications of the ACM*, 19(10):547–554, 1976.

[4] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice (2nd edition)*. Springer, London, 2007.

[5] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):478–500, 2010.

[6] Marc Levoy and Ross Whitaker. Gaze-directed volume rendering. *ACM SIGGRAPH Computer Graphics*, 24(2):217–223, 1990.

[7] Francisco Lopez, Ramon Molla, and Veronica Sundstedt. Exploring peripheral lod change detections during interactive gaming tasks. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 73–80. ACM, 2010.

[8] Lester Loschky, George McConkie, Jian Yang, and Michael Miller. The limits of visual resolution in natural scene viewing. *Visual Cognition*, 12(6):1057–1092, 2005.

[9] Lester C Loschky and George W McConkie. Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied*, 8(2):99, 2002.

[10] Lester C Loschky, Ryan V Ringer, Aaron P Johnson, Adam M Larson, Mark Neider, and Arthur F Kramer. Blur detection is unaffected by cognitive load. *Visual cognition*, 22(3-4):522–547, 2014.

[11] David Luebke and Carl Erikson. View-dependent simplification of arbitrary polygonal environments. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 199–208. ACM Press/Addison-Wesley Publishing Co., 1997.

[12] George W McConkie and Lester C Loschky. Perception onset time during fixations in free viewing. *Behavior Research Methods, Instruments, & Computers*, 34(4):481–490, 2002.

[13] Hunter Murphy and Andrew T Duchowski. Gaze-contingent level of detail rendering. *EuroGraphics 2001*, 2001.

[14] Toshikazu Ohshima, Hiroyuki Yamamoto, and Hideyulu Tamura. Gaze-directed adaptive rendering for interacting with virtual space. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*, pages 103–110. IEEE, 1996.

[15] Eli Peli, Jian Yang, and Robert B Goldstein. Image invariance with changes in size: The role of peripheral contrast thresholds. *JOSA A*, 8(11):1762–1774, 1991.

[16] Martin Reddy. *Perceptually modulated level of detail for virtual environments*. PhD thesis, University of Edinburgh. College of Science and Engineering. School of Informatics., 1997.

[17] Daniel R Saunders and Russell L Woods. Direct measurement of the system latency of gaze-contingent displays. *Behavior research methods*, 46(2):439–447, 2014.

[18] Benjamin Watson, Neff Walker, and Larry F Hodges. Managing level of detail through head-tracked peripheral degradation: a model and resulting design principles. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 59–63. ACM, 1997.