

User, Metric, and Computational Evaluation of Foveated Rendering Methods



Figure 1: Left: Our foveated resolution method running on a commercial video game engine. Right: Our foveated resolution, ambient occlusion, tessellation, and ray-casting (respectively) methods. Areas outwith the circles are the peripheral regions rendered in lower detail.

Abstract

Perceptually lossless foveated rendering methods exploit human perception by selectively rendering at different quality levels based on eye gaze (at a lower computational cost) while still maintaining the user's perception of a full quality render. We consider three foveated rendering methods and propose practical rules of thumb for each method to achieve significant performance gains in real-time rendering frameworks. Additionally, we contribute a new metric for perceptual foveated rendering quality building on HDR-VDP2 that, unlike traditional metrics, considers the loss of fidelity in peripheral vision by lowering the contrast sensitivity of the model with visual eccentricity based on the **Cortical Magnification Factor (CMF)**. The new metric is parameterized on user-test data generated in this study. Finally, we run our metric on a novel foveated rendering method for real-time immersive 360° content with motion parallax.

Keywords: Concepts: •Computing methodologies → Perception; Virtual reality;

1 Introduction

Providing high-quality image synthesis on high resolution displays in real-time is an ultimate goal of computer graphics. However, it remains a challenging problem even with full utilization of GPU hardware, as rendering operations are expected to perform in increasingly shorter time-frames (traditionally targeting 30 Hz to 60 Hz, the advent of commercial virtual reality has pushed the target to 90 Hz and higher). Controlling rendering quality to meet real-time requirements has been actively studied in past decades [Levoy and Whitaker 1990] broadly by reducing the number of rendering operations while minimizing the loss of quality.

Foveated rendering, a class of methods that vary the rendered quality across the image based on gaze, can be a fruitful approach to reduce the number of rendering operations. Human peripheral vision has lower spatial acuity than foveal vision (a small portion of the visual field centred at fixation), and so it is conceivable that a render could be degraded to provide computational benefit without any *perceivable* loss in quality. This is described as perceptual losslessness, an important feature of foveated rendering systems which justifies their adoption in the commercial realm.

To this end, we contribute four methods for and implementations of foveated rendering that can adaptively control peripheral quality in real-time. We also study the ideal quality-versus-computation bal-

ance for each method. We demonstrate that several computationally intensive features of modern real-time rendering pipelines can be adjusted for maximal computational gain with minimal perceivable quality loss. Three of our methods are evaluated against real users. We also introduce a perceptually-motivated extension of the HDR Visual Difference Predictor metric to account for foveation. Using this metric, we evaluate our final method specifically devised for 360° virtual reality content with motion parallax rendering, which we believe is one of the most suitable domains for these methods.

2 Background

2.1 Visual Perception

The spatial fidelity of human vision degrades as a function of visual eccentricity, which is in part explained by decreasing contrast sensitivity. Contrast sensitivity can be described as the minimal frequency and contrast required such that two distinct stimuli are perceived as separate. Geisler and Perry [1998] empirically derive the Contrast Sensitivity Function (CSF) to determine contrast detection thresholds as a function of eccentricity.

The CSF is present in some form in most perceptually informed quality metrics, such as SSIM [Wang et al. 2004] and HDR-VDP2 [Mantiuk et al. 2011] (see Section 2.3). Contrast sensitivity underlies our ability to perceive fine detail, texture, and contours which are typically some of the more computationally intensive aspects of rendering (e.g. shadows, resolution, surface texture).

More generally however, many perceptual phenomena may be encompassed and accounted for by the **Cortical Magnification Factor (CMF)**. The **visual cortex** is divided into several regions with varying structure and function. The primary visual cortex (V1) is the earliest visual cortex area, discriminating spatial frequencies, visual orientation, and other spatial and temporal factors [De Valois et al. 1988]. As we increase retinal eccentricity, the amount of visual cortex dedicated to each degree of visual field decreases. Prior studies have shown a strong relationship between the **Cortical Magnification Factor (CMF)** and the degradation of contrast sensitivity and visual acuity with visual eccentricity [Virsu and Rovamo 1979]. Using Equation 1, from Horton et al. [1991], we are able to calculate the **cortical magnification factor** for any given eccentricity:

$$M_e = \frac{A}{e + e_2} \quad (1)$$

Where A is the cortical scaling factor (mm) and e_2 is the eccen-

80 tricity at which a stimulus subtends half the cortical distance that
 81 it subtends in the fovea (degrees). Horton et al. supply the values
 82 $A = 17.3 \text{ mm}$, $e_2 = 0.75^\circ$. From Dougherty et al. [2003], we
 83 retrieve $A = 29.2 \text{ mm}$, $e_2 = 3.67^\circ$ for V1 and $A = 22.8 \text{ mm}$,
 84 $e_h = 2.54^\circ$ for V2.

2.2 Foveated Rendering

85 Levoy and Whitaker [Levoy and Whitaker 1990] varied image res-
 86 olution as a function of the Euclidean distance from the fovea's fixa-
 87 tion point using discrete levels of detail. Ohshima et al. [Ohshima
 88 et al. 1996] proposed a run-time selection method on sets of pre-
 89 computed object meshes at varying levels of details. Zha et al. [Zha
 90 et al. 1999] presented a gaze-directed mesh decimation model to re-
 91 duce the geometric complexity of a model. Murphy et al. [Murphy
 92 et al. 2009] designed a foveation method based on CSF, and varied
 93 image degradation according to the respective angular frequency,
 94 without modifying underlying scene geometry. Recently, Guenter
 95 et al. [2012] used three layers that include a different resolution and
 96 blended these layers to provide a high-quality foveated rendering
 97 result. For a broader lecture on level-of-detail rendering systems,
 98 we refer the reader to the excellent survey by Yoon et al [2008].

2.3 Image Quality Metrics

101 Traditionally, image quality metrics assume uniform quality per-
 102 ception at the foveal level across the entire image. Well known
 103 perceptually informed metrics such as Structural Similarity Index
 104 (SSIM) [Wang et al. 2004] and more recently HDR Visual Differ-
 105 ence Predictor (HDR-VDP2) [Mantiuk et al. 2011] perform signif-
 106 icantly better than other existing metrics for those scenarios. How-
 107 ever, foveated imagery (particularly in rendering) is meant to be
 108 appreciated at a single point in space and time, and are not meant to
 109 be appreciated entirely at foveal fidelity but instead at the varying
 110 level of fidelity across the visual field.

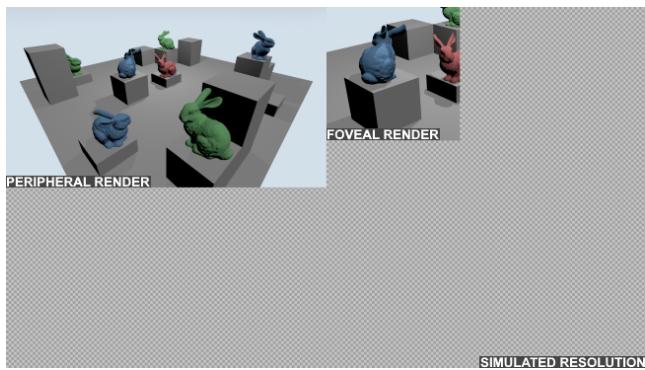


Figure 2: Annotated view of a foveated resolution render with moderate settings pre-composition. The checkerboard area represents the proportion of pixels saved for the targeted simulated resolution.

122 tural Similarity Index (SSIM) using image saliency. Similarly, they
 123 claim higher performance on tested databases, but their method re-
 124 lies on the selection of an appropriate saliency model.

3 Implementation

3.1 Foveated Rendering

127 Part of our aim in this study is the determination of adequate qual-
 128 ity settings for our methods that maintain perceptual losslessness.
 129 A perceptually lossless image is described as one that suffers imper-
 130 ceptible degradation such that to the average user it is indistinguish-
 131 able from the non-degraded, or reference, source image. Perceptual
 132 losslessness is an important feature for real-time rendering systems
 133 as it permits savings at compute time without a perceivable loss in
 134 quality.

135 We have implemented four methods which exploit quality degra-
 136 dation of resolution, Screen-Space Ambient Occlusion (SSAO), tes-
 137 sellation, and ray-casting steps with visual eccentricity. Increased
 138 quality in all three of these features of modern real-time rendering
 139 pipelines are associated with a large computational cost. Through
 140 this study, we aim to discern at what level of degradation do arte-
 141 facts become noticeable to the observer and determine the compu-
 142 tational savings that can be made at the limit of just-noticeable-
 143 difference. All four of our foveated rendering methods operate in
 144 real-time in their respective frameworks.

3.1.1 Foveal Window Size

145 The size of the high fidelity window in pixels on the screen is a
 146 function of the properties of the human visual system, the proper-
 147 ties of the screen, and the user's position in relation to the screen.
 148 Equation 2 provides the radius, in pixels, of the foveal window.

$$R_f = \rho_{pixel} d_u \tan\left(\frac{\alpha}{2}\right) + c + b_w \quad (2)$$

150 Where ρ_{pixel} is the pixel density of the display (pixels/mm), d_u is
 151 the user's distance from the screen (mm), and α is the angle sub-
 152 tended by the retinal region to test (in this case, the angle subtended
 153 by the fovea in radians). An error constant c is added to account
 154 for factors such as tracking error and highly off-axis fixations¹. Ad-
 155 ditionally, b_w specifies the width in pixels of an implementa-
 156 tion-specific blending border between the foveal and peripheral regions.

3.1.2 Peripheral Resolution

157 Our first method reduces the effective rendered pixel density of the
 158 peripheral region while maintaining the base density of the foveal
 159 window. Degraded peripheral resolution is a straightforward ap-
 160 proach to foveated rendering that has been explored previously (see
 161 Section 2.2). We render two views of the scene: first, the peripheral
 162 view, is a full field-of-view render at a fraction of the resolution
 163 we intend to simulate; second, the foveal view, is a limited field-
 164 of-view render at the intended pixel density (see Figure 2). The
 165 peripheral view is up-sampled to the target resolution with minor
 166 Gaussian blurring. Then, the foveal view is placed at the fixation
 167 point and a fraction of its outer radius is radially blended with the
 168 peripheral view to provide a smooth transition between each layer.

159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169

¹The foveal region on a flat surface such as a display (typically circular) becomes more elliptical with fixation eccentricity (requiring a larger rendered diameter). Although this could be modelled, highly eccentric fixations are not typical given the size of most modern displays, and so we rely on a simplification (the error constant).

170 **3.1.3 Screen-Space Ambient Occlusion**

171 Ambient occlusion [Pharr and Green 2004] is a well known technique
 172 in graphics to simulate the effect on diffuse lighting caused
 173 by occlusions created by objects present in the scene, including
 174 self-occlusions. It has been adopted to simulate a diffuse term that
 175 supports a complex distribution of incident light. Because ambient
 176 occlusion can be quite expensive to compute in real-time for
 177 dynamic scenarios, screen-space approaches are currently widely
 178 popular [Bavoil and Sainz 2008].

179 We exploit SSAO by varying the number of per-pixel depth-buffer
 180 samples in the foveal and peripheral fields of view. Although a very
 181 low number of per-pixel samples can cause banding (see Figure 3),
 182 we expect these differences to go unnoticed in the periphery due to
 183 the loss of visual acuity and contrast sensitivity. The scene we chose
 184 for this method is the Sibenik Cathedral populated with Stanford
 185 bunnies, as it provides a lot of occluding meshes with small details.



Figure 3: Strips from two foveated renders with the same fixation point but different peripheral sampling levels. Fixation point is at the bottom-right corner for each strip. Transition between foveal and peripheral regions are handled smoothly. At 4 samples there are noticeable artefacts, such as banding on the wall.

186 **3.1.4 Terrain Tessellation**

187 Our third method is a foveated implementation of a terrain ren-
 188 derer exploiting GPU-level tessellation. Geometry tessellation is
 189 a vertex processing stage that adaptively subdivides coarser geom-
 190 etry patches on-the-fly into smaller geometric primitives to gener-
 191 ate nicer and smooth-looking details. Tessellation has been incor-
 192 porated on modern GPU rasterization pipelines and is commonly
 193 driven by some view-dependent criteria. We chose this technique
 194 due to its wide adoption within the graphics industry.

195 Our foveated rendering method builds on an OpenGL framework
 196 exploiting tile-based tessellation. In order to determine the ap-
 197 propriate level of tessellation, we project the foveal window from
 198 screen coordinates into the scene. If a tile falls within either the
 199 foveal or peripheral field of view, the level of tessellation is set stat-
 200 ically to the appropriate level. If the tile falls between the two re-
 201 gions (on the blending border) the level of tessellation is linearly
 202 interpolated between the two levels. Figure 4 provides a wireframe
 203 view with exaggerated settings of our method in action.

204 **3.1.5 Foveated Real-time Ray-Casting**

205 Our fourth and final method, which we evaluate against the
 206 parametrized metric, employs foveally selective ray casting for
 207 360° immersive virtual reality content, rendered using a variant of
 208 multi-layer relief mapping originally developed by Pollicarpo and
 209 Oliveira [2006], which allows motion parallax within a limited en-
 210 velope of movement. The method normally casts rays to geometry

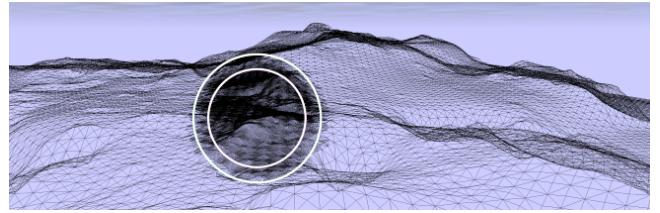


Figure 4: Wireframe view of a still from our foveated tessellation method. The foveal region is within the inner circle, the blending border between the inner and outer circles, and the peripheral region is outwith the outer circle.

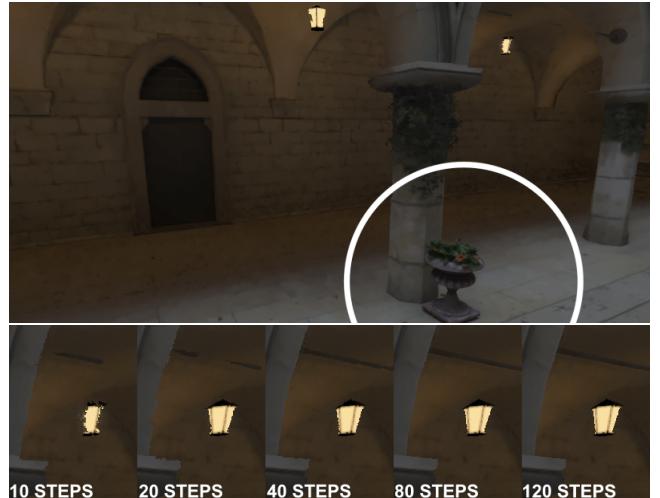


Figure 5: Top: Sample frame from our ray-casting method with 120 per-pixel steps in the foveal region (within circle) and 10 per-pixel steps in the peripheral region (outwith circle). Bottom: Close-up of right lamp showing artefacts across different quality levels.

211 and detects intersections with a given number of depth layers, rep-
 212 resented as a series of RGBA textures mapped on the geometry.
 213 We vary the number of per-pixel ray-casting steps across the field
 214 of view. This can cause significant dis-occlusion errors and stair-
 215 stepping artefacts if the number of steps is too low. Again, building
 216 on the lowered contrast sensitivity and visual acuity in peripheral
 217 vision, we expect there will be a balance between the severity of
 218 dis-occlusion and the number of per-pixel stepped samples that is
 219 sufficiently unnoticeable yet yield high performance.

220 **3.2 Foveated Image Metric**

221 We wish to develop a suitable image quality metric specifically for
 222 foveated imagery to assist with foveated rendering method eval-
 223 uation in the future. User trials are typically time consuming and
 224 costly, so their use should be reserved for methods that have reason-
 225 ably high chances of success. However, perceptually informed met-
 226 rics that take foveation into account are relatively unexplored (see
 227 Section 2.3). Instead of adopting and/or altering one of the afore-
 228 mentioned foveated metrics, we present a new metric that builds on
 229 an existing algorithm demonstrating a strong psychophysical back-
 230 ground but lacking consideration for loss of visual acuity with ec-
 231 centricity.

232 To this end, we extend HDR Visual Difference Predictor (HDR-
 233 VDP2) as it has a strong perceptual background, reports relatively
 234 good performance, is freely available, and is well documented. In

order to improve the algorithm meaningfully, we targeted the degradation of contrast sensitivity in peripheral vision. We introduce the CMF to the algorithm, as it describes the cortical surface area dedicated per degree of visual field with eccentricity, as a theoretically motivated parameter to calculating the extent of peripheral degradation.

There is a strong relationship between the CMF and the degradation of contrast sensitivity and visual acuity with visual eccentricity [Virsu and Rovamo 1979]. Difference between contrast sensitivity or visual acuity in central and peripheral vision could be accounted for by compensating stimulus size by the CMF. We scale the contrast sensitivity function by the CMF at a given pixel divided by the value of CMF at fixation. For HDR-VDP2, we target the neural contrast sensitivity function [Mantiuk et al. 2011] which discounts light scattering and luminance masking.

$$CSF_e^M = CSF_e - CSF_e \times (1 - \frac{M_e}{M_0})^{1+\alpha*(1-S)} \quad (3)$$

Where e is an eccentricity corresponding to a pixel position (x, y) , CSF_e is the Contrast Sensitivity Function at that eccentricity, M_e is the CMF at that position, and M_0 is the CMF at centre of vision. As HDR-VDP2 uses a multi-scale decomposition process, we increase sensitivity of detected contrast as scale decreases (S being 0.5, 0.25, etc) to allow the model to remain sensitive to large scale contrast changes over the visual field. Finally, α is a tunable parameter that we introduce to attenuate the effect of peripheral sensitivity.

3.3 Hypotheses

How perceptually lossless a foveated render appears to be can be determined by how reliably an average user would be able to distinguish the reference render as the higher quality render when also presented to the foveated render. Thus, to validate our methods and determine whether they are perceptually lossless, the average user should identify the reference render (uniformly high quality) over the foveal render (high quality window at fixation, lower quality elsewhere) worse than chance. The more significantly different from chance this value is, the more reliable is the foveated method/quality pairing. We advance the following hypotheses, such that when comparing a reference and a foveated render:

H₀ *The average viewer identifies the reference render as the high quality render at chance ($\approx 50\%$ of the time).*

H₁ *The average viewer identifies the reference render as the high quality render better than chance ($> 50\%$ of the time).*

H₂ *The average viewer identifies the reference render as the high quality render worse than chance ($< 50\%$ of the time).*

H_2 is our preferred hypothesis, as it indicates the reference render cannot reliably be identified as the higher quality render. A failure to reject the null hypothesis does not allow us to make any conclusions on the effectiveness of the method. If results favour H_1 , the method/quality pairing must be abandoned as the difference is reliably detectable.

4 Experimentation

4.1 Rendering Parameters

We use Equation 2 to calculate the foveal window size for our study. The fovea subtends the central 5° of radial area on the retina [Polyak 1941], however we increase the value used in our studies

to 9° to encompass the parafoveal area (approximately 7° of eccentricity) and to account for tracker error. This corresponded to a foveal window diameter of approximately 588.4 px (given the information in Section 4.4), which we round up to 600 px to account for minor accidental gaze drift.

The blending border between both regions is an additional 100 px, which is decided arbitrarily. Prior studies have shown that blending, or lack thereof, provides no significant user performance difference [Reingold and Loschky 2002]. However, the peripheral degradation in that study was noticeable and may have interfered with the results. As far as we are aware, there are no further studies that focus explicitly on this subject.

We select three levels of detail per method to experiment on and to ensure some coverage of the parameter space. These three levels of detail are described throughout this paper as **low**, **medium**, and **high**. Low settings were chosen to provide the largest computational gain, but the most likelihood of detection that could still justify foveation. Contrarily, high settings were chosen as very unlikely to be detected, but with the lowest computational gain that could still sufficiently justify the use of foveation. The medium setting was chosen as the middle point between the two, an intuitively ideal balance between likelihood of detection and performance. See Table 1 for exact values.

	Resolution (scaling)	SSAO (samples)	Tessellation (levels)
LOW	0.25	4	8
MED	0.50	16	16
HIGH	0.75	64	32
REF	1.00	128	64

Table 1: Peripheral quality parameter values used in our study. For the resolution method, we render the periphery at parameter value of the target resolution and then upscale. For the ambient occlusion method we vary the number of samples. For tessellation, we vary the refinement of the tessellated grid per tile.

4.2 Fixations

For our experiments we decided to focus exclusively on perceivable spatial artefacts for our methods. Although we understand the importance of evaluating our methods temporally, our work serves as a preliminary study in automated and subjective evaluation of gaze-contingent methods. As our extension to the HDR-VDP2 metric (and the base metric itself) does not take temporal factors of human vision into account, we would be unable to accurately evaluate the perceptibility of our modifications through the image quality metric in a temporal setting. Additionally, due to the tracking hardware available to us (see Section 4.4) we would not be able to isolate our experiments from external error, leading to potentially flawed conclusions about the methods' perceptibility. We instead adopt fixation-based testing and use our tracking hardware to validate fixations.

Fixation-based testing introduces a few problems when evaluating methods for user preferences, image quality metric results, and reported computational load. In terms of computation, the position of the foveal render can greatly affect rendering times depending on the method (e.g. tessellation on simple versus intricate surfaces). In terms of user preference, prior studies suggest that poor selection of the foveated region (such as random or brute-force selection) could lead to lower perceived image quality [Bailey et al. 2009]. In terms of image metrics, it must be general enough to provide realistic results for the phenomena it is modelling (in this case, the

336 human visual system), where simplifications can lead to excessive
 337 positive or negative performance. Temporal testing does not suf-
 338 fer from these specific issues as gaze is a direct reflection of user
 339 preference and real-world data (which would validate averaging for
 340 computational results, for example).

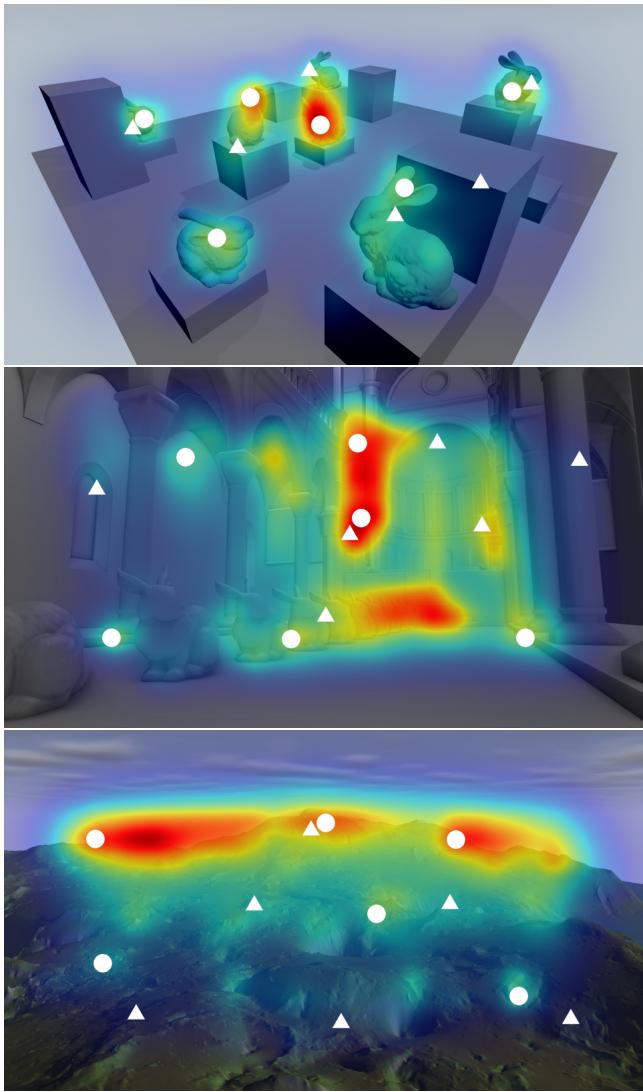


Figure 6: Reference renders for each method with respective Graph-Based Visual Saliency (GBVS) heat-map overlay, from top to bottom: resolution, SSAO, and tessellation. Circular marks denote fixations selected by GBVS. Triangular marks denote fixations selected subjectively by the authors.

341 In order to select plausible fixations we conducted a small pilot
 342 study, collecting gaze positions over a 10 second period during free-
 343 viewing sessions of our reference renders. We then ran Itti, Koch,
 344 and Neibur (ITTI) [Itti et al. 1998], GBVS [Harel et al. 2006], and
 345 Erdem and Erdem (CovSAL) [Erdem and Erdem 2013] on our ref-
 346 erence renders to select the saliency maps which fit closest to our
 347 collected free-viewing fixation data. The saliency model that most
 348 closely fit our data was GBVS, from which we select the centres of
 349 the 6 most salient, non-overlapping image regions. We also subjec-
 350 tively chose 6 additional fixation points which we found to demon-
 351 strate high detail variability or represented interesting regions of the

352 image. The fixation points for each method/reference render can be
 353 seen in Figure 6.

354 4.3 User Trials

355 The experiments consisted of a number of tests in randomized order
 356 comparing a foveated render to the reference render. For each trial,
 357 a foveated render was displayed before or after a reference render
 358 for the same amount of time. Once both images had been displayed,
 359 the subject would then have to decide whether the first image ap-
 360 peared higher quality, the second image appeared higher quality, or
 361 if both images appeared identical, and respond appropriately.

362 Each subject underwent three test blocks, one for each rendering
 363 method, in randomized order. A test block consisted of 81 trials in
 364 randomized order. Out of these 81 possible trials, 9 were control tri-
 365 als while the remaining 72 were test trials. The amount of test trials
 366 are divided equally among each of the three quality levels, lending
 367 to 24 test trials per quality level per method. Of these 24, there
 368 are 2 trials for each of the 12 fixation points; one trial in which the
 369 foveated render is presented first and one where the foveated ren-
 370 der is presented second. For the 9 control trials, 3 trials display the
 371 reference render against itself and 6 trials compare a fully periph-
 372 eral quality render against the reference (per quality level and per
 373 first/second order).

374 The procedure for a single trial was as follows. Firstly, a neutral
 375 grey screen would appear for two seconds. Then a small cross
 376 would appear on the grey screen indicating where the user was to
 377 maintain their fixation. Users were instructed to fixate at that pos-
 378 ition until the end of that specific trial. The eye tracker would ensure
 379 the user's gaze was fixated on the indicated area and would signal
 380 the start of the test. At this point, the first image in the trial would
 381 appear for two seconds, followed by the neutral grey screen with
 382 the cross at the same location for one second, followed by the sec-
 383 ond image in the trial for two seconds. If the user's gaze drifted
 384 away from the indicated fixation point at any time during the trial,
 385 the trial would not be interrupted but the results would be marked
 386 invalid. Finally, the neutral grey screen would return without the
 387 cross to await the user's response (first was better, second was bet-
 388 ter, or both appeared identical).

389 The user population consisted of 9 participants (1 female, mean
 390 population age of 32) who were computer graphics professionals
 391 with diverse backgrounds. All users had 20/20 or corrected to 20/20
 392 vision. The eye tracker (see Section 4.4) was calibrated for each
 393 user individually before their testing session. Users were allowed
 394 to take short breaks at any point during a block (provided this was
 395 done at the answer screen for a trial and they remembered their
 396 answer) to avoid fatigue. Between each block, breaks of any desired
 397 length were allowed and users could leave the testing area, also to
 398 prevent fatigue.

399 4.4 Equipment

400 We use an Acer CB280HK 4K UHD monitor with a display area ap-
 401 proximately 62 cm × 34.5 cm in size, corresponding to an approx-
 402 imate pixel density of 6.23 px mm^{-1} . For eye-tracking, we used
 403 Tobii's EyeX commercial level eye tracker with 9-point calibration,
 404 with no accuracy and precision reports² and no specified latency at
 405 time of purchase³, although internal testing yielded an approximate
 406 latency of 50 ms to 75 ms. Due to these specifications, we would be
 407 unable to reliably validate our methods temporally, and so our study
 408 focuses solely on spatial detectability. To easily accommodate the

²<http://archive.is/qWvMi>

³<http://archive.is/o7b1M>

409 eye tracker's tracking volume and increase tracking accuracy, users
 410 were secured on a head-rest at a distance of 600 mm from the mon-
 411 itor for all experiments. For our rendering and benchmark tests, we
 412 use a desktop computer equipped with an Intel Core i7 4820K CPU
 413 and an ASUS R9 290X GPU.

414 5 Results

415 5.1 User Trials

416 All subjects completed all trials for all three blocks. However, one
 417 subject's resolution block trial data had to be discarded due to a
 418 misunderstanding of testing procedure, which led to all responses
 419 being invalid. This data was removed from our results and there
 420 were no other changes made to the data set. All significance values
 421 are evaluated at the $\alpha = 0.01$ level.

422 The proportion of invalid responses to valid responses was similar
 423 across parameters within a given method, with $\approx 18\%$ invalid re-
 424 sponds for the resolution method and $\approx 16\%$ for the tessellation
 425 method. However, the ambient method demonstrated an overall
 426 higher proportion ($\approx 26\%$) of invalid responses when compared to
 427 the other two methods. Given that trial block order was randomized
 428 we exclude fatigue as a possible cause, and tracker error would have
 429 manifest itself across all trials. This suggests that the method may
 430 have caused distracting artefacts or the scene contained sufficiently
 431 distracting features to draw gaze. However, whether this difference
 432 is statistically significant is not determined.

433 Data for several quality/method settings demonstrate a "correct"
 434 (identified the reference render as the higher quality render of the
 435 pair) to "incorrect" (identified the foveated render as the higher
 436 quality render of the pair, or indicated that the quality of both were
 437 identical) ratio that was statistically significant in favour of H_2 ,
 438 thereby encouraging their adoption. These quality/method settings
 439 were **ambient high** ($P_{val} \approx 6.895 \times 10^{-9}$), **tessellation medium**
 440 ($P_{val} \approx 0.0011$), **tessellation high** ($P_{val} \approx 4.598 \times 10^{-7}$),
 441 **resolution low** ($P_{val} \approx 0.0023$), **resolution medium** ($P_{val} \approx$
 442 7.938×10^{-5}), and **resolution high** ($P_{val} \approx 5.822 \times 10^{-5}$). The
 443 remaining quality/method settings either favour H_1 (ambient low),
 444 thereby discouraging their use, or fail to reject H_0 (tessellation low
 445 and ambient medium).

446 Subjective responses from users suggest difficulties in distinguishing
 447 the images for the resolution trial block, with some subjects
 448 asking whether they were being shown different images at all. The
 449 users added that there were a few "obviously rough looking" images
 450 that they felt were easily distinguishable. These were most likely
 451 the control trials and a subset of the low quality trials. Subjects
 452 also reported the most confidence after the ambient tests, stating
 453 that the quality difference for many of the trials was clearly distin-
 454 guishable. For the tessellation trials, user confidence was mixed,
 455 but overall subjects believed that they had identified the reference
 456 correctly.

457 5.2 Quality Metric

458 Using the results from the user trials, we parametrize our met-
 459 ric. The metric will then be used to evaluate our fourth and final
 460 foveated rendering method for immersive content. We first deter-
 461 mine the ideal parameters for base HDR-VDP2, namely the peak
 462 sensitivity of the metric (p_{sens}), the excitation (p_{mask}), and inhib-
 463 ition (q_{mask}) of the visual contrast masking model. These are the
 464 tunable parameters provided by the base HDR-VDP2 metric.

465 HDR-VDP2 predicts the probability that the differences between
 466 two images are visible to the average observer (with 0 indicat-

467 ing impossibility and 1 indicating absolute certainty). To compare
 468 against the model's predictions, we derive our predictions from the
 469 data by comparing metric results against user testing results for the
 470 fully peripheral quality versus reference control trials. In this way,
 471 the base parameters for the HDR-VDP2 metric are calibrated for
 472 degradations at foveal fidelity (highest fidelity in the visual field).

473 We were unable to find a single set of base parameters that pro-
 474 vided detection probabilities close to our data for all three methods.
 475 Therefore, we provide parameters per method and evaluate our se-
 476 lective ray casting rendering model against each. For the resolution
 477 data we use $p_{sens} = 1.0$, $p_{mask} = 0.14$, and $q_{mask} = 0.19$. For
 478 SSAO we use $p_{sens} = 0.8$, $p_{mask} = 0.54$, and $q_{mask} = 1.50$. For
 479 Tessellation we use $p_{sens} = 0.8$, $p_{mask} = 0.54$, and $q_{mask} =$
 480 0.30.

481 We then calibrate our extended metric using the attenuation pa-
 482 rameter α from Equation 3, using the V1 cortex parameters from
 483 [Dougherty et al. 2003] for the CMF function. The detection prob-
 484 abilities output by our metric are compared against the foveated de-
 485 tection probabilities from our data; the number of valid and correct
 486 responses over the total number of valid responses. The attenu-
 487 ation values we found to have the best fit were $\alpha = 2.45$ for the
 488 resolution data, $\alpha = 4.45$ for the ambient data, and $\alpha = 0.43$ for
 489 the tessellation data. Using our metric, the average detection pre-
 490 dictions per quality setting per method (averaged over all foveated
 491 images in that class) can be seen in Table 2.

	Resolution ($\alpha = 2.45$)	SSAO ($\alpha = 4.45$)	Tessellation ($\alpha = 0.43$)
LOW	0.32 (0.27)	0.88 (0.80)	0.65 (0.57)
MED	0.02 (0.14)	0.29 (0.51)	0.12 (0.24)
HIGH	0.01 (0.14)	0.08 (0.07)	0.01 (0.11)

492 **Table 2:** Average predicted detection probabilities per setting per
 493 method (averaged over all foveated images in that class) from our
 494 extended metric, with probability values extracted from our data
 495 shown in parentheses.

496 5.3 Immersive Motion Parallax Rendering

497 We run our fully calibrated metric on our fourth and final method.
 498 For this dataset, we adjust the equipment and set-up specific base
 499 parameters of HDR-VDP2 to match values for a typical modern and
 500 commercial head-mounted display. In our case, we use the Oculus
 501 Rift DK2's resolution, screen dimensions, and typical eye distance
 502 from the screen. Renders from this dataset are then evaluated with
 503 our metric using the three parameter sets (one per method) derived
 504 in Section 5.2. The detection probabilities returned by our metric
 505 on this dataset are found in Table 3. Similarly to the other foveated
 506 rendering methods, we are only evaluating the method spatially at
 507 a single point in time. In this case, we use a single fixation point
 508 (in this case the flower pot in the scene, see Figure 5) and evaluate
 509 over a wider quality parameter space.

510 Out of the three parameter sets, the tessellation parameters seem to
 511 provide the most unrealistic results given the amount of degradation
 512 at lower steps. Since the artefacts produced by reduced peripheral
 513 resolution are similar to those produced by reduced sampling (loss
 514 of contour and texture fidelity, etc.) we use the resolution parameter
 515 set for our metric to determine the ideal balance between detectabil-
 516 ity and computational performance for this particular method in
 517 Section 5.4.

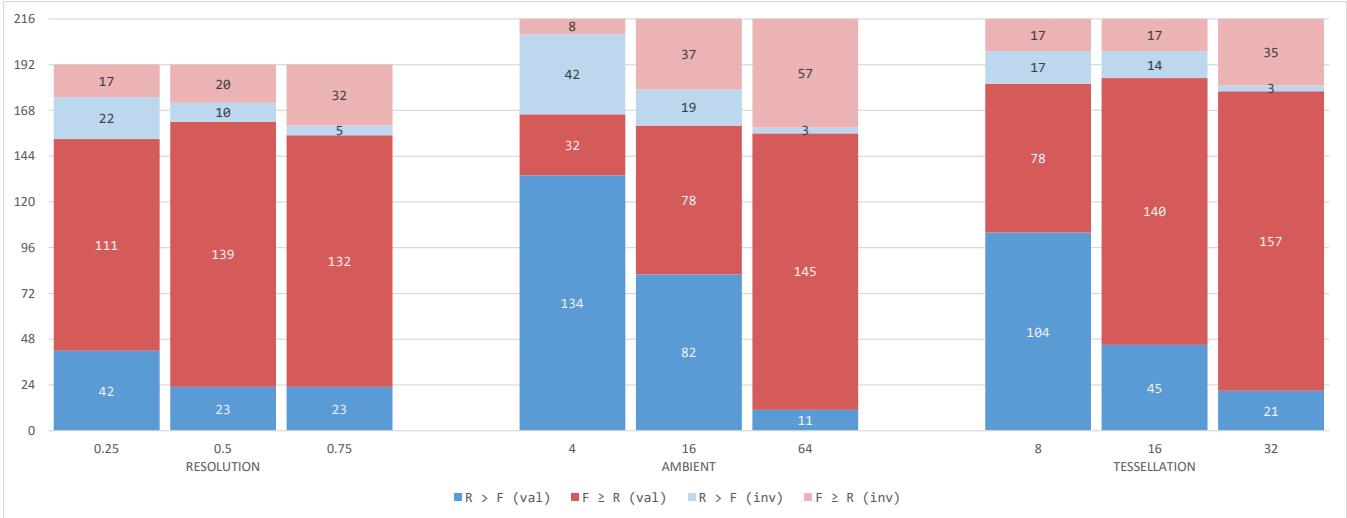


Figure 7: All trial results (excluding controls), split per method and per parameter setting. Valid instances where the reference was marked higher quality than the foveated render are in blue (invalid in light blue). Valid instances where the foveated render was marked equal or higher quality than the reference are in red (invalid in light red).

	Res. Settings	SSAO Settings	Tes. Settings
10 steps	0.50	0.37	0.03
20 steps	0.14	0.08	0.01
40 steps	0.04	0.02	0.01
80 steps	0.03	0.01	0.01

Table 3: Predicted detection probabilities for our fourth foveated rendering method, with foveal region rendered at 120 steps and periphery rendered at step rate listed in first column.

	Optimal Settings	Reference
Resolution	7.18 ms (7.01 ms / 7.27 ms)	14.69 ms
SSAO	22.31 ms (21.17 ms / 25.2 ms)	82.34 ms
Tessellation	5.88 ms (4.54 ms / 10.16 ms)	17.24 ms
Sampling	19.61 ms	28.57 ms

Table 4: Mean frame rendering time over all fixations per method/quality setting in milliseconds. Fixations with the best and worst (respectively) mean render time shown in parentheses. Resolution, SSAO, and tessellation methods are targeting 4K UHD while the Sampling method is targeting 1600×1018.

5.4 Performance Gains

To evaluate computational performance we settle on the lowest quality setting per method that favours H_2 , run our methods in real-time at each fixation point, and average the render time over 1000 frames. After which, we average across all fixation point times per method to provide the average rendering time for our method overall. We select **resolution medium**, **ambient high**, and **tessellation medium** for our quality settings. We chose the resolution medium over resolution low in order to be conservative with our estimates, as detection probabilities appear to plateau between the two.

The average render time over all fixation points, the fixation point with the worse average render time, and the fixation point with the best average render time compared against the average render time

for the reference *per method/quality setting* are show in Table 4. The table also includes the average rendering time for our foveated ray-casting method at the flower pot fixation point at the 20 step quality level.

6 Discussion

6.1 Analysis

Overall, all of our methods enjoyed some success. As expected, the low quality settings were the most easily detectable, but with the resolution method the difference between settings was much less substantial than initially expected. This may partially explain why resolution degradation remains a popular (and successful) method for foveation. Artefacts or perceivable foveation was much more prominent across the ambient method trials, but even within the tested sampling levels we found on which relatively imperceptible and provided substantial computational benefit. Our metric indicates that our ray-casting method is relatively undetectable at lower step rates (but not the lowest). These results may be the first paces towards motivating the use of real-time ray casting content for virtual reality. We expect the computational gains to be even more substantial once we are able to integrate multiple methods together.

We recognize a few limitations of our study. Firstly, we would like to conduct a larger exploration of the parameter space for our rendering methods to make more accurate inferences about the rate of change in terms of detectability. Additionally, we do not explore any temporal aspects of our methods and the detectability any temporal-specific aspects that may be introduced. We realize that temporal evaluation is critical to fully validate foveated methods, requiring accurate, fast, and reliable eye tracking.

6.2 Applications

We believe the largest application domain for perceptually lossless foveated rendering in the near future is in virtual reality. This is partly why we demonstrate our fourth foveated rendering method, foveally selective ray casting for immersive content. The current state of the virtual reality market demands expensive hardware that

561 puts living room virtual reality out of reach for the majority of the
 562 consumer entertainment market. For example, Oculus has recently
 563 announced that the minimum specifications for the consumer ver-
 564 sion of their head-mounted display requires a GPU equivalent to the
 565 GTX 970 or higher. A conservative estimate from the most recent
 566 Steam Hardware & Software survey in December 2015⁴, which col-
 567 lects hardware statistics for a major online game distribution com-
 568 pany, shows that less than 10% of users today fit that requirement.
 569 Companies like FOVE and StarVR have shown support foveated
 570 rendering by integrating eye-tracking in their head-mounted display
 571 models. Beyond head-mounted displays, immersive environments
 572 for very large scale real-time rendering (such as high quality CAVE
 573 installations) stand the most to gain from foveated rendering, as
 574 most of the rendered scene is never in view.

575 6.3 Future Work

576 We would also like to study problems specific to foveated render-
 577 ing in virtual reality, such as accounting for eye tracking failure and
 578 system latency in order to maintain perceptual losslessness. This
 579 may also involve exploring the effect of foveated rendering meth-
 580 ods in virtual reality and how they may affect motion sickness, or
 581 whether more active methods for foveation (such as explicitly di-
 582 recting gaze) are possible. This also extends to exploring novel
 583 foveated rendering methods that focus on, or integrate several, other
 584 aspects of the rendering pipeline. We would like to further refine
 585 our foveated metric to account for more spatial aspects of the hu-
 586 man visual system. Primarily, we would like to extend the method
 587 further by considering temporal factors as well. This will also re-
 588 quire a temporal evaluation with user trials for our existing and any
 589 future methods.

590 References

- 591 BAILEY, R., McNAMARA, A., SUDARSANAM, N., AND GRIMM,
 592 C. 2009. Subtle gaze direction. *ACM Transactions on Graphics*
 593 (*TOG*) 28, 4, 100.
- 594 BAVOIL, L., AND SAINZ, M. 2008. Screen space ambient occlu-
 595 sion. *NVIDIA developer information: http://developers.nvidia.*
 596 *com* 6.
- 597 DEVALOIS, R. L., BOTH PROFESSORS, B. K. K. D., ET AL.
 598 1988. *Spatial vision*. No. 14. Oxford University Press.
- 599 DOUGHERTY, R. F., KOCH, V. M., BREWER, A. A., FISCHER,
 600 B., MODERSITZKI, J., AND WANDELL, B. A. 2003. Visual
 601 field representations and locations of visual areas v1/2/3 in hu-
 602 man visual cortex. *Journal of Vision* 3, 10, 1.
- 603 ERDEM, E., AND ERDEM, A. 2013. Visual saliency estimation by
 604 non-linearly integrating features using region covariances. *Jour-*
605 nal of vision 13, 4, 11.
- 606 GEISLER, W. S., AND PERRY, J. S. 1998. Real-time foveated mul-
 607 tiresolution system for low-bandwidth video communication. In
 608 *Photonics West'98 Electronic Imaging*, International Society for
 609 Optics and Photonics, 294–305.
- 610 GUENTER, B., FINCH, M., DRUCKER, S., TAN, D., AND SNY-
 611 DER, J. 2012. Foveated 3d graphics. *ACM Transactions on*
612 Graphics (TOG) 31, 6, 164.
- 613 HAREL, J., KOCH, C., AND PERONA, P. 2006. Graph-based visual
 614 saliency. In *Advances in neural information processing systems*,
 615 545–552.
- 616 HORTON, J. C., AND HOYT, W. F. 1991. The representation of
 617 the visual field in human striate cortex: a revision of the classic
 618 holmes map. *Archives of ophthalmology* 109, 6, 816–824.
- 619 ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-
 620 based visual attention for rapid scene analysis. *IEEE Trans-*
621 actions on Pattern Analysis & Machine Intelligence, 11, 1254–
 622 1259.
- 623 LEE, S., PATTICHIS, M. S., AND BOVIK, A. C. 2002. Foveated
 624 video quality assessment. *Multimedia, IEEE Transactions on* 4,
 625 1, 129–132.
- 626 LEVOY, M., AND WHITAKER, R. 1990. Gaze-directed volume
 627 rendering. *ACM SIGGRAPH Computer Graphics* 24, 2, 217–
 628 223.
- 629 MANTIUK, R., KIM, K. J., REMPEL, A. G., AND HEIDRICH, W.
 630 2011. Hdr-vdp-2: a calibrated visual metric for visibility and
 631 quality predictions in all luminance conditions. In *ACM Trans-*
632 actions on Graphics (TOG), vol. 30, ACM, 40.
- 633 MURPHY, H. A., DUCHOWSKI, A. T., AND TYRRELL, R. A.
 634 2009. Hybrid image/model-based gaze-contingent rendering.
ACM Transactions on Applied Perception (TAP) 5, 4, 22.
- 635 OHSHIMA, T., YAMAMOTO, H., AND TAMURA, H. 1996. Gaze-
 636 directed adaptive rendering for interacting with virtual space.
 637 In *Virtual Reality Annual International Symposium, 1996*, Pro-
 638 ceedings of the IEEE 1996, IEEE, 103–110.
- 639 PHARR, M., AND GREEN, S. 2004. Ambient occlusion. *GPU*
 640 *Gems 1*, 279–292.
- 641 POLICARPO, F., AND OLIVEIRA, M. M. 2006. Relief mapping
 642 of non-height-field surface details. In *Proceedings of the 2006*
643 symposium on Interactive 3D graphics and games, ACM, 55–62.
- 644 POLYAK, S. L. 1941. The retina.
- 645 REINGOLD, E. M., AND LOSCHKY, L. C. 2002. Saliency of
 646 peripheral targets in gaze-contingent multiresolutional displays.
 647 *Behavior Research Methods, Instruments, & Computers* 34, 4,
 648 491–499.
- 649 RIMAC-DRLJE, S., VRANJEŠ, M., AND ŽAGAR, D. 2010.
 650 Foveated mean squared error novel video quality metric. *Mul-*
651 timedia tools and applications 49, 3, 425–445.
- 652 TSAI, W.-J., AND LIU, Y.-S. 2014. Foveation-based image quality
 653 assessment. In *Visual Communications and Image Processing*
 654 Conference, 2014 IEEE, IEEE, 25–28.
- 655 VIRSU, V., AND ROVAMO, J. 1979. Visual resolution, contrast
 656 sensitivity, and the cortical magnification factor. *Experimental*
657 Brain Research 37, 3, 475–494.
- 658 WANG, Z., BOVIK, A. C., LU, L., AND KOULOHERIS, J. L.
 659 2001. Foveated wavelet image quality index. In *International*
660 Symposium on Optical Science and Technology, International
 661 Society for Optics and Photonics, 42–52.
- 662 WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI,
 663 E. P. 2004. Image quality assessment: from error visibility to
 664 structural similarity. *Image Processing, IEEE Transactions on*
 665 13, 4, 600–612.
- 666 YOON, S.-E., GOBBETTI, E., KASIK, D., AND MANOCHA, D.
 667 2008. Real-time massive model rendering. *Synthesis Lectures*
 668 *on Computer Graphics and Animation* 2, 1, 1–122.
- 669 ZHA, H., MAKIMOTO, Y., AND HASEGAWA, T. 1999. Dynamic
 670 gaze-controlled levels of detail of polygonal objects in 3-d envi-
 671 ronment modeling. In *3-D Digital Imaging and Modeling, 1999*.
 672 *Proceedings. Second International Conference on*, IEEE, 321–
 673 330.
- 674

⁴<http://store.steampowered.com/hwsurvey/videocard/>