Perception of Highlight Disparity at a Distance in Consumer Head-Mounted Displays

Robert Toth[1] Jon Hasselgren[1] Tomas Akenine-Möller[1,2]

[1]Intel Corporation [2]Lund University

Abstract

Stereo rendering for 3D displays and for virtual reality headsets provide several visual cues, including convergence angle and highlight disparity. The human visual system interprets these cues to estimate surface properties of the displayed environment. Naïve stereo rendering effectively doubles the computational burden of image synthesis, and thus it is desirable to reuse as many computations as possible between the stereo image pair. Computing a single radiance for a point on a surface, to be used when synthesizing both the left and right images, results in the loss of highlight disparity. Our hypothesis is that absence of highlight disparity does not impair perception of surface properties at larger distances. This is due to an ever decreasing angular difference between the surface and the two view points as distance to the surface is increased. The effect is exacerbated by the limited resolution of consumer head-mounted displays. We verify this hypothesis with a user study and provide rendering guidelines to leverage our findings.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture I.3.m [Computer Graphics]: Miscellaneous—Cognitive science

Keywords: computer graphics, virtual reality, stereoscopic rendering, psychophysical user study

1 Introduction

Consumer-class virtual reality (VR) and augmented reality (AR) devices have received much attention lately. Several consumer head-mounted displays (HMDs) from multiple vendors are expected to be commercially available within a few years. Rendering to these devices is a large computational burden since the virtual environment must be rendered in stereo and at a high spatial and temporal resolution. This burden will likely be placed on mainstream consumer desktops, laptops, and even phones, which may be underpowered for this task. It is therefore important to reduce the computational cost of synthesizing stereo images as far as possible while maintaining high visual fidelity. In addition, for high-performance cloud-based rendering services, improving image synthesis efficiency can translate into cost savings.

There are several technologies which are designed to, or can be adapted to, compute the radiance of surfaces only once while rasterizing two images from separate viewpoints. REYES [Cook et al. 1987] computes radiance separately from rasterization, allowing reuse of computed radiance. On contemporary graphics hardware, radiance can be computed once, stored in textures, and be reused during subsequent rasterization of multiple viewpoints [Nehab et al. 2007; Sitthi-amorn et al. 2008a; Sitthi-amorn et al. 2008b; Liktor and Dachsbacher 2012; Clarberg and Munkberg 2014; Andersson et al. 2014]. GPUs could also be modified to allow efficient reuse of radiance between viewpoints during rasterization [Hasselgren and Akenine-Möller 2006; Ragan-Kelley et al. 2011; Clarberg et al. 2013; Clarberg et al. 2014].

Figure 1: *A test subject is wearing a head-mounted display and is classifying whether two stimuli are identical or different using the mouse buttons. The goal of the study was to find the circumstances under which highlight disparity could no longer be perceived. Our findings could be used to optimize VR rendering.*

While generally less accurate, performance can also be significantly improved by rendering the scene from a single viewpoint only (e.g., the user's left eye) and then synthesizing images of other viewpoints (e.g., the right eye). This is accomplished by warping the image from the first viewpoint using an associated depth map [McMillan and Bishop 1995; Didyk et al. 2010].

Regardless of which technique is employed, shading reuse implies that radiance is only computed for a single exitant angle. This effectively precludes rendition of *binocular luster*, where the surface luminance is different for each eye. The human visual system interprets such luminance contrast as the material being lustrous [Dove 1851], and removal of this visual cue may cause observers to underestimate surface glossiness [Sakano and Ando 2010]. Furthermore, sharing radiance also causes specular highlights to appear at the same convergence angle as the reflecting object's surface. This is in contrast with physical reality, where specular reflections appear at a depth different than that of the reflecting surface [Kirschmann 1895; Muryy et al. 2013], an effect called *highlight disparity*. Again, this may cause observers to underestimate surface glossiness [Wendt et al. 2010] and even reduce authenticity [Wendt et al. 2008]. Recent studies indicate that physically-correct depiction may not be desired when synthesizing stereo images. Templin et al. [2012] found that viewers may prefer a non-physical highlight disparity in some cases. This was reinforced by Dąbała et al. [2014] who manipulate disparity to make reflections and refractions easier to fuse.

Our hypothesis is that highlight disparity is not perceptible at sufficiently large distances. By conducting a user study [Ferwerda et al. 2002] (see Figure 1), we verify this hypothesis under different shading circumstances and provide rendering guidelines based on our

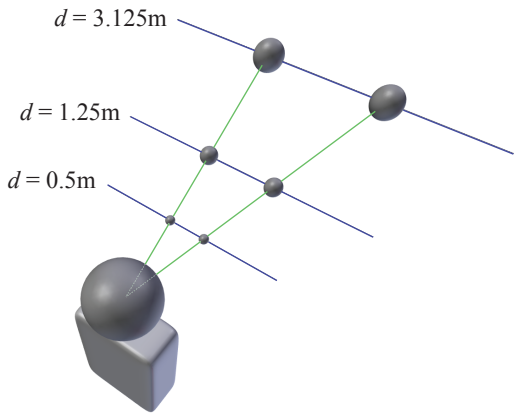


Figure 2: *Test subjects are presented with one pair of stimulus at a time in virtual reality, where each pair appears at one of three distances as shown above. The angle between the stimuli is* 0.5 *radians and the sphere radius is* 0.1m *per meter distance. The task is to determine if the two presented stimuli looked identical or different in a two-alternative forced choice setting.*

findings.

2 Experiment

Our virtual experiment setup is illustrated in Figure 2. The test subjects were tasked with determining whether a pair of stimuli had *identical* or *different* material appearance. Each stimulus pair is presented at one of three depths, level with the subject's eyes and with a size chosen to maintain a constant solid angle from the viewer's nominal position. The subjects were able to lean approximately 0.3m in any direction from this position, and their movements were mirrored in the virtual environment. We chose simple spheres as stimuli since it is harder to judge reflection correctness of complex objects [Ramanarayanan et al. 2007].

For shading, we use a Blinn-Phong BRDF with a Gaussian specular approximation [Wang et al. 2009][1] where we varied two parameters: *specular intensity*, k_s, and *glossiness* (the standard deviation of the specular lobe), σ. The material parameter configurations used in the experiment are shown in Figure 3. The spheres have a diffuse marble texture to give visual cues about the surface position, without distracting from the reflection. An image-based light probe is used to illuminate the scene. The probe[2] was chosen due to its distinct light sources and is visualized as the scene background to allow the test subjects to look around and compare the reflection with the environment. For efficiency reasons, we compute lighting based on high-quality pre-integrated cube maps for diffuse and specular components, generated using ATI's CubeMapGen tool. The diffuse component is stored in a separate cube map and specular reflection is computed by performing a mipmapped texture lookup based on the glossiness of the material.

As illustrated in Figure 4, the stimuli are either *monoscopically* lit, with specular reflections for both eyes computed from a common viewpoint located between the subject's eyes, or *stereoscopically* lit, with specular reflections for each eye rendered using the correct viewpoint. For each configuration of depth and material parameters, we show the test subjects two pairs of stimuli: one pair of

[1] A Gaussian approximation of the Blinn-Phong specular lobe is found in the supplemental material of Wang et al. [2009].

[2] Dining room of the Ennis-Brown House, Los Angeles. Courtesy of University of Southern California.

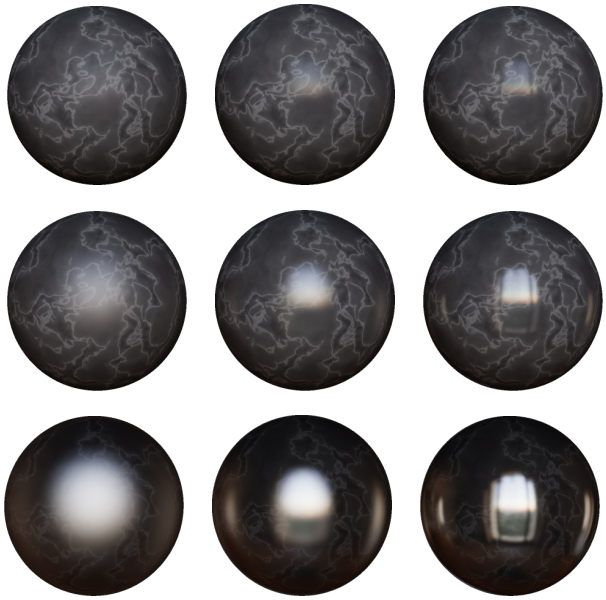


Figure 3: *The different materials used in the test. Glossiness* (σ)*, from left to right:* 37.8°, 12.6°, 4.2°. *Specular intensity* (k_s)*, from top to bottom:* 0.02, 0.1, 0.5. *Diffuse intensity* k_d *is decreased with increasing* k_s *to keep total intensity approximately constant.*

control stimuli with identical spheres (either monoscopic or stereoscopic lighting picked at random) and one pair of *test stimuli* with one monoscopically lit and one stereoscopically lit sphere (with the order of the spheres picked at random).

In total, we thus have

$$\underbrace{3}_{\substack{d\\ \text{distance}}} \times \underbrace{3}_{\substack{k_s\\ \text{spec. intensity}}} \times \underbrace{3}_{\substack{\sigma\\ \text{glossiness}}} \times \underbrace{2}_{\text{control/test}} = 54$$

different stimulus pairs, shown to each participant in random order. The set of parameters spans the most relevant range of values, based on a pilot study among the researchers using a wider variety of parameters. The chosen parameter set is also reasonably small, keeping the test time low (~10 minutes) to avoid fatigue and nausea.

Physical setup For all our tests, we have used an Oculus Rift DK2 HMD, which has a single low-persistence OLED display with 960×1080 pixels per eye, a diagonal field of view of approximately 100°, and a display refresh rate of 75Hz. The DK2 has a 6-DOF tracking system detecting both orientation and position. Our application renders the scene at $2\times$ more pixels per degree than the center of the display, for a total of $2364 \times 2922 \times 2$ pixels per frame. The rendered image is warped to compensate for the HMD optics using a high-quality filter with 5 bilinear texture lookups per color channel and display pixel. Our physical setup is shown in Figure 1.

In our application, the subject uses the mouse buttons to vote whether two presented stimuli are identical or different. A virtual mouse with buttons labeled "same" and "different" is rendered below and slightly in front of the subject, such that she can glance down at it at any time for a reminder. When the subject registers a vote, the stimuli fades to black, moves to a new distance, and fades back in with a new configuration over a period of 2 seconds. This process is repeated for all 54 stimulus pairs.

Limitations We had to make several restrictions in order to make the study practical. While varying depth and material, we limited

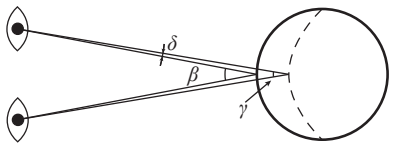
Figure 4: *Test subjects were shown a training example in VR of monoscopic lighting (top row) and stereoscopic lighting (bottom row), and the researchers explained which visual cues to look for when distinguishing between the two. This figure shows a fusible image of the presented objects, arranged for crossed viewing.*

ourselves to a fixed object shape, size, texture, and environment light source. We therefore made choices that we believe facilitate the subjects in the task of perceiving highlight disparity. Since it is difficult to perceive reflection correctness for complex objects [Ramanarayanan et al. 2007], we chose a sphere as it is a simple object with familiar shape. The light probe was chosen as it depicts a simple indoor environment, and contains an intense and distinct window light source with high-contrast, straight edges. Similarly, the marble texture was used to provide visual cues of the object surface, but is deliberately low contrast in order to avoid distracting the viewer from the lighting.

Perhaps the most significant limitation of our study comes from the physical limitations of the HMD itself. Table 1 shows the maximum convergence angle for a point on the object's surface and a point from the reflected environment for a viewer with an inter-pupillary distance (IPD) of 63mm, which is the mean IPD of our test subjects. With stereoscopic lighting, the reflection appears to lie behind the surface, which means that it has a slightly lower convergence angle (refer to Figure 4 for a visual example). The highlight disparity is the difference between the surface convergence angle and the reflection convergence angle. When observed in an Oculus DK2 (the device used in the experiment), the disparity amounts to only a few pixels as indicated in Table 1.

Test subjects The test was performed by 36 participants: 26 male and 10 female. The subjects are a mix of students and faculty members of Lund University, Sweden. Before testing, the IPD of each subject was measured.

Subjects were educated about the visual artifacts they would encounter before performing the test. Written information (available as supplemental material) was handed out, and the researchers were available for answering questions and explaining further. In addition, the subjects were presented with a training stimulus pair of one monoscopically and one stereoscopically lit sphere in VR, configured to accentuate the differences. The researchers guided them until they felt comfortable in what visual cues to look for, before proceeding with the actual test. Figure 4 shows a fusible image, similar to the training stimuli. The introductory spheres were presented at a distance of 0.25m, with a radius of 0.05m (2× the relative size of the other stimuli) and using a highly glossy BRDF



Distance	Surface vergence (β)	Reflection vergence (γ)	Highlight disparity (δ)	Highlight disp. pixels
0.5m	3.989°	3.780°	$\pm 0.209^\circ$	± 2.01px
1.25m	1.598°	1.514°	$\pm 0.084^\circ$	± 0.81px
3.125m	0.639°	0.606°	$\pm 0.034^\circ$	± 0.32px

Table 1: *Highlight disparity $\delta = \pm(\beta - \gamma)/2$, is the angular difference between the convergence angle for a point on the object surface, β, and the convergence angle for a point at the focus plane of the reflection (the dashed line), γ. At the distances used in our test, and the resolution of our HMD, this translates to 0.32–2.01px disparity between a point on the surface and the reflection.*

σ	k_s	$d = 0.5$m	$d = 1.25$m	$d = 3.125$m
4.2°	0.02	T: 7 C: 28	T: 8 C: 25	T: 15 C: 24
	0.10	T: 7 C: 25	T: 8 C: 26	T: 15 C: 26
	0.50	T: 7 C: 28	T: 9 C: 26	T: 21 C: 26
	all	*T: 21 C: 81*	*T: 25 C: 77*	*T: 51 C: 76*
12.6°	0.02	T: 10 C: 24	T: 21 C: 18	T: 25 C: 23
	0.10	T: 4 C: 23	T: 12 C: 21	T: 21 C: 21
	0.50	T: 9 C: 28	T: 16 C: 23	T: 26 C: 26
	all	*T: 23 C: 75*	*T: 49 C: 62*	*T: 72 C: 70*
37.8°	0.02	T: 32 C: 30	T: 26 C: 31	T: 31 C: 28
	0.10	T: 18 C: 20	T: 25 C: 24	T: 27 C: 26
	0.50	T: 11 C: 25	T: 20 C: 24	T: 26 C: 29
	all	*T: 61 C: 75*	*T: 71 C: 79*	*T: 84 C: 83*
all	0.02	*T: 49 C: 82*	*T: 55 C: 74*	*T: 71 C: 75*
	0.10	*T: 29 C: 68*	*T: 45 C: 71*	*T: 63 C: 73*
	0.50	*T: 27 C: 81*	*T: 45 C: 73*	*T: 73 C: 81*
	all	**T: 105 C: 231**	**T: 145 C: 218**	**T: 207 C: 229**

Table 2: *Number of "identical stimuli" votes by 34 participants. T is test stimuli (objectively different), C is control stimuli (objectively identical). Italic indicates subtotal for one parameter (glossiness or specular intensity), out of 102 votes. Boldface indicates total for both glossiness and specular intensity, out of 306 votes.*

($\sigma = 1.4^\circ$, $k_s = 0.2$).

Two out of the 36 participants could not perceive any difference between the test and control stimuli even at the introductory setting with guidance from the researchers. These two data points are treated as outliers, and statistical analysis is performed only for the remaining 34 subjects.

3 Results

The votes are summarized in Table 2. For each of the 27 configurations of $\{d, k_s, \sigma\}$, we compute a per-subject *score*, $x_i \in [-1, 1]$, where one point is awarded for identifying the difference in the test stimuli and one point is awarded for identifying the similarity in the control stimuli, and we apply a bias of -1. Random guessing therefore has an expected score of $\mathrm{E}[x] = 0.0$, as does providing the same answer to the tests and controls. Perfect perception has an expected score of $\mathrm{E}[x] = 1.0$. Average scores $\bar{x}$ and standard errors of means, $\mathrm{SE}_{\bar{x}}$, are tabulated in Table 3, where

$$\mathrm{SE}_{\bar{x}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

In Figure 5, a plot of the mean score of each participant is shown. The test subjects gravitate towards classifying the stimuli as be-

σ	k_s	$d = 0.5$m	$d = 1.25$m	$d = 3.125$m
4.2°	0.02	0.618± 0.095	0.500± 0.106	0.265± 0.114
	0.10	0.529± 0.097	0.529± 0.121	0.324± 0.092
	0.50	0.618± 0.085	0.500± 0.106	0.147± 0.105
	all	*0.588± 0.053*	*0.510± 0.063*	*0.245± 0.060*
12.6°	0.02	0.412± 0.096	-0.088± 0.115	-0.059± 0.103
	0.10	0.559± 0.086	0.265± 0.106	0.000± 0.103
	0.50	0.559± 0.105	0.206± 0.110	0.000± 0.112
	all	*0.510± 0.055*	*0.127± 0.065*	*-0.020± 0.061*
37.8°	0.02	-0.059± 0.059	0.147± 0.075	-0.088± 0.088
	0.10	0.059± 0.133	-0.029± 0.108	-0.029± 0.089
	0.50	0.412± 0.086	0.118± 0.092	0.088± 0.088
	all	*0.137± 0.059*	*0.078± 0.053*	*-0.010± 0.051*
all	0.02	*0.324± 0.056*	*0.186± 0.062*	*0.039± 0.061*
	0.10	*0.382± 0.065*	*0.255± 0.068*	*0.098± 0.057*
	0.50	*0.529± 0.053*	*0.275± 0.061*	*0.078± 0.059*
	all	**0.412± 0.034**	**0.239± 0.037**	**0.072± 0.034**

Table 3: *Voting score of 34 participants, $\bar{x} \pm 1\mathrm{SE}_{\bar{x}}$. Random guessing yields $\bar{x} = 0$, and perfect perception yields $\bar{x} = 1$.*

ing identical when not being able to see any difference, and having more correct than incorrect answers. Some subjects, however, had a bias towards classifying identical stimuli as "different", and two test subjects had more incorrect than correct answers.

The data set contains 54 configurations, with many possible correlations between independent variables and the measured response. If one were to perform many t-tests to find these correlations, there is a high risk of finding seemingly significant differences where there actually are none, as each individual test has a small probability of indicating significance due to chance. We therefore employ repeated measures ANOVA (analysis of variance), which is designed with this in mind and is commonly used in perceptual user studies. Its purpose is to reliably identify true correlations, while taking the increased probability of false positives into account. ANOVA is based on the F-test, which is used to quantify the ratio of variance explained by a parameter to the remaining variance. While the F-test assumes values to be normally distributed, ANOVA has been shown to be fairly robust when violating this assumption with a moderate number of samples [Donaldson 1966]. We therefore deem the normality assumption of ANOVA to be satisfied due to the reasonably large number of participants in our study.

Statistical analysis of the scores using three-way ANOVA (using MATLAB's `anovan` implementation) shows two significant main effects: the glossiness σ ($F_{2,915} = 33.11$, $p < 10^{-13}$) [3] and the distance d ($F_{2,915} = 25.96$, $p < 10^{-10}$). There is also a significant interaction between σ and d ($F_{4,909} = 3.78$, $p < 0.004$).

The main trends of mean score as functions of σ and d are shown in Figure 6, where it is clear that the score decreases with both σ and d. Figure 7 shows the interaction between σ and d, in which the detection distance increases with surface glossiness (decreasing σ). For instance, with $\sigma = 4.2^\circ$ there was no drastic difference between 0.5m and 1.25m, but with a large drop in score to 3.125m, while the largest drop with $\sigma = 12.6^\circ$ occurs closer, between 0.5m and 1.25m.

Figure 8 hints at a possible connection between specular intensity and glossiness. For surfaces with high glossiness, observers may detect highlight disparity even with low-contrast reflections (low k_s). As glossiness decreases, higher contrast (larger k_s) may be required to perceive the highlight disparity. This interaction is not significant at the $p = 0.05$ level, with $F_{4,909} = 2.37$, $p < 0.051$, so further studies with a larger sample are needed to establish whether there is a real correlation.

[3] F is a measure of the effect strength. p is the probability that there is no effect at all, and that the observed results are just due to chance.

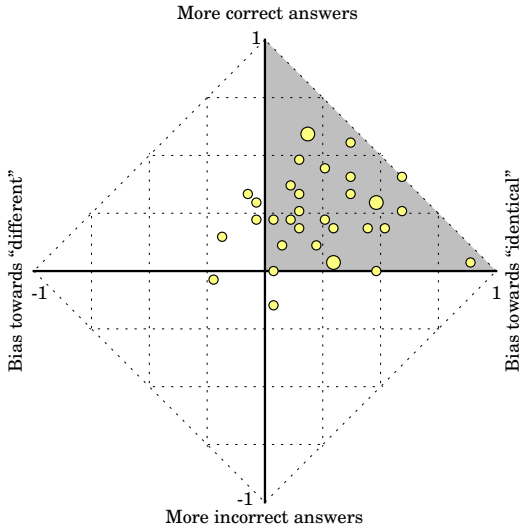


Figure 5: *A plot of the test scores of all participants (datapoints are accentuated for duplicates). The x-axis shows the test subject's bias towards answering "different" or "identical" when guessing, and the y-axis shows the average test score of each subject with zero indicating random guessing. The gray triangle shows where we initially expected the results to lie.*

4 Discussion

Score should not be interpreted as "share of viewers who can detect monoscopic lighting." The same score can be obtained in different scenarios: every viewer detects the difference some of the times, some viewers always detect the difference while others never do, all viewers detect the differences but sometimes think there may be one where none exists. A better interpretation is "share of correct detection unaccounted for by random guessing." A score of 0 does not necessarily mean that participants could not tell any difference, as a participant may deem different stimuli to appear very different and identical stimuli to appear slightly different. Voting "different" for both of these yields a score of 0.

Our initial hypothesis was that monoscopic lighting is not discernible from stereoscopic lighting at large distances. As can be seen in Figures 6 and 7, our data confirm this hypothesis, and at shorter distances than we first anticipated. The short detection threshold distance is most probably caused by the low resolution of the HMD used in the study (see Table 1). Our results should therefore not be interpreted as the overall limits of human perception, but rather the limits of human perception with such a low-resolution display device.

We find it remarkable that highlight disparity of only ± 0.32 pixels (see Table 1) had scores above random guessing with the glossiest materials (see Table 3). As HMD resolutions increase, the distance threshold at which highlight disparity becomes undetectable will likely also increase. Assuming that detection rate is proportional to highlight disparity in pixels, results for a 4K panel can be extrapolated by scaling the distances by a factor of 2. However, reliable detection distances are still short relative to typical scene scales. It is also worth noting that detection distances for surfaces with low gloss may not necessarily be limited by resolution, as scores were near zero for low-gloss surfaces even with ± 2 pixels of highlight disparity.

Monoscopic lighting can theoretically reduce the shading burden by up to 50% with a variety of different techniques, some of which

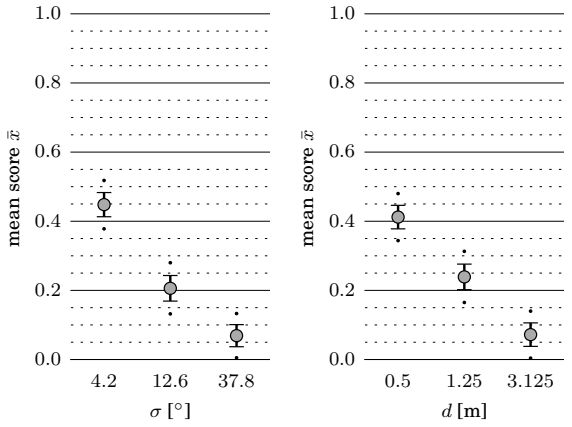


Figure 6: *Plot of score averages and standard errors for different levels of glossiness σ (left) and different distances d (right). The circles indicate where the average ($\bar{x}$) is located, while the horizontal lines indicate one standard error of means ($\mathrm{SE}_{\bar{x}}$) and small black circles indicate the corresponding two standard error of means limit.*

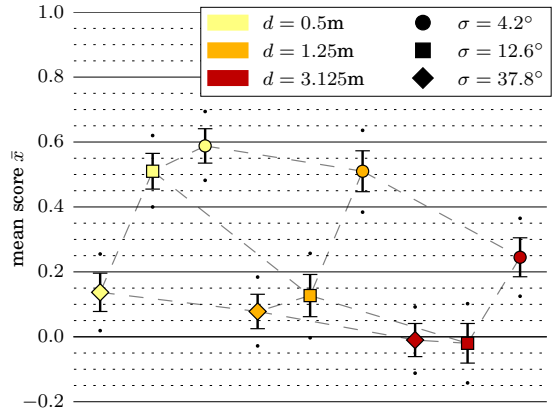


Figure 7: *Plot of score for different levels of glossiness (σ) and distances (d). The square, circles, and diamonds indicate where the averages are located and the rest of the indicators are the same as in Figure 6.*

were listed in Section 1. To leverage this opportunity, application developers will need to determine when and where to apply monoscopic lighting. Figure 9 shows an interpolation of the data in Figure 7, and can be used as a rough guideline. The score threshold used as monoscopic lighting criteria should ideally depend on the user's quality settings, with a reasonable range of 0.1 – 0.3. Alternatively, artists could tag their authored objects with detection thresholds. Further studies are required to understand the resolution dependence of the detection threshold, but a conservative estimate is simply scaling the distance axis by the HMD resolution (compared to the DK2 used in this study).

Our test is a side-by-side comparison, with the task to determine whether there is any difference between the two stimuli. Many participants did not see the difference on the introductory stimuli without further guidance from the researchers on what to look for, despite being educated about the nature of the artifacts in advance. In addition, two participants did not perceive any difference in the training scene even with extensive guidance. We therefore believe that the detection threshold without a reference may be much closer

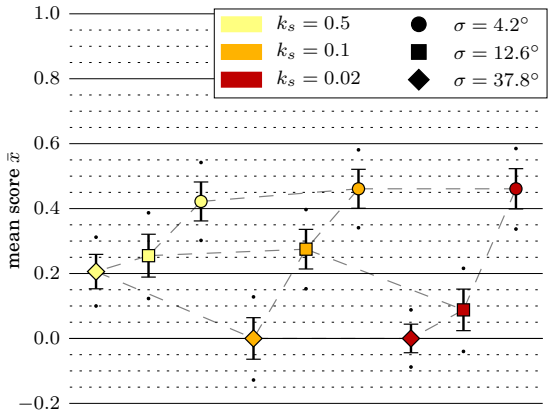


Figure 8: *Plot of score for different levels of glossiness σ and specular intensities k_s. For more glossy surfaces, lower contrast (specular intensity) may be sufficient to reliably detect the differences. As $p < 0.051$, further studies are needed to confirm this correlation.*

to the viewer, especially for more complex geometries, and that even total absence of highlight disparity might be an acceptable approximation in performance-constrained scenarios. Furthermore, the user's preference is not accounted for in our study. One participant commented that he preferred the monoscopically lit stimuli over the stereoscopically lit ones.

For applications that dynamically choose between monoscopic and stereoscopic lighting, the authors recommend using a stereoscopically lit transition region in which the IPD value used for lighting is progressively lowered to zero. This can prevent spatio-temporal discontinuities in appearance near the threshold.

Applications usually distinguish between large reflective flat surfaces (e.g., mirrors and polished floors) and other objects. Reflections in the former are usually rendered by re-drawing the entire scene reflected around the plane, or by some image-space ray marching technique. Reflections in the latter are most often rendered less accurately using environment cube maps. The authors recommend retaining this distinction, as monocular lighting of a flat mirror-like surface would be more easily detectable than monocular lighting of more complex objects, and should probably be avoided.

Future Work For VR, it is important to generate knowledge from user studies so that informed decisions can be made about which rendering methods to use for a particular platform and how to trade performance for image quality. It would be valuable to implement an adaptive algorithm, combining monoscopic and stereoscopic lighting as appropriate, in a modern game engine and study the effects on a wide range of scenes. We would also like to complement our results with an HMD with a higher resolution display in order to evaluate resolution dependence. It would also be very interesting to design a study without reference stimuli (in contrast to our side-by-side test). In addition, studying the psychophysical effects of level-of-detail techniques with gaze tracking is also an interesting direction for future work.

Acknowledgements

We thank the anonymous reviewers for their feedback. We thank David Blythe for supporting this research. Tomas Akenine-Möller is a *Royal Swedish Academy of Sciences Research Fellow* supported by a grant from the Knut and Alice Wallenberg foundation.

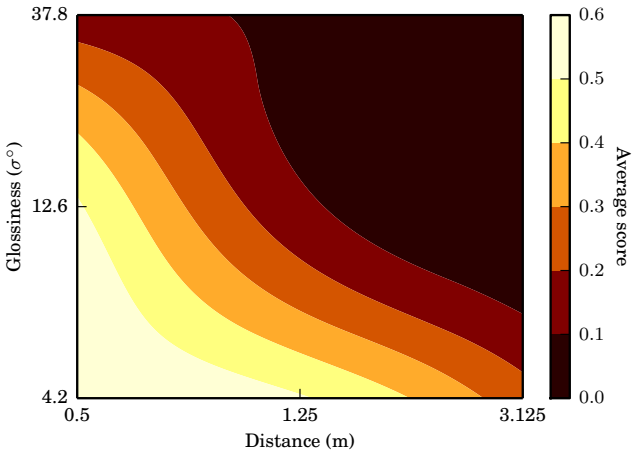


Figure 9: *A contour plot of the data from Figure 7, interpolated bicubically in logarithmic space. The score indicates probability of detection with a reference present, and can be used to guide adaptive LOD algorithms.*

References

ANDERSSON, M., HASSELGREN, J., TOTH, R., AND AKENINE-MÖLLER, T. 2014. Adaptive Texture Space Shading for Stochastic Rendering. *Computer Graphics Forum, 33*, 2, 341–350.

CLARBERG, P., AND MUNKBERG, J. 2014. Deep Shading Buffers on Commodity GPUs. *ACM Transactions on Graphics, 33*, 6, 227:1–227:12.

CLARBERG, P., TOTH, R., AND MUNKBERG, J. 2013. A Sort-Based Deferred Shading Architecture for Decoupled Sampling. *ACM Transactions on Graphics, 32*, 4, 141:1–141:10.

CLARBERG, P., TOTH, R., HASSELGREN, J., NILSSON, J., AND AKENINE-MÖLLER, T. 2014. AMFS: Adaptive Multi-Frequency Shading for Future Graphics Processors. *ACM Transactions on Graphics, 33*, 4, 141:1–141:12.

COOK, R. L., CARPENTER, L., AND CATMULL, E. 1987. The Reyes Image Rendering Architecture. In *Computer Graphics (Proceedings of SIGGRAPH 87)*, ACM, vol. 21, 95–102.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Adaptive Image-space Stereo View Synthesis. In *Vision, Modeling and Visualization Workshop*, 299–306.

DĄBAŁA, Ł., KELLNHOFER, P., RITSCHEL, T., DIDYK, P., TEMPLIN, K., MYSZKOWSKI, K., ROKITA, P., AND SEIDEL, H.-P. 2014. Manipulating Refractive and Reflective Binocular Disparity. *Computer Graphics Forum, 33*, 2, 53–62.

DONALDSON, T. S. 1966. *Power of the F-test for Nonnormal Distributions and Unequal Error Variances*. Rand Corporation.

DOVE, H. W. 1851. Über die Ursachen des Glanzes und der Irradiation, abgeleitet aus chromatischen Versuchen mit dem Stereoskop. *Annalen der Physik, 159*, 5, 169–183.

FERWERDA, J. A., RUSHMEIER, H., AND WATSON, B. 2002. Psychometrics 101: How to Design, Conduct, and Analyze Perceptual Experiments in Computer Graphics. In *ACM SIGGRAPH Courses*.

HASSELGREN, J., AND AKENINE-MÖLLER, T. 2006. An Efficient Multi-View Rasterization Architecture. In *Eurographics Symposium on Rendering*, 61–72.

KIRSCHMANN, A. 1895. *Der Metallglanz und die Parallaxe des indirecten Sehens*. Verlag von Wilhelm Engelmann.

LIKTOR, G., AND DACHSBACHER, C. 2012. Decoupled Deferred Shading for Hardware Rasterization. In *Symposium on Interactive 3D Graphics and Games*, 143–150.

MCMILLAN, L., AND BISHOP, G. 1995. Head-Tracked Stereoscopic Display Using Image Warping. In *Proceedings of SPIE 2409, Stereoscopic Displays and Virtual Reality Systems II*, 21–30.

MURYY, A., WELCHMAN, A., BLAKE, A., AND FLEMING, R. 2013. Specular Reflections and the Estimation of Shape from Binocular Disparity. *Proceedings of the National Academy of Sciences, 110*, 6, 2413–2418.

NEHAB, D., SANDER, P. V., LAWRENCE, J., TATARCHUK, N., AND ISIDORO, J. R. 2007. Accelerating Real-time Shading with Reverse Reprojection Caching. In *Graphics Hardware*, 25–35.

RAGAN-KELLEY, J., LEHTINEN, J., CHEN, J., DOGGETT, M., AND DURAND, F. 2011. Decoupled Sampling for Graphics Pipelines. *ACM Transactions on Graphics, 30*, 3, 17:1–17:17.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual Equivalence: Towards a New Standard for Image Fidelity. *ACM Transactions on Graphics, 26*, 3 (July), 76:1–76:12.

SAKANO, Y., AND ANDO, H. 2010. Effects of Head Motion and Stereo Viewing on Perceived Glossiness. *Journal of Vision, 10*, 9, 15:1–15:14.

SITTHI-AMORN, P., LAWRENCE, J., YANG, L., SANDER, P. V., AND NEHAB, D. 2008. An Improved Shading Cache for Modern GPUs. In *Graphics Hardware*, 95–101.

SITTHI-AMORN, P., LAWRENCE, J., YANG, L., SANDER, P. V., NEHAB, D., AND XI, J. 2008. Automated Reprojection-based Pixel Shader Optimization. *ACM Transactions on Graphics, 27*, 5, 127:1–127:11.

TEMPLIN, K., DIDYK, P., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. Highlight Microdisparity for Improved Gloss Depiction. *ACM Transactions on Graphics, 31*, 4, 92:1–92:5.

WANG, J., REN, P., GONG, M., SNYDER, J., AND GUO, B. 2009. All-frequency Rendering of Dynamic, Spatially-varying Reflectance. *ACM Transactions on Graphics, 28*, 5, 133:1–133:10.

WENDT, G., FAUL, F., AND MAUSFELD, R. 2008. Highlight Disparity Contributes to the Autenticity and Strength of Perceived Glossiness. *Journal of Vision, 8*, 1, 14:1–14:10.

WENDT, G., FAUL, F., EKROLL, V., AND MAUSFELD, R. 2010. Disparity, Motion, and Color Information Improve gloss constancy performance. *Journal of Vision, 10*, 9, 7:1–7:17.