Sensitivity to Peripheral Artifacts in VR Display Systems

David M. Hoffman, Zoe Meraz, Eric Turner
Google, Mountain View CA

Abstract

We evaluated the visual system's sensitivity to different classes of image impairments that are closely associated with rendering in VR systems. Even in the far periphery, the visual system was highly sensitive to volatile downsampling solutions. Temporally stable downsampling in the periphery was generally acceptable even with sample spacing up to half a degree.

Author Keywords

VR; peripheral acuity; foveated rendering; downsamping; spatial-temporal artifacts

1. Introduction:

Depending on the acuity metric, people can visually resolve details with spatial frequencies of close to 60 cycles per degree. For a pixelated display system, at half this frequency with 60 pixels per degree, the display panel can render 20/20 lettering and is accepted for a target resolution for a premium experience. However, the visual system is only sensitive to such fine details in the central 0.5° region of the retina (the foveola), with sensitivity to details falling off rapidly with eccentricity [1]. It is possible to take advantage of these properties for high-resolution display systems by developing solutions to render and transmit imagery with lower information bandwidth in the periphery. In such systems, ideally the image degradation in the periphery will go unnoticed by the viewer but yield meaningful savings in power and compute.

Based on letter acuity studies, the fall-off in letter acuity with eccentricity is fairly linear for high contrast text exemplified by the minimum feature size in an acuity task increasing nearly 14 times at 30° eccentricity compared to 1° eccentricity. The fall-off can be even more pronounced with other factors such as crowding and reduced contrast [2].

For modern VR systems, even with real imagery that is much lower contrast and has more crowding features than a letter acuity stimulus, resolution reduction in the periphery has been limited to levels far more modest than we might expect from the letter acuity experiments reported in the literature. Studies in VR systems have explored system level constraints on rendering including Albert and colleagues [3] on latency and transition boundaries, but there has been little work on explicit sensitivity to image artifacts in localized regions of peripheral vision.

A VR system has several fundamental differences from conventional direct-view displays. It is reasonable that one might know approximately where the eye is fixated by active eye tracking, or priors about what types of eye movements are comfortable or typical in a wide field-of-view system [4]. Also, in a head-mounted display, imagery is always moving; accordingly, world-static imagery must be counter-shifted to the head motion to avoid discomfort. The head moves constantly with every breath so there is a constant requirement to adjust the pixel data, even when there is no overt movement in the scene or deliberate neck motion. VR systems, with the demands of stereoscopic rendering, compensation for lens distortions and high frame rates to mitigate flicker and motion blur, push real-time graphics to their limit; any savings possible are eagerly seized to reduce power and weight, or to permit greater scene complexity.

This work is designed to explore the visual system's sensitivity to different types of impairments representative of graphics or optical downsampling at different locations in the retinal field of view. Differing from classical psychophysical literature on letter acuity, these tests emphasize the use of natural imagery with image resampling impairments to probe the sensitivity of the visual system to display artifacts and loss of high frequency image content rather than details that are part of the original image such as letter identification.

2. Experiment:

Our experiment explores the sensitivity of the visual system to the types of artifacts that could be introduced by a VR system at different parts of the visual field.

We test three specific classes of artifacts for visibility: blur, temporally stable aliases, and volatile aliases.

Each of these artifacts can be present in a typical VR user experience. We can apply these distortions to different classes of imagery to see how the effects manifest when viewed in different parts of the visual field. In this experiment, we tested two types of imagery: photography and computer graphics. The test image selected to represent each type is shown in the top row of Figure 1. Photography can be detail heavy and is widely found in VR as 360 and 180 degree image formats and video capture gain popularity. For these experiments we use an image of a gathering called *Crowd*. Photographic content is often captured in high fidelity but must be downsampled to meet storage and transmission goals. The second type of imagery is computer graphics which can be used as part of games and other real-time interactive experiences. It is exemplified with *Forest*. For real time rendered content, there is a tradeoff between fidelity and the rendering time. This tradeoff is particularly prominent for mobile VR experiences in which the polygon count is deliberately limited to work within the bounds of existing compute resources. This can lead to straight edges and simplified textures.

Blur: Blur in VR is often the result of either an optical process or from a deliberate graphics manipulation. With wide field of view headsets, blur is inadvertently introduced by losses in optical quality due to optics limitations towards the periphery of the field of view. In this case, the amount of blur is not a direct attribute of the display panel itself, but rather the lens and it increases with eccentricity. Near the center of the field of view, the system resolution is often limited by the display panel and blur would originate almost exclusively as part of the digital imaging pipeline.

One of the key situations in which blur could be introduced digitally is that image content is available in rasterized full-resolution, but then needs to be compressed, such as for cloud-computing and transmission [5]. An example of image blur is shown in the second row of Figure 1. Whether via optical or graphics, blur removes the details without aliasing acting as a low-pass filter. A similar type of blur is introduced at a pixel scale for anti-aliasing processing.

SID 2018 DIGEST ISSN 0097-996X/18/4702-0858-$1.00 © 2018 SID

Temporally stable aliased: The second class of artifacts tested are temporally-stable aliases. Spatial aliasing occurs when the virtual content detail is higher than the rendering resolution. This type of image resampling can lead to discontinuities, jaggies and gaps that are position-locked to the content. These artifacts may be introduced by the use of artificially reduced texture detail, such as when an application forces textures to the next mipmap level. Applications can also dynamically adjust the geometry complexity to reduce computation when objects are far away or in the peripheral vision [6].

Although these aliasing artifacts are present, they are aligned to the world geometry rather than the panel coordinates, meaning they don't produce flickering or scintillation effects during head movement. An example of aliasing effects due to reduced resolution is shown in the third row of Figure 1. Note that the slanted branches in *Forest* become vertical and there is a break between the tree trunk and the branch.

Volatile aliased: The third class of artifacts are temporally-volatile aliasing effects. Much like the spatial aliasing mentioned above, this class of artifacts occurs due to image sampling being performed at a lower resolution than the native fidelity of the content. In this case, aliasing artifacts are aligned with the pixel grid of the output display, rather than the geometry itself. This effect can be a result of having a relatively low-resolution panel or by forcing the rendering to be performed at an artificially low resolution in certain parts of the screen, such as for foveated rendering [7]. The advantage of such methods are that they can improve the efficiency of the rendering pipeline with little overhead. An unfortunate result is that additional spatio-temporal aliasing artifacts are introduced. As the user moves their head, these artifacts move relative to the virtual content. This relative movement causes flickering or scintillation effects, resulting in temporally-volatile artifacts.

Figure 2 shows an example of how temporal artifacts can change the rendered image from frame-to-frame. As the heading angle of the VR headset changes relative to the virtual content coordinate system, the pixel alignment moves with respect to the virtual content. As shown in each column, looking at the same content but with slight motion can lead to dynamic aliasing patterns. This effect causes the tree to change shape each frame, producing spatio-temporal flicker artifacts, even though the original content is static. Subjectively, these dynamic resampling artifacts manifest as a scintillation that is typical of image reprocessing on frame by frame basis [8].

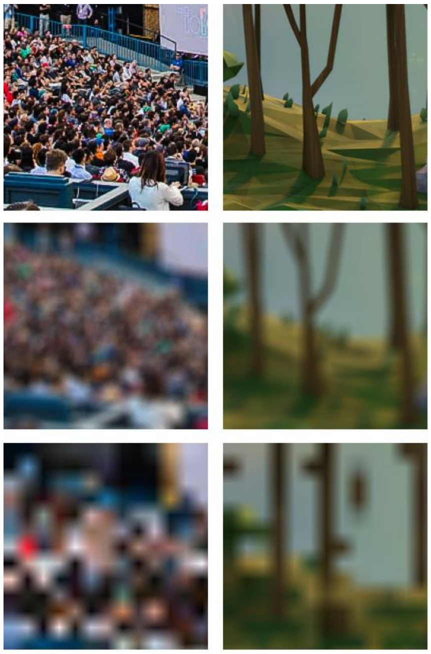

Figure 1. Test images under different visual artifact conditions: original full-resolution imagery (first row); filtered via cylindrical blur (second row); aliased due to nearest-neighbor downsampling and bilinear upsampling (third row). (Downsampling exemplified with 20 arcmin sampling spacing. 1/20th downsampling based on original 1 arcmin spacing)

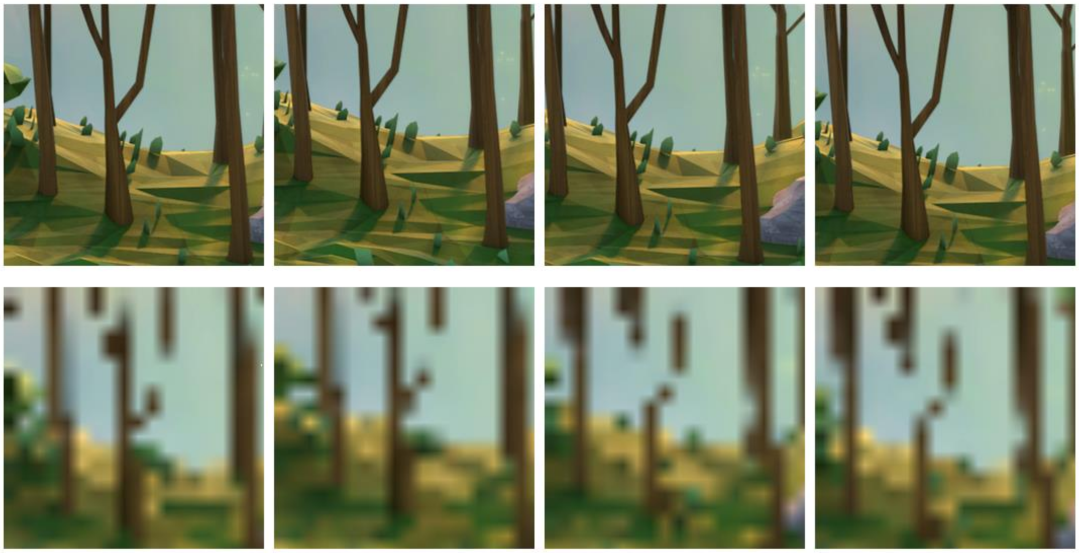

Figure 2. Example of temporal fluctuation. The same content is observed with different shifts, and even though the content is unchanged, the aliasing pattern is dynamic frame-to-frame (4 frames here showing volatile aliasing).

As shown later in this paper, subjects are much more sensitive to these temporally-volatile artifacts than temporally-stable artifacts of the equivalent size. Existing foveation approaches have attempted to reduce the severity of these effects in a number of ways. Post-processing spatial filtering can be applied to reduce the magnitude or detectability of these artifacts [9]. Temporal filtering can also be applied post-render to reduce the scintillation effect [10]. Methods can also reduce the initial generation of scintillation artifacts prior to rendering by either randomizing the rasterization sampling [11] or by aligning the pixel grid to world coordinates [12]. Each of these methods introduce trade-offs of extra computation for improved visual quality.

Method: To avoid the limitations in resolution and optics with the current HMD systems, we presented imagery on a Sony BVM 300 OLED monitor with 4K resolution and 60 Hz update. When viewed from 55 cm, pixels subtended 1 arcmin at the center.

Observers were given a forced-choice task of selecting the superior rendering (i.e. 'the reference') via button press. The reference and test image sequences consisted of a static aperture, behind which an image orbited (in-plane translation) in a 15 pixel radius circular path to emulate head movement. The test and reference sequences were displayed simultaneous with a 5° vertical offset (An example screen is shown in Figure 3). The position of the test image (top or bottom in the pair) was randomized per trial. A fixation target remained on the display throughout the test. The horizontal eccentricity in the nasal visual field of the image pair was set to one of the following settings with respect to the fixation target: 0, 10, 20, 30 or 40°. Observers were instructed to maintain fixation on the target to stimulate different parts of the retina. There was no time limit, but observers typically made a decision in 1-2 seconds. If they could not perceive a difference between the images they were asked to guess. If they were conscious of having a fixation failure, they were encouraged to reset the trial.

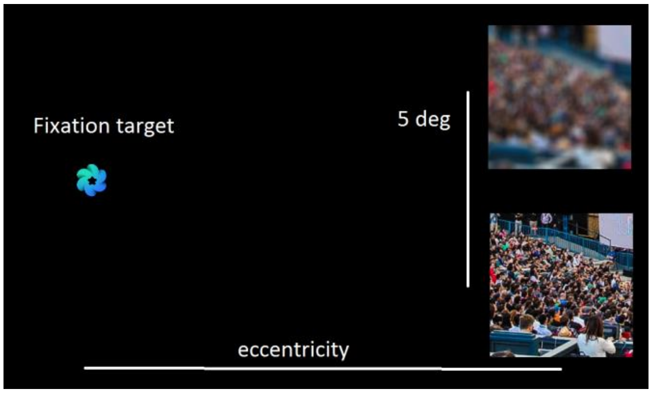

Figure 3. Example stimulus presentation target with force choice presentation and fixation target.

Based on their response, the level of downsampling/blur was adjusted via a 2 down/1 up staircase procedure (with a clamped range of 2 to 30 arcminutes). There were a total of 30 combinations of stimuli with five eccentricities, three types of impairments, and the two images shown in Figure 1. All of the conditions were randomized and each observer completed a minimum of 30 trials per condition.

Five observers, with a letter acuity of 20/20 or corrected to 20/20, completed each test. Ages ranged from 25 to 40 years old.

Results: The experiment was designed to look for differences in visual sensitivity to downsampling for the different approaches and at different eccentricities. We fit the staircase data with a cumulative Gaussian function to estimate the 75% correct point as a threshold for detecting the downsampling. These thresholds are plotted as a function of eccentricity in Figure 4 with the blur, stable alias and volatile alias conditions represented by different colors. The *Crowd* and *Forest* images are plotted separately. The dominant effect was that the visual system was highly sensitive to detecting even small amounts of downsampling with the volatile alias condition. In these conditions the image scintillated and the temporal changes were highly noticeable, especially in the *Crowd* photo. This extended out to 40° eccentricity.

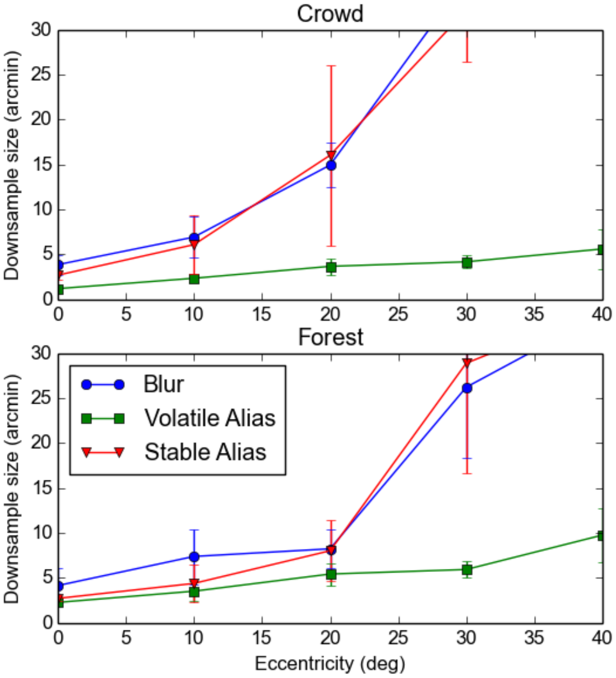

Figure 4. Average threshold for downsampling across all observers with standard deviation represented as error bars. The temporally stable downsampling approaches of blur and stable alias are shown as blue circles and red triangles respectively. The frame-by-frame resampling that introduces spatio-temporal scintillation is shown as green squares.

For the temporally stable conditions of blur and world-based downsampling, there was strong sensitivity to these artifacts up to 20° eccentricity, with the downsampling visible at 10 arcmin spacing. Sensitivity became much lower beyond 20° eccentricity. There was little difference between the blur and stable-aliased conditions. This suggests that in the periphery, the visual system is indifferent to the loss or presence of the high spatial frequency details and the associated aliases.

At the mid eccentricities there was a trend towards the visual system being more sensitive to the temporally stable impairments for *Forest* image than *Crowd* image. This may be related to the cluttered nature of *Crowd* which hides some of the artifacts, compared to the clean nature of the *Forest* image with isolated edges and obvious breaks in the tree branches.

For the volatile aliasing downsample approach, the visual system may have been slightly more sensitive to artifacts in the *Crowd* image compared to the *Forest* image. The detail-heavy nature of crowd offers features throughout that can readily scintillate with every position update.

3. Impact:

Current generation HMDs have panels with pixel sizes that are approximately 5-8 arcminutes. At this type of sampling spacing, we would expect any foveated rendering inside of a 20^o radius to potentially lead to artfacts to which the visual system is sensitive. Within this central zone, even if the image were rendered without explicit downsampling, anti-aliasing at the native panel resolution would be expected to improve image quality by reducing alias volatility. At 30^o retinal eccentricity and beyond, modest downsampling could be possible if care is taken to ensure that the imagery is drawn with stable resampling. Frame-by-frame resampling could introduce visible scintillation even in the far periphery. Even with downsampling at 30 arcmin, observers were unable to discriminate the downsampled from the full resolution image at 30^o and beyond. This applied both to blurred as well as world-aligned (stable) aliased.

If it becomes possible to construct panels and optical systems capable of resolving 1 arcmin details, aggressive downsampling to reduce bandwidth and computational load becomes a much more practical solution. At 10, 20 and 30^o retinal eccentricity, we found downsampling thresholds greater than 2, 3, and 4 arcmin respectively, even with volatile aliasing. With stabilized aliases, at 30^o retinal eccentricity and beyond, downsampling to 15 arcmin or more would be possible. With a high PPI display system with pixel sizes that are on the order of 3 arcmin or less, such savings could be substantial.

Reaching the ideal pixel pitch and panel size to address acuity and field-of-view targets opens up an opportunity for more aggressive downsampling in the periphery and this will be essential as the pixel count increases by up to 16X, and we begin to approach bandwidth limitations for internal interfaces[13]. In such a system, the potential savings from foveated rendering becomes too large to ignore, and if designed well, could only require a modest increase in the number of pixels rendered compared to current systems.

4. References:

[1] SM. Anstis. A chart demonstrating variations in acuity with retinal position. Vision research. 31;14(7):589-92. (1974)

[2] H. Strasburger, I. Rentschler, M. Jüttner Peripheral vision and pattern recognition: A review. Journal of vision. 1;11(5). (2011)

[3] R. Albert, A. Patney, D. Luebke, J. Kim. Latency Requirements for Foveated Rendering in Virtual Reality. ACM Transactions on Applied Perception 14(4):25. (2017)

[4] M. Haynes, and T. Starner, Effects of Lateral Eye Displacement on Comfort While Reading from a Video Display Terminal. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(4), p.138.2018.

[5] B. Bastani, E. Turner, C. Vieri, H. Jiang, B. Funt, and N. Balram. Foveated Pipeline for AR/VR Head-Mounted Displays. Information Display, 33(6). (2017)

[6] H. Hoppe. Progressive meshes. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM. (1996)

[7] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder. Foveated 3d graphics. ACM Transactions on Graphics (TOG), 36(6). (2012)

[8] D.M. Hoffman, and D. Stolitzka. A new standard method of subjective assessment of barely visible image artifacts and a new public database. Journal of the Society for Information Display, 22(12), pp.631-643 (2014).

[9] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn. Towards foveated rendering for gaze-tracked virtual reality. ACM Transactions on Graphics (TOG), 35(179), (2016).

[10] B. Karis. High-Quality Temporal Supersampling. SIGGRAPH 2014.

[11] M. Stengel, S. Grogorick, M. Eisemann, and M. Magnor. Adaptive Image-Space Sampling for Gaze-Contingent Real-time Rendering. EUROGRAPHICS, 35(4). (2016)

[12] E. Turner, H. Jiang, D. Saint-Macary, and B. Bastani. Phase-Aligned Foveated Rendering for Virtual Reality Headsets. IEEE VR 2018, the 25th IEEE Conference on Virtual Reality and 3D User Interfaces. (2018)

[13] A. Amirkhany, M. Hekmat, S. Sankaranarayanan, A. Jose, V. Abramzon, N. Jaffari, K. Saito, M. Elzeftawi, M. Wang, S. Moballegh, G. Malhotra. 9-5L: Late-News Paper: 6Gb/s Ultra Definition Display Interface (UDDI) for Large-size 8K Displays. InSID Symposium Digest of Technical Papers. 48(1) pp. 108-111. (2017).