Modeling and Optimizing Human-in-the-Loop Visual Perception Using Immersive Displays: A Review

Qi Sun*, Budmonde Duinkharjav*, Anjul Patney**

***New York University**

****Nvidia Research**

Abstract

New and rapidly-evolving classes of display devices bridge the gap between us and the immersive experiences of the future. The most intimate of these displays are the Virtual- and Augmented-Reality (VR and AR) ones, because they are capable of presenting synthetic environments that rival those in the real world. This ecosystem of personal and highly-immersive displays offers new challenges for research in computer graphics, display technologies, and human visual perception. While the extensive advancements in the areas of display and computer graphics technologies traditionally end at the on-screen "image," there are several untapped opportunities for advances that exploit the interplay between the display characteristics and how our visual system perceives them.

In this article, we review recent progress in understanding and modeling the perception of immersive displays, as well as perceptually optimizing display technologies for immersive experiences. We present this review in the form of a taxonomy that maps the various properties of modern displays with the perceptual phenomenon that most closely interacts with them. From this taxonomy, we deduce several unsolved challenges in understanding human perception of displays, as well as perceptually-optimal characteristics of future displays.

Author Keywords

Human Perception; Virtual Reality; Augmented Reality; Computational Displays; Vision Science.

1. Introduction

Immersive displays in virtual and augmented reality (VR and AR) aim to recreate a realistic visual world. To achieve this goal, significant efforts have been deployed to increase spatial [1], angular (i.e., 3D capabilities) [2], temporal [3] resolutions, and dynamic ranges [4], etc. A large proportion of recent research in the areas of display optics and computer graphics terminates at the "screen" and only considers simplistic models of the end-user's perceptual faculties.. However, the final receiver of the displayed content, our visual system, is a complex biological and neural system. With growing diversity and ubiquity of display systems, especially near-eye head-mounted displays, it is increasingly insufficient to overlook their interactions with the human visual system.

In this article, we review recent research that studies and models human visual perception on immersive displays. We primarily focus on how modern and upcoming display technologies interact with different biological and neurological behaviors of visual perception. Specifically, we classify the behaviors into spatial, temporal, and behavioral categories. We also discuss the possibilities of inversely optimizing human perception with various applications. As a step further, we envision several future research directions that improve the immersive displays in a human-centered fashion.

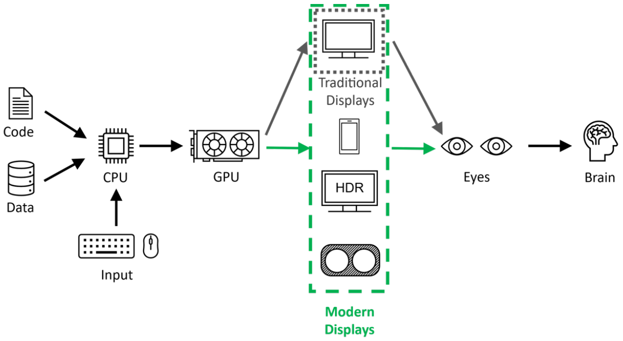


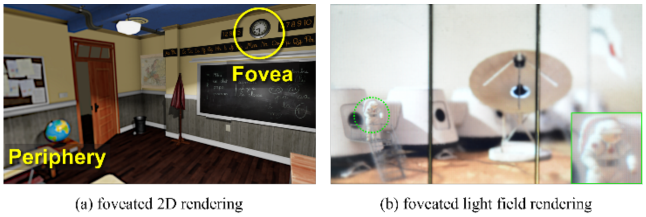
Figure 1. Modern view of displays in computer graphics. We interact with an ever-increasing set of digital displays, each of which presents novel challenges to understanding perception.

2. Immersive Displays Meet Human Perception

Table 1 summarizes how display properties connect to the perceived visual content, as well as the most prominently-influenced perceptual behaviors

Table 1. Our taxonomy of the correlations between perceptual characteristics (first column) display properties (second column). Example research being reviewed is listed in the third column. Note that the actual correlations are not one-dimensional but instead consist of various interplays.

Perceptual Characteristic		Display Property	Example Research
Spatial	Retinal Acuity	Spatial Resolution, Subpixel layout	[25]
	Stereo Vision	Stereo Displays	[12]
	Accommodation	Varifocal and Light Field Displays	[2, 6]
	Peripheral Vision	Display Foveation	[24]
Temporal	Flicker & judder perception	Refresh rate & persistence	[5]
Behavioral	Simulator sickness	Display latency, VAC	[11]
	Change blindness	Display flickering	[2.13]



(a) foveated 2D rendering

(b) foveated light field rendering

Figure 2. Spatial perception. (a) shows a foveated rendering approach on 2D stereo head-mounted displays [12]. (b) shows a foveated rendering tailored for light field displays by preserving both spatial and depth perception [2].

2.1 Spatial Properties

Humans perceive space primarily by integrating multi-modal visual cues. The retinal receptors, which act as visual sensors for our eyes, are finite in number and are distributed in a highly non-uniform fashion across the visual field [12]. The discrepancy leads to the mismatching information density between most existing displays and their users.

(a) Resolution vs. visual acuity. Immersive displays are close to the eye, resulting in the demand for ultra-high resolution. The high resolution not only brings challenges of optical design but also high computational and energy costs. Further, unlike most screens, human vision is foveated, with its acuity being highest at the retinal center, a.k.a. fovea, and decreasing progressively toward peripheral vision. Several recent approaches in the area of foveated rendering have proposed to utilize this effect to improve performance and quality of graphics on near-eye displays [12, 13, 20, 28, 30]. An example is shown in Figure 2(a). However, due to the complexity of foveated rendering algorithms, realizing practical performance enhancement in FPS demands novel GPU architectural support [30].

(b) 3D displays vs. depth perception Perceiving depth from the physical world is a completed procedure that integrates multiple cues such as occlusion, motion parallax, accommodation, and binocular disparity. Compared to traditional displays, head-mounted displays offer stereo cues, but do not always present images at the correct focal distance. This causes the well-known vergence-accommodation conflict (VAC). Light field and holographic displays optically create natural defocus blur thus showing their potential to be the future consumer-level immersive displays [31,32].

Once again, efficiency of practical systems that avoid VAC remains an open and ongoing research challenge. The main roadblocks are their high computation load and noisy quality. By leveraging the eccentricity-based depth acuity [7] towards a closed-form sampling theory, the computation of a light field was demonstrated to be reduced to 16%-30% without the loss of either quality or depth cues (as in Figure 2(b) and [2]). A similar method can also be applied to holographic displays with perceptually reduced speckle noise [6, 33].

2.2 Temporal Properties

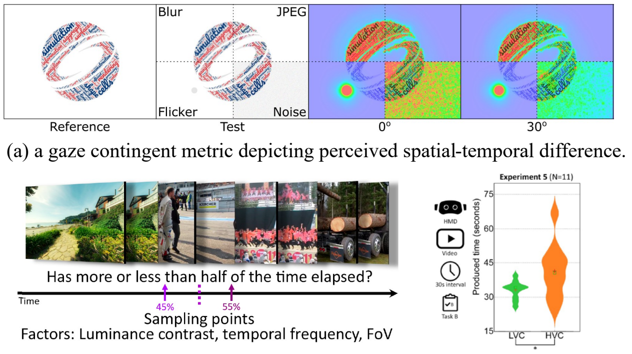
(a) Latency In the vision of metaverse, the virtual content is typically stored in the cloud and dynamically stream to the user end. However, remote data transmission inevitably causes latency regardless of network condition. In [14], Albert et al. suggested the minimal latency as 80-100ms for a foveated immersive display. The study provides general guidance of the imperceivable delay for both local computation and remote transmission.

(b) Refresh rate vs. flicker/motion perception Displays with low refresh rate can cause several visual artifacts, e.g. judder, flicker, and blur, which significantly deteriorate user experience and perceived realism of moving content [34]. The experience is much more objectionable for near-eye displays [3]. Consequently, high refresh-rate displays on desktop, mobile as well as VR devices are becoming increasingly common [35]. However, rendering high-quality visual content at ultra-high framerates is challenging, and has led to explorations of variable and dynamic refresh rate systems [30, 34]. Further, compared to spatial models, computational models of how humans perceive temporal changes in images are relatively new and underdeveloped [26, 27].

(c) Image characteristics vs. time perception. Given a stimulus, the precise perception of timing significantly affects our performance in critical tasks such as e-sports, defense, or long-term usability/fatigue. In [10, 15], a series of studies reveal how displayed and visual content affects our perception of time. Surprisingly, from low-level visual features such as frequency/contrast to mid-level display properties such as FoV, various visual content may alter our judgment of how long a stimulus lasts. As the first study on the correlation between display properties and long-scale time perception (up to minutes), we discovered a remarkably consistent pattern: larger displayed featured all shortened perceived time in intervals of up to 3min. Here, "larger" is defined in broad semantics such as higher FoV/contrast, and denser image content, as shown in Fig. 3(b). The discoveries suggest the possibilities of optimizing display properties and content to alter temporal perception at practice, towards the ultimate goal of reducing fatigue and improving performance.

2.3 Behavioral Effects

(a) Simulator sickness is a core roadblock of immersive displays to replace our current mixed-reality devices. Completely eliminating sickness will allow for long-term usage. Commonly, a major cause of sickness is the sensory conflict between the incomplete simulation in displays and the physical world. For instance, beyond the extensively studied vergence-accommodation conflict [2], our study revealed the remarkable role of visual-



(a) a gaze contingent metric depicting perceived spatial-temporal difference.

(b) display properties alter our perception of time in second-to-minute scales

Figure 3. Display properties and temporal perception. (a) shows a high field-of-view (FoV) video metric that predicts spatial and temporal visual similarities, considering several factors such as motion and flicker [26]. At long-term (seconds-minutes) scale, (b) shows how the display contrast/frequency/FoV all significantly affect our perceived time. [10] Here H-/L-VC indicates higher/lower-level features, e.g., larger/smaller FoV.

vestibular conflict in VR displays [11]. Due to the spatial limitation, the navigation in a virtual environment is typically implemented with the controller. Consequently, in highly dynamic virtual environments, the vestibular-sensed (stationary) conflicts with the visually sensed (moving) motions.

(b) Change blindness. Humans rapidly move the eye (a.k.a. saccades) 3-4 times per second, more frequent than our heartbeats (see Fig. 4). This behavior shifts the visual area of interest (i.e., targets) and the high-acuity foveal region. In highly interactive and intensive scenarios such as aircraft operation, saccade timing and landing accuracy determine our performance in searching peripheral targets. The displayed image flickering may change our understanding of peripheral vision, thus triggering saccades [16]. It is well-known that our visual acuity is significantly suppressed during, and even before/after saccades. Recent studies have shown that the change blindness may also change our locomotive actions without noticing [17].

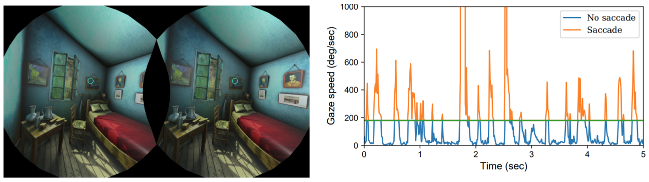


Figure 4. Eye motion and change blindness behaviors while observing a stereo-displayed stimulus from [17]. The right image shows the gaze motion speed (Y) along time (X). The orange areas indicate those periods with saccade occurrence.

3. Challenges and Future Directions

(a) A generalizable display-perception metric Metrics that predict perceived image quality are essential for evaluating modern displays and to identify the development/manufacturing directions. Most existing quality metrics, such as SSIM, PSNR, and [26], have limited scenarios in which they are useful and reliable. There are several modern use cases where usage is impractical or impossible. E.g., comparing quality between two ray-traced images often requires generating a reference image which might be prohibitively expensive to compute, especially at a large scale. Also, almost no existing metrics fully account for temporal behavior, including the impact of various eye motions (saccades, fixation, smooth pursuit). These are other limitations present some key d for future image metrics:

- Metrics that account for temporal artifacts including those caused by eye motion
- Metrics that do not require perfect references
- Metrics that consider stereo vision and accommodation

(b) Micro-scale visual-motor behaviors. Throughout Table 1, we've summarized the relationship between various display properties, and their effects on various aspects of visual perception. However, the study of human vision is not only limited to perception and visual acuity but also accounts for motor behavior in response to visual stimuli. Although the vision science community has produced a plethora of research studying the relationship between visual stimuli and motor behavior ([18], [19]), there is a gap in this area of research from the computer graphics side. With the emergence of various gaze-contingent computer graphics pipelines in AR/VR systems ([12], [20]), there is an increasing need to study motor behaviors such as gaze motion accuracy and timing, and how the properties of various display technologies affect these behaviors.

(c) Multimodal perception. We perceive the world by integrating multiple sensors beyond vision, e.g., audio [21] and haptics [22]. The future hyper-realistic immersive environment hopes to accurately replicate sensory signals such that the user perceives themselves as being in another physical or virtual world. That said, considering the cross-effect in a multi-modal fashion is a critical future research direction that optimizes display systems.

(d) Eye-in-the-loop optimization. Hardware-in-the-loop computation has been demonstrated to be an effective means in the era of deep learning. However, it still remains an open challenge to compute how the retina and visual systems perceive the light rays, despite the growing knowledge of statistical neuroscience discoveries on ventral stream [23]. Developing a differentiable model that predicts the optical retinal image would inversely optimize the display properties with regard to the human-perceived quality.

4. Conclusion

The ongoing wave of displays continues to improve the variety, quality and immersion of our digital visual experiences. In this review, we identify a notable gap between development of novel displays and our understanding of human visual perception. We discuss how this gap is an effective opportunity to further optimize the efficiency and quality of displayed pixels, and the various behaviors which are key parts of our perceptual interaction with digital displays. We also identify specific problems and directions of future research which would help accelerate the advances in future displays. Proposed investigations attempt to optimizing user-oriented experience in immersive displays. We call for attention and participation in new research in this growing area.

5. Acknowledgements

Author Q. Sun is partially supported by the DARPA PTG program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

6. References

1. C. Papadopoulos, K. Petkov, A. E. Kaufman, and K. Mueller, “The reality deck–an immersive gigapixel display,” IEEE computer graphics and applications, vol. 35, no. 1, 33–45, 2014.
2. Q. Sun, F.-C. Huang, J. Kim, L.-Y. Wei, D. Luebke, and A. Kaufman, “Perceptually-guided foveation for light field displays,” ACM Trans. Graph., vol. 36, Nov. 2017.
3. P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs, “From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality,” IEEE transactions on visualization and computer graphics, vol. 22, no. 4, 1367–1376, 2016.
4. H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, “High dynamic range display systems,” in ACM SIGGRAPH 2004 Papers, 760–768, 2004.
5. T. Rhodes, G. Miller, Q. Sun, D. Ito, and L.-Y. Wei, “A transparent display with per-pixel color and opacity control,”

in ACM SIGGRAPH 2019 Emerging Technologies, SIGGRAPH ’19, (New York, NY, USA), Association for Computing Machinery, 2019.

6. P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs, “Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays,” IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 11, 4194–4203, 2021.
7. Q. Sun, F.-C. Huang, L.-Y. Wei, D. Luebke, A. Kaufman, and J. Kim, “Eccentricity effects on blur and depth perception,” Opt. Express, vol. 28, 6734–6739, Mar 2020.
8. K.-E. Lin, Z. Xu, B. Mildenhall, P. P. Srinivasan, Y. Hold-Geoffroy, S. DiVerdi, Q. Sun, K. Sunkavalli, and R. Ramamoorthi, “Deep multi-depth panoramas for view synthesis,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16, 328–344, Springer, 2020.
9. Z.-C. Dong, W. Wu, Z. Xu, Q. Sun, G. Yuan, L. Liu, and X.-M. Fu, “Tailored reality: Perception-aware scene restructuring for adaptive vr navigation,” ACM Transactions on Graphics (TOG), vol. 40, no. 5, 1–15, 2021.
10. S. Malpica, B. Masia, L. Herman, G. Wetzstein, D. Eagleman, D. Gutierrez, Z. Bylinskii, and Q. Sun, “Has half the time passed? investigating time perception at long time scales,” Journal of Vision, vol. 20, no. 11, 489–489, 2020.
11. P. Hu, Q. Sun, P. Didyk, L.-Y. Wei, and A. E. Kaufman, “Reducing simulator sickness with perceptual camera control,” ACM Trans. Graph., vol. 38, Nov. 2019.
12. 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 Trans. Graph., vol. 35, nov 2016.
13. O. T. Tursun, E. Arabadzhiyska-Koleva, M. Wernikowski, R. Mantiuk, H.-P. Seidel, K. Myszkowski, and P. Didyk, “Luminance-contrast-aware foveated rendering,” ACM Transactions on Graphics (Proc. SIGGRAPH), vol. 38, no. 4, 2019.
14. R. Albert, A. Patney, D. Luebke, and J. Kim, “Latency requirements for foveated rendering in virtual reality,” ACM Transactions on Applied Perception (TAP), vol. 14, no. 4, 1–13, 2017.
15. D. M. Eagleman, “Human time perception and its illusions,” Current opinion in neurobiology, vol. 18, no. 2, 131–136, 2008.
16. R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, “Subtle gaze direction,” ACM Transactions on Graphics (TOG), vol. 28, no. 4, 1–14, 2009.
17. Q. Sun, A. Patney, L.-Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman, “Towards virtual reality infinite walking: dynamic saccadic redirection,” ACM Transactions on Graphics (TOG), vol. 37, no. 4, 1–13, 2018.
18. M. Lisi, J. A. Solomon, and M. J. Morgan, “Gain control of saccadic eye movements is probabilistic,” Proceedings of the National Academy of Sciences, vol. 116, no. 32, 16137–16142, 2019.
19. J. Laubrock, A. Cajar, and R. Engbert, “Control of fixation duration during scene viewing by interaction of foveal and peripheral processing,” Journal of Vision, vol. 13, 11–11, 10 2013.
20. D. R. Walton, R. K. D. Anjos, S. Friston, D. Swapp, K. Ak¸sit, A. Steed, and T. Ritschel, “Beyond blur: Real-time ventral metamers for foveated rendering,” ACM Trans. Graph., vol. 40, jul 2021.
21. A. J. Ecker and L. M. Heller, “Audio-visual cue combination in depth perception,” Journal of Vision, vol. 4, no. 8, 699–699, 2004.
22. M. O. Ernst and M. S. Banks, “Humans integrate visual and haptic information in a statistically optimal fashion,” Nature, vol. 415, no. 6870, 429–433, 2002.
23. J. Freeman and E. P. Simoncelli, “Metamers of the ventral stream,” Nature neuroscience, vol. 14, no. 9, 1195–1201, 2011.
24. J . Kim, Y. Jeong, M. Stengel, K. Ak¸sit, R. Albert, B. Boudaoud, T. Greer, J. Kim, W. Lopes, Z. Majercik, et al., “Foveated ar: dynamically-foveated augmented reality display,” ACM Transactions on Graphics (TOG), vol. 38, no. 4, 1–15, 2019.
25. Y.-C. Chiang, S.-S. Cheng, H.-S. Chen, L.-J. Wei, L.-M. Huang, and D. K. Chu, “Retinal resolution display technology brings impact to vr industry,” in ACM SIGGRAPH 2018 Posters, 1–2, 2018. 7
26. R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, “Fovvideovdp: A visible difference predictor for wide field-of-view video,” ACM Transactions on Graphics (TOG), vol. 40, no. 4, 1–19, 2021.
27. Krajancich, Brooke, Petr Kellnhofer, and Gordon Wetzstein. "A Perceptual Model for Eccentricity-dependent Spatio-temporal Flicker Fusion and its Applications to Foveated Graphics." ACM Transactions on Graphics (TOG), 2021
28. X. Meng, R Du, M. Zwicker, & A. Varshney (2018). Kernel foveated rendering. Proceedings of the ACM on Computer Graphics and Interactive Techniques, 1(1), 1-20.
29. B. Guenter, M. Finch, S. Drucker, D. Tan, & J. Snyder (2012). Foveated 3D graphics. ACM Transactions on Graphics (TOG), 31(6), 1-10.
30. A. Jindal, K. Wolski, K. Myszkowski, & R. K. Mantiuk, (2021). Perceptual model for adaptive local shading and refresh rate. ACM Transactions on Graphics (TOG), 40(6), 1-18.
31. A. Maimone, A. Georgiou, & J. S. Kollin (2017). Holographic near-eye displays for virtual and augmented reality. ACM Transactions on Graphics (Tog), 36(4), 1-16.
32. D. Lanman, & D. Luebke (2013). Near-eye light field displays. ACM Transactions on Graphics (TOG), 32(6), 1-10.
33. V. Bianco, P. Memmolo, M. Leo, S. Montresor, C. Distante, M. Paturzo, P. Picart, B. Javidi and P. Ferraro, 2018. Strategies for reducing speckle noise in digital holography. Light: Science & Applications, 7(1), pp.1-16.
34. G. Denes, A. Jindal, A. Mikhailiuk and R.K. Mantiuk, 2020. A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution. ACM Transactions on Graphics (TOG), 39(4), pp.133-1.
35. L. Kugler (2021). The state of virtual reality hardware. Communications of the ACM, 64(2), 15-16.