Editor: Jim Foley, foley@cc.gatech.edu

DEPARTMENT: DISSERTATION IMPACT

Leveraging Human Visual Perception for an Optimized Virtual Reality Experience

Qi Sun, *New York University, New York, NY, 10013, USA*

FROM THE EDITOR

Qi Sun received the 2019 VGTC Virtual Reality Best Dissertation Award.

We explore perceptually motivated research through virtual reality (VR) with the human as an experimental medium. We track, study, and model our spatio-temporal gaze behaviors to understand cognitive responses. Furthermore, we leverage this knowledge to advance computer graphics techniques in VR with more efficient rendering, increased display comfort, and safer locomotion.

Immersive technologies such as virtual reality (VR) and augmented reality (AR) have the advantage of full field-of-view, stereo depth cues, and rich interaction through positionally tracked hands or controllers. However, head-mounted displays (HMDs) fully replace the visual input provided to our brain, making paramount the latency, accuracy, and fidelity of such systems. Considering these unique challenges, implementing immersive platforms requires innovative graphics technologies, including efficient real-time rendering, natural interaction, and visually comfortable output. The research presented in this article focuses on understanding and leveraging human perceptual factors to further improve VR systems. Principally, we aim to reduce the onset and severity of simulator sickness and to describe an innovative way to locomote in a large virtual environment despite limited physical space.

As a unique optical system, human vision has a rich and complex set of biophysical behaviors, such as rapidly refocusing at multiple depths and cooperating with the vestibular (balance) system. In real-time rendering pipelines, human eyes are often approximated with the pinhole camera model (using a fixed focal plane). In VR, sensory conflicts introduced by this overly simple optical model have been shown to be a major contributor to simulator sickness.

Research in light-field display technology is a significant step beyond the pinhole camera approximation (see Figure 1). Light fields afford the ability to view multiple focal depths across their surface, naturally imitating the way our eyes expect to view the physical world. A major challenge in using light-field displays in VR is the extraordinary computational power required to drive the system and the resulting rendering latency. Our work contributes to the state of the art by investigating the intersection of light fields and the human visual system. We show how the use of gaze-aware algorithms can provide substantial speedups in light-field rendering.

Another long-standing challenge in VR is the intrinsic disconnect between virtual space and physical space. While clever techniques allow us to teleport or otherwise artificially locomote in VR, most experiences do not afford the most basic and natural movement: walking. The challenge here is that virtual scenes are typically larger than the physical space available to the user. Consequently, users wearing HMDs are restricted by walls and other obstacles such as people, tables, and cats. To encourage safe movement during natural locomotion, a body of research proposes redirecting the users' physical

0272-1716 © 2021 IEEE
Digital Object Identifier 10.1109/MCG.2021.3113392
Date of current version 10 December 2021.

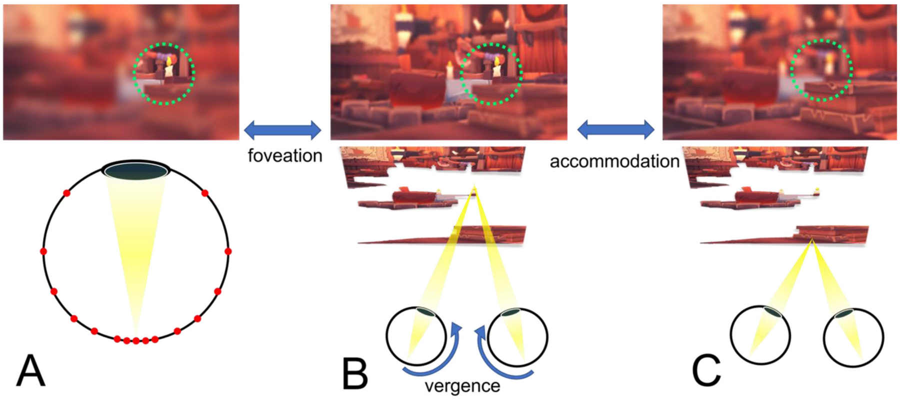

FIGURE 1. (A) With the centralized retinal receptor (red dots) distribution (bottom), we only perceive foveated (high-resolution image only in the center of vision) image in the retinal (top). (B) While trying to observe an object at a given depth (top), the eyeballs first rotate to converge at the target (bottom). (C) While refocusing without changing gaze (top), the corresponding accommodation/defocus blur gives us cues of depth as well.

walking paths through subtle—ideally unnoticeable—modifications of the virtual scene. An example of redirected walking is physically moving along a curved trajectory while believing you are walking in a straight line.[8] Unfortunately, many redirected walking techniques suffer from distracting visual artifacts. Similar to our work on light-field rendering, we employ gaze-aware algorithms to model perception in an effort to mitigate the visual artifacts necessary for comfortable redirected walking.

In summary, the human visual system provides a wide entryway to advance the state of the art when combined with gaze-aware modeling and computation. This research encourages the notion that there are a remarkable number of opportunities in our field when it shares ideas across fields, such as neuroscience, optics, and mechanical engineering.

COMPUTER GRAPHICS MEETS VR FROM A PERCEPTUAL PERSPECTIVE

A major concern of computer graphics is the accurate modeling and simulation of visual phenomena for presentation on a display. The resulting output takes a complex path through the eye on its way toward the brain's visual cortex. This section briefly introduces eye models and explains why and how gaze behaviors connect to computational systems.

Spatial Retina Patterns

Light rays emitted from screens are sensed on the retina after modulation by the pupil. The individual color receptors (cones) in the eye are concentrated on the retina around the fovea at the rear center of the eye. The density of receptors dramatically drops beyond the $\sim 5^\circ$ foveal cone. A VR technique, known as foveated rendering [see Figure 1(A) and (B)],[2] makes use of this anatomical feature as a way of reducing image fidelity in areas *outside* the fovea. It is accomplished through the use of gaze-tracking and modern GPU architectural support.

Temporal Gaze Behaviors

Our eyes naturally move in all three dimensions as we observe the world. In viewing objects at various depths, our eyes simultaneously rotate in all axes to converge on the target. Such eye movement is called *vergence* [see Figure 1(B)]. Stereo displays such as VR headsets can naturally support vergence by showing different perspective-correct views to each eye. Meanwhile, each eye may also adjust the focus of its lens. This optically results in a blur at depths outside of the focal plane. This effect is called *accommodation* [see Figure 1 (C)]. Both vergence and accommodation support perceptual cues of depth. However, ordinary nonlight-field display panels only cast rays from one depth. In this mechanical arrangement of display–lens–eye, a natural conflict arises between vergence and the inability to accommodate beyond the fixed focal plane of the display. This vergence–accommodation conflict has been experimentally shown to be a major contributor to simulator sickness.

Beyond spatio-temporal depth cues, the human visual system also exhibits a quirk that results in temporary blindness under rapid eye movement. Due to foveated vision, the brain can only receive the full visual information within the fovea. Thus, the pupil moves rapidly while entering a new environment to scan through the whole scene (see Figure 2). These rapid movements are called

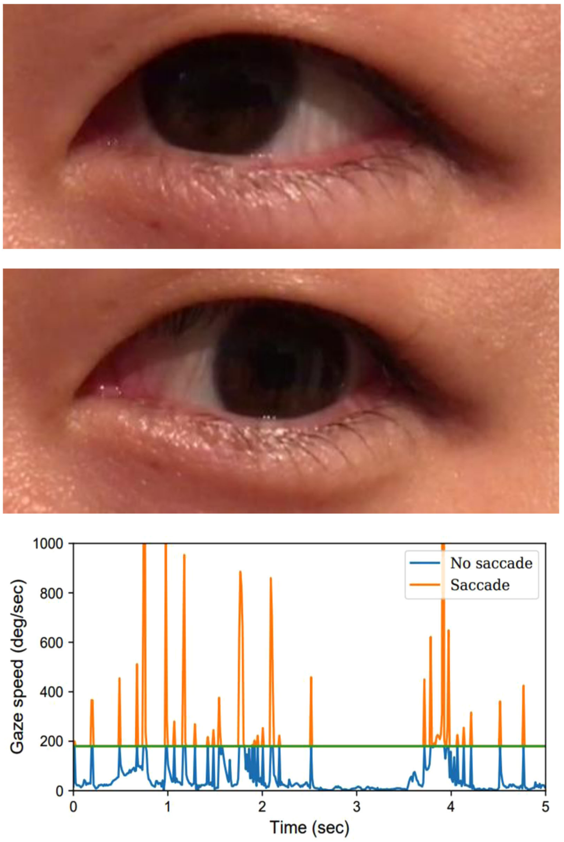

FIGURE 2. Illustration of saccades. The first two rows show the rapid movement before/after a saccade (~200 ms). The last row shows the frequent occurrence of saccade (orange) during VR game playing. It can be visualized that saccade occurs consistently and frequently but with a short duration.

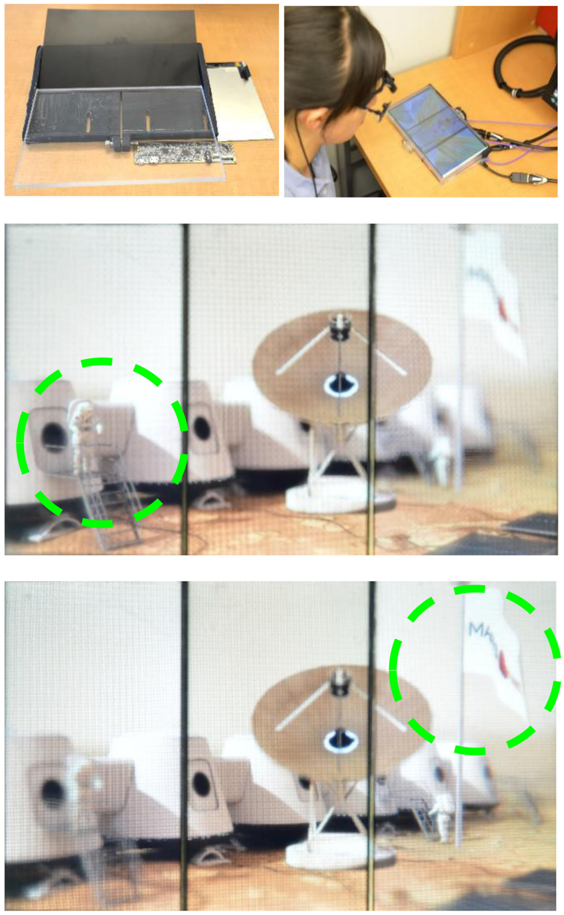

FIGURE 3. Light-field display hardware and camera-captured defocus blur. By assembling off-the-shelf components (first row), the display can support users' gaze from focusing on varied depths. (Second/third row: DSLR camera captured results with its lens focusing on near/far depths.)

saccades. Because of the high-speed motion (up to 900–1000°/s), the brain temporarily suppresses signals from the eye. This saccadic blindness phenomenon has been previously explored in the graphics community as a way of altering scene content below a perceptual threshold.[9]

ACCELERATING DISPLAY SYSTEMS WITH EYE-TRACKING

Solving the vergence–accommodation conflict in VR is an active area of research in the display community.[6] As previously mentioned, light-field displays unlock the power of accommodation, enabling focal adjustments that are not possible in 2-D display systems.[10] Despite year-over-year advancements in panel density and GPU rendering architectures, light fields remain out of reach for real-time consumer VR platforms, especially given the trend toward wire-free VR systems powered by mobile-phone chipsets, such as the Oculus Quest.

Inspired by prior work on 2-D gaze-contingent image rendering,[1,2] we implemented the first documented foveated light-field rendering and display system that supports low-latency without sacrificing visual quality or accommodation (see Figure 3). A key to this research was a rigorous study examining the effects of depth detection and discrimination thresholds within 15° from the foveal center. Similar to the previous research in color perception, we demonstrated significantly reduced depth sensitivity in the far periphery of the visual field.

Leveraging these psychophysical findings, we further analyze the entire spatio-temporal light-field system across content, display, and eye.

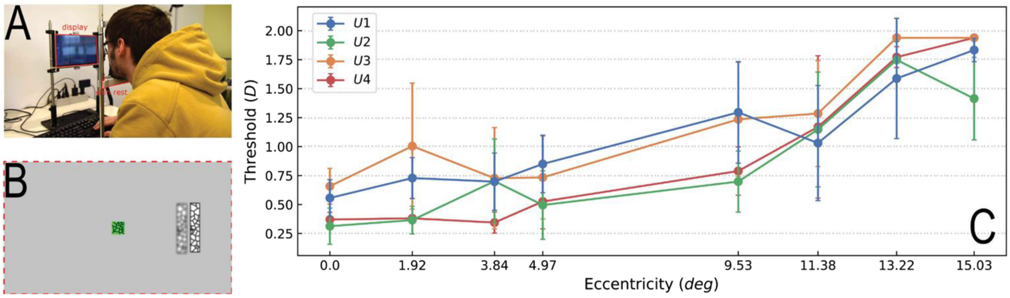

FIGURE 4. Psychophysical experiment to measure the depth perception across the visual field. (A) Participants fixed their heads, (B) observed 3-D stimuli on a light-field display, and determined whether two objects appear at different depths. (C) Minimal discriminable depth disparities (*y*-axis) along the visual field (*x*-axis). A significantly decreased depth perception (i.e., higher depth discrimination thresholds) can be observed at the farther visual field.

Analytical Connections Among Content, Display, and the Eye

Foveated vision demonstrates a significant reduction of spatial (color sensitivity) and angular (depth sensitivity) retinal sampling rate in the far periphery. Based on our pilot psychophysical experiment and data (see Figure 4), we modeled the effect in the frequency domain.[4] In ray space, the perceived retinal image is an integration of the retinal light-field across the pupil.

After converting it to the frequency domain with a Fourier transform, the corresponding spectrum is then obtained through the Fourier slice theorem. We discovered that the display capability, the eye lens, and the retinal foveated sampling could all be regarded as frequency bounds. They limit how much of the display-emitted rays actually activate retinal receptive cells.

Perception-Aware Real-Time Sampling

Following our blur and depth perception studies and the display–eye–retina frequency analysis, we derived a spatially-varying ray sampling strategy aimed at answering the scientific question, "what is the minimal number of rays needed to support defocusing cues identical to the natural world?"

Intuitively, the sampler presumes a constant focusing distance and two objects. We consider all the visible objects within a ray and compute the corresponding importance indicator for sampling. However, a core challenge in importance-based rendering is the tedious atomic add operations in the GPU to obtain the overall weight (denominator), limiting its use to offline cases only. To address this problem, we rederived the model through closed-form integration, which enabled GPU sampling at real-time rates. Our final study shows that with an eye-tracker and our gaze-content light-field sampling, subjects perceive a ~45 FPS (16%–30% computation of a full-resolution rendering) foveated light-field display as perceptually identical to a full-resolution display that runs at about 15 FPS, as in Figure 5.

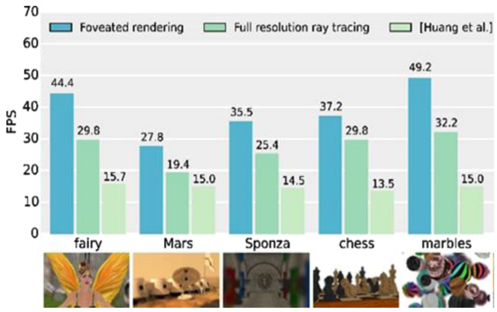

FIGURE 5. Performance benefits of foveated light fields. A full-resolution rendering typically results in very high latency (~15 FPS). Our foveated remarkably improve it to up to 45 FPS.

COLLISION-FREE VR LOCOMOTION

Existing systems such as projected rooms (CAVEs) and HMDs provide fast and realistic rendering, but often require users to remain stationary due to navigation concerns (real environments not visible through HMD). Thus, users have to employ less natural means such as devices to control their movements, which may negatively impact a sense of presence compared with real walking. However, virtual and real worlds often differ significantly in terms of size and layout. Ideally, a mapping between the two is required to offer a believable presence in the virtual world and feasible navigation in the real world (i.e., users are present and engaged in VR without bumping into walls, obstacles, or other users in reality).

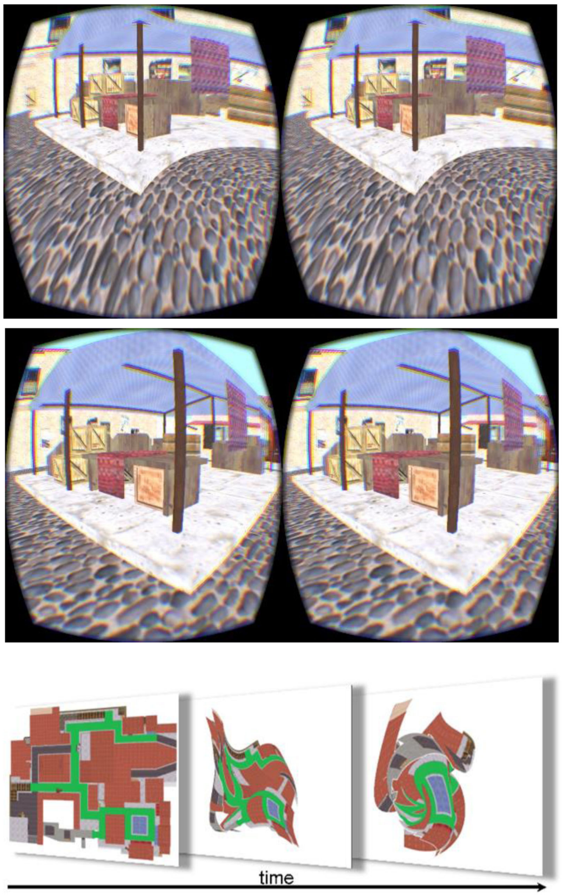

FIGURE 6. Distorting geometry appearance in VR to redirect user's physical walking path. The first/second row shows the HMD view on the same street before/after warping. The third row shows the temporal progress on the warping of the floor plan. It can be noticed that the warped scene (thus walking path) takes smaller areas by reusing the space.

Existing redirected walking techniques suffer from missing a general mapping between any given pair of virtual and physical environments. In building a new general-purpose method, we first propose a geometric technique.[3] Specifically, we represent both worlds as planar floor plans and aim at finding their least distorted mapping. This mapping defines a warp in the virtual world that is then rendered inside the HMD. With this system, users are guided to stay within the boundary walls and away from obstacles such as furniture in the real world. Figure 6 shows an overview of this system and its corresponding user experience. This solution consists of two key components: 1) a planar mapping between virtual and real-world floor plans, and 2) a real-time rendering system to visualize the mapped scene. With a radial basis function, the mapping aims to preserve both shapes and distances between the virtual and real worlds for visual consistency and obstacle avoidance. Our method seeks a many-to-one folding of the virtual world to reuse the limited real space. Next, our system also relies on the mapping to properly guide user locomotion and disguise the perceptual discrepancies that result as a part of the correction. Overall, the correction/warping and rendering system balances visual realism, geometric consistency, and perceptual comfort. We validated this system through subjective reports with practical VR applications across visualization scenarios in architecture and medical imaging and traditional games.

The geometric perspective of altering users' walking has inspired other innovations, such as the work by Dong *et al.*[12] However, one major limitation in the series of approaches is the heavy precomputation to obtain a proper mapping: it may take 15–20 minutes to converge to an optimal solution. In practicality, physical surroundings may change dynamically and rapidly. For instance, other users may walk in the room, or a cat believes it's playtime. The precomputed mapping may cause users to bump into each other. On the other hand, the introduced distortion may compromise the original functionality from the designers' intentions.

Cooperating With Eye Movements

To address the limitations above, we use saccadic suppression to dynamically redirect users without visible geometric distortion. The main idea is that—during saccades—we can subtly alter the virtual cameras' direction to nudge the users to walk in the desired paths without them noticing the manipulation.

Different users may show significantly varied sensitivity to the camera redirection. To guarantee imperceptibility, in the work by Sun *et al.*,[5] we conducted a pilot study to find the most conservative threshold. We discovered that participants could not detect camera rotation less than 12.6°/s (0.14°/frame at 90 FPS) when their gaze velocity was above 180°/s. We increase the redirection scale linearly with regard to the saccade duration.

The camera manipulation described above relies on detecting saccades in real time. However, a saccade, as a perceptual phenomenon, has unpredictable and short occurrences. This fleeting uncertainty requires a fast response. We thus developed a fast, GPU-based path planning system. It senses physical surroundings, including stationary (such as furniture) and dynamic (such as other users) obstacles. The goal of collision avoidance is modeled as a loss function whose minima predicts the optimal redirection. This optimization can be completed within about 5 ms.

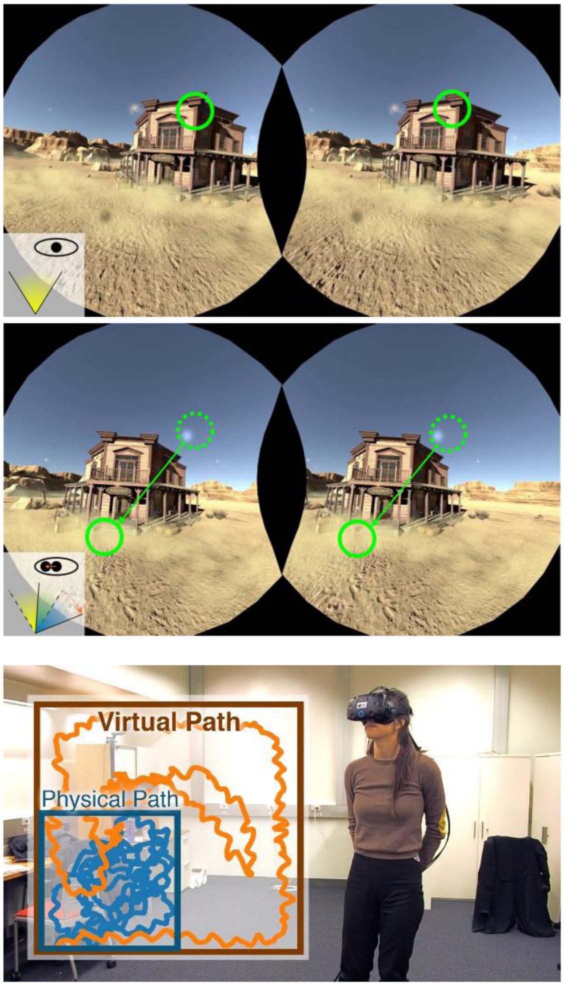

FIGURE 7. Saccadic redirection. The first/second row shows the HMD view before/after saccadic redirection. The third row shows the user's virtual and physical walking path from the free exploration.

Benefiting from the real-time performance, the redirection can be optimally achieved whenever the eye-tracker detects saccades. The fast optimization is computed with 1) a parallelized GPU line searching algorithm, and 2) the relatively limited searching range due to the highly limited redirection range (12.6°/s). The system can be seen in Figure 7. Path planning can also be combined with other perceptual suppression phenomena, such as blinks.[11]

Beyond publishing the system in academic venues, we also developed the technology as a live conference exhibition. For the GTC 2018 conference (https://blogs.nvidia.com/blog/tag/gtc-2018/), we created a chessboard game that allows participants wearing an eye-tracked HMD to freely walk around to place chess pieces on a virtual board (see Figure 8). The system was demonstrated to conference attendees with varied backgrounds and familiarity with immersive media. We observed a surprising consistency during the demo that almost none of our users noticed any redirection in the gameplay.

NOW AND THE FUTURE

For a long period, a major focus of computer graphics techniques has been on machine-oriented computational techniques, such as building higher quality models/textures/materials, crafting photorealistic ren-dering algorithms, or accurately simulating natural phenomena. When it meets the unique challenges of VR, we attempt to introduce the "computational human" perspective. We have demonstrated that mathematically modeling how humans perceive virtual visual stimuli can bring new perspectives and broaden applications to the community, such as safer, faster, more natural, and friendly interactions. We foresee substantially new research opportunities and challenges in both the near and far future.

Future Immersion

VR and AR have shown their groundbreaking roles and impacts on assisting real-world applications, such as sports, medical treatment, assistive driving, and remote education. Their main advantages in these scenarios are the full immersion, thus better engagement.

The research on perception-aware graphics attempts to initiate our prediction of future immersion: a passive, comfortable,[7] and automated system that does not require users' explicit (yet commonly unnatural) intervention. Instead, we use human nature by sensing behavioral actions such as eye-tracking. With passive sensing, the systems may reduce the cognitive and physical loads of VR/AR users. The corresponding applications, such as redirected walking, demonstrated a plausible user experience.

We believe the always-refreshed definition of "immersion" would introduce new concepts and innovative research directions in the long term. For instance, more light-weighted and faster holographic displays that realistically reproduce the visual world. With the recent advancement of GPU designs, high-quality, high-resolution, and real-time graphics have moved with significant steps. Meanwhile, for full immersion, the capability of becoming a 24$\times$7 assistant is the core factor in replacing monitors. That said, the graphics components should also consider energy consumption in future immersive media. This need can be observed from the significantly shortened battery life while running heavy computations in current mobile applications, such as *Pokémon Go*.

FIGURE 8. Live demo of saccadic redirection. We implemented the idea as a VR chess board game. The figures show the demo booth, live capture of user experience, and the final path (red/green for virtual/physical), respectively. Note that the physical path (inside the demo both) takes a significantly smaller area than the virtual one.

Future Sensory Modeling

One central theme in this article is eye-tracking and the means of further leveraging it to address practical issues in VR. However, humans are complicated systems with various sensors. Several recent projects have introduced multisensory research, such as visual–audio synthesis and haptics rendering of complex geometries.

Future graphics and VR will also enable perceivers (users?) to fully engage in a multimodal manner. We have also been witnessing its impact on machine learning and assistive technologies. By introducing sensing technologies to the graphics pipeline, we envision fresh opportunities to have a broad impact. Such efforts would involve additional nonvisual information, such as visual-haptic blending and advanced biomedical devices (e.g., electrocorticography) to bridge the gap.

Future Interdisciplinary Graphics

Today, computer graphics is far beyond entertainment (visual effects, games, etc.): It has been serving as the visualization platform for biology (microscope), neuroscience, physics/photonics, etc. In this article, we build computational models of the fundamental scientific discoveries to use in practical applications. Human factors may have a leading role in the future progress of traditional graphics problems. That said, helping to accelerate scientific discoveries could be another critical mission of human-centered computer graphics.

REFERENCES

1. B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, “Foveated 3D graphics,” *ACM Trans. Graphics*, vol. 31, no. 6, pp. 1–10, 2012.
2. A. Patney *et al.*, “Towards foveated rendering for gaze-tracked virtual reality,” *ACM Trans. Graphics*, vol. 35, no. 6, pp. 1–12, 2016.
3. Q. Sun, L.-Y. Wei, and A. Kaufman, “Mapping virtual and physical reality,” *ACM Trans. Graphics*, vol. 35, no. 4, pp. 1–12, 2016.
4. Q. Sun, F.-C. Huang, J. Kim, L.-Y. Wei, D. Luebke, and A. Kaufman, “Perceptually-guided foveation for light field displays,” *ACM Trans. Graphics*, vol. 36, no. 6, pp. 1–13, 2017.
5. Q. Sun *et al.*, “Towards virtual reality infinite walking,” *ACM Trans. Graphics*, vol. 37, no. 4, pp. 1–13, 2018.
6. G. Wetzstein, A. Patney, and Q. Sun, “State of the art in perceptual VR displays,” in *Real VR – Immersive Digital Reality (Lecture Notes in Computer Science)*. Cham, Switzerland: Springer-Verlag, 2020, pp. 221–243.
7. P. Hu, Q. Sun, P. Didyk, L. Y. Wei, and A. E. Kaufman, “Reducing simulator sickness with perceptual camera control,” *ACM Trans. Graphics*, vol. 38, no. 6, pp. 1–12, 2019.
8. R. Sharif, Z. Kohn, and M. C. Whitton, “Redirected walking,” in *Proc. Eurographics*, Manchester, U.K., 2001, pp. 5–7.
9. B. Bolte and M. Lappe, “Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression,” *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 4, pp. 545–552, Apr. 2015.
10. F. Huang, K. Chen, and G. Wetzstein, “The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues,” *ACM Trans. Graphics*, vol. 34, no. 4, Aug. 2015, Art. no. 60.
11. E. Langbehn, F. Steinicke, M. Lappe, G. F. Welch, and G. Bruder, “In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality,” *ACM Trans. Graphics*, vol. 37, no. 4, pp. 1–11, 2018.
12. Z. C. Dong, X. M. Fu, Z. Yang, and L. Liu, “Redirected smooth mappings for multiuser real walking in virtual reality,” *ACM Trans. Graphics*, vol. 38, no. 5, pp. 1–17, 2019.

QI SUN is with New York University, New York, NY, USA. Contact him at qisun0@gmail.com.

Contact department editor Jim Foley at foley@cc.gatech.edu.