Single-Shot VR

Yingsi Qin
yingsiq@andrew.cmu.edu
Carnegie Mellon University
USA

Wei-Yu Chen
weiyuc@andrew.cmu.edu
Carnegie Mellon University
USA

Matthew O'Toole
motoole2@andrew.cmu.edu
Carnegie Mellon University
USA

Aswin C. Sankaranarayanan
saswin@andrew.cmu.edu
Carnegie Mellon University
USA

Figure 1: *Left*: Our display prototype. *Middle* and *Right*: Captured images from our prototype using a Nikon Z5 with a *near (mid)* and *far (right)* camera focus, at $F/5.6$, 100 ISO, and 1.3 sec exposure. (Scene courtesy of "Entity Designer" at Blender Market)

ABSTRACT

The physical world has contents at varying depths, allowing our eye to squish or relax to focus at different distances; this is commonly referred to as the accommodation cue for human eyes. To allow a realistic 3D viewing experience, it is crucial to support the accommodation cue—the 3D display needs to show contents at different depths. However, supporting the native focusing of the eye has been an immense challenge to 3D displays. Commercial near-eye VR displays, which use binocular disparity as the primary cue for inducing depth perception, fail this challenge since all contents they show arise from a fixed depth—ignoring the focusing of the eye. Many research prototypes of VR displays do account for the accommodation cue; however, supporting accommodation cues invariably comes with performance loss among other typically assessed criteria for 3D displays. To tackle these challenges, we present a novel kind of near-eye 3D display that can create 3D scenes supporting realistic accommodation cues in a single shot, i.e., without using time multiplexing or eye tracking. This display, which we present in our demo, can stream 3D content over a large depth range, at 4K spatial resolution, and in real-time. Our display offers an exciting step forward towards a truly immersive real-time 3D experience. Participants will get to enjoy 3D movies and play interactive games in their demo experience.

ACM Reference Format:
Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C. Sankaranarayanan. 2023. Single-Shot VR. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Emerging Technologies (SIGGRAPH '23 Emerging Technologies), August 06–10, 2023, Los Angeles, CA, USA*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3588037.3595396

SIGGRAPH '23 Emerging Technologies, August 06–10, 2023, Los Angeles, CA, USA

ACM ISBN 979-8-4007-0154-2/23/08.
https://doi.org/10.1145/3588037.3595396

1 OVERVIEW

When using a commercial VR display, we gain depth perception from the content disparity between the images shown to each eye. Our eyes, however, cannot accommodate their focus on the depth that the vergence of the eyes suggests, since all virtual content arrives from the one depth where the 2D display is placed. Commercial VR displays today suffer from conflicting vergence-accommodation cues and impose visual fatigue and discomfort on the user.

In the research community, there have been varying types of near-eye 3D displays that support accommodation cues. However, each type suffers from its own subset of shortcomings that hinders their practical deployments. Stereoscopic displays by default do not support accommodation cues. Varifocal displays [Padmanaban et al. 2017] require eye tracking to show adaptively rendered content at the eye's focused depth, making them fragile to unpredictable eye movements. Time-multiplexed multifocal displays [Zhan et al. 2020] show content at all physical depths within the persistence of human vision, demanding high bandwidth electronics that challenge real-time streaming. Light field displays [Lanman and Luebke 2013] bear the inherent tradeoff between angular and spatial resolution. Holographic displays [Maimone et al. 2017] suffer from small etendue (product of field of view (FOV) and eyebox) limited by today's commercially available spatial light modulators (SLMs).

To tackle the above challenges, we introduce a new kind of near-eye 3D display that supports accommodation cues in a single shot, i.e., without requiring time multiplexing or eye tracking. Our display features a large depth range from 0.25 m to infinity, at high spatial and depth resolutions, and a large étendue. Our lightweight computational footprint also enables streaming interactive 3D content in real-time. See Figure 2 for the attribute comparison.

2 IMPLEMENTATION

Our display is achieved using a novel programmable lens that provides high spatial selectivity of its focal length. The design modifies a Lohmann lens [Lohmann 1970]—a focus-tunable lens whose focal length is determined by the relative translation between the

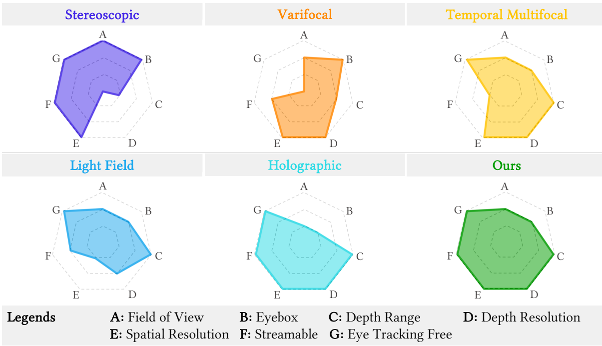

Figure 2: Comparison of common types of VR displays. Our display (in green) has a clear overall advantage in the most commonly assessed criteria of a VR display.

two cubic phase plates that it is composed of. To remove the need for *mechanical translation*, we optically collocate two cubic phase plates at the two ends of an optical 4f system and place a prism in the Fourier/pupil plane. Controlling the amount of prism tilt allows us to optical-equivalently translate a cubic plate relative to the other, thereby controlling the focal length of the equivalent lens. To enable *spatially-varying focal length*, we replace the prism with an SLM and collocate an OLED display onto it. Displaying spatially varying phase ramps on the SLM is then equivalent to changing the focal length for each pixel, thereby achieving content optically focused at varying physical depths. Please visit our project website [Qin et al. 2023] for detailed information.

Prototype. We show our monocular prototype in Figure 1. We use a 1.03"-diagonal micro-OLED that has a 2560×2560 RGB resolution with a pixel pitch of 7 µm. We use a 0.7"-diagonal phase-only SLM that has a 2464×4000 resolution with a pixel pitch of 3.747 µm. The smaller area of the SLM makes it the limiting factor for displaying content and hence with a 40 mm eyepiece, we get a 25° FOV and a 5 mm eyebox. The OLED displays RGB colors simultaneously.

Overall, our prototype enjoys a high spatial resolution, largely retaining the OLED's native resolution, at a depth resolution that is greater than 1/8–th of a diopter over a working depth range from 0.25 m to infinity. Our prototype features a temporal resolution equal to the native frame rate of the display and SLM. It also has a large étendue limited only by the choice of lenses used.

3 DEMONSTRATION

Participants will enjoy a gallery of 3D scenes and play interactive 3D games such as Minecraft and GTA on our prototype. We show example captures of video and interactive scenes in Figure 3. Participants will experience different depth range, depth resolution, and the option to view 3D content without accommodation cues—to better understand where current VR experiences stand compared with the immersive 3D experience our innovative display provides.

ACKNOWLEDGMENTS

We thank Daohan Lu for acting as the player during the interactive game demo capture. This work was supported by the National Science Foundation under the awards 1730147 and 2008464.

Figure 3: Captured images of video and interactive VR results. ***Left:*** We show three *video* snapshots at time $t_0 < t_1 < t_2$ with the camera manually focused at the green inset regions. In this video, the viewer moved forward to the center of the scene, and the camera focuses on the foreground. For example, the bananas go from out-of-focus at t_0 to in-focus at t_1. ***Right:*** We show the *interactive scene* snapshots of a player navigating Minecraft. The camera was focused on the close white plant at t_0 and focused on the distant yellow flower at t_1. The player then continued walking towards the in-focus distant region at t_2. (Left: scene courtesy of "Entity Designer" and animation courtesy of Mohamad Zeina; right: game courtesy of Michael Fogleman)

REFERENCES

Douglas Lanman and David Luebke. 2013. Near-eye Light Field Displays. *ACM Transactions on Graphics* 32, 6, Article 220 (Nov. 2013), 10 pages. https://doi.org/10.1145/2508363.2508366
Adolf W Lohmann. 1970. A new class of varifocal lenses. *Applied Optics* 9, 7 (1970), 1669–1671.
Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. *ACM Transactions on Graphics* 36, 4 (July 2017), 85:1–85:16.
Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. 2017. Optimizing Virtual Reality for All Users Through Gaze-contingent and Adaptive Focus Displays. *Proceedings of the National Academy of Sciences* 114 (2017), 9.
Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C. Sankaranarayanan. 2023. Project Page: Split-Lohmann Multifocal Displays. https://imaging.cs.cmu.edu/split_lohmann.
Tao Zhan, Jianghao Xiong, Junyu Zou, and Shin-Tson Wu. 2020. Multifocal displays: review and prospect. *PhotoniX* 1, 1 (2020), 1–31.