Reverse Pass-Through VR

Nathan Matsuda
Facebook Reality Labs Research

Brian Wheelwright
Facebook Reality Labs Research

Joel Hegland
Facebook Reality Labs Research

Douglas Lanman
Facebook Reality Labs Research

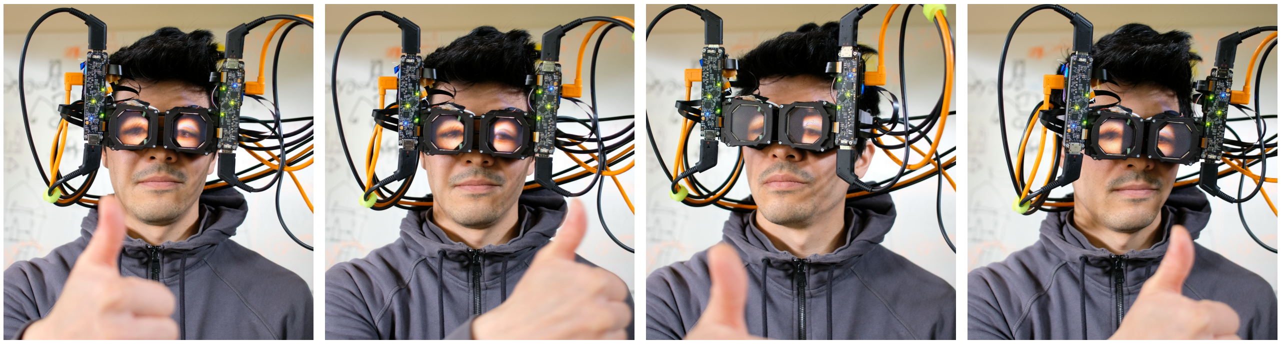


Figure 1: Our reverse pass-through VR prototype depicts a live, three-dimensional reconstruction of the wearer's eyes to any number of external viewers. Here, wearer gaze (centered on the raised thumb) is interpretable for different gaze directions and head rotations. View our supplementary video to see the prototype in operation.

ABSTRACT

We introduce *reverse pass-through VR*, wherein a three-dimensional view of the wearer's eyes is presented to multiple outside viewers in a perspective-correct manner, with a prototype headset containing a world-facing light field display. This approach, in conjunction with existing video (forward) pass-through technology, enables more seamless interactions between people with and without headsets in social or professional contexts. Reverse pass-through VR ties together research in social telepresence and copresence, autostereoscopic displays, and facial capture to enable natural eye contact and other important non-verbal cues in a wider range of interaction scenarios.

CCS CONCEPTS

• **Computing methodologies** → **Virtual reality**.

KEYWORDS

reverse pass-through, light field displays, virtual reality

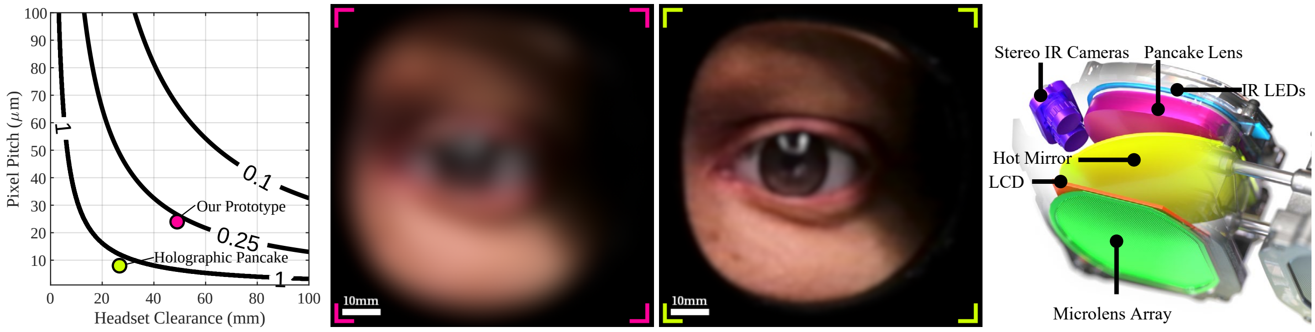
ACM Reference Format:
Nathan Matsuda, Brian Wheelwright, Joel Hegland, and Douglas Lanman. 2021. Reverse Pass-Through VR. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Emerging Technologies (SIGGRAPH '21 Emerging Technologies), August 9–13, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3450550.3465338

1 INTRODUCTION

In recent years, virtual reality (VR) headsets have become a consumer technology, with the hope that their uniquely immersive display and interaction systems will lead to more compelling entertainment, productivity, and telepresence applications. Yet, as emphasized by Gugenheimer et al. [2019], little attention has been paid to resolving a core deficiency: VR displays isolate the user from their environment and, in doing so, limit VR use and acceptance in shared and public spaces [Mai and Khamis 2018; Schwind et al. 2018]. Eliminating this isolation is a key motivation for the development of *video pass-through VR*, wherein the VR headset user sees a reproduction of their external environment and the individuals within it. Yet, a crucial gap remains: External viewers cannot hold a natural conversation with a VR headset user, whose upper face and eyes remain occluded.

Several efforts have been made to depict the occluded features of a VR headset user's face on external, world-facing displays. Chan and Minamizawa [2017] depict an eye-tracked illustration of the user's eyes to give a sense of the user's gaze direction and attention. Their approach stops short of depicting the surrounding facial regions and eliminates all perspective depth cues. To partially address these limitations, Mai et al. [2017] depict a hand-tuned face model that is aligned to the perspective of a single external viewer and supports a manually controlled gaze direction. Yet, across these and related works, we identify remaining capabilities that are necessary to deliver authentic social copresence using external, world-facing displays, including faithful reproduction of the occluded periocular region of the face, accurate depiction of three-dimensional depth, and support for multiple external viewers. We introduce *reverse pass-through VR* to meet these needs.

SIGGRAPH '21 Emerging Technologies , August 9–13, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8364-6/21/08.
https://doi.org/10.1145/3450550.3465338



(a) Light Field Display Resolution **(b) Our Prototype (Modeled)** **(c) Holographic Lens (Modeled)** **(d) Prototype Layout**

Figure 2: The apparent resolution of a reverse pass-through light field display depends on the underlying 2D display resolution, microlens array optical properties, and the distance from the light field display to the user's eyes. (a) Here, the display pixel pitch and headset thickness (with 15mm eye relief) are varied as the MLA parameters remain fixed. Contours denote the effective light field resolution (in cycles/mm) at the depth coinciding with the user's face and eyes. The modeled apparent resolution for our current prototype is shown as the pink dot, while the green dot shows the apparent resolution for a future hypothetical headset using a recent high-density microdisplay [Kopin 2021] and the holographic pancake lens architecture proposed by Maimone and Wang [2020]. (b) A simulated eye image at the resolution of our light field display. (c) A simulated eye image possible at the resolution of the holographic pancake lens design. (d) A cutaway view showing our prototype architecture with external light field display, compact pancake optics, and folded eye imaging path using IR LED illumination, an IR hot-mirror, and stereo cameras.

We advocate for user-facing cameras, real-time reconstruction of facial geometry, and autostereoscopic world-facing displays to deliver an accurate recreation of the user's hidden face and eyes for an arbitrary number of external viewers. While the underlying technologies have been under development for decades, we are aware of no effort to combine them in this manner to unlock social VR. Furthermore, our proposed system is timely, leveraging recent research and industry trends in VR facial capture [Lombardi et al. 2018], high-resolution autostereoscopic displays [Martínez-Corral and Javidi 2018], and more compact VR headsets [Maimone and Wang 2020] (which further improve resolution with world-facing autostereoscopic displays). If successfully developed, reverse pass-through VR devices may function more like augmented reality (AR), which already allows direct eye contact via *optical see-through* displays. In this manner, users in shared social spaces may benefit from the wider fields of view and better occlusion cues currently delivered with VR displays.

2 REVERSE PASS-THROUGH DISPLAYS

Prior work in social telepresence establishes the need for autostereoscopic displays to accurately reproduce a user's eye gaze from the vantage point of every observer. Reprojection to a 2D display, previously demonstrated by Mai et al. [2017], is insufficient due to incorrect binocular and motion parallax depth cues. These prior findings lead us to set a design requirement that reverse pass-through VR systems must use an autostereoscopic display. Many autostereoscopic displays have been proposed, but we advocate that light field displays based on microlens array (MLA) technology are most compatible with the current research and industry trends toward thinner headset form factors.

Recent advances in polarization-based optical folding or "pancake" viewing optics establish a path toward significantly reduced headset volumes [Geng et al. 2018; Wong et al. 2017]. Recently, Maimone and Wang [2020] have shown that holographic optics may further decrease headset thickness, approaching sunglasses-like form factors.

We observe that light field displays have not been widely adopted, in part, due to their inherent spatio-angular resolution trade-off, which effectively limits the "depth of field" (i.e., the range over which a high-resolution image can be depicted). However, given current state-of-the-art display panel resolutions (whose densities are being driven by the VR industry itself), emerging VR viewing optics, and the limited depth of field needed to support an image plane at a headset-user's eye relief, light field displays are well suited for constructing compelling reverse pass-through VR systems today.

In Figure 2, we assess how thinner headsets and higher-density displays will impact the resolution of reverse pass-through VR. Here we have fixed the microlens parameters to the values outlined in Section 3. This analysis predicts a spatial resolution approaching 0.25 cycles/mm. Figure 2 further includes visualizations of the predicted resolution, aligning with our experimental results. We also predict the resolution for a headset using much thinner holographic pancake optics [Maimone and Wang 2020] and a high-density display with around 8μm pixel pitch that should be available in the near future [Kopin 2021]. This combination could support a spatial resolution around 1 cycle/mm. If these significant gains can be achieved in the near-term, light field reverse pass-through VR could become a practical, visually compelling solution that can be realized with technology that is already being developed within the AR/VR research community.

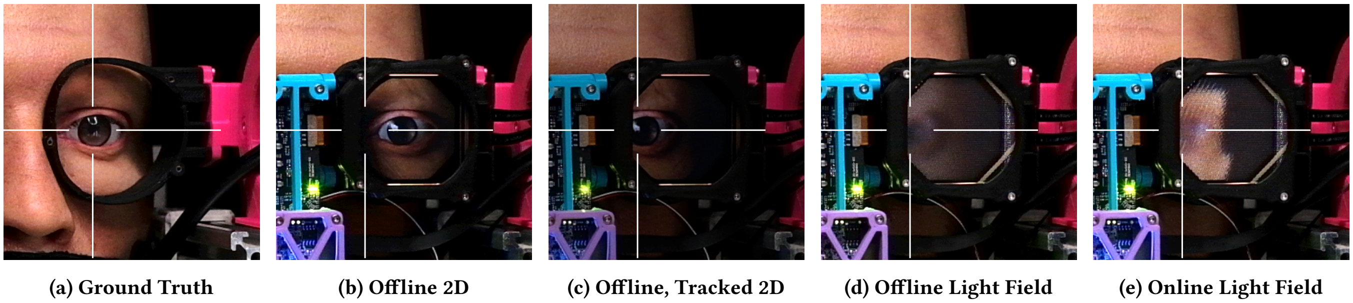

(a) Ground Truth (b) Offline 2D (c) Offline, Tracked 2D (d) Offline Light Field (e) Online Light Field

Figure 3: (a) Ground truth view of a mannequin head with display assembly removed. (b) Baseline approach showing an offline photogrammetry reconstruction of the face on an external 2D display. (c) Offline photogrammetry reconstruction of face reprojected to tracked viewer position on 2D display, supporting one viewer. (d) Proposed light field architecture displaying offline photogrammetry reconstruction, supporting multiple viewers. (e) Proposed architecture displaying live reconstruction, supporting multiple viewers. The white crosshairs are aligned to the ground truth pupil position.

3 IMPLEMENTATION

Our aim in designing a physical prototype was to use components that are readily available. We designed hardware subsystems (the light field display and stereo camera systems) and software subsystems (eye image synthesis, camera and display calibration, and light field rendering) following this principle. We hope that this motivates others to pursue research on reverse pass-through topics in the near term.

We designed a reverse pass-through module around an overall track length typical for pancake lenses. Each module, shown in a cutaway view in Figure 2d, contains a light field display subsystem and an eye capture subsystem. The headset contains two such modules, one for each eye, rigidly mounted to each other via carbon fiber rods over the nose bridge and affixed to the head with the strap from an Oculus Go headset.

3.1 Light Field Display

We established in Section 2 that MLA-based light field displays are uniquely well suited to this application. We back up this argument by implementing a high performance MLA design using an off-the-shelf LCD and a commercially mature resin-molding manufacturing technique. We then show that this display can be driven at real-time rates with conventional game rendering techniques.

To maximize field of view for a given focal length, the MLA lenslets need to be as fast as possible (high numerical aperture) while maintaining imaging performance. We designed a hex-packed, dual-sided MLA that distributes the optical power over two surfaces, reducing aberrations compared to single-sided MLAs. The design was fabricated by Holographix LLC using a resin casting process on a 200μm glass substrate (Corning EagleXG).

The MLA has the following specifications:

- 42° field of view
- 520μm focal length
- 500μm pitch
- F/0.90 measured corner-to-corner (F/1.04 edge-to-edge)

The microlens array is backed by a BOE 1600x1600 color LCD with 24μm pixel pitch (8μm RGB stripe). We drive this display with a Synaptics VXR7200-based display bridge. We also placed a 2° engineered diffuser from Luminit at the front of the display stack to low-pass filter the output to prevent single red, green, or blue pixels from being sharply imaged to the viewer.

3.2 Eye Image Capture and Synthesis

We designed an eye capture system that is compatible with existing eye tracking architectures. We used a pair of near-IR Omnivision OV9281 CMOS sensors driven by an Omnivision OV580 USB stereo capture board. A Chroma Technology T700 IR-reflective hot mirror provides a relatively on-axis view of the eye (17.5deg off-axis). See Figure 2 for a cutaway view of the camera geometry. The stereo pair were calibrated with a conventional OpenCV pipeline, using a small printed circle grid pattern.

The requirement for high-quality, low-latency 3D reconstruction and colorization from infrared images for a limited domain (eyes) leads naturally to deep learning techniques. We surveyed leading candidates for stereo depth inference and colorization and selected AnyNet [Wang et al. 2019] and CycleGan [Zhu et al. 2017], respectively. We produced a training dataset with a custom textured and rigged face model derived from the Digital Emily project [Alexander et al. 2009] implemented in Blender [Blender Foundation 2021]. The resulting dataset contains 10,000 stereo image pairs for color and IR textures, over varying head positions, eye gaze directions, and eyelid poses. This dataset was used to refine the AnyNet model, pre-trained on the SceneFlow Driving Dataset [Mayer et al. 2016], over 300 epochs using default parameters. To train the colorization network, we captured 300 real IR/color pairs by affixing a color camera (Arducam IMX298) to the headset eye cup such that the entrance pupil was co-located with the left IR camera entrance pupil. The CycleGan model was trained on these images for 200 epochs.

Improved facial reconstruction for reverse pass-through remains a challenge for future research. Models optimized for telepresence, such as the one described by Lombardi et al. [2018], could generate facial reconstructions for both the local reverse pass-through view and remote VR views of the headset-wearer.

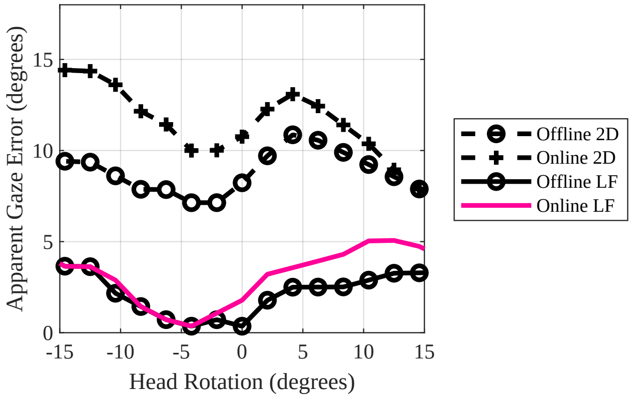


Figure 4: Estimated pupil reprojection error (in degrees) plotted as a function of horizontal head rotation angle for offline (photogrammetry) and online (live stereo) modes. 2D tracked reprojection consistently shows angular errors that are much higher than the light field display.

4 RESULTS

Images of the prototype headset running the live reconstruction are shown in Figure 1. A subject gazes toward an outstretched hand in several poses to show the perspective-correct reproduction of eye images.

In Section 1, we reviewed past research that utilizes tracked 2D reprojection of the eyes instead of an autostereoscopic display. These approaches do not support multiple external viewers, but also have limited fidelity in the single-user case due to the lack of stereoscopy; the reprojected view can be correct for one eye, or the average position of the eyes, but not both eyes simultaneously.

We implemented the 2D tracking approach for comparison with our prototype hardware by swapping in the same LCD, but without the MLA. We mounted one full display pod to a two-axis gimbal comprised of two Thorlabs HDR50/M rotation stages and 3D printed support assemblies. We also mounted a silicone mannequin head fabricated by Legacy Effects to use as a static, and thus repeatable, reference face. Using the known gimbal angle and camera location, we render the correct perspective for a view 32mm to the right of the camera, as if the camera were the left eye on a person with 64mm interpupillary distance. We later compare to the light field output for that same angle. To control for reprojection errors due to the online, live stereo facial reconstruction pipeline, we also produced an offline, precomputed photogrammetric model of the mannequin.

Figure 3 depicts these display modes, along with a ground truth image without the display pod mounted, all from a viewing angle of 12.5° azimuthal rotation from the axis of the display. The offline 2D display shows significant displacement of the pupil due to the parallax induced by the LCD's approximately 50mm offset from the plane of the pupil. The tracked 2D output reprojection falls within the 32mm range mentioned earlier, but only works for a single viewer assuming perfect face tracking. The offline and online light field displays show correct perspective for any number of viewers.

We estimate eye rotation angles by tracking the displayed pupil position, then projecting this position back to the eyeball surface, which is fixed relative to the mannequin head. These estimates, for different head rotation angles, are shown in Figure 4. The light field display produces more accurate view perspectives than the tracked 2D output. Not only do the 2D pupil rotation errors exceed the light field output everywhere, those approaches only work for a single person and are thus not viable for social or professional settings.

The experimental results show that light field displays are suitable for reverse pass-through VR. This display choice is aligned with research and industry trends toward thinner VR headsets, thereby enhancing the effective resolution of the facial reconstruction. As research into believable digital avatars for telepresence continues to make progress, these techniques can be used to generate images for light field reverse pass-through displays, eventually approaching a high fidelity experience for external observers that is indistinguishable from looking at a pair of glasses.

ACKNOWLEDGMENTS

The authors thank Tracy Campbell, Bruce Cleary, Salah Nouri, Sam Pepperwood, and Daniel Taylor for their contributions.

REFERENCES

Oleg Alexander, Mike Rogers, William Lambeth, Matt Chiang, and Paul Debevec. 2009. The Digital Emily Project: Photoreal Facial Modeling and Animation. In *ACM SIGGRAPH Courses*. 1–15.

Blender Foundation. 2021. *Blender - a 3D modelling and rendering package*. Blender Foundation. http://www.blender.org

Liwei Chan and Kouta Minamizawa. 2017. FrontFace: Facilitating Communication between HMD Users and Outsiders Using Front-Facing-Screen HMDs. In *ACM Human-Computer Interaction with Mobile Devices and Services*. Article 22.

Ying Geng, Jacques Gollier, Brian Wheelwright, Fenglin Peng, Yusufu Sulai, Brant Lewis, Ning Chan, Wai Sze Tiffany Lam, Alexander Fix, Douglas Lanman, et al. 2018. Viewing Optics for Immersive Near-Eye Displays: Pupil Swim / Size and Weight / Stray Light. In *Digital Optics for Immersive Displays*, Vol. 10676. International Society for Optics and Photonics.

Jan Gugenheimer, Christian Mai, Mark McGill, Julie Williamson, Frank Steinicke, and Ken Perlin. 2019. Challenges Using Head-Mounted Displays in Shared and Social Spaces. In *ACM Conference on Human Factors in Computing Systems Extended Abstracts*.

Kopin. 2021. Kopin's 2.6k x 2.6k OLED Display Incorporated in Panasonic's New VR Glasses. (Jan 2021). https://kopin.irpass.com/profiles/investor/NewsPrint.asp?v=6&b=2379&ID=96490&m=rl&g=1207

Stephen Lombardi, Jason Saragih, Tomas Simon, and Yaser Sheikh. 2018. Deep Appearance Models for Face Rendering. *ACM Transactions on Graphics* 37, 4, Article 68 (July 2018).

Christian Mai and Mohamed Khamis. 2018. Public HMDs: Modeling and Understanding User Behavior around Public Head-Mounted Displays. In *ACM Pervasive Displays*. Article 21.

Christian Mai, Lukas Rambold, and Mohamed Khamis. 2017. TransparentHMD: Revealing the HMD User's Face to Bystanders. In *ACM Mobile and Ubiquitous Multimedia*.

Andrew Maimone and Junren Wang. 2020. Holographic Optics for Thin and Lightweight Virtual Reality. *ACM Transactions on Graphics* 39, 4 (2020).

Manuel Martínez-Corral and Bahram Javidi. 2018. Fundamentals of 3D Imaging and Displays: A Tutorial on Integral Imaging, Light-Field, and Plenoptic Systems. *OSA Advances in Optics and Photonics* 10, 3 (Sep 2018), 512–566.

Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. 2016. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Valentin Schwind, Jens Reinhardt, Rufat Rzayev, Niels Henze, and Katrin Wolf. 2018. Virtual Reality on the Go? A Study on Social Acceptance of VR Glasses. In *ACM Human-Computer Interaction with Mobile Devices and Services*. 111–118.

Yan Wang, Zihang Lai, Gao Huang, Brian H Wang, Laurens Van Der Maaten, Mark Campbell, and Kilian Q Weinberger. 2019. Anytime Stereo Image Depth Estimation on Mobile Devices. In *IEEE International Conference on Robotics and Automation*.

Timothy L Wong, Zhisheng Yun, Gregg Ambur, and Jo Etter. 2017. Folded Optics with Birefringent Reflective Polarizers. In *Digital Optical Technologies*, Vol. 10335. International Society for Optics and Photonics.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision*.