Foveated AR: Dynamically-Foveated Augmented Reality Display

JONGHYUN KIM, YOUNGMO JEONG*, MICHAEL STENGEL, KAAN AKŞIT, RACHEL ALBERT, BEN BOUDAOUD, TREY GREER, JOOHWAN KIM, WARD LOPES, ZANDER MAJERCIK, PETER SHIRLEY, JOSEF SPJUT, MORGAN MCGUIRE, and DAVID LUEBKE, NVIDIA, United States

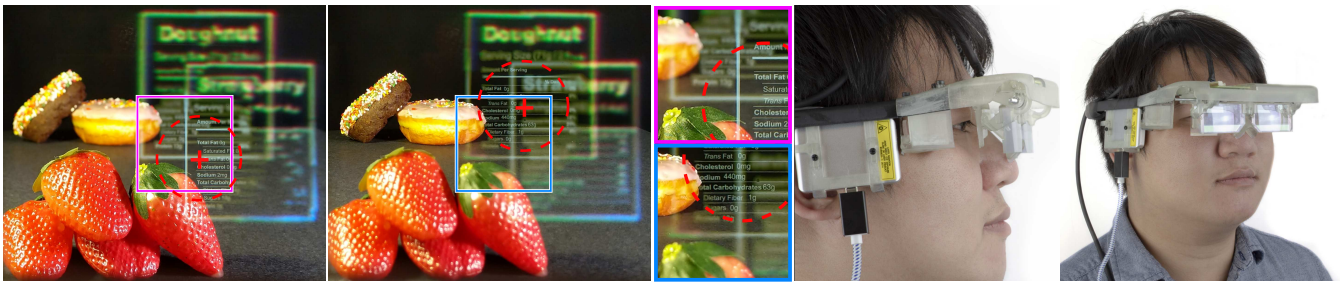


Fig. 1. Display results from our Foveated AR prototype. By tracking the user's gaze direction (red cross), the system dynamically provides high-resolution inset images to the foveal region and low-resolution large-FOV images to the periphery. The system supports accommodation cues; the magenta and blue zoom-in panels show optical defocus of real objects together with foveated display of correctly defocus-blurred synthetic objects. Red dashed discs highlight the foveal vs peripheral display regions. A monocular wearable prototype (functional but manually actuated) illustrates the compact optical path.

We present a near-eye augmented reality display with resolution and focal depth dynamically driven by gaze tracking. The display combines a traveling microdisplay relayed off a concave half-mirror magnifier for the high-resolution foveal region, with a wide field-of-view peripheral display using a projector-based Maxwellian-view display whose nodal point is translated to follow the viewer's pupil during eye movements using a traveling holographic optical element. The same optics relay an image of the eye to an infrared camera used for gaze tracking, which in turn drives the foveal display location and peripheral nodal point. Our display supports accommodation cues by varying the focal depth of the microdisplay in the foveal region, and by rendering simulated defocus on the "always in focus" scanning laser projector used for peripheral display. The resulting family of displays significantly improves on the field-of-view, resolution, and form-factor tradeoff present in previous augmented reality designs. We show prototypes supporting 30, 40 and 60 cpd foveal resolution at a net $85^\circ \times 78^\circ$ field of view per eye.

CCS Concepts: • **Hardware → Communication hardware, interfaces and storage**; **Displays and imagers**.

*Also with Seoul National University.

Authors' address: Jonghyun Kim, jonghyunk@nvidia.com; Youngmo Jeong, youngmo.snu@gmail.com; Michael Stengel, mstengel@nvidia.com; Kaan Akşit, kaksit@nvidia.com; Rachel Albert, ralbert@nvidia.com; Ben Boudaoud, bboudaoud@nvidia.com; Trey Greer, tgreer@nvidia.com; Joohwan Kim, sckim@nvidia.com; Ward Lopes, wlopes@nvidia.com; Zander Majercik, amajercik@nvidia.com; Peter Shirley, pshirley@nvidia.com; Josef Spjut, jspjut@nvidia.com; Morgan McGuire, mcguire@nvidia.com; David Luebke, dluebke@nvidia.com, NVIDIA, 2788 San Tomas Expressway, Santa Clara, CA, United States, 95051.

0730-0301/2019/7-ART99 $15.00
https://doi.org/10.1145/3306346.3322987

Additional Key Words and Phrases: foveated,varifocal,augmented,display

ACM Reference Format:
Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akşit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, Peter Shirley, Josef Spjut, Morgan McGuire, and David Luebke. 2019. Foveated AR: Dynamically-Foveated Augmented Reality Display. *ACM Trans. Graph.* 38, 4, Article 99 (July 2019), 15 pages. https://doi.org/10.1145/3306346.3322987

1 INTRODUCTION

Augmented reality (AR) aims to present virtual objects at real-world positions and orientations, without compromising the viewer's natural vision. These objects may be rendered photorealistically, or more practically, like the emissive ghosts depicted by science fiction "holographic" projections. Also called mixed reality, AR allows richer and more natural interaction than displays such as Google Glass, which simply superimpose 2D content in a "heads-up display" style akin to aviation and automotive windshield displays, or the video-passthrough smartphone AR applications.

The enabling technology for widespread consumer AR is a high-fidelity head-mounted display (HMD), sometimes termed a near-eye display (NED), that is comfortable both in visual properties and form factor. Such AR displays offer a tantalizing replacement for smartphone and computer screens, but in practice all designs involve tradeoffs between field of view (FOV), resolution, eye box (the "sweet spot" within which a viewer's pupil will perceive a correct image), correct focus cues, and form factor. The greatest challenge in HMD design is not in optimizing any individual metric, but instead *simultaneously* providing a wide FOV, variable focus, high resolution, a wide eye box, and a slim form factor.

Meeting this challenge requires significant new advances in the underlying technology. Inherent physical constraints preclude evolving traditional, larger displays to hit AR targets. The requirement that an AR display be see-through constrains the form factor and

materials involved. The desired product of resolutions, FOV, eye box, and eye relief for HMDs pushes the boundaries of diffraction for visible light wavelengths. Fundamental to this challenge is the *optical invariant*, which limits the spatial bandwidth of light and forces a trade-off between eye box and FOV. This leads to necessary co-optimization across the design, and requires leveraging diffraction in some places via holography rather than avoiding/combating it throughout the design. Eye box size is not just a comfort factor in positioning the display. The eye box must be at least the size of the pupil to minimize diffractive ringing (blur). The eye box must also be aligned with the pupil to avoid vignetting. A larger eye box tolerates more error in gaze tracking, but increasing the eye box size alone without also increasing the brightness of the display lowers the total effective brightness.

Classic geometric optical elements such as the lenses in Virtual Reality (VR) HMDs cannot meet the weight or size requirements of an AR display seeking the form factor of corrective eyeglasses/sunglasses. For size and weight, Holographic Optical Elements (HOEs) provide a promising direction. FOV is highly constrained by form factor in addition to the optical invariant, as well as by display pixel density and power constraints.

Using today's display interface standards, it is impractical to transmit video data at the bandwidth required for a uniform resolution, high update rate, wide FOV display that matches user visual acuity across its full FOV. Even with a belt-mounted compute pack or remote rendering system and custom display interface the power required for driving such a display would be problematic for both battery life and thermal management. Thus, we suggest that high resolution AR displays must take advantage of foveation.

To enable natural focus depth, either a light field or multifocal/varifocal display is required. Light fields massively increase rendering demands, resolution, and optical complexity; varifocal requires robust, low-latency gaze tracking and miniaturized moving elements; and multifocal is a blend of these challenges.

For context, the Microsoft Hololens [Kress and Cummings 2017] is the best-documented commercial AR display. It has a fixed focal depth, $30^\circ \times 17.5^\circ$ monocular FOV, and 21 cycles per degree (cpd) resolution in a 579g package, including the processor and battery. The Magic Leap One [2019a] and nReal light [2019b] developer kits have slightly better specifications and form factors. nReal has the best field of view at 52° (diagonal, monocular). These impressive devices approach the limits of engineering improvements on conventional display technology; however they still fall well short of the desired properties for AR. They primarily achieve their form factors by severely limiting field of view, resolution, and focus.

The research state of the art is the experimental prototype displays by Maimone et al. [2017]. These address each of the design concerns *in isolation*, primarily by innovating on holographic imaging and trading off different properties in each prototype. Maimone et al. observed that the grand challenge is now providing the desirable properties simultaneously, which motivates our work.

Our specific contribution is the optical and systems design, and analysis, of the first wearable AR display architecture to simultaneously provide:

- High resolution (30, 40, and 60 cpd) inset display for foveal region.
- An eyebox and field of view that exceed the optical invariant for a Maxwellian display by dynamic positioning of HOEs.
- A simple and fast rendering pipeline using on-axis gaze tracking, foveated varifocal rendering, and calibration across the geometric distortion, intensity, and color of differing optical paths.

In addition we present the following prototype systems:

- A bench-top, full-color, dynamically-foveated, and varifocal prototype with resolution matching human visual acuity over a wide, $85^\circ \times 78^\circ$ monocular FOV (100° diagonal).
- A monochrome, wearable form factor (with external power and processing) through compact optics for both foveal and peripheral regions with a $77^\circ \times 53^\circ$ monocular FOV (86.4° diagonal).

2 RELATED WORK

Our work proposes a new class of gaze-contingent accommodation-supporting see-through NEDs that exploits the non-uniform visual acuity of the human visual system (HVS). Therefore, we review relevant literature across foveated displays, accommodation-supporting see-through near-eye displays, near-eye gaze tracking, and multifocal rendering. We also provide a comparison among the state-of-the-art see-through NEDs as in Table 1.

2.1 Foveated Displays

Foveated Display Hardware. The earliest gaze-contingent graphics was presented by Reder [1973]. Baldwin et al, [1981] created a single variable resolution display that was the first work in the spirit of our NED, where a high resolution inset is presented to the fovea and a larger area at lower resolution is presented to the rest of the retina. Spooner et al. [1982] presented a combination of two displays. Shenker et al. [1987] was the first to combine two different displays in a NED, leading to a steerable foveal inset with a 20 cpd resolution using fiber optics and pancake optical relays. Rolland et al. [1998] also combined two displays using a beam-splitter in a NED, in which a high-resolution inset with 24 cpd resolution is relayed to a fovea using microlenses with a FOV of $13.30^\circ \times 10.05^\circ$, while a lower resolution display at 6 cpd spans an FOV of $50^\circ \times 39^\circ$ through a magnifier lens. Recently, VR display prototypes with fixed foveation (i.e. StarVR, Varjo) have been shown. Most recently, Tan et al. [2018] showed dynamically foveated VR display by combining two identical displays using a beam-splitter with the different magnifications. They steered the foveal inset with the liquid crystal director. Lee et al. [2019] also showed a time-multiplexed see-through fixed foveated holographic display using a beam splitter and a tunable lens, whose foveal FOV was 1.04° and peripheral FOV was 22.6°. We are not aware of any previous dynamically foveated AR displays, or rendering algorithms for such a display.

Foveated rendering. uses the knowledge of the gaze direction to choose the lowest cost 3D graphics sufficient for each part of the spatially varying visual field [2012]. Our foveated display does not require a special technique as the visual acuity provided by the foveal and peripheral display mechanism matches the HVS.

Our work distinguishes itself from other foveated displays by providing superior optical qualities (resolution, FOV, eyebox, brightness) and supporting accommodation with a form-factor approaching conventional prescription glasses.

Table 1. A comparison of see-through accommodation-supporting near-eye displays modeled after those in Dunn et al. [2017] and Akşit et al. [2017]. Note that we define a moderate FOV as 40-60 degrees, moderate resolution as 10-20 cpd, and a moderate eyebox as 5-10 mm. Moderate values are adapted from [Cakmakci and Rolland 2006]. A moderate transparency corresponds to 50 – 80% of the ambient light arriving at a viewer's eye. Our work distinguishes itself as a foveated varifocal display that provides a wide FOV, large eyebox, and high resolution while maintaining a thin form factor.

Display technique	Focus mechanism	Transparency	FOV	Resolution	Eyebox	Form factor	Computation	Gaze tracking
Pinlight displays [Maimone et al. 2014]	light fields	low	wide	low	small	thin	high	no
Freeform optics [Hua and Javidi 2014]	light fields	high	narrow	low	moderate	moderate	high	no
HOE [Jang et al. 2017]	light fields	high	moderate	low	large	moderate	high	yes
HOE [Maimone et al. 2017]	holographic	high	wide	moderate	small	very thin	high	yes
HOE [Jang et al. 2018]	holographic	high	moderate	moderate	large	very thin	high	yes
Multifocal plane display [Hu and Hua 2014]	multifocal	high	narrow	moderate	moderate	bulky	high	yes
Membrane [Dunn et al. 2017]	varifocal	moderate	wide	low	large	bulky	low	yes
Varifocal HOE [Akşit et al. 2017]	varifocal	moderate	wide	moderate	large	moderate	low	yes
Multifocal display [Lee et al. 2018a]	multifocal	moderate	narrow	low	large	thin	high	no
This work	varifocal	moderate	wide	high	large	thin	low	yes

2.2 Transparent Accommodation-Supporting Displays

Unlike consumer-grade NEDs, most recent see-through NEDs found in the literature provide focus mechanisms to generate virtual images at various depths. Displays supporting such focus mechanisms are also known as accommodation-supporting displays. Our design is one such accommodation-supporting display. Here, we classify and review the most recent accommodation-supporting see-through NEDs found in the literature [Hua 2017].

Light Field Displays. Maimone et al. [2014] created a monochrome NED prototype with a 110° FOV and 2 – 3 cpd resolution using a transparent sparse backlight.

Varifocal Displays. Liu et al.'s [2008] design uses a tunable lens system combined with a spherical mirror, and demonstrates 28° of diagonal FOV with 10 – 14 cpd resolution, which switches depth from one extreme to an another within 74 ms. The design of Dunn et al. [2017] provides a monocular > 60° FOV and a varifocal mechanism switching in 300 ms. Work of Akşit et al. [2017] proposes HOEs as a part of an AR varifocal NED system, offering a FOV of 60° with 18 cpd and varifocal mechanism switching at 410 ms.

Multi-focal plane displays. The work of Lee et al. [2018a] proposes a compact AR NED composed of a waveguide and a holographic lens which demonstrates a FOV of 38° × 19°. Most recently, Zhan et al. [2018] proposed the use of a stack of switchable geometric phase lenses to create a multi-focal, additive light field NED providing approximate focus cues over a 80° FOV. Most recently, Lee et al. [2018c] demonstrate a multi-layer structure on a large optical bench using a time-multiplexed multi-planar structure with a FOV of 30° FOV and a resolution of 8 cpd.

Holographic Displays. Static holograms encoded into HOEs have been used in various NED types as optical combiners [Jang et al. 2017; Kim and Park 2018; Lee et al. 2018a; Maimone et al. 2017] or projection surfaces [Akşit et al. 2017], although the static holograms in these displays do not provide 4D light fields. Dynamic holographic AR NEDs can be achieved using phase-only SLMs which encode holograms [Maimone et al. 2017; Shi et al. 2017], promising a wide FOV (~80 degrees) but with a limited eye box. Recently, the work of Jang et al. [2018] demonstrate a holographic display using a novel beam combiner HOE, a pupil-shifting HOE, and a phase modulating SLM with a thin form factor, achieving 45° × 40° FOV, and a resolution of 12 cpd. Most recently, work of Lee et al. [Lee et al. 2018b] showed that meta-lenses, optical components with sub-wavelength features, can also pave a way to a compact AR NED.

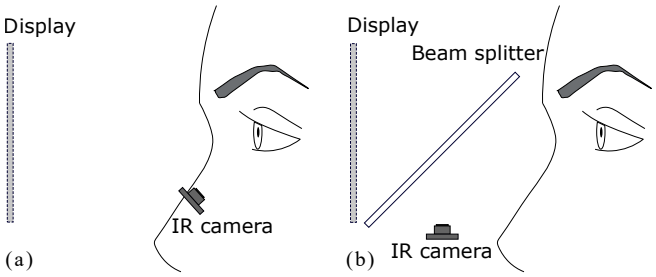


Fig. 2. **Near-eye display gaze tracking camera configurations.** Off- and on-axis placements of gaze tracking cameras inside near-eye displays. (a) Off-axis placement occupies less space, at the cost of non-uniform accuracy in gaze estimation. (b) Using a beam splitter provides an on-axis frontal view of the eye, which allows for more accurate gaze estimation.

Our proposal provides a varifocal mechanism similar to tunable focus displays for the foveal region and uses static holograms as an optical combiner for the periphery. Building on the literature of accommodation-supporting see-through near-eye displays reviewed above, we describe a foveated varifocal display that simultaneously provides a wide FOV (85° ×78°) and high resolution (60 cpd in fovea) in a thin form factor.

2.3 Near-Eye Gaze Tracking

A recent survey focusing on near-eye displays can be found in the work of Koulieris et al. [2019]. Due to our AR use case we focus on related research in the near-eye camera setting. Recently, methods using GPU-accelerated deep learning have enabled highly robust pupil localization and gaze estimation at high sampling rates [Fuhl et al. 2018, 2017; Kim et al. 2019; Lemley et al. 2018]. Video-based eye tracking systems can be categorized by different attributes such as near-eye vs remote tracking, on-axis vs off-axis, model-based vs regression-based tracking, and single camera vs multi-camera input.

Near-eye tracking can be divided into *on-axis* and *off-axis* configurations depending on the camera position viewing the eye as shown in Fig. 2. Our design enables on-axis eye tracking for uniform accuracy in gaze estimation.

We use a regression-based gaze estimation method to locate the pupil and then map pupil location to a screen location using a polynomial mapping function along with user-specific calibration [Fuhl et al. 2015; Santini et al. 2017; Tonsen et al. 2017]. Although calibration-free methods exist for multiple-camera settings [Miki et al. 2016], the highest accuracy for a single camera is usually provided with user-specific calibration [Mansouryar et al. 2016].

2.4 Multi-focal Rendering

To trigger the accommodation-related visual cues, AR systems attempt to display objects with levels of blur appropriate for their depth relative to the viewer. For displays with a fixed (or set of fixed) optical distances, this blur must be introduced computationally with some kind of depth of field (defocus-blur) algorithm. This blur must be accomplished quickly as it directly adds to the rendering time.

When a single virtual object that lies within a narrow depth range [Liu et al. 2010], or a planar object orthogonal to the view axis, defocus blur is simple because only a single blur needs to be computed for the entire display. For more complex situations various discrete decompositions have been proposed (e.g., [Mercier et al. 2017]) for approximating content as a series of planes that can be individually blurred by Gaussian post-processing. However, Narain et al. [2015] show that discrete decompositions create artifacts at depth boundaries that can affect perception of depth, reducing the value of the focus cue. Real-time post-processing defocus filters were first introduced by Shinya [1994] and have a modern form that avoids discrete decomposition with other heuristics [Abadie 2018; Bukowski et al. 2013; Sousa 2013; Yang et al. 2016]. These were designed for aesthetic appeal in entirely virtual scenes and have other quality and calibration limitations, notably that they require exact depth maps (which is hard for real-world scenes) and hallucinate detail near occlusion boundaries. The main alternative is correct in-camera simulation of focus by relatively slow techniques such as stochastic ray tracing [Cook et al. 1984] or accumulation buffers [Haeberli and Akeley 1990]. Some point filtering methods such as the recent one by Selgrad et al. [2015] combines accuracy with relative performance, but are still about two orders of magnitude slower than the fastest real-time techniques and thus inappropriate for AR. Xiao et al. [2018] introduced the high-quality DeepFocus deep neural net filter that is the state of the art for virtual simulated focus in AR. It is significantly faster and more accurate than previous methods. However, with linear resolution scaling from their performance measurements, it takes 150 ms/frame for a 3 Mpix display. We are optimistic about the future of their approach, given the trend of increasing hardware optimizations for machine learning.

However, for the near term, a 100× faster solution is required in the near term. The rendering budget for defocus is a small fraction of the ≈5 ms frame time per eye in AR. So, we introduce a simple and fast defocus algorithm in Sec. 4.2.2. It is based on the observation that *additive-only* rendering matches the capabilities of AR displays and greatly simplifies accurate depth of field simulation, and thus yields a large speedup by limiting the application case to those in which virtual objects compose but do not occlude one another.

As with some prior work, our approach also uses an array of render targets. However, it uses these as a novel focus binning strategy instead of a depth layer decomposition. A bin may accumulate contributions from disparate depths, and a single depth may contribute to multiple adjacent bins. This decomposition of the net defocus operator into multiple, uniform point-spread functions produces fast, order-independent, and nearly perfect compositing of emissive surfaces when the number of bins is proportional to the maximum circle of confusion.

3 A FOVEATED AR DISPLAY DESIGN

Our foveated AR display combines light from two elements: a high-resolution, small FOV *foveal display* and a large FOV, low-resolution *peripheral display*. We designed the foveal optical path with a planar image combiner (IC) and also embedded a reverse optical path for on-axis gaze tracking. In the periphery, an HOE refracts light rays from a laser projector to create a Maxwellian viewpoint. These two displays move as with the user's gaze.

3.1 Foveal display

We adopt a high resolution, small FOV, varifocal, see-through near-eye display for the foveal inset in our design. A micro OLED display of size $w_d \times h_d$ and resolution $N_{dx} \times N_{dy}$ is employed. As shown in Fig. 3, light rays from the display are reflected from a 45-degree planar half mirror onto a concave half mirror inside a transparent planar IC and delivered to the observer's eye. The concave half mirror (located within the IC) with radius of curvature r acts as a magnifier to produce an enlarged virtual image of size $w_f \times h_f$ at distance d_f by the Gaussian thin lens formula:

$$\frac{2}{r} = \frac{1}{n(a+t)+u} - \frac{1}{nd_I} \quad \textit{and} \quad d_f = d_I + e + t + u/n, \tag{1}$$

where t is half mirror thickness measured from side view, u is the thickness of the concave half mirror, a is the distance to the micro display from the IC in the side view, d_I is the distance to the image plane from the IC, and n is the refractive index of the IC. For this configuration, the micro display appears magnified by a factor M,

$$M = \frac{w_f}{w_d} = \frac{r}{r - 2(na + nt + u)}. \tag{2}$$

Instant field of view θ_f and the average angular resolution $\overline{c_f}$ are

$$\theta_f = 2\tan^{-1}\left(\frac{w_f}{2d_f}\right) \text{ cpd and } \overline{c_f} = \frac{N_{dx}}{2\theta_f} \text{ cpd}. \tag{3}$$

Eq. 3 can be used to assess the two-dimensional trade-off space, and the parameters can be chosen as plotted in Fig 4.

To always provide a high resolution inset to the fovea, the micro display travels along its horizontal axis in correspondence with the user's gaze direction angle. The display travels within the width of the IC w_m, and the maximum gaze angle α_{max} is

$$\alpha_{max} = \tan^{-1}\left(\frac{Mw_m}{2d_f + D_e}\right), \tag{4}$$

where D_e is the eye diameter.

In addition, the foveal display provides focal cues. The virtual image distance can be changed by moving the micro display back and forth in the relay path. Eq. 1 shows how the focal distance d_f is modified according to the micro display position a. The display can

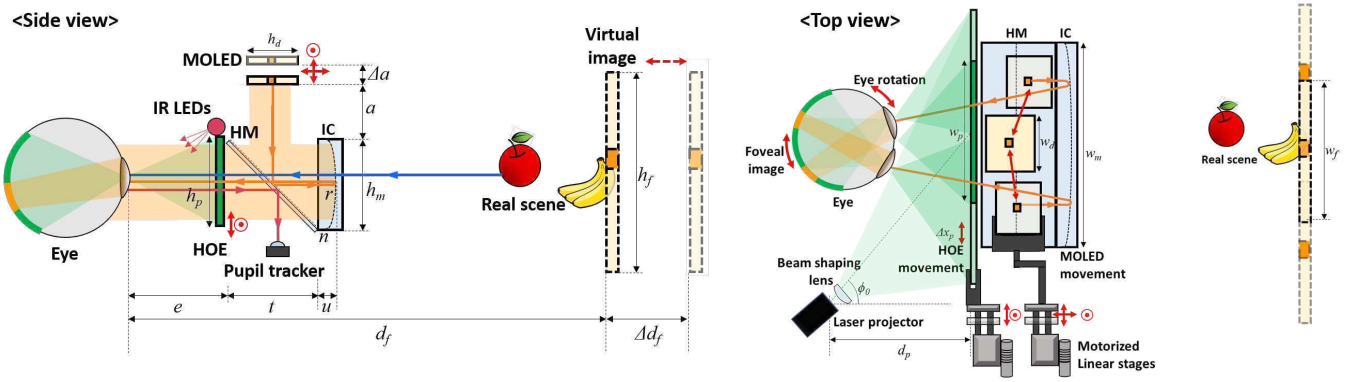


Fig. 3. Schematic diagram of Foveated AR. Note that the light paths for fovea (orange), gaze tracking (pink), and real scene (blue) are all integrated in a half mirror (HM) and a image combiner (IC). Light rays from a micro OLED (MOLED) are reflected by HM and IC and create a magnified virtual image located a distance d_f from the eye. The virtual image depth can be dynamically changed from 0D to 2.5D by moving MOLED vertically (Δa). The on-axis pupil image is obtained for gaze tracking through a separate light path (pink) by using the other side of half mirror. (right) the top view of the display. The laser projector positioned a distance d_p projects an image onto holographic optical elements (HOE) with an incident angle ϕ, and the diffracted light is converged to the pupil center. The linear actuator controlled by the gaze angle signal follows pupil swim and always provides visual acuity resolution over eccentricity.

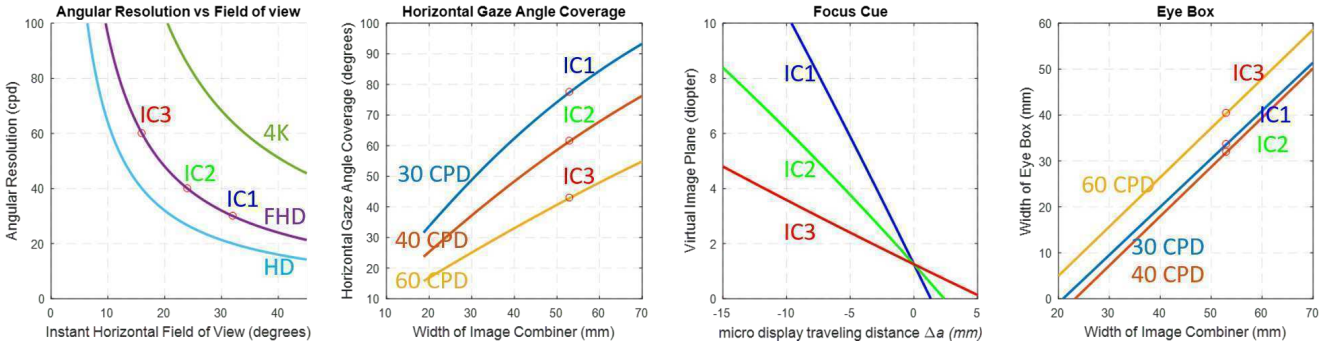


Fig. 4. Design trade-off space for the foveal display. The micro display width w_d = 18.7 mm and pixel pitch p = 9.6 μm, virtual image plane d_f = 1.25D, and the Image Combiners (ICs) width w_m = 52.0 mm correspond to the ones in the prototype (red circles). (left) Angular resolution vs. instant FOV. Their product is determined by the number of pixels. (center-left) The gaze angle coverage is determined by the width of the IC. The micro display travels laterally and axially as gaze angle changes. (center-right) Focus cue. The virtual image plane can be moved back and forth with the axial movement of micro display. (right) Large enough eye box can be easily achieved because of the simple magnifier structure.

cover all focal depths with 10 to 15 mm range of travel as shown in the center-right of Fig. 4.

Since it is a simple magnifier system, the foveal display has a wide eye box. This eye box is determined by the concave half mirror size and display size. The width of the eye box w_e is given by:

$$w_e = \frac{d_f w_m - e' w_f}{d_f - e'} \tag{5}$$

where $e' = e + t + u/n$.

Figure 4 shows the design trade-off spaces for this foveal display. Based on these trade-offs and their relationships, one can design a foveal display to meet a particular set of application requirements. One particularly relevant trade-off is that the angular resolution or FOV can be improved, but one at the cost of the other for a given display resolution (pixel count). Furthermore, it is physically impractical to achieve an instantaneous foveal FOV larger than 40° as a compromise to the real-world scene FOV and form factor of the final design. If one were to build the largest possible instantaneous FOV foveal display system, the beam splitter thickness t would be chosen to be equal to the height of the micro display h_d. In this case, the micro display would be moved only horizontally not vertically. One can choose this feature to minimize the stackup thickness and make the overall system wearable while still preserving a large horizontal gaze angle coverage by choosing a short focal length IC (IC1, wearable prototype). To achieve a higher angular resolution system, one can choose larger t, larger a, and larger r to secure a longer optical path length. In this case, the angular resolution is higher but the overall system is bulkier and less wearable (IC3).

3.2 Peripheral display

For the peripheral display, we propose a large FOV, always-in-focus, see-through near-eye virtual retinal (i.e., Maxwellian-view) display which is composed of a laser scanning projector of resolution $N_{px} \times N_{py}$, a beam shaping lens of focal length f, and a reflective HOE as shown in Fig. 3. The HOE is manufactured to be reflective only to the wavelengths used by the projector and creates a Maxwellian

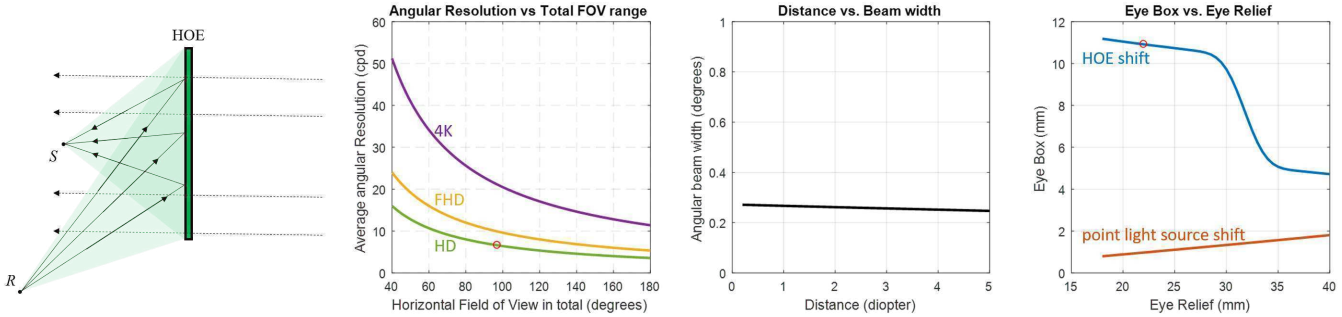


Fig. 5. Design trade-off space for the peripheral display. The beam width, HOE size w_p, and the projector resolution 1280×720 correspond to the prototype in Section 4 (red circle). (left) The angular selectivity of HOE. The HOE is only reflective to the light rays from the projector (R) and creates a Maxwellian viewpoint (S). (center-left) Angular resolution vs. FOV. Note that the FOV here is the total FOV including eye rotation, 97 degrees in our prototype (red circle). (center-right) The beam shaping of the laser scanning projector. The beam width is consistent over several diopters, so the observer perceives an always-in-focus image regardless of accommodation. (right) Eye box vs. Eye Relief when horizontal FOV at the center is 85 degrees.. The eye box is enlarged by laterally translating the HOE, which exceeds the previous point light source shift method.

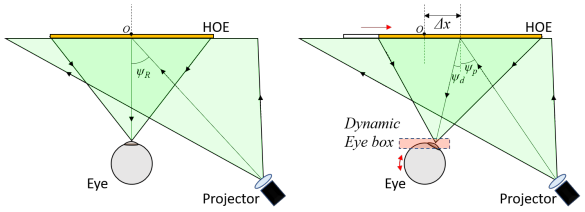


Fig. 6. Eye box expansion of the peripheral display using linear actuators. The dynamic eye box is generated by laterally translating the HOE. The projector is static because it covers the extent of this motion.

viewpoint at the wearer's pupil center. Note that HOEs have angular selectivity, so this element is transparent to most of the light rays incident from other directions as shown in the left of Fig. 5.

The width of the HOE is w_p and it is moved along with the foveal motion stage in a related (though not 1:1) ratio according to the user's gaze angle. The total HOE travel region $w_{p,tot}$ is covered by the projector. Here, the instantaneous horizontal FOV θ_p and the average angular resolution $\overline{c_p}$ is given by

$$\theta_p = 2\tan^{-1}\left(\frac{w_p}{2e}\right) \quad \text{and} \quad \overline{c_p} = \frac{N_{px}}{2\theta_{p,tot}}. \tag{6}$$

respectively, where $\theta_{p,tot}$ is the total FOV covered by the projector.

By using a laser scanning projector together with a beam shaping method, the peripheral display can provide an always-in-focus image despite its off-axis projection path. That is to say, individual pixel rays converge to the Maxwellian-view pupil center and the observed beam diameter doesn't change much over focal distance. As shown in the center-right of Fig. 5, the peripheral display gives constant angular beam waist, angle of the beam width from the eye, from 0D to 5D in the given parameters. Thus, the peripheral display provides a low resolution always-in-focus image, while the foveal display provides a high-resolution in-focus image over many depths using the varifocal design. Note that computational defocus blur (see Sec. 4.2.2) can be used in both the fovea and periphery to simulate objects in other focal planes.

3.2.1 Dynamic eye box using a Linear Actuator. The eye box of the static Maxwellian view displays is usually the same as the beam diameter, as all chief rays intersect at the center of the pupil. So there is a small viewpoint, or eye box, from which the observer can see a large FOV image. In this design, by laterally translating the HOE the eye box can be expanded. The HOE has an angular tolerance for the incident beam decided by the material parameters [Kogelnik 1969]. We exploit these tolerances to effectively translate the Maxwellian viewpoint by changing the HOE position. Figure 6 shows the principle of eye box expansion. When the linear actuator moves the HOE with Δx_p distance, the angle of the reconstructed light rays ψ_d at the center of HOE is given by

$$\psi_d = \sin^{-1}\left(\frac{1+a_n}{1+a_l}\frac{\lambda_r}{\lambda_p}(\sin\psi_r) - \sin\psi_p\right) \tag{7}$$

where λ_r and λ_p are wave length of recording and probe beams respectively, a_n is index of refraction modulation coefficient , a_l is shrinkage coefficient, and ψ_r and ψ_p are incident angles of recording and probe beams respectively [Hsieh and Hsu 2001; Jang et al. 2017].

With Eq. (7) and diffraction efficiency of off-Bragg reconstruction, the maximum eye box can be calculated (See Supplementary A.3). Previously, a similar dynamic eye box method was introduced by Jang et al.[Jang et al. 2018, 2017]. They changed the incident angle or moved the point light source to shift the Maxwellian viewpoint. Compared to this method, our linear actuation approach can provide a larger eye box (12 mm by 8 mm) at a given large FOV (85° horizontal) with a closer eye relief (22 mm) as shown in the right of Fig. 5 and Supplementary A.4. Note that the proposed method can achieve a sufficient eye box with even smaller eye relief, which makes the wearable prototype feasible.

3.3 Matching visual acuity

We evaluate the resolution of our foveated AR display by comparing it against two kinds of psychophysical visual acuity data measured as a function of visual eccentricity (Fig. 7). The solid black line represents letter acuity starting at 20/20 (30 cpd, the clinical standard acuity at the fovea) and falling monotonically as a function

of eccentricity [Anstis 1974]. This 20/20 standard was the basis for display formats for many years, but it can potentially underestimate the requirement as vision is often corrected to be better than 20/20 [Elliott et al. 1995]. Therefore we also include the more demanding requirement (solid gray line), which is detection acuity of grating pattern directly formed on the retina using a laser interference technique [Thibos et al. 1996], which bypasses the imperfect eye optics as an imaging device. This second line serves as the conservative criterion required by a perfectly corrected eye.

Our foveal display provides 1920 x 1200 pixels, and the IC design determines foveal resolution and FOV. IC1 was optimized for FOV, yet providing resolution higher than letter acuity at the fovea. IC3 was optimized for resolution, which was higher than the conservative requirement at the fovea. IC2 was a compromise of the two - its resolution was significantly higher than the 30 cpd standard, and FOV was wider than that of IC3.

The resolution of the peripheral display varies because the projected size of each pixel is decreased as projection distance is shortened near the temple, and the number of allocated pixels per degree is higher at larger eccentricity (see Supplementary A.5). The blue solid line shows resolution of the peripheral view when looking straight ahead, gradually increasing from 3.4 cpd at the fovea to 9.2 cpd at 42.5 deg of eccentricity. Note that the resolution of peripheral display has the minimum value at the fovea, which allows additional efficiency in the given pixel numbers.

The overall resolution of the system is the combination of the foveal resolution where available and the peripheral resolution otherwise. In all cases, we provide a display resolution higher than normal visual acuity across most eccentricities. Use of IC1 provides display resolution that is higher than normal acuity for all eccentricities. IC3 provides a superior image quality for the foveal view, but causes a small range of eccentricity to become display-limited.

We expect 4k MOLEDs and FHD micro projectors to be available soon. The pale red and blue lines illustrate the resolution distribution that these next generation displays will enable. Display resolution can be kept higher than the conservative requirement while keeping the same FOV. Even with advances in display technology, the concept of foveation will remain crucial to match visual acuity across a wide FOV.

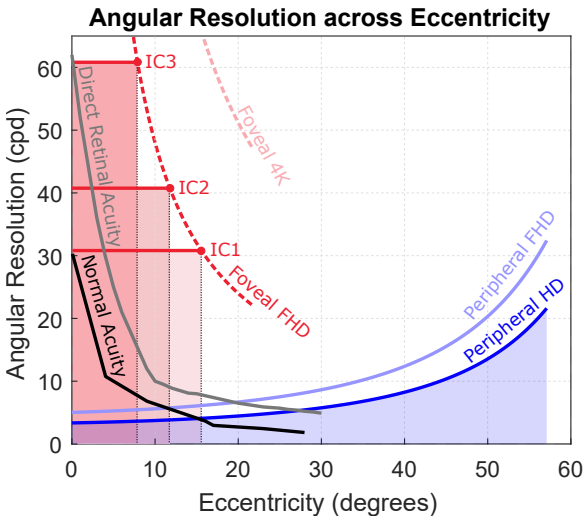


Fig. 7. Resolution as a function of eccentricity in foveated AR. Our ability to visually resolve spatial patterns decreases monotonically as a function of visual eccentricity. The solid black line represents the visual acuity under normal viewing conditions. By integrating two display views in a foveated manner, we keep the display resolution higher than the normal visual acuity for most eccentricities.

3.4 On-axis gaze-tracking

Our design implements on-axis gaze tracking by placing the eye tracking cameras at the bottom of the display so that the eye is observed through the beam splitter (Fig. 2b and Fig. 3). This camera location does not occlude the view of user's hands and feet and thereby ensures unconstrained navigation and hand-eye coordination. The on-axis, folded eye view allows for high and uniform gaze accuracy. Unfavorably, the beam splitter reduces the received light intensity from the eye. In addition, some reflections from incoming environmental light sources can appear in the eye tracking camera image. We deal with these challenges in software by using a robust pupil localization algorithm (see Sec.4.2.1). We use PupilLabs cameras producing monochrome infrared images under active infrared LED illumination from the camera position. The images contain corneal reflections (*glints*). However, the tracking algorithm being used is not dependent on the tracking of individual glints.

3.5 Gaze angle change

As the eye moves, it is necessary to adjust the position of the displays to maintain proper angular resolution over the entire FOV. In the proposed design, two lightweight components, a micro display ($3g$) and an HOE film ($<1g$), are moving based on the gaze. When the gaze angle α changes, the micro display and HOE should travel x_f and x_p, respectively, to provide proper angular resolution over eccentricity ϵ. The travel distance of the foveal display x_f is

$$x_f = \left(b + \frac{D_e}{2}\right) \times \frac{\tan(\alpha)}{M}, \tag{8}$$

where D_e is the diameter of the eyeball. The displacement of the peripheral display x_p cannot be derived in an analytic form due to the non-linearity of the HOE. Instead, we can numerically compute x_p with the HOE simulator (see Supplementary A.3). The travel distance as a function of gaze angle for our prototype is shown on the left of Fig. 8. The ratio of the two distances x_f/x_p is consistent over the range of gaze angles (the dashed line in the left of Fig. 8). It is therefore possible for two independent linear actuators to be used to achieve perfect calibration (optical bench prototype), or a single linear actuator with a dual-threaded assembly (wearable prototype) could be used for a smaller form factor and lower cost.

The angular resolution over gaze angle and eccentricity is given by a geometric relationship. The angular resolution of the foveal display $c_f(\alpha, \epsilon)$ within the FOV θ_f is given by

$$c_f(\alpha, \epsilon) = K_f b \sec^2(\alpha + \epsilon) \tag{9}$$

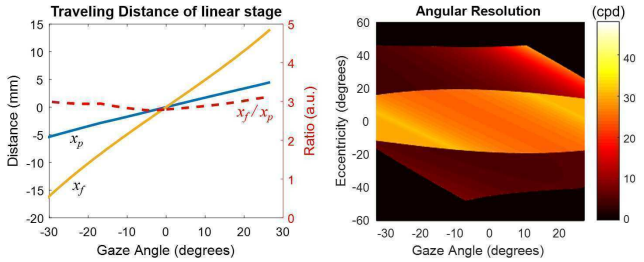


Fig. 8. Angular resolution vs gaze angle for the prototype with IC1. (left) The required travel distance for the micro display (x_f, yellow line) and the HOE (x_p, blue line). The micro display travels around 3 times more than the HOE as their relative positions are different (dashed line). (right) The angular resolution over eccentricity as a function of gaze angle. The high resolution inset is always located in the foveal region. Detailed resolution analysis is demonstrated in Supplementary A.5.

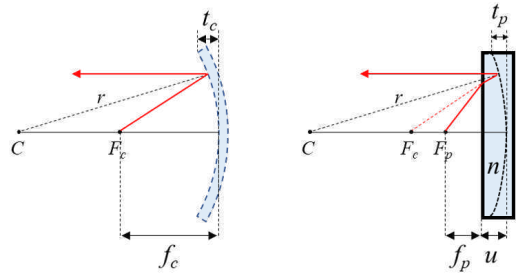


Fig. 9. (a) A spherical image combiner and (b) a planar image combiner with the same radius of curvature r. Note that the focal length of planar image combiner f_p is shorter than that of spherical image combiner f_c with the same thickness ($t_c = t_p$)

while the angular resolution of the peripheral display $c_p(\alpha, \epsilon)$ within the FOV θ_p is given by

$$c_p(\alpha, \epsilon) = \frac{K_p e \sec^2(\alpha + \epsilon)}{d_p \sec^2(\phi_{\alpha,\epsilon})} \tag{10}$$

where K_f and K_p are resolution constants determined by the total pixel numbers, $\phi_{\alpha,\epsilon} = \tan^{-1}(\tan\phi_0 - e\tan(\alpha + \epsilon)/d_p)$, and ϕ_0 is the incident angle of the projector. The total angular resolution provided by the prototype with IC1 is shown in the right of Fig. 8. The implemented system can provide ~30 cpd resolution in the fovea and >3 cpd in the periphery, for ±20° gaze angle.

4 IMPLEMENTATION

Here we summarize many subsystem-level design decisions not captured in the high-level design summary above.

4.1 Hardware implementation

4.1.1 Planar image combiners. The foveal display uses a planar IC. Compared to a spherical combiner, the planar one has three advantages. First, it leads to a thinner display because reflection occurs inside a glass medium. As shown in Fig. 9, the focal length of a spherical IC is $f_s = r/2$, where r is radius of the spherical surface.

Table 2. Specifications of the manufactured image combiners for the foveal display. The wearable prototype was implemented with IC1, while the optical-bench prototype tested all of them.

Specification	IC1	IC2	IC3
Angular Resolution (cpd)	30	40	60
Horizontal FOV (degrees)	32	24	16
r (mm)	100.14	132.87	205.75
t (mm)	11.75	30	30
a (mm)	16.66	8.25	29.25
u (mm)	4	4	4
Wearable	Y	N	N

The focal length of the planar IC with radius r is

$$f_p = \frac{r - 2u}{2n}, \tag{11}$$

where n is the refractive index and u is the inner thickness of the IC. Moreover, the thickness of the IC t_i,

$$t_i = r - \sqrt{r^2 - (D/2)^2}, \tag{12}$$

where D, the maximum diameter of the IC, decreases with r. Second, ghost images due to double reflections at the surface are reduced by using single magnifying surface and an anti-reflective coating. Third, planar ICs enable a relatively compact and flat profile (Fig. 1), avoiding the bug-eyed look of curved ICs.

We fabricated three ICs: small form-factor IC1, high angular resolution IC3, and compromise IC2. Table 2 shows their specifications. We reflectance-matched the IC and half-mirror at 30% to realize 50% transparency and anti-reflection coated all other optical surfaces.

4.1.2 Full-color holographic optical elements. In the peripheral display, full-color HOEs were recorded on photopolymer films (Litiholo, Covestro). Three lasers (red (Cobolt Flamenco, 660*nm*), green (Cobolt Samba, 532*nm*) and blue (Coherent Genesis MX, 460*nm*)) were used to record the HOEs with a phase-conjugated method (see Supplementary B.1.2). For better uniformity and higher diffraction efficiency, three layers of photopolymer films were stacked to record R, G and B HOEs (see Supplementary B.1.3).

4.1.3 Linear stage requirements for saccade and accommodation. An ideal foveated AR display would instantly react to gaze direction changes, thereby keeping the foveal and peripheral views constantly aligned with the gaze direction. This is extremely challenging because the eyeball can change gaze direction very quickly. This fast and abrupt eye movement is referred to as a saccade and its velocity can be as high as 500°/s [Rodieck 1998]. The linear stage for our peripheral display needs to closely support such rapid changes, as misaligning the peripheral beam with the user's pupil results in a noticeable change in global brightness.

The reaction requirements for the foveal view, however, are substantially relaxed since the visual system becomes somewhat insensitive to the visual input before, during, and after saccades [Ibbotson and Cloherty 2009; Matin 1974], making it very hard to notice the brief transition between low-resolution peripheral and high-resolution foveal views. Recent studies on foveated rendering suggest that human observers do not notice transitions within 50

ms of the completion of a saccade [Albert et al. 2017; Loschky and Wolverton 2007]. We can estimate the required average actuation velocity based on the saccade amplitude S and the typical duration Δt associated with the given saccade amplitude ($\Delta t = 2.7ms \times S + 37ms$) [Baloh et al. 1975]. The required average (angular) velocity for a 20° saccade, which is larger than most saccades [Bahill et al. 1975], is $141°/s$ and $205°/s$ for a saccade across the full 40° (±20°) of gaze directions supported by our system. We implemented our prototype with these requirements taken into account.

In accommodation, a reaction time in the range of 300−500 ms has typically been observed before the actual change in the lens shape is initiated [Bharadwaj and Schor 2005; Campbell and Westheimer 1960; Heron et al. 2001; Phillips et al. 1972]. The duration of actual lens accommodation of 500 − 800 ms has been reported [Bharadwaj and Schor 2005; Campbell and Westheimer 1960; Heron et al. 2001; Phillips et al. 1972], which means that the complete accommodation cycle, including the latency, typically requires around 1 second [Campbell and Westheimer 1960], which is a much relaxed condition than for saccades.

In order to maintain alignment between real and virtual content, it is important that the display show the right image at the right location. The bench-top prototype has limited mechanical travel, and uses *soft-steering* to compensate. *Soft-steering* is using software to adjust the rendered image within the available display area, which is larger than the actual foveal region (Fig. 7). For example in the prototype with IC1 (instant horizontal FOV : ±16°), the foveal display can cover most eye motions (except large saccades) with soft-steering. When mechanical actuation is required, the display only needs to reach the region where soft-steering works.

4.2 Software implementation

4.2.1 Deep-learning gaze tracking. We used the most recent deep learning based gaze estimation approach provided by Kim et al. [2019]. This neural network for pupil center estimation consists of seven convolutional layers is pre-trained on 16k synthetic eye images and 7k eye images from real people. Eye images are captured at a resolution of 640 × 480 pixel and uploaded to the GPU for inference. The network achieves an accuracy of 5 pixels in inference inputs with a success rate of 95% resulting in 2.06 ± 0.44° accuracy across a 30° × 40° FOV. The 9 ms system latency for gaze tracking at a sampling rate of 120 Hz includes 8 ms for image capture and 1 ms for GPU-based gaze estimation using cuDNN on a NVIDIA GeForce RTX 2080 Ti. The approach by Kim et al. shows excellent robustness against reflections and image noise. In our tests, the gaze tracking approach has been robust to glints or other visible reflections from user's glasses revealing this approach to be well-suited to our on-axis, gaze tracking design. Using an initial default calibration a user-specific gaze calibration is performed using a 7-point ring pattern. The regression-based polynomial mapping function is derived using the 2D-to-2D approach of Mansouryar et al. [2016].

We do not estimate pupil size although it can be estimated from the pupil region. We also do not extract vergence information. A binocular version of the eye tracker can be used to compute vergence from both eyes. The vergence point would allow for automatic focus adaptation of the foveated display.

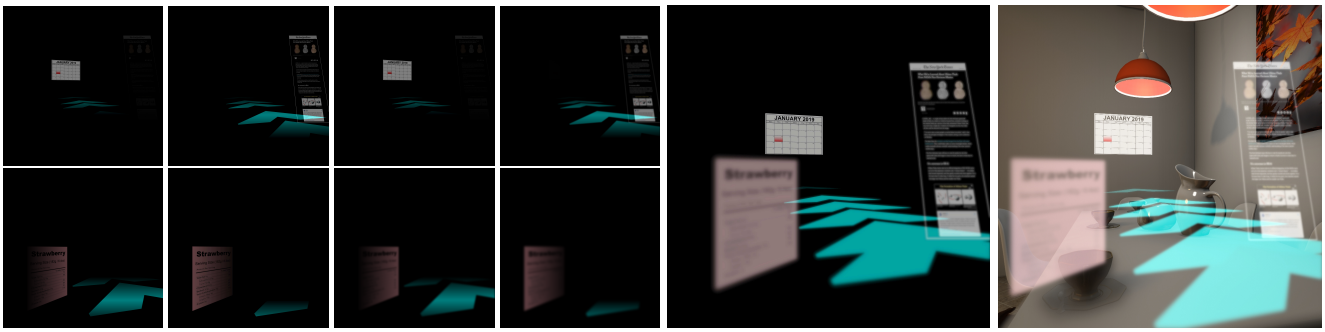
4.2.2 Real-time rendering.

Foveated image synthesis. We render separate 2 Mpix foveal and 1 Mpix peripheral images based on the current display position and precalibrated intrinsics. There is some redundancy between these images, so we apply a stencil mask to the outer parts of the foveal image to reduce the rendering workload. One could also mask the foveal region of the peripheral image, but the relative savings are so small that it costs about the same as the additional draw calls.

Our renderer supports two kinds of content: unshaded 3D meshes and HTML applications. We use the open-source Chromium project as a library to render GPU-accelerated HTML content to a texture asynchronously from the AR rendering. This amortizes the cost of rendering across both images, as well as across multiple frames. JavaScript and CSS provide animation capability for our custom AR HTML application mockups, and the system can of course render any existing web page that chromium can render.

Defocus simulation. We introduce a novel defocus algorithm for AR that renders in 1-2 ms and gives high-quality results, by limiting the renderer to *additive* virtual content. All current see-through displays, including ours, are additive: they can only add light over the real world, not occlude real objects. Because virtual content already cannot occlude real content, we introduce the constraint that virtual content also cannot occlude other virtual content–it must compose in the same way that it will compose with the real world. This requires that content have the character of an emissive ghost, which we consider aesthetically acceptable for many applications, and perhaps preferable to photorealistic rendering with imperfect lighting and registration. Additive rendering allows order-independent virtual compositing and eliminates depth discontinuity effects, which we exploit to build a fast defocus shader. Additive compositing and our efficient defocus simulation for it are well-suited to applications such as AR notifications, "hologram" video conferencing as depicted in *Star Wars* films, labels on the real world, and virtual screens. It is not a viable restriction for other applications such as 3D games or design preview with real-world shading in which occlusion between virtual objects is an important queue. In those cases, we approximate defocus for opaque objects using a prior method from the games industry [Bukowski et al. 2013; Sousa 2013]. There unfortunately is no currently-known method for efficient defocus simulation for a mixture of opaque and translucent objects, so our system must be run exclusively in either fully opaque or fully additive-transparent mode.

The input to the rendering pipeline is backface-culled meshes (to increase readability) with no depth buffer (for the no-occlusion, additive-only model). The framebuffer contains n render targets (we use $n = 4$ for all results). Each render target corresponds to bin for a contiguous range of circles of confusion, i.e., a *focus bin*. These are not depth layers and this is not a depth decomposition, although for an ideal single lens system, the focus bins can be thought of as *pairs* of depth bands in front and behind the plane of focus. For multi-lens optics or non-ideal optics in which the circle of confusion varies across the lens, that analogy breaks down as focus is no longer a simple function of depth, but the core mathematical notion of binning by focus remains.



(a) (b) (c) (d)

Fig. 10. Focus-matching the virtual content to the real world. (a) $n = 4$ render targets in a single frame buffer with sharp images segmented into circle-of-confusion bands. In this case, the viewer's focus is directed towards the label in the distance at 2.0 m. (b) Render targets after separated 2D Gaussian blur simulating defocus. (c) Final AR image for the periphery after compositing. (d) Simulated real-world view with overlaid AR content.

The pixel shader for the virtual content rendering pass computes the circle of confusion for each fragment of content and additively blends its contribution linearly into the two closest focus bins for that circle of confusion. In the case where there are an equal number of bins to the pixel width of the largest point spread function, then each fragment would only contribute to a single bin that matches its kernel. Blending into two bins allows a smaller number of bins to span a large focus range by approximating each point spread function with a linear combination of a slightly-larger and slightly-smaller kernel. This approximation is similar to trilinear MIP-mapping for managing texture scale in rendering. Figure 10(a) visualizes the peripheral framebuffer after this stage (the foveal framebuffer is similar).

The system then performs a Gaussian blur of each focus bin (Fig. 10(b)) in a single compute pass over multiple targets. This approximates the eye's point spread function for each render target's circle of confusion. Because both convolution and additive compositing are linear operators, we are able to perform the defocus blur as a post-process without introducing error from the order of operations. Note that this exploits linearity and separability in two ways: due to the additive and Gaussian constraints, blurring and compositing can be performed in any order, and can also be performed independent of fragment depth-order.

A Gaussian blur is a common, if imperfect approximation of optical blur. However we chose the Gaussian kernel for efficiency, not quality. The 2D Gaussian blur is separable into two 1D passes. Combined with the linearity, separation allows efficient blurring even in the fovea where the blur kernel may have a radius of 40 pixels and thus require an intractable 6.5k samples per pixel if performed naively in 2D. The second pass of the Gaussian blur also composites all bins into a single output image (Fig. 10(c)), avoiding an additional pass.

Multiple display matching. To register the two displays, we measure the following properties once per display and then apply correction at runtime per frame. Please refer to our supplement for implementation details. For geometric registration, we find corresponding points across the displays by displaying a checkerboard and then warp the fovea to the peripheral projection (Fig. 14). For intensity and color, we measure each display's output with a Gamma-Sci GS-1160 spectroradiometer and scale the output to the overlapping gamut. We feather the transition between the foveal and peripheral displays using a Gaussian mask to minimize any residual miscalibration and eliminate a spatial frequency discontinuity due to the differing resolutions (see Supplementary B.2.3).

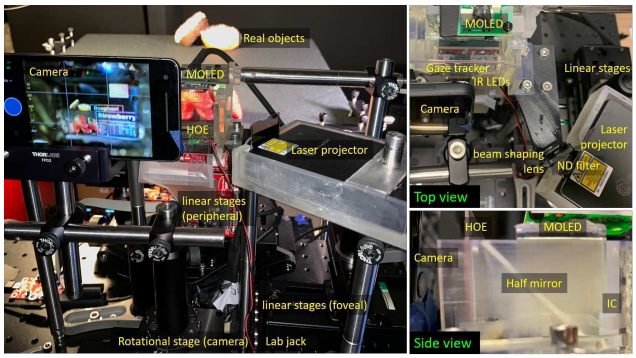


Fig. 11. Optical bench prototype. (left) Experiment configuration for Fig. 1. The center of the rotational stage is 12 mm from the camera pupil to simulate human eye rotation. (right) Close-ups from the top and side.

4.3 Prototypes

We implemented two Foveated AR designs: an optical bench prototype for controlled experiments, and a wearable prototype to evaluate form factor and explore miniaturization. The optical bench prototype includes gaze-contingent motorized 2D movement of the micro display and the HOE and the manual focal plane change, while the wearable prototype includes horizontal-only movement of the micro display and the HOE by a hand-screwed dual-threaded actuator to minimize the form factor.

4.3.1 Optical bench prototype. Fig. 11 shows the optical bench prototype. Most of the display results were from this setup except for Fig. 17. As described in Sec. 4.1.1, three customized ICs (manufactured by ILLUCO) were used in the optical bench prototype. We

used the eMagin (WUXGA LUT) micro OLED for the foveal display. Each 18.7×11.75 mm, 3 g micro display has 1920×1200 24-bit color pixels at >1000 cd/m^2 brightness. Overall transparency was 50% and the final brightness of the foveal display was around 70 nits. Our peripheral display is powered by the Celluon 1280×720 laser projector (PicoBit), with a 10 mm diameter 75 mm focal length plano-convex lens (Edmund Optics 84-281) for beam shaping. A full-color HOE film was recorded with our custom recording setup (see Supplementary B.1). To maximize the efficiency of all three wavelengths, we stacked three HOEs (Supplementary B.3). Two perpendicular linear polarizers were used in front of the laser projector to match the brightness of the peripheral display to that of foveal display. In the optical bench prototype, the micro OLED travels horizontally and axially via two high-speed motorized linear stages (DDSM50) from Thorlabs, Inc. The stages' maximum travel distance is 50 mm and their maximum speed is 500 mm/s. For the foveal view, this converts to $845°/s$ when using IC1 and $429°/s$ when using IC3, which far exceeds the $205°/s$ requirement for a 40° saccade in the foveal view (see 4.1.3). The HOE travels horizontally and vertically via two smaller motorized stages (ELL7/M) from Thorlabs, Inc., the stages' maximum travel distance is 25 mm and maximum speed is 180 mm/s. For the peripheral view, this converts to $1053°/s$, exceeding the typical speeds of the fastest saccades (500°). Due to the heavy, high-speed motorized stages, the vertical position of the micro OLED must be manually adjusted by changing the height of a jack, which can be driven by a motor. For capturing eye images in real time we use a PupilLabs camera working at a framerate of 120 Hz.

4.3.2 Wearable prototype. The wearable prototype consists of a modular, 3D printed frame that houses and aligns all of the optical/mechanical components used in the system (see Supplementary C) including a compact laser projector (MEGA1, MEGA1-F1), a beam shaping lens (Edmund Optics, 84-281), a right angle prism (Thorlabs, PS908), a micro OLED (eMagin, WUXGA LUT), optical front-end (combiners and half mirrors for the fovea and peripheral optical path), and the motion stage used to translate the foveal and peripheral displays in relation to each other. The basic optical structure of the wearable prototype is identical to the benchtop prototype, but it uses a smaller projector and the folded optical path for periphery. Note that the HOE was delaminated from the glass substrate to minimize the form factor and weight. The micro OLED display driver is located remotely and relayed to the display using a 5-feet cable provided by the manufacturer.

The wearable prototype uses a dual-threaded actuator to carefully control displacement of the foveal display's micro OLED relative to the peripheral display's HOE. This actuator consists of a custom-manufactured, threaded part that uses a #6-32 and #00-96 standard thread (turned on the same shaft) to create a 3:1 ratio of linear motion per turn (32 vs 96 threads per inch). The use of standard threads allows off-the-shelf nuts to be mated to the custom shaft and then to the micro OLED and HOE respectively to allow a single turn of the shaft to move each part by the appropriate ratio. This dual-threaded assembly can either be turned by hand or using an electromechanical source such as a DC or stepper motor. The weight of all components building the wearable prototype excluding attached cables is 235g.

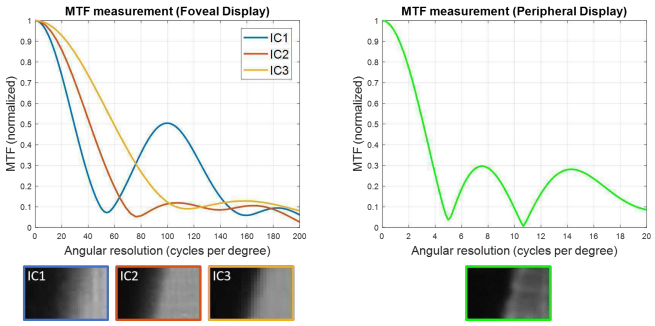


Fig. 12. MTF measurement results with a slanted edge method (left) MTF of the foveal display for each image combiner (IC1, IC2 and IC3), and (right) MTF of the peripheral display.

5 DISPLAY ASSESSMENT

We evaluate the optical properties of individual elements and the assembled display through photographs and videos. We captured through-the-display images on a 4032 × 3024 Pixel2 phone camera with a minimum f-number of 1.8.

5.1 Optical bench prototype

5.1.1 Resolution. The angular resolution of foveal and peripheral displays were measured using a slanted edge method. Figure 12 shows the measured MTF graphs and close-up photo of slanted edges. The normalized MTF of the foveal display is greater than 0.5 at 29 (IC1), 42 (IC2), and 59 (IC3) cpd at the center of FOV, while MTF of the peripheral display is greater than 0.5 at 3 cpd at the center. Note that the foveal display was captured by a 5184 × 3456 Canon EOS Rebel T6 DSLR camera with a zoom lens (Canon EF 75-300 mm) for its higher resolution while the peripheral display is captured using a Pixel 2 smart phone camera for its large FOV.

5.1.2 Field of view and eye box. Since the foveal display is a simple, magnified micro display with on-axis optical components, it provides a rectangularly shaped FOV and a large static eye box (47 mm × 15 mm, see Supplementary D). In order to measure the achieved foveal display FOV, the whole display was illuminated with a green image and was captured with a background FOV panel located at a 15 cm distance. IC1, IC2 and IC3 provided 33°, 22° and 16° horizontal FOVs, respectively.

The peripheral display FOV couldn't be measured using the same method as its FOV exceeded the maximum camera FOV of Pixel 2 and added fish-eye lenses blocked the beam path of the prototypes. Instead, we captured a top view of the peripheral display system with a 1mm grid paper in the plane of the user's pupil, as shown in Fig. 13. A white image from the laser projector was then projected onto the green HOE and only the green light was diffracted at the HOE plane as recorded. This diffracted light created a footprint image on the paper and the FOV was measured with the angles of the footprint. The peripheral display provided 85° horizontal FOV and 78° vertical FOV (101.4° diagonal) at the center position. The peripheral display eye box was measured similarly by moving the HOE along the horizontal and the vertical direction (see Supplementary video and Supplementary D.4). The peripheral display provided

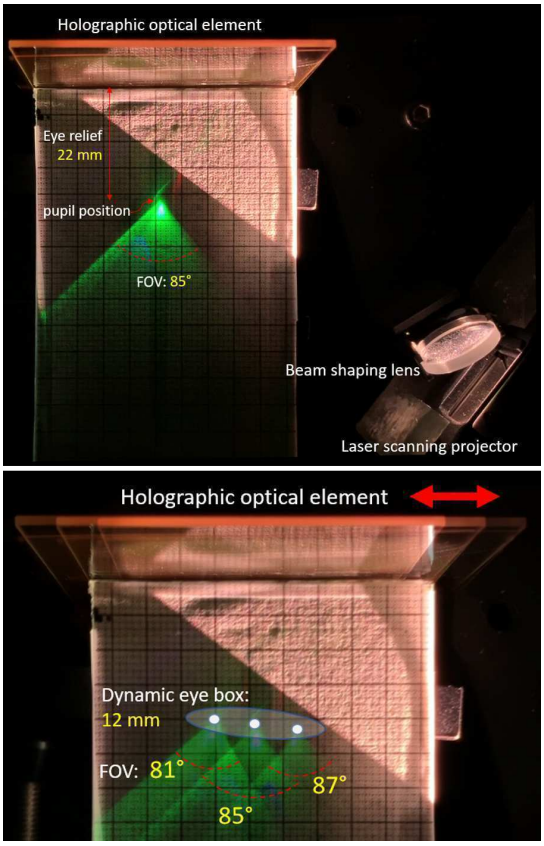


Fig. 13. FOV and eyebox measurement by extracting the peripheral display and intercepting its output on a white grid. We projected a green field, which appears as a triangle on the measurement grid. (top) Peak 85° horizontal FOV Maxwellian viewpoint at 22 mm eye relief. (bottom) Superposition of captured images for left-most, center and right-most translated viewpoints, showing a net 12 mm horizontal eye box.

12mm of horizontal and 8mm of vertical dynamic eye box while preserving 81 ~ 87° of horizontal FOV and 74 ~ 78° of vertical FOV (97 ~ 103° diagonal).

5.1.3 Multiple display calibration results. Figure 14 shows the monochrome Foveated AR display results for the given fovea location. The display results show that the foveal and peripheral images were well-calibrated in geometry, color and intensity. The foveal region can be moved anywhere in the FOV by translating the micro OLED within its plane (see Supplementary video). The resolution difference between the foveal and peripheral region is clearly shown in the close-up photos. The transition between the foveal and peripheral regions is quite subtle because of the blending algorithm (see Supplementary B.2.3).

5.1.4 Gaze angle coverage. Figure 15 shows the display results for a gaze angle coverage experiment. A Pixel 2 camera was located on a rotational stage and the distance between the lens and the rotational center was set to 12 mm to simulate the human eye rotation. The base images were generated with consideration of the image transformation between the center view and the given gaze direction, so that the virtual scene can be located regardless of the observer's gaze direction (see Supplementary B.2.1). The optical bench prototype covered ±20° of gaze angle in the horizontal direction with the foveal region always located at the center of camera FOV (see Supplementary video).

5.1.5 Focus cue change. Figure 16 shows the Foveated AR display results with various camera focal length and foveal display planes. A (real) police car, a lion doll, and a dinosaur doll are located at the 30, 80, and 250 cm respectively in the real world. In the top figure, both the foveal display plane and the camera focal distance were set at the 80 cm from the observer, and all the foveal and peripheral content, as well as the lion doll, were in focus as expected. When the camera focus was changed to 30 cm from the observer, the foveal display and the lion doll were out of focus while the police car was in focus. Note that the peripheral display shows the same size blur because of the always-in-focus characteristic of a Maxwellian view display. As the micro OLED was moved closer to the image combiner along the vertical axis, the foveal display comes back into focus again while the lion remains out of focus as shown in the bottom of Fig. 14.

5.2 Wearable prototype

Figures 1 and 17 show the wearable prototype and its display results. The foveal and peripheral display, a gaze tracker, IR illumination, and a dual-threaded actuator were implemented into a wearable design. The traveling distances of micro display and the HOE were coupled to 3:1 ratio and controlled by a single hand screw. The wearable prototype also provided high resolution images for the foveal region and low resolution, large FOV images for periphery. The geometric registration and blending algorithm were applied correctly. The FOV of the wearable prototype was limited to 77° horizontal FOV and 53° vertical FOV (86.4° diagonal) due to the laser projector orientation for miniaturization as shown in Fig. 1. The color calibration was not applied because the compact projector's wavelengths didn't well-matched with the recording wavelengths and the peripheral blue was out of foveal display color gamut.

6 CONCLUSION

We have introduced and analyzed a new AR display system that advances the state of the art for simultaneous wide FOV (100° diagonal), compact form factor, high foveal resolution (60 cpd), variable focus display and rendering, and large eyebox (12 mm × 8 mm). Several new design ideas allow our system to exceed the capabilities of previous displays. A key innovation is the use of a holographic element with dynamic position driven by gaze tracking, sidestepping the optical invariant for a static element. Integrated low-latency gaze tracking, motors, and rendering enable the dynamic position and varifocal system. The combination of an HOE, projector, and OLED allows a wide FOV and high resolution, and our prototype optical design demonstrates that a compact foveated AR headset is feasible. We now reflect on the remaining constraints and prospective directions for future work.

Eye relief is a key constraint on FOV. Our phase-conjugate HOE method can record an element suitable for a 130° horizontal, monocular FOV. However, under our design, the projector would then be partly occluded by the wearer's eyelashes and the HOE would

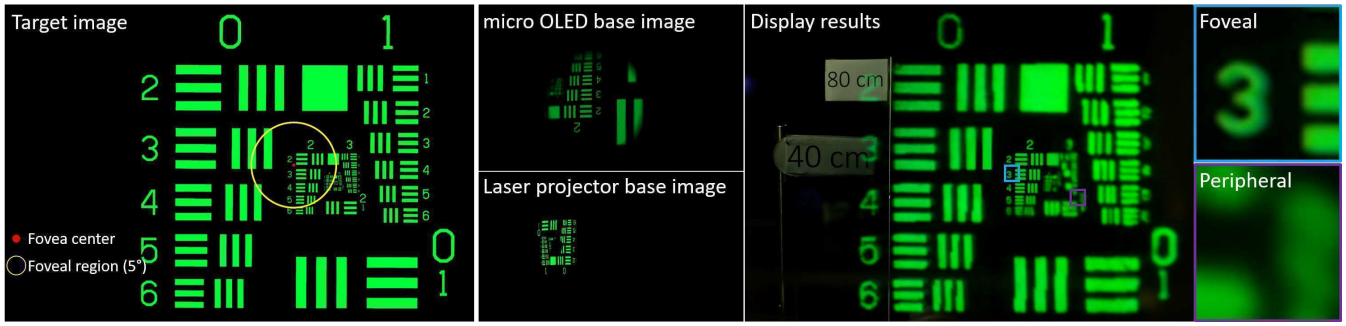


Fig. 14. Display results for the multiple display calibration. An example of target image, generated base images for foveal and peripheral displays and display results for given fovea position are shown. Note that the micro OLED covers most of the field of view so the foveal region can be located anywhere.

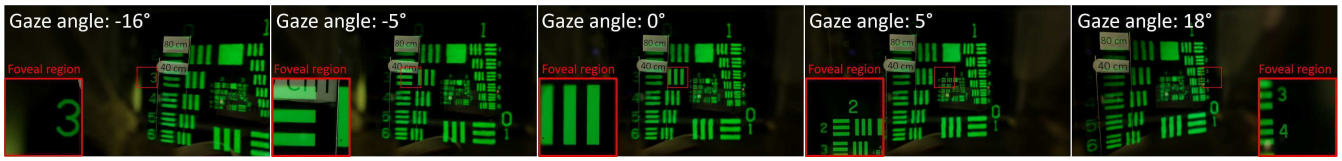


Fig. 15. Display results of different gaze angles with the optical bench prototype. Note that the virtual scene was located at the center view and independent of the gaze angle, and that the foveal region is always located at the center.

be uncomfortably close to the eye. Relocating the projector to the forehead and introducing a waveguide may be a viable solution.

Mechanical complexity is the biggest disadvantage to our approach. Many consumer devices with dynamic elements demonstrate high reliability, such as ink jet printers and blu-ray disk drives. However, those are mature modern systems representing decades of engineering work. Some of the interesting brave ideas never have been practically commercialized. Nevertheless, this research explores an interesting region of the design space and will stimulate more work in the area. Given the benefits of dynamic elements for AR, we look to microrobotics, camera design, and other areas for techniques for making such AR designs viable for consumer devices. Of particular interest for future work are voice coils and piezoelectric motors. These currently have drawbacks for responsiveness, torque, and power draw, but still-lighter HOEs and advances in device design may enable further miniaturization.

Robust estimation of accommodative response (i.e., focus depth) remains a challenging problem. Previous studies support the feasibility of estimating focus depth solely from binocular vergence, but also report inaccuracy on the scale of half a diopter or more [Mlot et al. 2016]. We contemplate a specialized gaze tracking network, such as Kim et al. [2019], to estimate focus depth directly rather than predict it from separately-tracked pupils. A coarse depth map from outward cameras and from the rendered scene would provide an important additional input for such a framework.

Our rendering system efficiently simulates defocus blur in both the foveal area and the periphery. For the far periphery, the eye's own limited resolution is unlikely to capture this effect as a depth or accommodation-driving cue [Gu and Legge 1987; Kim et al. 2017; Wang et al. 2006]. However, the chromatic aberration in the periphery may be a significant factor [Cholewiak et al. 2017], and we are investigating methods for simulating it accurately and efficiently.

Looking forward, the natural role of a near-eye device is to combine prescription corrective optics, adaptive sunglasses, and AR display into a single accessory that is personal and always worn. Many challenges remain. We must create solutions that integrate seamlessly with vision correction. Power and form factor determine viability for continuous use: for professional use cases these must meet certain thresholds for safety and usability, but for widespread consumer use we must address comfort and aesthetics. In this work, we created a novel display that surpassed design constraints of prior approaches through the combination of mechanical and optical elements with GPU computation for machine learning and rendering. We hope our work will encourage others to explore such hybrid devices that lead to future breakthroughs in AR display.

REFERENCES

G. Abadie. 2018. A Life of a Bokeh. SIGGRAPH Course: Advances in real-time rendering in games part 1.
K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke. 2017. Near-Eye Varifocal Augmented Reality Display using See-Through Screens. In *Proc. of SIGGRAPH Asia.*
R. Albert, A. Patney, D. Luebke, and J. Kim. 2017. Latency requirements for foveated rendering in virtual reality. *ACM TAP* 14, 4 (2017).
S. Anstis. 1974. A chart demonstrating variations in acuity with retinal position. *Vision research* 14, 7 (1974).
A. Bahill, M. Clark, and L. Stark. 1975. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* 24, 3-4 (1975).
D. Baldwin. 1981. Area of interest: Instantaneous field of view vision model. In *Image Generation/Display Conference.*
R. Baloh, A. Sills, W. Kumley, and Vi. Honrubia. 1975. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology* 25, 11 (1975).
S. Bharadwaj and C. Schor. 2005. Acceleration characteristics of human ocular accommodation. *Vision Research* 45, 1 (2005).
M. Bukowski, P. Hennessy, B. Osman, and M. McGuire. 2013. The Skylanders SWAP Force Depth-of-Field Shader. In *GPU Pro 4: Advanced Rendering Techniques.*

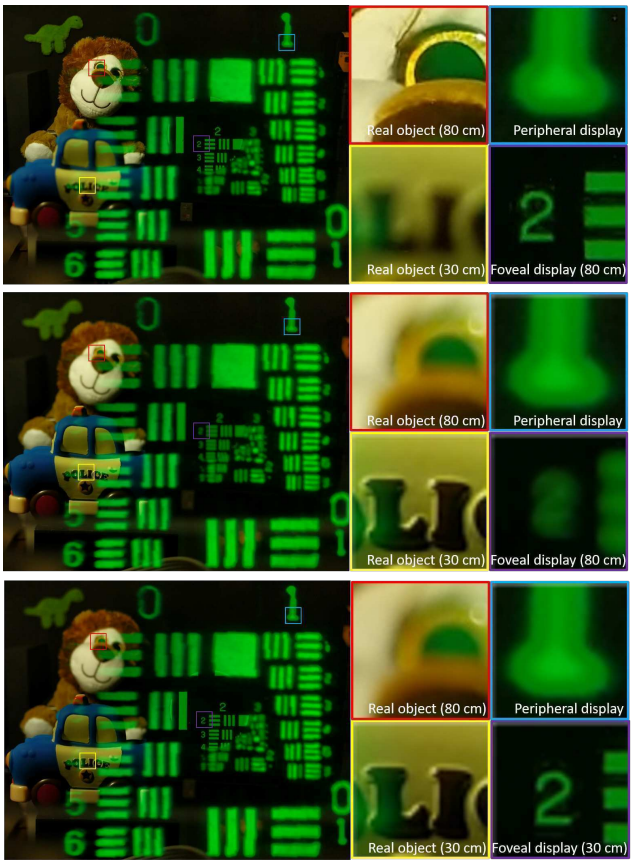


Fig. 16. Focus changing experiment results. A police car, lion doll, and a dinosaur doll are located at the 30, 80, and 250 cm respectively. (top) foveal display plane: 80 cm and camera focal plane : 80 cm (middle) foveal display plane: 80 cm and camera focal plane: 30 cm (bottom) foveal display plane: 30 cm and camera focal plane: 30 cm.

Fig. 17. Display photographed through the display of the wearable prototype. The top, red inset shows a detail from the foveal region and the lower, green inset is from the peripheral region. Note that the smaller projector used in the wearable prototype has narrower light cone angle and worse color representation which led to smaller FOV (77° × 53 °) and severe color mismatch.

O. Cakmakci and J. Rolland. 2006. Head-worn displays: a review. *Journal of display technology* 2, 3 (2006).

F. Campbell and G. Westheimer. 1960. Dynamics of accommodation responses of the human eye. *J. Physiol.* 151, 2 (1960).

S. Cholewiak, G. Love, P. Srinivasan, R. Ng, and M. Banks. 2017. ChromaBlur: Rendering chromatic eye aberration improves accommodation and realism. *ACM TOG* 36, 6 (2017).

R. Cook, T. Porter, and L. Carpenter. 1984. Distributed Ray Tracing. In *Proc. of SIGGRAPH.*

Magic Leap Corportation. 2019a. Magic Leap One Creator Edition. https://www.magicleap.com/magic-leap-one. Accessed: 2019-01-12.

nReal Corportation. 2019b. nReal Light. https://www.nreal.ai/. Accessed: 2019-01-12.

D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs. 2017. Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE TVCG* 23, 4 (2017).

D. Elliott, K. Yang, and D. Whitaker. 1995. Visual acuity changes throughout adulthood in normal, healthy eyes: seeing beyond 6/6. *Optometry and vis. science* 72, 3 (1995).

W. Fuhl, D. Geisler, T. Santini, T. Appel, W. Rosenstiel, and E. Kasneci. 2018. CBF: Circular Binary Features for Robust and Real-time Pupil Center Detection. In *ACM Symposium on Eye Tracking Research & Applications.*

W. Fuhl, T. Kübler, K. Sippel, W. Rosenstiel, and E. Kasneci. 2015. Excuse: Robust pupil detection in real-world scenarios. In *Computer Analysis of Images and Patterns.*

W. Fuhl, T. Santini, G. Kasneci, W. Rosenstiel, and E.Kasneci. 2017. PupilNet v2.0: Convolutional Neural Networks for CPU based real time Robust Pupil Detection. *CoRR* abs/1711.00112 (2017). http://arxiv.org/abs/1711.00112

Y. Gu and G. Legge. 1987. Accommodation to stimuli in peripheral vision. *JOSA A* 4, 8 (1987).

B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder. 2012. Foveated 3D Graphics. *ACM TOG* 31, 6 (2012).

P. Haeberli and K. Akeley. 1990. The Accumulation Buffer: Hardware Support for High-quality Rendering *(Proc. of SIGGRAPH).*

G. Heron, W. Charman, and C. Schor. 2001. Dynamics of the accommodation response to abrupt changes in target vergence as a function of age. *Vision Res.* 41, 4 (2001).

M.-L. Hsieh and K.Y. Hsu. 2001. Grating detuning effect on holographic memory in photopolymers. *Optical Engineering* 40, 10 (2001).

X. Hu and H. Hua. 2014. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics. *Optics Express* 22, 11 (2014).

H. Hua. 2017. Enabling Focus Cues in Head-Mounted Displays. *Proc. IEEE* 105, 5 (2017).

H. Hua and B. Javidi. 2014. A 3D integral imaging optical see-through head-mounted display. *Optics Express* 22, 11 (2014).

M. Ibbotson and S. Cloherty. 2009. Visual perception: saccadic omission, suppression or temporal masking? *Current Biology* 19, 12 (2009).

C. Jang, K. Bang, G. Li, and B. Lee. 2018. Holographic near-eye display with expanded eye-box. In *Proc. of SIGGRAPH Asia.*

C. Jang, K. Bang, S. Moon, J. Kim, S. Lee, and B. Lee. 2017. Retinal 3D: Augmented Reality Near-eye Display via Pupil-tracked Light Field Projection on Retina. In *Proc. of SIGGRAPH Asia.*

J. Kim, M. Stengel, A. Majercik, S. De Mello, S. Laine, M. McGuire, and D. Luebke. 2019. NVGaze: Low-Latency, Near-Eye Gaze Estimation with an Anatomically-Informed Dataset. In *Proc. of CHI.*

J. Kim, Q. Sun, F. Huang, L. Wei, D. Luebke, and A. Kaufman. 2017. Perceptual Studies for Foveated Light Field Displays. *arXiv preprint arXiv:1708.06034* (2017).

S.-B. Kim and J.-H. Park. 2018. Optical see-through Maxwellian near-to-eye display with an enlarged eyebox. *Optics Letters* 43, 4 (2018).

H. Kogelnik. 1969. Coupled wave theory for thick hologram gratings. *Bell System Technical Journal* 48, 9 (1969).

A. Koulieris, M.and Mantiuk R. Akşit, K.and Stengel, K. Mania, and C. Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. In *Computer Graphics Forum*, Vol. 38.

B. Kress and W. Cummings. 2017. Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. In *SID Symp. Digest of Technical Papers.*

G. Lee, J. Hong, S. Hwang, S. Moon, H. Kang, S. Jeon, J. Kim, H.and Jeong, and B. Lee. 2018b. Metasurface eyepiece for augmented reality. *Nature comm.* 9, 1 (2018).

J. S. Lee, Y. K. Kim, M. Y. Lee, and Y. H. Won. 2019. Enhanced see-through near-eye display using time-division multiplexing of a Maxwellian-view and holographic display. *Optics Express* 27, 2 (2019).

S. Lee, J. Cho, B. Lee, Y. Jo, C. Jang, D. Kim, and B. Lee. 2018a. Foveated retinal optimization for see-through near-eye multi-layer displays. *IEEE Access* 6 (2018).

S. Lee, Y. Jo, D. Yoo, J. Cho, D. Lee, and B. Lee. 2018c. TomoReal: Tomographic Displays. *arXiv preprint arXiv:1804.04619* (2018).

J. Lemley, A. Kar, A. Drimbarean, and P. Corcoran. 2018. Efficient CNN Implementation for Eye-Gaze Estimation on Low-Power/Low-Quality Consumer Imaging Systems. *arXiv preprint arXiv:1806.10890* (2018).

S. Liu, D. Cheng, and H. Hua. 2008. An optical see-through head mounted display with addressable focal planes. In *Mixed and Augmented Reality.*

S. Liu, H. Hua, and D. Cheng. 2010. A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *IEEE TVCG* 16, 3 (2010).
L. Loschky and G. Wolverton. 2007. How late can you update gaze-contingent multiresolutional displays without detection? *ACM TOMM* 3, 4 (2007).
A. Maimone, A. Georgiou, and J. Kollin. 2017. Holographic Near-eye Displays for Virtual and Augmented Reality. In *Proc. of SIGGRAPH*.
A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. 2014. Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses Using Defocused Point Light Sources. *ACM Trans. Graph.* 33, 4 (2014).
M. Mansouryar, J. Steil, Y. Sugano, and A. Bulling. 2016. 3d gaze estimation from 2d pupil positions on monocular head-mounted eye trackers. In *Proc. of the Symp. on Eye Tracking Research & Applications*.
E. Matin. 1974. Saccadic suppression: a review and an analysis. *Psychological bulletin* 81, 12 (1974).
O. Mercier, Y. Sulai, K. Mackenzie, M. Zannoli, J. Hillis, D. Nowrouzezahrai, and D. Lanman. 2017. Fast Gaze-contingent Optimal Decompositions for Multifocal Displays. *ACM TOG* 36, 6 (2017).
K. Miki, T. Nagamatsu, and D. Hansen. 2016. Implicit user calibration for gaze-tracking systems using kernel density estimation. In *Proce. of the Symp. on Eye Tracking Research & Applications*.
G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci. 2016. 3D Gaze Estimation using Eye Vergence.. In *HEALTHINF*. 125–131.
R. Narain, R. Albert, A. Bulbul, Gr. Ward, M. Banks, and J. O'Brien. 2015. Optimal presentation of imagery with focus cues on multi-plane displays. *ACM TOG* 34, 4 (2015).
S. Phillips, D. Shirachi, and L. Stark. 1972. Analysis of accommodative response times using histogram information. *American Journal of Optometry* 49, 5 (1972).
S. Reder. 1973. On-line monitoring of eye-position signals in contingent and noncontingent paradigms. *Behavior Research Methods & Instrumentation* 5, 2 (1973).
R. Rodieck. 1998. *The first steps in seeing*. Sinauer Associates Sunderland, MA.
J. Rolland, A. Yoshida, L. Davis, and J. Reif. 1998. High-resolution inset head-mounted display. *Applied optics* 37, 19 (1998).
T. Santini, W. Fuhl, and E. Kasneci. 2017. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction. In *Proc of CHI*.
K. Selgrad, C. Reintges, D. Penk, P. Wagner, and M. Stamminger. 2015. Real-time Depth of Field Using Multi-layer Filtering. In *Proc of I3D*.
M. Shenker. 1987. Optical design criteria for binocular helmet-mounted displays. In *Display System Optics*.
L. Shi, F. Huang, W. Lopes, W. Matusik, and D. Luebke. 2017. Near-eye Light Field Holographic Rendering with Spherical Waves for Wide Field of View Interactive 3D Computer Graphics. In *Proc. of SIGGRAPH Asia*.
M. Shinya. 1994. Post-filtering for Depth of Field Simulation with Ray Distribution Buffer. In *Proc. of Graphics Interface*.
T. Sousa. 2013. CryEngine3 Graphics Gems. SIGGRAPH Course: Advances in Real-Time Rendering Course.
A Spooner. 1982. *The trend towards area of interest in visual simulation technology*. Technical Report. Naval Training Equipment Center Orlando FL.
G. Tan, Y.-H. Lee, T. Zhan, J. Yang, S. Liu, D. Zhao, and S.-T. Wu. 2018. Foveated imaging for near-eye displays. *Optics Express* 26, 19 (2018).
L. Thibos, D. Still, and A. Bradley. 1996. Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision research* 36, 2 (1996).
M. Tonsen, J. Steil, Y. Sugano, and A. Bulling. 2017. InvisibleEye: Mobile Eye Tracking Using Multiple Low-Resolution Cameras and Learning-Based Gaze Estimation. In *Proc. Interact. Mob. Wearable Ubiquitous Technol.*
B. Wang, K. Ciuffreda, and T. Irish. 2006. Equiblur zones at the fovea and near retinal periphery. *Vision Research* 46, 21 (2006).
L. Xiao, A. Kaplanyan, A.r Fix, M. Chapman, and D. Lanman. 2018. DeepFocus: learned image synthesis for computational displays. In *Proc. of SIGGRAPH Asia*.
Y. Yang, H. Lin, Z. Yu, S. Paris, and Ji. Yu. 2016. Virtual DSLR: High Quality Dynamic Depth-of-Field Synthesis on Mobile Platforms. In *Digital Phot. and Mob. Imaging*.
T. Zhan, Y. Lee, and S. Wu. 2018. High-resolution additive light field near-eye display by switchable Pancharatnam–Berry phase lenses. *Optics Express* 26, 4 (2018).