



## High-performance integral-imaging-based light field augmented reality display using freeform optics

Item Type	Article
Authors	Huang, Hekun; Hua, Hong
Citation	Hekun Huang and Hong Hua, "High-performance integral-imaging-based light field augmented reality display using freeform optics," Opt. Express 26, 17578-17590 (2018)
DOI	<a href="https://doi.org/10.1364/OE.26.017578">10.1364/OE.26.017578</a>
Publisher	OPTICAL SOC AMER
Journal	OPTICS EXPRESS
Rights	© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement.
Download date	27/08/2022 13:30:55
Item License	<a href="http://rightsstatements.org/vocab/InC/1.0/">http://rightsstatements.org/vocab/InC/1.0/</a>
Version	Final published version
Link to Item	<a href="http://hdl.handle.net/10150/628650">http://hdl.handle.net/10150/628650</a>



# High-performance integral-imaging-based light field augmented reality display using freeform optics

HEKUN HUANG AND HONG HUA\*

3D Visualization and Imaging Systems Laboratory, College of Optical Sciences, University of Arizona, 1630 East University Boulevard, Tucson 85721, Arizona, USA

\*[hhu@optics.arizona.edu](mailto:hua@optics.arizona.edu)

**Abstract:** A new integral-imaging-based light field augmented-reality display is proposed and implemented for the first time, to our best knowledge, to achieve a wide see-through view and high image quality over a large depth range. By using custom-designed freeform optics and incorporating a tunable lens and an aperture array, we demonstrated a compact design of a light field head-mounted-display that offers a true 3D display view of 30° by 18°, maintains a spatial resolution of 3 arc minutes across a depth range of over 3 diopters, and provides a see-through field of view of 65° by 40°.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

**OCIS codes:** (120.2040) Displays; (120.2820) Heads-up displays; (330.7322) Visual optics, accommodation; (220.0220) Optical design and fabrication; (220.3620) Lens system design.

## References and links

1. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *J. Vis.* **5**(10), 7 (2005).
2. J. Wann, S. Rushton, and M. Mon-Williams, "Natural problems in the perception of virtual environments," *Vision Res.* **35**, 2731–2736 (1995).
3. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science* **265**(5173), 749–752 (1994).
4. G. E. Favalora, J. Napoli, D. M. Hall, R. K. Dorval, M. G. Giovinco, M. J. Richmond, and W. S. Chun, "100 million-voxel volumetric display," *Proc. SPIE* **4712**, 300–312 (2002).
5. A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360° light field display," *ACM Trans. Graph.* **26**(3), 40 (2007).
6. J. P. Rolland, M. W. Krueger, and A. Goon, "Multifocal planes head-mounted displays," *Appl. Opt.* **39**(19), 3209–3215 (2000).
7. K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.* **23**(3), 804–813 (2004).
8. S. Liu and H. Hua, "A systematic method for designing depth-fused multi-focal plane three-dimensional displays," *Opt. Express* **18**(11), 11562–11573 (2010).
9. X. Hu and H. Hua, "Design and Assessment of a Depth-Fused Multi-Focal-Plane Display Prototype," *J. Disp. Technol.* **10**(4), 308–316 (2014).
10. Y. Taguchi, T. Koike, K. Takahashi, and T. Naemura, "TransCAIP: A Live 3D TV system using a camera array and an integral photography display with interactive control of viewing parameters," *IEEE Trans. Vis. Comput. Graph.* **15**(5), 841–852 (2009).
11. H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," *Opt. Lett.* **26**(3), 157–159 (2001).
12. A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "An interactive 360° light field display," *ACM SIGGRAPH emerging technologies* **13**, (2007).
13. Z. Zhang, Z. Geng, M. Zhang, and H. Dong, "An Interactive Multiview 3D Display System," *Proc. SPIE* **8618**, 86180P (2013).
14. H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Opt. Express* **22**(11), 13484–13491 (2014).
15. W. Song, Y. Wang, D. Cheng, and Y. Liu, "Light field head-mounted display with correct focus cue using micro structure array," *Chin. Opt. Lett.* **12**(6), 060010 (2014).
16. D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.* **32**(6), 220 (2013).
17. A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs, "Pinlight displays: wide field of view augmented reality eyeglasses using defocused point light sources," *ACM Trans. Graph.* **33**(4), 1–11 (2014).
18. J. Hong, S. W. Min, and B. Lee, "Integral floating display systems for augmented reality," *Appl. Opt.* **51**(18), 4201–4209 (2012).

19. H. Huang and H. Hua, "An integral-imaging-based head-mounted light field display using a tunable lens and aperture array," *J. Soc. Inf. Disp.* **25**(3), 200–207 (2017).
20. J. Han, J. Liu, X. Yao, and Y. Wang, "Portable waveguide display system with a large field of view by integrating freeform elements and volume holograms," *Opt. Express* **23**(3), 3534–3549 (2015).
21. D. Cheng, Y. Wang, H. Hua, and M. M. Talha, "Design of an optical see-through head-mounted display with a low f-number and large field of view using a freeform prism," *Appl. Opt.* **48**(14), 2655–2668 (2009).
22. H. Huang and H. Hua, "Systematic characterization and optimization of 3D light field displays," *Opt. Express* **25**(16), 18508–18525 (2017).
23. J. Hughes, A. Dam, M. McGuire, D. Sklar, J. Foley, S. Feiner, and K. Akeley, *Computer Graphics: Principles and Practice* (Addison-Wesley, 2013).
24. S. Liu, D. Cheng, and H. Hua, "An Optical See-Through Head-Mounted Display with Addressable Focal Planes", *Proc. IEEE/ACM Int'l Symp. Mixed and Augmented Reality (ISMAR '08)*, 33–42 (2008).

## 1. Introduction

Conventional stereoscopic displays, which present a pair of stereoscopic images with binocular disparities and other pictorial depth cues on a fixed image plane to stimulate the perception of 3D space and shapes, lack the ability to correctly render focus cues, including accommodation and retinal blur effects. These displays thus force an unnatural decoupling of the accommodation and convergence cues and induce a fundamental problem sometimes referred to as vergence-accommodation conflict (VAC), which might lead to various visual artifacts such as distorted depth perception and visual fatigue [1,2]. In recent years, several display methods that are potentially capable of resolving the VAC problem have been demonstrated, including holographic displays [3], volumetric displays [4,5], multi-focal plane displays [6–9], and light field displays [10–18]. Among these methods, integral-imaging-based (InI-based) light field display allows the reconstruction of the full parallax of a 3D scene seen from a predesigned viewing window and have been conventionally demonstrated for its use in direct-view 3D display systems [10,11]. Because of its relatively low requirement upon the amount of hardware complexity compared to multi-view display systems involving either a mechanically rotating element [12] or projector array [13], it is possible to integrate the InI-based light field display with an optical see-through head-mounted-display (OST-HMD) system and create a wearable true 3D augmented-reality (AR) display. Some of the pioneering works have already demonstrated the promising potential of such architecture. Hua and Javidi presented a proof-of-concept monocular prototype system by integrating an off-the-shelf microlens array (MLA) for the creation of a micro-InI unit with a freeform eyepiece [14], while Song et al. presented a similar setup in which the MLA of the micro-InI unit was replaced with a pinhole array [15]. Both prototypes showed their abilities to render reconstructed 3D targets at their corresponding depths. However, alike other InI-based display and imaging technologies [10,11], the conventional micro-InI display method applied to HMD systems [14–16], depending on the display scheme adopted, suffers from several major limitations such as a narrow depth of field (DOF) for maintaining a decent spatial resolution of the 3D scene [13,14], or constant but low spatial resolution over a long DOF [15] due to the diffraction and defocus effect of the modulating element (MLA or pinhole array), and a small viewing window due to the crosstalk between the neighboring elemental images (EIs) on the display panel. For instance, the prototype system in our prior work demonstrated about a low spatial resolution of about 10 arc minutes per pixel in the visual space, low longitudinal resolution of about 0.5 diopters, a narrow DOF of about 1 diopter for a 10-arcminute resolution criteria, noticeable crosstalk within a 4-mm view window, and low view density which can barely affords 2 different views to fill a 3-mm eye pupil [14]. These limitations will inevitably compromise the viewing experience and it is paramountly important to overcome these limitations with innovative optical architecture.

Recently, using off-the-shelf components, we demonstrated a proof-of-concept prototype that improves the performance of an InI-based light field HMD by incorporating a tunable lens to extend the DOF without sacrificing the spatial resolution and an aperture array to reduce the crosstalk, or equivalently expand the viewing window [19]. Although the

prototype successfully demonstrated the potential capability of improving the optical performance of an In-HMD system, its off-the-shelf optics can only offer acceptable image quality of spatial resolution around 10 arc minutes over a very narrow field of view (FOV) of less than 15 degrees diagonal, is very bulky and thus not suitable for a wearable device, and is unable to support see-through view.

In this paper, we focus on developing new optical solutions that help to overcome the above-mentioned limitations of the state-of-the-art 3D light field displays based on integral imaging method and significantly improve the overall optical performance of this type of 3D displays. To achieve this goal, we propose a new optical architecture for a high-performance InI-based light field OST-HMD system, consisting of three key sub-units, a micro-InI unit integrated with a custom aspherical MLA and a custom aperture array, a tunable relay group that is able to dynamically tune the axial position of the reconstructed light field of a 3D miniature scene, and a custom freeform eyepiece (Sec. 2). We also proposed two new rendering methods (Sec. 3) that utilize the unique capability of our tunable relay optics to render the light field of a large depth volume from very near to very far to the viewer without compromising the spatial resolution. Combining the fully-optimized custom optics with our new light field rendering methods enables us to significantly improve the performance of InI-based 3D light field displays in terms of its spatial resolution, DOF, FOV, and crosstalk. Based on the proposed optical design architecture and custom design of the optical system, we experimental demonstrated a binocular, wearable prototype system that is capable of rendering a true 3D light field at a constant spatial resolution of 3 arc minutes across a very large depth range of over 3 diopters (from optical infinity to as close as 30 cm to the viewer) and removed crosstalk with an eye box around than 6mm by 6mm in an OST-HMD system (Sec. 3). The prototype, with a total weight of about 450 grams and with a volume of about 210mm (width) by 80mm (depth) by 40mm (height), not only maintains high quality imagery across a depth range of over 3 diopters for the virtual display path but also achieves high visual resolution and wide FOV for the see-through view. To our best knowledge, there is no prior art that has ever demonstrated a true 3D AR display using integral imaging with comparable optical performance and form factor.

## 2. System design and specifications

Figure 1(a) shows the cross-section view of the display path of the optical system, which mainly consists of three key parts, a micro-INI unit, a tunable relay group, and a freeform waveguide-like prism eyepiece. The micro-InI unit, including a high-resolution microdisplay, a custom-designed aspherical MLA, and a custom aperture array, renders the 3D light fields of a reconstructed scene. Specifically, a set of 2D EIs, each representing a different perspective of a 3D scene, are displayed on the high-resolution microdisplay (e.g. the three red rays of point *A* represent the chief rays from three EIs). Through the MLA, each EI works as a spatially-incoherent object and the conical ray bundles emitted by the pixels in the EIs intersect and integrally create the perception of a 3D scene (e.g. point *A*) that appears to emit light and occupy the 3D space. By changing arrangement of the EIs, targets at different depth can thus be rendered. The aperture array on the other hand helps reduce the crosstalk for any EI by blocking the rays from its neighboring EIs entering into its corresponding lenslet in the MLA. The tunable relay group, mainly made of 4 stock spherical lenses with an Optotune EL-10-30 tunable lens sandwiched inside, relays and adjusts the axial positions of the intermediate images of the reconstructed targets for DOF extension without compromising spatial resolution. The waveguide-like prism formed by 4 freeform surfaces denoted as S1 to S4, magnifies the intermediate images of the reconstructed targets and projects the light toward the exit pupil, or the viewing window, at which a viewer sees the magnified 3D scene reconstruction. The light transmittance of the display path maintains as high as 50% in the current implementation because the system is only subject to a 50% loss at S4 of the freeform prism as the result of the see-through optical combiner with 50/50 coating. It could be

potentially improved by adjusting the coating. It is worth pointing out several factors contributed to our choice of a freeform waveguide-like eyepiece for our design. First of all, freeform prism eyepieces have been demonstrated with much higher optical performance over the existing holographic or geometric waveguides [20], while the holographic or geometric waveguides typically function as a pupil expander and are subject to compromised image quality due to stray light, light leakage, or color artifacts. Secondly, holographic and geometric waveguides typically require a collimated source to be coupled in, which is not compatible with our design of an InI-based light field display without additional optics. Finally, though not as compact as holographic or geometric waveguides, freeform prism eyepieces are much more compact than other conventional eyepiece designs.

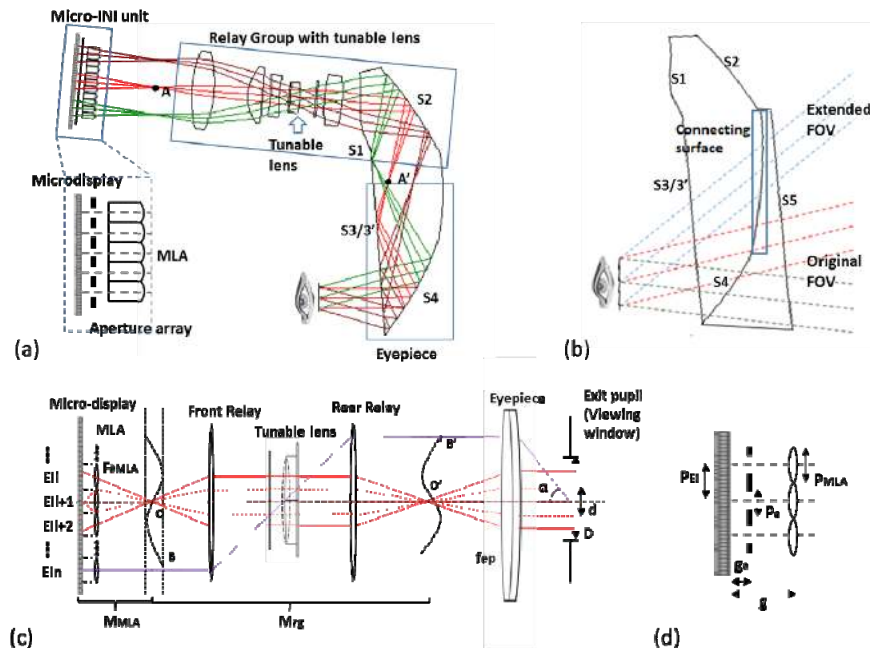


Fig. 1. The optical layout of (a) the display path, (b) the see-through path of the proposed design, (c) unfolded display path with parameters labeled and (d) micro-INI unit with parameters labeled.

Unlike most of the previous freeform eyepiece designs where the freeform prism solely acts as the function of eyepiece for magnifying a 2D image [17,18,21], in the proposed design we integrate part of the rear relay lenses with the eyepiece as a whole piece. Under such circumstances, S1 and S2 can be thought of being a part of the relay group and the intermediate images (e.g. point  $A'$ ) of the reconstructed targets (e.g. point  $A$ ) will be formed inside the prism, which will then be magnified through multiple reflections and refraction by S3 and S4 to create virtual images of the reconstructed targets overlaying the see-through view. Such a configuration significantly lowers the total number of the optical elements needed for the relay group, which helps to reduce the complexity of mechanical mount of the whole system, and improves the overall optical performance. More importantly, by extending the overall folded optical path inside of the freeform prism, the see-through FOV could be greatly expanded. Figure 1(b) shows the cross-section view of the see-through path of the HMD with a matching compensator attached to the aforementioned prism. By properly generating the connecting surface between S2 and S4, we can achieve an undistorted, high-fidelity see-through view that almost doubles the horizontal FOV, when compared to the

conventional prism design simply using S4 as the see-through window, without increasing the overall thickness of the prisms.

Figure 1(c) further shows the unfolded optical layout for the display path in which we label the key parameters of the system. As illustrated in Fig. 1(c),  $d$  stands for the footprint diameter of the ray bundle from a single pixel on the microdisplay projected on the viewing window through the display optics. By setting the viewing window at the back focal length of the eyepiece, its paraxial approximation is expressed by Eq. (1) as:

$$d = \frac{f_{ep}}{(M_{MLA} + 1)F_{\#MLA}M_{rg}}, \quad (1)$$

where  $f_{ep}$  is the focal length of the eye piece, and  $F_{\#MLA}$  is the F-number of the lenslet in the MLA.  $M_{MLA}$  is the lateral magnification of the lenslet in the MLA and is expressed as  $f_{MLA}/(g - f_{MLA})$  where  $g$  is the gap between the microdisplay and the MLA, and  $f_{MLA}$  is the focal length of the lenslet.  $M_{rg}$  is the lateral magnification of the tunable relay group including both the relay lens and the tunable lens. The view density,  $\sigma_{view}$ , of a light field display can then be obtained by calculating the reciprocal of the footprint area of a view projected on the view window and expressed by Eq. (2) as,

$$\sigma_{view} = \frac{4}{\pi d^2}. \quad (2)$$

Also,  $D$  stands for the footprint dimension of all the rays from a reconstructed point (e.g. point  $O'$ ) and defines the dimension of the viewing window or eyebox of the display in which the light field of a 3D reconstructed scene can be observed. It can be obtained by integrating the ray bundles from the different EIs corresponding to the reconstructed point at CDP and estimated as

$$D = N \cdot d = M_{MLA} \cdot d, \quad (3)$$

where  $N$  is the number of views in either horizontal or vertical direction used for reconstructing a 3D point, which equals to the lateral magnification of the lenslet in the MLA on the CDP.

By placing the tunable lens at the back focal length of the front relay lens and making it optically conjugate to the exit pupil of the eyepiece, the compound optical power of the relay group is maintained constantly independent of the optical power of the tunable lens due to constant chief ray directions owing to the property of object-space telecentricity. Consequently, varying the optical power of the tunable lens leads to a lateral displacement of the intermediate images of the reconstructed 3D scene, but the virtually reconstructed 3D scene through the eyepiece maintains a constant field of view regardless of the focal depths of the scene.  $\alpha$  stands for the half FOV of the display system, which corresponds to the chief ray angle of the center pixel of the edge EI through the display optics with respect to the viewing window (e.g. point  $B'$ ). Under the above-mentioned telecentricity condition and setting the viewing window at the back focal length of the eyepiece,  $\alpha$  can be estimated by Eq. (4) as,

$$\alpha = \tan^{-1} \left[ \frac{h_{MD} M_{rg}}{f_{ep}} \right], \quad (4)$$

where  $h_{MD}$  is the half-height of the microdisplay in its diagonal direction corresponding to the center of the edge EI. Finally, under the assumption that all the elemental views of a reconstructed point are perfectly aligned and converged, the spatial resolution,  $\eta$ , of the proposed system on the virtual CDP could be calculated by the angular separation of a single pixel in the visual space through the display optics with respect to the viewing window and is estimated by Eq. (5) as

$$\eta = \tan^{-1} \left[ \frac{p_{MD} M_{MLA} M_{fg}}{f_{ep}} \right], \quad (5)$$

where  $p_{MD}$  is the pixel pitch of the microdisplay.

On the other hand, as shown in Fig. 1(d), in order to make the aperture array function as designed to block only the crosstalk from the neighboring elemental view, the location of the aperture array and its open aperture needs to be constrained by Eqs. (6) and (7), respectively,

$$g_a \leq g_{a-\max} = \frac{g p_{EI}}{p_{EI} + p_{MLA}}, \quad (6)$$

$$p_a \leq p_{a-\max} = \frac{(g_{a-\max} - g_a) p_{EI}}{g_{a-\max}}, \quad (7)$$

where  $g_a$  and  $p_a$  are the real gap between the aperture array and the microdisplay and open aperture size for each of the apertures in the aperture array, respectively, while  $g_{a-\max}$  and  $p_{a-\max}$  are the maximum allowable gap and open aperture size, respectively,  $p_{EI}$  is the dimension of the elemental image, and  $p_{MLA}$  is the pitch of the MLA.

Based on the optical layout in Fig. 1 and the analytical relationships described above, we designed a fully custom prototype system. The key specifications of the prototype system are summarized in Table 1. As discussed in [22], one of the key parameters to determine in our design process is the view density of the system defined as the number of views per unit area on the entrance pupil of the eye in order to characterize the view sampling property of a light field 3D display. The effects of the view density on the retinal image quality, DOF, and the eye accommodative response of a light field display were systematically characterized in [22]. For instance, a light field 3D display constructed with a view density of  $0.57\text{mm}^{-2}$  and a fill factor of 1, which corresponds to a view pitch of 1.5mm at the exit pupil of an HMD, offers a total of 4 elemental views over a 3-mm eye pupil. As suggested in [22], without considering the resolution limit of a microdisplay and eye accommodation error, such a display can yield a DOF of approximately  $\pm 1$  diopters to achieve a limiting spatial resolution of 1.2 arc minutes or threshold spatial frequency of 25 cycles/degree. In the meantime, when the reconstruction depth is shifted away from the CDP of the light field 3D display by 1 diopter, the eye accommodation error, which is defined as the axial displacement between the actual eye accommodative distance and the depth dictated by the rendering to characterize the residual VAC conflict in a light field 3D display, can be as large as 0.2 diopters for targets with a spatial frequency of 5 cycles/degree or lower [22]. Although a larger view density potentially offers a larger DOF and smaller accommodation error, it comes at the cost of a lower image resolution and also lower image contrast due to increasing diffraction effects. Based on the trade-off analysis in [22], we firstly set the view density of our light field 3D system to be  $0.38\text{mm}^{-2}$ , which corresponds to a total of 2 by 2 views encircled by a 3.5 mm pupil, to obtain a good balance among the factors of the limiting spatial resolution, DOF, and eye accommodation error. To further allow for the eye movement in viewing the display, a larger viewing window of 6mm by 6mm was created. As a result, a total of 3 by 3 views will be rendered by the prototype system for each point of a 3D scene, though at any time about 4 views are received by a 3.5mm eye pupil.

The remaining optical specifications are primarily driven by the choices of the view density and view window size described above. For instance, the MLA, consisting of 17x9 lenslets, was designed to offer an equivalent focal length of 3.5mm, a lens pitch of 1mm, and a transverse magnification of 3 which equals to the number of views in one direction. When combined with a 0.7" Sony organic light emitting display (OLED) offering an 8- $\mu\text{m}$  pixel pitch and a total of 1920x1080 color pixels, each lenslet of the MLA renders an EI consisting of 125x125 color pixels and yields a spatial resolution of about 24 $\mu\text{m}$  for the micro-InI unit.

When the micro-InI unit is combined with the designed relay group of 24mm focal length and the eyepiece of 27.5mm focal length, each EI yields an FOV of  $6.25^\circ$  by  $6.25^\circ$  overlaying with 3 by 3 adjacent EIs including itself, and all of the EIs resulted from the 17x9 MLA yield a total FOV of about  $30^\circ$  by  $18^\circ$ . In other words, the display renders a total of 3 by 3 views for each point of the reconstructed scene across the display FOV of  $30^\circ$  and  $18^\circ$ . Furthermore, the cutoff resolution of the system, corresponding to the visual angle of one pixel on the microdisplay, is around 3 arc minutes (arcmins) in visual space. As later demonstrated, the system is capable of maintaining a spatial resolution of 3 arcmins across a depth range of over 3 diopters, which is a significant improvement over the state-of-the-art [14–17,19]. To further improve the spatial resolution of the system while maintaining the same or achieving even larger FOV or more views, a microdisplay with smaller pixel pitch and more pixels or a larger display panel with an eyepiece of lower optical magnification is desired. For example, to fully match the fovea resolving ability of human eye which is around 1 arcmin, a microdisplay with the same panel size of 0.7" but offering a  $2.67\mu\text{m}$  pixel pitch, or one offering the same pixel pitch of  $8\mu\text{m}$  pixel but with a panel size of 2.1" combined with an eyepiece of focal length around 10mm, or one with the specifications in between of previous two examples satisfying the trade-off between the FOV and spatial resolution is needed.

Table 1. Specifications of the System

Parameters	Value
Display FOV	$\sim 30^\circ(\text{H}) * 18^\circ(\text{V})$
See-through FOV	$\sim 65^\circ(\text{H}) * 40^\circ(\text{V})$
Size of viewing window	6mm(H)*6mm(V)
Display MTF	$>0.2 @ 3\text{arcmins}$
Display DOF	$\sim 3$ diopters
See-through MTF	$>0.2 @ 0.5\text{arcmin}$ for central $25^\circ(\text{H}) * 20^\circ(\text{V})$ ; $>0.2 @ 3\text{arcmins}$ for full FOV
Microdisplay	Sony 0.7" OLED panel, 8um pixel pitch, 1920(H)*1080(V)
MLA	F/3.5, 1mm pitch, 17(H)*9(V)
Magnification of MLA	3
Aperture array	1mm pitch, 0.6mm open aperture
Relay group	24mm focal length for both front and rear group
Tunable lens	Optotune EL-10-30
Freeform eyepiece	27.5mm focal length

The optical system design of the proposed InI-HMD system was very challenging due to not only the complexity of the optics layout but also due to the requirements for optimizing the optical performance of light field rendering across a large FOV, depth range, and viewing window. In order to achieve a high performance of the display path, the micro-InI unit should be optimized along with the relay group and the eyepiece. Under such circumstances, the whole design needs to be split into a series of sub configurations, each of which represents the ray paths of different sampled field points from a single EI on the microdisplay passing through the corresponding lenslet in the MLA as well as the remaining optical elements.

The multi-configuration design will not only largely change the forms of the constraints but also inevitably increase the overall complexity of the optical design due to the large number of rays traced. A ‘divide and conquer’ scheme therefore was adopted, where we first designed the micro-InI unit separately from the relay group and the eyepiece, and then optimized the integrated system together by fine-tuning the freeform surfaces. To maximize the spatial resolution of the MLA without adding optical elements, the surfaces of each lenslet in the MLA was set as asphere up to 6th order and repeated the lens shape for the entire MLA. At this stage, therefore, only one of the lenslets needs to be set up and optimized. For the relay group, as mentioned above, a part of it was incorporated with the eyepiece, leaving only one lens between the tunable lens and the freeform prism for aberration control, mainly for the chromatic ones. Each of the freeform surfaces was set as XY-polynomials up



to 10th order with symmetry respect to the horizontal direction. After obtaining designs with the reasonable performance for the different parts through the initial optimizations, we then combined the micro-InI unit, the relay group, and the eyepiece together by configuring a number of zooms in each of which the field points and lenslet of the MLA were decentered accordingly to properly model the light paths of the elemental views. The integrated system was then optimized as a whole by fine-tuning mainly the freeform surfaces of the eyepiece. As a result, Fig. 1 shows the optical layout of the optimization results. The modulation transfer function (MTF) of the virtual display path evaluated for any of the EIs is above 0.2 at the Nyquist frequency of the micro-InI unit (3 arcmins in visual space) for the full display field. Also, the prototype system yields a see-through view with an MTF above 0.2 at the spatial frequency of 60 cycles/degree for the central  $25^\circ$  by  $20^\circ$  area and of 10 cycles/degree for full see-through field of  $65^\circ$  by  $40^\circ$ , with a distortion smaller than 1% for full see-through field. The detailed test results for the display and see-through paths will be shown in the following section.

### 3. Prototype and test results

Figure 2(a) shows an integrated prototype of a monocular system according to the design layout shown in Fig. 1, Fig. 2(b) shows a 3D model of a binocular system fitting to an average-size human head model, and Fig. 2(c) shows a photo of our integrated binocular prototype. A tolerance analysis indicates that the design itself has relatively low demand upon the optomechanics, so most of the mechanical mounts were 3D printed with stereolithography technology. An aperture array was printed on a paper-like thin film and carefully attached to the MLA. The freeform eyepiece, the compensator and the 3D printed lens cell were clamped by customized metal pieces. Most of the weight is contributed from the metal clamps, off-the-shelf tunable lens assembly, and the stock glass lenses used in our relay lens group. The overall size as well as the total weight of the proposed system could be significantly reduced by optimizing the mechanic design and customizing most of the optical elements such as the tunable lens and stock lenses.

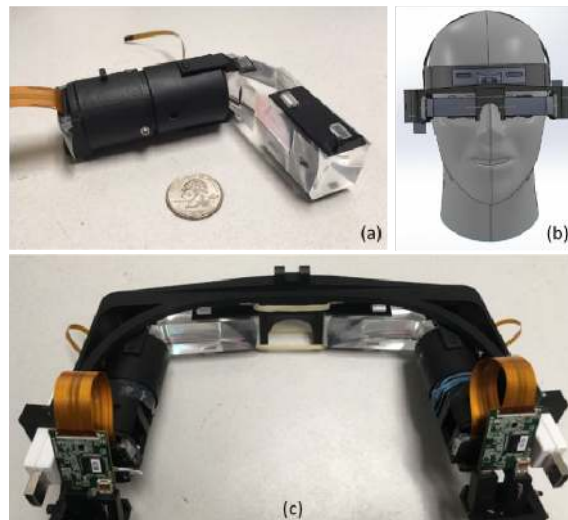


Fig. 2. (a) The image of bench-top prototype with a quarter coin; (b) the 3D model of the binocular system worn on human head; and (c) the photograph of an integrated binocular prototype.

To render the light field of a 3D target scene, we start with creating a virtual 3D scene. For the convenience of reference, we place a virtual camera at the position corresponding to the eye position of a viewer and reference the depths of the scene objects with respect to the

viewer or the virtual camera. Each 2D EI of the 3D light field rendering represents a slightly different perspective of the 3D scene seen by the virtual camera, which is achieved by pointing the viewing axis of the virtual camera in a direction matching the corresponding elemental view. Following the conventional rendering pipeline of 3D computer graphics [23], an array of 15x9 elemental images of a 3D target scene are simulated, each of which consists of 125x125 color pixels. These EIs are then mosaicked to create the full-resolution image of 1920x1080 pixels for the microdisplay. During the process of rendering the 2D EIs, the projection plane of the virtual camera is set to coincide with the virtual CDP of the display optics seen by the viewer, which is the optical conjugate plane of the microdisplay in the visual space and where usually the highest contrast of the 3D light field could be achieved. As demonstrated below, the depth of the virtual CDP may vary under the different rendering modes of the display. Therefore, the projection plane of the virtual camera is reconfigured accordingly.

To demonstrate the optical performance of the light field display, a virtual 3D target scene consisting of three depth planes located at 3, 1 and 0.5 diopters away from the viewing window was created. On each depth plane three groups of Snellen letter 'E's with different spatial resolutions (3, 6, and 10 arcmins for the individual strokes or gaps of the letters) and orientations (horizontal and vertical) as well as the depth indicators ('3D', '1D' and '0.5D') were rendered. Figure 3(a) shows the mosaic of 11x5 EIs of the virtual 3D target generated for the microdisplay where the virtual CDP was set at 1 diopter.

We started the demonstration by fixing the optical power of the tunable lens so that the CDP of the display system was set at a fixed distance of 1 diopter to the viewer, which simulates the display properties of a conventional InI-based HMD. For qualitative assessment of focus cues, three spoke resolution targets were physically placed at the corresponding depths of three depth planes of the virtual 3D target. A camera with a 2/3" color sensor of 2448x2048 pixels and a 16mm lens was used. The camera system overall yields a spatial resolution of 0.75 arcmins per pixel, which is substantially better than that of the display optics. The entrance pupil diameter of camera lens was set to about 4mm such that it is similar to that of the human eye. Figure 3(b) shows the captured images of the reconstructed virtual 3D target overlaying with the real-world targets where the camera was focusing on 1 diopter. It can be observed that only the targets, both the real (indicated by the arrow) and virtual (indicated by the box) ones, located at the same depth of the focus plane of the camera are correctly and clearly resolved, which suggests the ability of the InI-based HMD to render correct focus cues to the viewer. The ability to resolve the smallest Snellen letters on the top row of the 1 diopter targets further suggests the spatial resolution of the prototype matches with the designed nominal resolution of 3 arcmins. In this configuration of fixed lens focus, it can be further observed that the EIs of the virtual targets at the depths (e.g. 3D and 0.5D) different from the focus plane of the camera do not converge properly, causing multiple copies of the letters being captured in Fig. 3(b). These targets can properly converge when the camera focus is adjusted to focus on their corresponding depths, as demonstrated in Figs. 3(c) and 3(d), which show the captured images of the same virtual and real-world scene with camera being focused at 0.5 and 3 diopters, respectively. The targets corresponding to the camera focus depth were marked by a yellow and blue box, respectively. However, unlike a traditional InI-based HMD, the image contrast and resolution of the targets reconstructed at the depth plane other than the CDP can only maintain in a relatively short, limited DOF and degrade severely beyond that, even though the EIs of these targets converge correctly and located at the same depth as the focus plane of the camera. For instance, the captured images in Fig. 3(c) can still resolve the letters corresponding up to 6 arcmins while that in Fig. 3(d) can only resolve the letters corresponding to 10 arcmins and the EIs start to converge improperly.

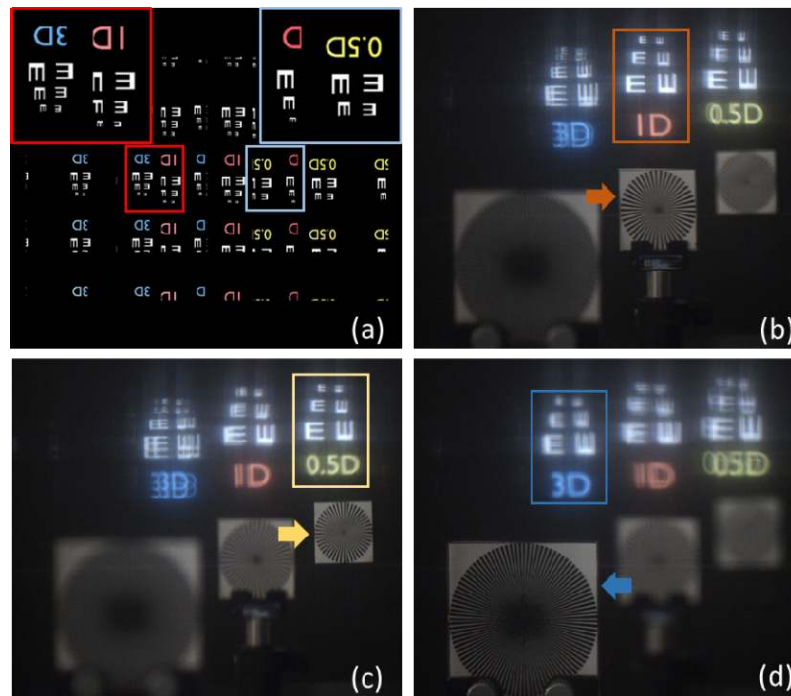


Fig. 3. (a) Array of the EIs on microdisplay and captured image of both real and virtual targets with the camera focusing on (b) 1 diopter, (c) 0.5 diopters and (d) 3 diopters, respectively (see [Visualization 1](#) for video demonstration).

With the assistance of tunable lens, the depth of the CDP can be dynamically adjusted. This capability allows the system to operate in two different modes: vari-depth mode and time-multiplexed multi-depth mode. In the vari-depth mode, the depth of the CDP is adaptively varied according to the average depth of the displayed contents or the depth of interest. In multi-depth mode, the power of the tunable lens is rapidly switched among several states corresponding to several discrete CDP depths, while in synchronization the light field rendering is updated at the same speed such that the contents of different depth are time-multiplexed and viewed as an extended volume if the switching occurs at flickering-free rate. For the purpose of demonstrating the vari-depth mode, we re-rendered the EIs for targets at 3 and 0.5 diopters with the CDP adjusted to match the depth of 3 diopters instead of 1 diopter in Fig. 3. We also varied the optical power of the tunable lens so that the CDP of the display system matches with its adjusted depth. Figures 4(a) and 4(b) show the captured images through the HMD with the camera focused at the depth of 3 and 0.5 diopters, respectively. By correctly adjusting the optical power of the tunable lens as well as regenerating the contents on the microdisplay, the system is able to maintain the same level of the spatial resolution of 3 arcmins and image quality for the targets located at the depth of 3 diopters, shown in Fig. 4(a), as for the targets located at 1 diopter in Fig. 3(b). The vari-depth mode, however, only achieves high-resolution display for targets near the specific depth dictated by the CDP of the display hardware. As shown in Fig. 4(b), the targets at the depth of 0.5 diopters show more severely degraded resolution than in Fig. 3(c) due to its increased separation from the given CDP, even when the camera is focused at the depth of these 0.5-diopter targets.

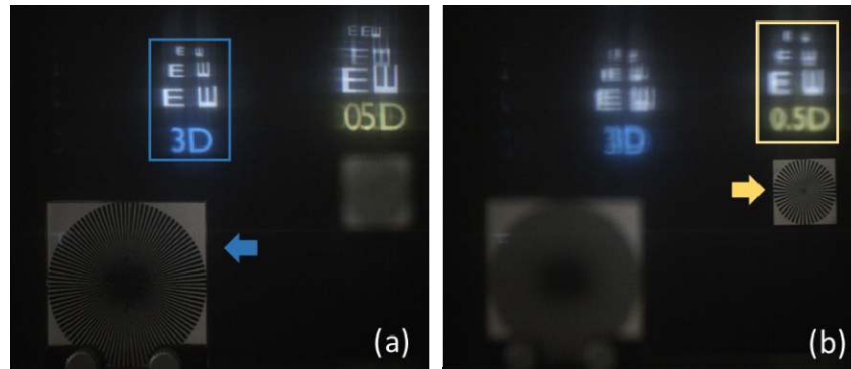


Fig. 4. Captured images of both real and virtual targets in vari-depth mode with the CDP set at 3 diopters while the camera focusing on (a) 3 diopters and (b) 0.5 diopters, respectively (see [Visualization 1](#) for video demonstration).

To further demonstrate the multi-depth mode, we re-rendered the EIs for the two targets at 3 and 0.5 diopters separately with the CDP adjusted to match the depth of the corresponding objects. The separately-rendered EIs were displayed in a time-multiplexing fashion at a frame rate of about 30Hz while in synchronization the CDP of the display was rapidly switch between the depths of 3 and 0.5 diopters. The refresh speed of 30Hz is due to the limit of the highest 60Hz refresh rate of the OLED microdisplay. Figures 5(a) and 5(b) show the captured images through the HMD with the camera focused at the depth of 3 and 0.5 diopters, respectively. Along with the virtual display, two spoke resolution targets were physically placed at the corresponding depths of the letters. As shown in Fig. 5(a), when the camera was focused at the near depth of 3 diopters, both of the virtual and real objects at the near depth (the letters and the spoke on the left) appears to be in sharp focus, while the far objects (the letters and the spoke on the right) show noticeable out-of-focus blurring as expected. Figure 5(b) demonstrates the case when the camera focus was switched to the far depth of 0.5 diopters. It can be clearly observed that both of the letters at far and near depths are comparably sharp at the corresponding focus of the camera. By driving the display in a dual-depth mode, the system achieves high-resolution displays of targets with a large depth separation of nearly 3 diopters while rendering focus cues comparable to their real counterparts.

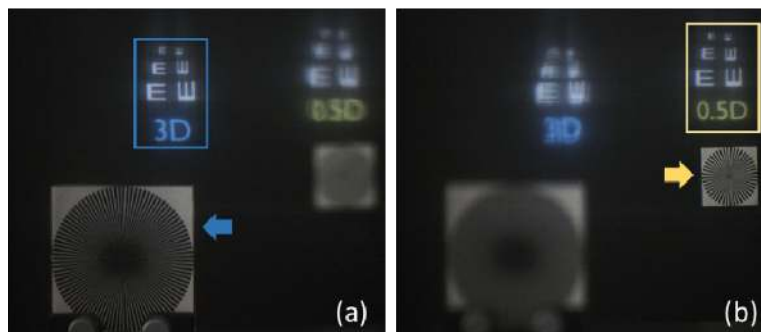


Fig. 5. Captured images of both real and virtual targets in multi-depth mode with the camera focusing on (a) 3 diopters and (b) 0.5 diopters (see [Visualization 1](#) for video demonstration).

The vari-depth and multi-depth modes of the InI-based light field rendering method share some similarity to the conventional vari-focal [24] and multi-focal plane HMD works [6–9] in the sense that the depth of the CDP is either adaptively varied according to the depth of interest in the vari-depth mode or is rapidly switched among several discrete depths in the

multi-depth mode. However, their visual effects and implications on focus cues are noticeably different. For instance, as demonstrated in Fig. 4, in the vari-depth mode of an InI-HMD, the contents away from the CDP are rendered with correct blurring cues, though in potentially degraded resolution, due to the nature of light field rendering, while in a conventional vari-focal HMD the contents away from its focal plane can be as high resolution as the contents on the focal depth unless artificially blurred but do not show proper focus cues due to its 2D rendering nature. In the multi-depth mode, a significant advantage over the traditional multi-focal plane HMD approach is the requirement of much less number of depth switch to render correct focus cues in the same depth range, while depth blending is necessary in a multi-focal system to render focus cues for contents away from the physical focal planes. As demonstrated in [9], 6 or even more focal planes and their associated blending functions are needed to extend the depth range of 3D volume to 3 diopters, which is a huge overtake on display speed, tunable optics speed, and graphics rendering speed. On the other hand, in the case of InI-based light field rendering, covering a depth range of 3 diopters only requires 2 focal depth and the focus cues generated in this case are also more accurate and continuous, as suggested in [22].

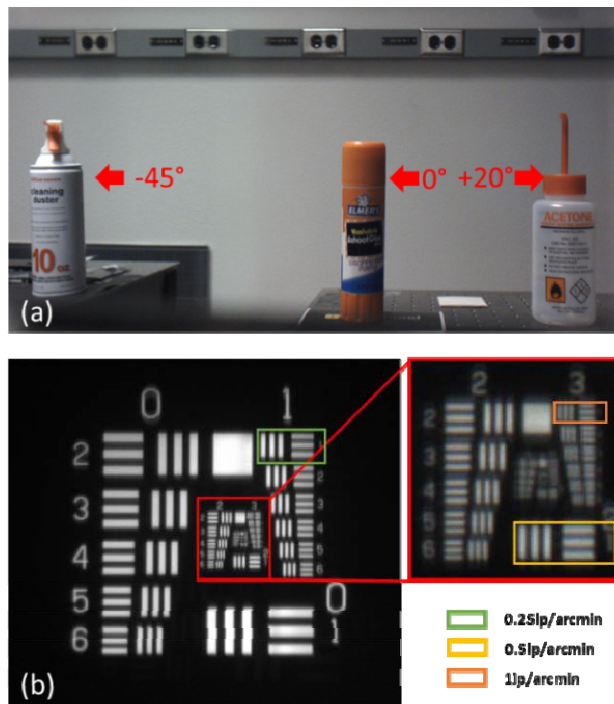


Fig. 6. Captured images of the see-through view by (a) short focal length lens and (b) long focal length lens.

Figures 6(a) and 6(b) show the celmercaptured images of the real-world scene through the HMD. Figure 6(a) was taken by a short focal length camera lens (8mm) to capture of the entire see-through view, while Fig. 6(b) was taken by a long focal length camera lens (50mm) that was focusing on a 1951 USAF resolution test chart to demonstrate the resolving limitation of the see-through view. The HMD apparently supports a large horizontal see-through FOV from  $20^\circ$  (inwards) to  $-45^\circ$  (outwards) and a vertical FOV of  $\pm 20^\circ$  without significantly dropping the viewing quality. It can resolve fine details up to 0.5 arcmins for the centered  $25^\circ$  by  $20^\circ$  region of the see-through FOV that outperforms the fovea visual acuity of a 20/20 normal vision observer. These results clearly demonstrated that little degradation and negligible distortion on the real-world view is introduced by the see-through optics.

#### 4. Conclusion

In conclusion, we proposed a novel design of an InI-based light field HMD system using freeform optics and tunable lens for optical performance improvements and an aperture array for crosstalk reduction. Based on the new architecture and optimization method, we demonstrated a prototype system that offers a true 3D display FOV of 30° (horizontal, H) and 18° (vertical, V), maintains a spatial resolution of 3 arcmins over a depth range of 3 diopters, and provides a see-through FOV of 65° (H) by 40° (V).

#### Funding

National Science Foundation (grant award 14-22653).

#### Acknowledgment

Dr. Hong Hua has a disclosed financial interest in Magic Leap Inc. The terms of this arrangement have been properly disclosed to The University of Arizona and reviewed by the Institutional Review Committee in accordance with its conflict of interest policies.