



Generalized methods and strategies for modeling and optimizing the optics of 3D head-mounted light field displays

HEKUN HUANG  AND HONG HUA* 

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

*hhua@optics.arizona.edu

Abstract: An integral-imaging based light field head-mounted display, which typically renders a 3D scene by reconstructing the directional light rays apparently emitted by the scene via an array optics, is potentially capable of rendering correct or nearly correct focus cues and therefore solving the well-known vergence-accommodation conflict problem plaguing conventional stereoscopic displays. Its true 3D image formation nature, however, imposes significant complications and the well-established optical design process for conventional head-mounted displays becomes inadequate to address the design challenges. To our best knowledge, there are no existing methods or framework that have been previously proposed or demonstrated to address the challenges of modeling and optimizing an optical system for this type of display systems. In this paper, we present novel and generalizable methodology and framework for designing and optimizing the optical performance of integral-imaging based light field head-mounted displays, including methods of system configurations, user-defined metrics for characterizing the performance of such systems, and optimization strategies unique in light field displays. A design example is further given based on the proposed design methodology for the purpose of validating the proposed design method and framework.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Conventional stereoscopic three-dimensional displays (S3D) stimulate the perception of 3D spaces and shapes from a pair of two-dimensional (2D) perspective images at a fixed distance, one for each eye, with binocular disparities and other pictorial depth cues of a 3D scene seen from two slightly different viewing positions. A key limitation to the S3D-type displays is the well-known vergence-accommodation conflict (VAC) problem, which stems from the inability to render correct focus cues, including accommodation and retinal blur effects, for 3D scenes [1,2]. It causes several cue conflicts and is considered as one of the key contributing factors to various visual artifacts associated with viewing S3D displays [3,4].

In recent years, several display methods that are capable of rendering focus cues and thus potentially overcoming the VAC problem have been demonstrated [5], including holographic displays [6], volumetric displays [7,8], vari-focal-plane displays [9–12], multi-focal-plane displays [13–16], and light field displays [17–22]. Among these approaches, an integral-imaging-based (InI-based) light field 3D (LF-3D) display allows the reconstruction of a 3D scene by rendering the directional light rays apparently emitted by the scene via an array optics seen from a predesigned viewing window [17]. Due to its relatively low requirements upon the amount of hardware complexity, it is possible to implement the InI-based LF-3D method in a head-mounted-display (HMD) system and create a wearable InI-HMD that can potentially render correct focus cues. Some of the pioneering works have already demonstrated the promising potential of such an architecture. Lanman and Luebke proposed a near-eye light field display by placing a microdisplay and microlens array (MLA) in front of viewer's eye [19]. Such a direct-view configuration yields

a compact and thin-profile package, but due to the limited degrees of freedom provided, it suffers from relatively poor image quality of the reconstructed 3D scene. Hua and Javidi proposed an alternative architecture for an optical see-through system by combining a micro-InI unit with a freeform magnifying eyepiece to enable see-through capability and improve the overall depth of reconstruction and image quality [20]. Song et al. presented a similar setup in which the MLA of the micro-InI unit was replaced with a pinhole array [21]. Such a magnified-view architecture, however, still suffers from several major limitations such as a narrow depth of field (DOF) for maintaining a decent spatial resolution of the 3D scene [20] or constant but low spatial resolution over a long DOF [21], and a small viewing window due to the crosstalk between the neighboring elemental images on the display panel. Recently, we further improved the magnified-view architecture by incorporating an electrically tunable lens to extend the DOF without sacrificing the spatial resolution and utilizing an aperture array to reduce the crosstalk [23]. The prototype successfully demonstrated the potential capability of improving the optical performance of an InI-HMD, but its off-the-shelf optics can only offer an overall spatial resolution of around 10 arc minutes over a very narrow field of view (FOV) of less than 15 degrees diagonally, is very bulky and thus not suitable for a wearable device.

Despite the aforementioned pioneering works, there lacks a systematic investigation for optical design methods suitable for the optimization of InI-HMDs. Designing a high-performance InI-HMD system remains to be a challenge. To effectively address the VAC problem, an InI-HMD system requires that different elemental views created by multiple elements of an MLA are rendered to be seen through each of the eye pupils [24]. Therefore, the light rays emitted by multiple spatially-separated pixels on these elemental views are received by eye pupil and integrally sum up to form the perception of a 3D reconstructed point, which essentially is the key difference of an InI-HMD from a conventional HMD. Due to this inherent difference of the image formation process, the well-established optical design methods for conventional HMDs [16,25] become inadequate for designing a true 3D InI-HMD system. A new design method that is able to (1) set up a 3D InI-HMD design to precisely execute real ray tracing, and (2) optimize the design to precisely sample and render the light field of a reconstructed 3D scene, which is key to drive the accommodation status of the viewer's eye and thus solve the VAC problem, is hence required. Without a well-optimized design, an InI-HMD will not be able to correctly render the depth and accommodation cue of the reconstructed 3D scene but yield images of compromised quality and comfort.

In this paper, we concentrate on the various critical aspects described above and propose a novel methodology for designing a high-performance InI-HMD. Along with a design example and experimental demonstration of an InI-HMD system presented under "Result", we summarize the basic image formation process of an InI-HMD, the method of obtaining the key relationships among the system specifications and the process of configuring an InI-HMD system in optical design software, and the optimization method and process to achieve minimal aberration-induced light field rendering taking both positional and directional sampling into consideration. Although the design example presented in the paper assumes a magnified-view configuration, the method can be easily adapted for the design of InI-HMDs based on a direct-view configuration without a magnifying eyepiece and other types of LF-3D systems.

2. Modeling image formation process of InI-HMD

The light field of a 3D scene can be represented by the well-known 4-D light field function [26], $L(s, t, u, v)$, which characterizes the radiance of a light ray as a function of a ray position (s, t) and direction (u, v) . An InI-HMD reconstructs the 4-D light field of a 3D scene by angularly sampling the directions of the light rays apparently emitted by the 3D scene. As illustrated in Fig. 1(a), in an InI-HMD an array of 2D elemental images, each of which represents a different perspective of a 3D scene, is rendered on a microdisplay (e.g. $A1$ to $A3$ for a reconstructed point

A , $B1$ to $B3$ for a reconstructed point B). Each pixel on these elemental images is considered as the image source defining the positional information, (s, t) , of the 4-D light field function. Associated with the array of elemental images is an array optics, such as an MLA, each of which defines the directional information, (u, v) , of the light field function. To reconstruct the light field of a 3D point the ray bundles emitted by multiple pixels, each on a different elemental image are modulated by their corresponding MLA elements to intersect at the 3D position of reconstruction. As a result, light rays from these spatially separated pixels integrally create the 3-D point that appears to emit light in different directions. In a magnified-view configuration, an eyepiece is inserted to further magnify the miniature 3D scene into a large 3D volume with an extended depth in virtual space (e.g. A' and B'). When the eye is accommodated at the depth of the reconstructed point A' , as shown in Fig. 1(a), the rays from its corresponding elemental pixels will overlap with each other and naturally form a sharply focused image on the retina, while the rays reconstructing point B' , which are located at a different depth from point A , will be spatially displaced from each other and create a retinal blur varying with the difference between the depths of reconstruction and eye accommodation. As illustrated in Fig. 1(b), when the eye accommodation depth is switched to point B' , the retinal image of point B' becomes in-focus while the retinal image of point A' become blurry. Under such circumstances, the retinal image of viewing the reconstructed 3D scene by an InI-HMD will approximate the visual effects of viewing a natural 3D scene.

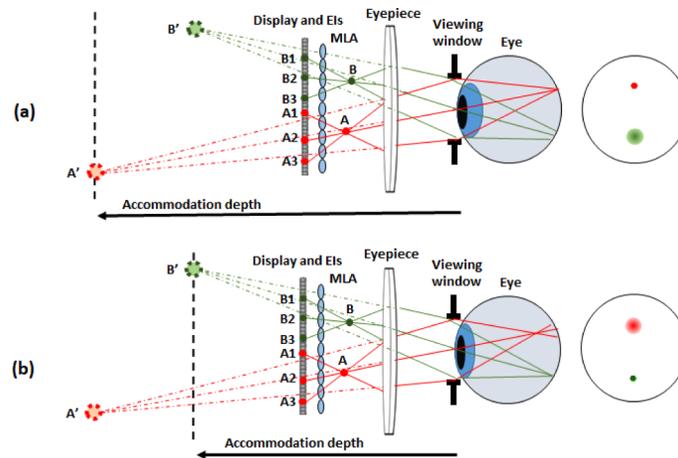


Fig. 1. Illustrations of basic optical principle of reconstructing 3D scene of InI-based LF-3D HMD system.

There exist two very unique and important conceptual reference planes in InI-HMDs that are widely recognized and worth pointing out. The first one is the virtual central depth plane (CDP) on which the light rays emitted by a point source on the microdisplay converge to form an image point after propagating through the MLA and eyepiece, as illustrated in Fig. 2(a). It is viewed as the plane of reference in the visual space optically conjugate to the microdisplay. The second reference plane is the viewing window defining the area within which a viewer observes the reconstructed 3D scene. It coincides with the entrance pupil plane of the eye optics by design and is commonly known as the exit pupil or eye box of the system in a conventional HMD. Distinctively different from a conventional HMD, however, the ray bundles emitted by the elemental pixels that reconstruct a 3D point are projected onto different locations at the view window as illustrated in Fig. 2(b).

As suggested in our previous works [24,27], by referring to the two reference planes described above, the 4-D light field function, $L(s, t, u, v)$, physically sampled by the pixels on the microdisplay

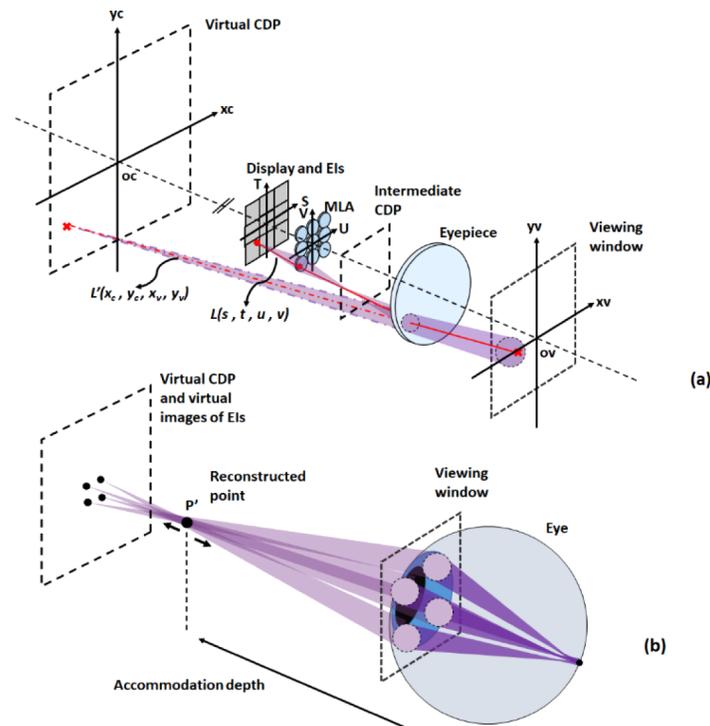


Fig. 2. Illustrations of (a) mapping the light field function and (b) basic optical principle of reconstructing 3D scene of InI-based LF-3D display.

and the lenslets of the MLA in the object space of an InI-HMD, can be conveniently mapped to the light field function $L'(x_c, y_c, x_v, y_v)$ defined in the virtual image or visual space, as if the ray bundles of the elemental views are emitted by pixels located on the virtual CDP toward the viewing window. Figure 2(a) illustrate the simplified process of light field rendering in the visual space, where the ray positions of the light field function are sampled by the projected virtual pixels (x_c, y_c) on the virtual CDP and the ray directions are defined by the projected coordinates (x_v, y_v) of the array elements on the viewing window. The new set of projected coordinates in fact directly relates to the spatial resolution of the apparent display in the visual space and the view density of the viewing window, respectively.

Theoretically with the ideal light field function $L'(x_c, y_c, x_v, y_v)$ rendered by an InI-HMD, the corresponding elemental images reconstructing a light field point will be imaged as point sources on the virtual CDP, and their ray bundles will perfectly intersect at the lateral and longitudinal position of the point and projected at their corresponding viewing zones on the viewing window as illustrated in Fig. 2(b). Any deviation in rendering the light field function due to the imperfection of the imaging process of an InI-HMD, however, will lead to compromised elemental images on the virtual CDP and/or wrong directions and footprints of the ray bundles projected onto the viewing window. Consequently, such deviation leads to compromised reconstruction of the 3D scene and reduced perceived image quality of an InI-HMD by the viewer. Therefore, in designing an InI-HMD system, we need to ensure the light fields, $L(s, t, u, v)$, physically rendered by a display panel and MLA in the object space to be mapped accurately into the light fields, $L'(x_c, y_c, x_v, y_v)$, viewed by the eye in the visual space. To achieve a good mapping, it is critical to obtain (1) a good control of the positional sampling mapping from (s, t) to (x_c, y_c) of the light field function so that each of the elemental images rendered on the display panel is well imaged

onto the virtual CDP, and (2) a good control of the directional sampling mapping from (u, v) to (x_v, y_v) of the light field function so that the ray bundles from each of the imaged elemental images are projected onto the viewing window with the correct directions and footprints and thus the elemental views are well integrated without displacement from each other. It requires a completely new design metrics that is capable of precisely evaluating the quality of the directional sampling of the reconstructed light field of a 3D scene, yet shall still be easily obtainable from any conventional lens design software so that we can further optimize the display system based on such a metrics.

It is noteworthy that in some of latest InI-HMD systems, more optical elements that are capable of further improving the overall display performance, such as tunable relay group [23,28], are further added in the display path, which adds more complexity in designing InI-HMDs but can still follow the same design methodology as a system consisting of only an eyepiece. Without the loss of generality, we assume these added optical elements are combined with the eyepiece, if applicable, referred to as “eyepiece group” in the following sections.

3. Design setup and key parameters

To account for the unique image formation process of an InI-HMD and accurately sample the ray positions and directions, we divide the overall system into M by N sub-systems, where M and N are the total number of lenslets in the MLA or equivalently the number of elemental images rendered on the microdisplay in the horizontal and vertical directions, respectively. Each of the sub-systems represents one single imaging path from an elemental image through its corresponding lenslet of MLA and a shared eyepiece group. It is worth noting that we choose to configure the sub-systems such that the ray tracing starts from the microdisplay or equivalently from the elemental image toward the viewing window which effectively avoids ray tracing failure due to the fact that the projections of the array of apertures of the lenslets of MLA on the viewing window do not form a commonly-shared exit pupil as in conventional HMDs. Under such circumstances, an ideal lens emulating the eye optics of a viewer or an established eye model such as Arizona eye model [29] is further inserted with its entrance pupil coinciding with the viewing window for better optimization convergence.

Due to the array nature of the MLA, the imaging path of each sub-system is off-axis, non-rotational symmetric to the main optical axis as illustrated in Fig. 3. Each of the sub-systems may be configured as a zoom configuration in optical design software. Depending on the types of surface symmetry of the MLA and the eyepiece group, the number of zoom configurations may be reduced to improve the convergence speed during optimization. For ease of further optimization, one may even further reduce the number of zoom configurations along each direction by sampling the lenslets at larger step for faster convergence. Without loss of generality, we hereby assume that all the lenslets in the MLA have identical surface shapes and are arranged in a rectangular array with equal lens pitch in both horizontal and vertical directions.

The zoom configurations among the sub-systems mainly differ from each other by the lateral positions of the corresponding lenslet and elemental image with respect to the optical axis of the eyepiece group. The lateral position of each lenslet, (u, v) , is solely determined by the displacement between the neighboring lenslet, Δp_{MLA} , or equivalently the lens pitch p_{MLA} , and the arrangement of the lenslets. Following the same coordinate system as shown in Fig. 2(a), for a given zoom configuration indexed as (m, n) , the lateral coordinates of its corresponding lenslet can be expressed as

$$\begin{cases} u_0(m, n) = \frac{(2m-M-1)p_{MLA}}{2} \\ v_0(m, n) = \frac{(2n-N-1)p_{MLA}}{2} \end{cases} \quad (1)$$

Although the microdisplay is also divided into an M by N array of elemental images, one for each lenslet, the lateral position and size of each elemental image is more complex and dependent

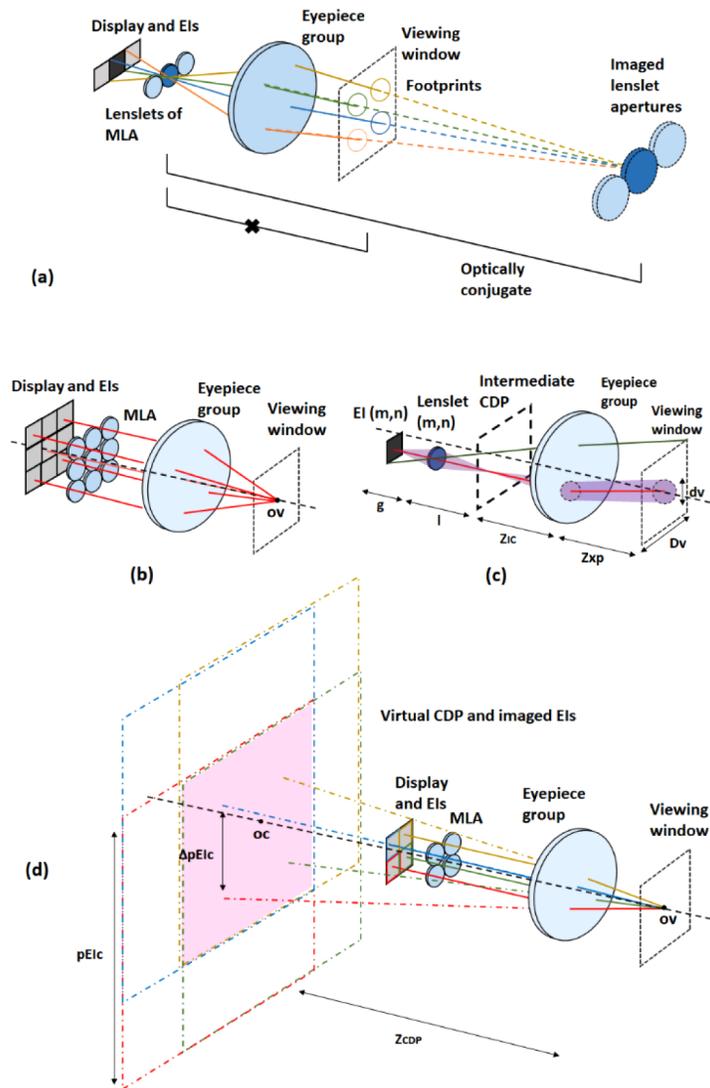


Fig. 3. Illustrations of (a) footprints on viewing window and imaged apertures of lenslets of MLA (real exit pupil), (b) the divide of the elemental images on the microdisplay, (c) sub-system including elemental image, lenslet of MLA and shared eyepiece group, and (d) partially overlapped virtual images of elemental images on the virtual CDP.

on several other specifications of the display system. For instance, the viewing window, which is not necessarily the optical conjugate of the MLA and can be shifted longitudinally along the optical axis according to a design requirement as shown in Fig. 3(a), plays a vital role in dividing and configuring the whole system into a series of sub-systems. An array of imaged apertures is actually formed in the visual space, which is considered as being the exit pupil in the sense of a conventional display or imaging system. Placing a viewer's eye at this exit pupil location would however indicate that the viewer cannot simultaneously see all the elemental images as the imaged apertures are spatially separated. To ensure that all of the elemental images can be properly seen at the viewing window, the microdisplay is divided in such a way that the chief ray of the center of each elemental image through the whole display optics, including the

corresponding lenslet of MLA and the shared eyepiece group, will intersect with the optical axis at the center of the viewing window, O_v (e.g. red line in Fig. 3(b)). Under such constraint, the displacement between the adjacent elemental images, Δp_{EI} , and the size of the elemental images, p_{EI} , on the microdisplay are equal and can be written as

$$\Delta p_{EI} = p_{EI} = p_{MLA} \left(1 + \frac{g}{l + z_{IC} + z'_{xp}} \right), \quad (2)$$

where g is the gap between the display panel and MLA, l is the gap between the MLA and the intermediate CDP, z_{IC} is the gap between the intermediate CDP and the eyepiece group, and z'_{xp} is introduced to refer to the distance between the eyepiece group and the imaged viewing window by the eyepiece group, which can be further given as

$$z'_{xp} = \frac{f_{ep} z_{xp}}{z_{xp} - f_{ep}}, \quad (3)$$

where f_{ep} is the equivalent focal length of the eyepiece group. Therefore, for a given sub-system unit indexed as (m, n) , the lateral coordinates of the center of the corresponding elemental image can be expressed as

$$\begin{cases} s_0(m, n) = \frac{(2m-M-1)p_{EI}}{2} \\ t_0(m, n) = \frac{(2n-N-1)p_{EI}}{2} \end{cases}. \quad (4)$$

The footprint size, d_v , of the ray bundle from a pixel of an elemental image projected through the whole optics on the viewing window, which determines the view density or equivalently the total number of views encircled by the eye pupil [24], can be determined by tracing the ray bundles emitted by the center point of the elemental image (e.g. the purple-shaded ray bundles in Fig. 3(c)), and the overall size of the viewing window, D_v , can be determined by tracing the chief ray of the edge point of the elemental image through the whole display optics (e.g. green line in Fig. 3(c)). The ray footprint diameter and the overall viewing window size can be expressed as

$$d_v = p_{MLA} \frac{z_{IC}}{l} \left| 1 - \frac{z_{xp}}{f_{ep}} + \frac{z_{xp}}{z_{IC}} \right|, \quad (5)$$

$$D_v = p_{EI} \frac{l + z_{IC}}{g} \left| 1 - \frac{z_{xp}}{f_{ep}} + \frac{z_{xp}}{l + z_{IC}} \right|, \quad (6)$$

where $||$ stands for the absolute value symbol.

According to the paraxial geometry, both the footprint diameter and the viewing window size are the same for any of the sub-systems, and the footprints corresponding to the same field object of different sub-systems will intersect on the viewing window so that they share the same coordinates (x_v, y_v) . For example, as mentioned above, the chief ray of the center of each elemental image will intersect with the optical axis at the center of the viewing window so that $x_{v0}(m, n)$ and $y_{v0}(m, n)$ both equal to 0 for any of the sub-system.

In an InI-HMD system, the elemental images are seen as an array of spatially displaced virtual images observed from the viewing window. Figure 3(d) illustrates a simple example where four neighboring elemental images rendered on the microdisplay, each illustrated with a different color, are imaged through their corresponding lenslets of MLA and the shared eyepiece and are projected on the virtual CDP as four partially overlapping virtual elemental images illustrated by the dashed box of the same corresponding colors. The displacement between centers of the neighboring virtual elemental images on the virtual CDP, Δp_{EIC} , is no longer equal to the size of

the virtual elemental images, p_{Elc} , and their paraxial values are further expressed, respectively, as

$$\Delta p_{Elc} = \Delta p_{EI} \frac{z'_{xp}}{g + l + z_{IC} + z'_{xp}} \frac{z_{CDP}}{z_{xp}}, \quad (7)$$

$$p_{Elc} = p_{EI} \frac{l(z_{CDP} - z_{xp})}{gz_{IC}}. \quad (8)$$

where Z_{CDP} is the distance between the virtual CDP and the viewing window. For a given sub-system unit indexed as (m, n) , the lateral coordinates of the paraxial center of the corresponding virtual elemental image can be expressed as

$$\begin{cases} x_{c0}(m, n) = \frac{(2m-M-1)\Delta p_{Elc}}{2} \\ y_{c0}(m, n) = \frac{(2n-N-1)\Delta p_{Elc}}{2} \end{cases}. \quad (9)$$

Equations (7) and (8) essentially provide the paraxial calculations of the image dimension and center position for each of the sub-systems. The image size of the virtual elemental images on the virtual CDP is usually much greater than the displacement of the neighboring virtual elemental images so that the virtual elemental images on the virtual CDP would partially overlap with each other. As illustrated in Fig. 3(d), the pink-shaded area on the virtual CDP corresponds to the region where all the four virtual elemental images overlap with each other, and is where four distinctive elemental views, one from each of the elemental images, can be rendered to reconstruct the light fields of a sub-volume of a scene seen from the viewing window. Due to this nature, the overall FOV of an InI-HMD is mosaiced by the sub-volumes rendered by each of the individual elemental images and thus could not be straightforwardly calculated as in conventional HMDs. For simplicity, by referring to the centers of the virtual images of the edge elemental images on the virtual CDP, the diagonal FOV of the display system could be estimated as,

$$FOV_D = 2 \tan^{-1} \left(\sqrt{\frac{(M^2 + N^2)}{2}} \frac{\Delta p_{Elc}}{Z_{CDP}} \right). \quad (10)$$

4. Optimizing ray positional sampling of light field

Optimizing ray positional sampling of the light field function can be achieved by obtaining well-imaged elemental images on the virtual CDP from the display panel through their corresponding lenslets of the MLA and eyepiece group. Following the setup procedures introduced, we are able to optimize the imaging process of each elemental image individually by each of the sub-systems. From this perspective, the well-established optimization constraints and performance metrics available in optimizing the 2D image-conjugates for conventional HMDs can be adequately utilized except that the entire FOV composed by the individual elemental images are optimized separately instead of being treated as a whole as in conventional HMD designs.

Such an individual optimization for each of the elemental images, however, overlooks the corresponding connection between the neighboring elemental images and more importantly the relative positions and sizes of virtual elemental images with respect to the total FOV. For an ideal InI-HMD, as shown in Fig. 3(d), the imaged elemental images on the virtual CDP would be aligned perfectly and partially overlapping with neighboring ones, with the paraxial relationships given by Eqs. (7)–(9). However, image aberrations induced by both the MLA and the eyepiece group will distort the virtual elemental images on the virtual CDP and the distorted virtual elemental images will cause potential failure of view convergence when reconstructing the light fields of 3D scenes of different depths.

To account for the effects of distortions induced to the elemental images locally and globally, we propose to apply two different types of constraints during optimization. The first constraint is

the control of the local distortion aberrations for each of the sub-systems representing a single elemental image, which can be readily implemented by adopting the distortion-related optimization constraints already available in the optical design software to each zoom configuration. These local controls of distortion in each sub-system ensure the dimensions and shapes of the virtual elemental images remain within a threshold level in comparison to their paraxial non-distorted images. The second constraint is the control of the global distortion, which is related to the lateral positions of the virtual elemental images with respect to the whole FOV of the reconstructed 3D scene. To optimize for this global distortion, the chief ray of the center object field of each elemental image on the microdisplay is ought to be specially traced and its interception on the virtual CDP needs to be extracted and constrained within a threshold level comparing to its paraxial positions in global coordinates. For a given sub-system indexed as (m, n) , the global deviation of the center position of virtual elemental image on the virtual CDP from its paraxial position can be quantified by a possible metrics, GD , along with the corresponding constraints, which are expressed, respectively, as

$$GD(m, n) = \tan^{-1} \left(\frac{\sqrt{(x_c(m,n) - x'_c(m,n))^2 + (y_c(m,n) - y'_c(m,n))^2}}{Z_{CDP}} \right), \quad (11)$$

where $x'_c(m,n)$ and $y'_c(m,n)$ are the real positions of the chief ray of the center object field of the given elemental image indexed as (m, n) on the virtual CDP obtained by real ray tracing horizontally and vertically, respectively, while $x_c(m,n)$ and $y_c(m,n)$ are their paraxial positions of the chief ray on the virtual CDP given by Eq. (9).

The GD metric in Eq. (11) examines the angular deviation between the real and theoretical position of the chief ray of the center object field measured from the viewing window. Compared to the conventional definition of distortion ratio, we expect that the GD metric works better in off-axis lens design, especially in systems with freeform surfaces, where on- and near axis fields will possibly encounter distortion of same magnitude to those at the full FOV so that the relative ratio as in conventional definition of distortion ratio will not work efficiently.

A constraint corresponding to the metric is therefore created by obtaining the maximum values of the metric for all the elemental images through all the sub-systems. By adding constraint to the optimization process and modifying the value of the constraint, the maximally allowed global distortion can be adjusted and optimized. Figure 4 further demonstrates the overall correlation between the global distortion and the value of GD by utilizing examples of barrel distortion and keystone distortion simulated in a 40° by 40° InI-HMD system with the depth of CDP at 1 diopter. In each of the sub-figures, both the full theoretical FOV grid free from global distortion (black) and the distorted FOV grid (red) corresponding to the specific type and value of distortion were plotted, and the numbers stand for the maximum and average value of GD calculated from Eq. (12) for a total of 11 by 11 sampled centers of elemental images (the intersection points of the grids). For example, while 1% barrel distortion yields a maximum GD of only 0.36° , 5% barrel distortion yields a maximum GD as large as 1.78° . Clearly, the value of GD provides a good control of the global distortion of the center positions of the elemental images, either as conventional distortion pattern (e.g. barrel or pincushion distortion) or unconventional one (e.g. keystone distortion), in optimizing ray positions of the light field function.

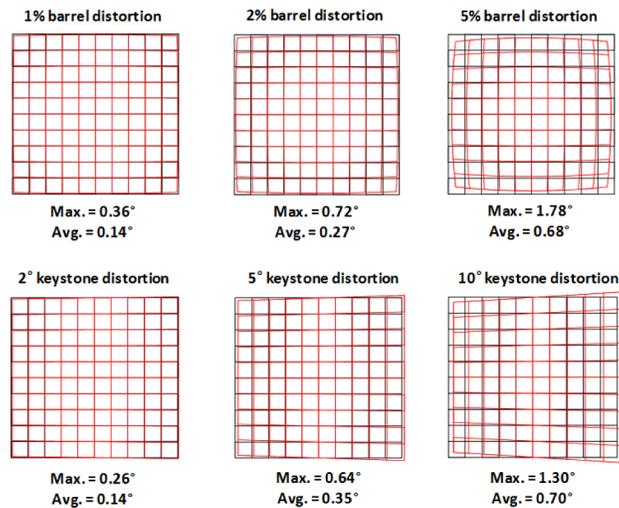


Fig. 4. Simulations of effects of *GD* regarding the global distortion in ray positions upon full field distortion grid.

5. Optimizing ray directional sampling of light field

In conventional display systems, the directions of the ray are usually not strictly constrained (unless sometimes telecentricity is required). In some extreme cases, pupil aberration, which is a set of aberrations that would be observed at the exit pupil with respect to the entrance pupil of an optical system, or vice versa, and responsible of deformed, mismatched directions of the ray bundles, is even deliberately introduced to help compensate the image aberrations. However, due to the unique property of an InI-HMD, the ray directions of the light fields play a very important role in designing such a display system. As discussed above, incorrect sampling of ray directions will not only affect the integration of the elemental images but also potentially lead to uneven number of elemental views for reconstructed light field targets and thus misrepresented focus cues [24,27]. In the case of severe pupil aberration, it is even possible that the number of elemental views encircled by a viewer's eye pupil reduces to be less than two so that it makes no difference from a conventional stereoscopic display system and fail to properly render true light fields.

As suggested above, the viewing window is where all the chief rays through the center pixels of all the elemental images intersect with the optical axis, as shown in Fig. 3(a), to ensure all of the elemental images can be properly seen simultaneously. The footprints of the ray bundles projected from each of the elemental views, characterized by Eq. (5), collectively define the viewing window with its paraxial dimension characterized by Eq. (6). To optimize the ray directional sampling of light fields in designing InI-HMDs, we need to develop proper constraints for the footprints of each elemental views projected on the viewing window rather than directly optimizing for the exit pupil(s) of the elemental views. For an InI-HMD with ideal ray directions of the light field function or equivalently free from pupil aberration, the merged footprint diagram at the viewing window should have two notable characteristics. First of all, the chief rays from different object fields on a single elemental image passing through the corresponding lenslet of the MLA as well as the eyepiece group should converge at the center of the imaged aperture, and intersect with the viewing window in a regular grid with uniform spacing, resembling the pixel array of the elemental image. Secondly, the chief rays from the same object field (with respect to their own elemental images) passing through their corresponding lenslets and eyepiece should converge at the viewing window, and the footprints of the ray bundles from these pixels should form the same shape and overlap perfectly with each other on the viewing window. It is worth

noting that to clearly differentiate the shapes of elemental images from lenslets, for the purpose of illustrations, we used circular shapes in Fig. 3 to represent the apertures of the lenslets and their corresponding footprints on the viewing window, while square apertures have been utilized for the lenslets in the examples of simulation and actual design.

To account for the effects of pupil aberration induced to the ray footprints and directions on the viewing window during optimization, we would need to (1) extract the exact footprints of the ray bundles from any given pixel of a given elemental image on the viewing window; and (2) establish metric functions that properly quantify any deviations of the ray footprints from their paraxial shapes and positions so that constraints can be applied during the optimization process to control the deviations within a threshold level. In practice for each give object field we only sampled four marginal rays through the lenslet aperture to avoid exhaustive computation time during the optimization process. The coordinates of these marginal rays on the viewing window define the envelop of the ray footprint of a sampled field on a given elemental image in a given sub-systems. For a sampled object field indexed as (i, j) on a given elemental image corresponding to a sampled sub-system indexed as (m, n) , the deformation of the ray footprints from its paraxial shape can be quantified by a metric function, PA , expressed as

$$PA(m, n, i, j) = \frac{1}{4} \sum_{k=1}^4 \frac{\sqrt{(x_{v,k}(m,n,i,j) - x'_{v,k}(m,n,i,j))^2 + (y_{v,k}(m,n,i,j) - y'_{v,k}(m,n,i,j))^2}}{\sqrt{2}d_v}, \quad (12)$$

where x'_v and y'_v are the real positions of the marginal rays on the viewing window obtained via real ray tracing horizontally and vertically, respectively, while x_v and y_v are their corresponding paraxial positions on the viewing window. k is the index of four marginal rays for a sampled object on a given elemental image corresponding to a sampled sub-system.

The metric PA in Eq. (12) quantifies the deformation of the ray footprint of a given ray bundle from its paraxial shape by examining the relative ratio of the average deviated distance between the real and theoretical positions of the marginal rays on the viewing window to the diagonal width of the paraxial footprint. A single constraint is then created by obtaining the maximum value of the metric form all the sampled object fields on each of the sampled sub-systems. By adding the constraint to the optimization process and modifying the value, the maximally allowed deviation and deformation of the footprint, or equivalently, the pupil aberration affecting the ray directions of the light field of an InI-HMD can be adjusted and optimized. Figure 5 further demonstrates the overall correlation between the footprint diagrams on the viewing window, pupil aberrations, and the metric function values of PA . The figures plot simulated ray footprint diagrams for the center field point of the elemental image centered with the optical axis of an InI-HMD, with and without pupil aberration. For simplicity in the simulation the lenslets were treated as ideal lenses, and the eyepiece group was modeled with different aberration terms and magnitudes (e.g. spherical aberration from 0.25 to 1 waves peak to valley (λ PV), and tilt from 0.25 to 1°) applied as pupil aberration. Specifically, the diameter of the theoretical footprint, d_v , was set as 1 mm. In each of the sub-figures, both the theoretical footprint diagram free from pupil aberration (black) and the deformed or displaced footprint diagram (red) corresponding to the specific type and value of pupil aberration were plotted. The number beneath each of the sub-figures stands for the value of PA calculated from Eq. (12) for each case.

It can be easily observed that due to the presence of pupil aberration, the actual footprint diagrams can be significantly deformed (e.g. by pupil spherical aberration) or displaced (e.g. by tilt) from their theoretical ones. On the other hand, Eq. (12) makes a good estimation of the severity of the pupil aberration in terms of PA based on the footprint diagram. For example, 0.25 λ PV pupil spherical aberration yields a PA of only 0.059, while 1 λ PV pupil spherical aberration yields a PA as large as 0.23.

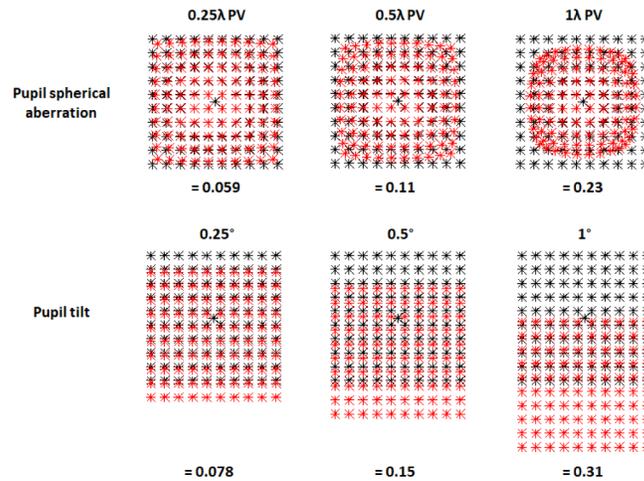


Fig. 5. Simulations of effects of *PA* regarding the pupil aberration in ray directions upon footprint diagram on the viewing window.

6. Result

Based on the design setup and optimization methods described above, we hereby present a design example of an InI-based light field optical see-through HMD using state-of-the-art freeform optics for the purpose of validating our proposed methods for designing and optimizing this type of display optics. Figure 6(a) shows the schematic layout of a binocular setup of our custom-designed InI-HMD system with respect to the viewer's head, and Fig. 6(b) further elaborates the detailed optical layout of its monocular setup (right eye) with key elements labelled [28]. As illustrated, the optics for the display path, which is of our prime interest, mainly consists of three key parts: a micro-INI unit including a high-resolution microdisplay, a custom-designed aspherical MLA, and a custom aperture array, a tunable relay group mainly made of 4 stock spherical lenses with an Optotune EL-10-30 tunable lens sandwiched inside, and a freeform waveguide-like prism. The waveguide-like prism, mainly formed by 4 freeform surfaces denoted as $S1$ to $S4$, further magnifies the reconstructed intermediate miniature scene and projects the light toward the exit pupil, or the viewing window, at which a viewer sees the magnified 3D scene reconstruction. More detailed descriptions regarding the systematical level of the design example, including working principle, specifications, experimental verification and evaluation, can be further found in [28]. Here in this section, we mainly focus on demonstrating the utility of the design and optimization methods described above for optimizing the optics of the display path.

We started the design process by initially optimizing the MLA and the relay-eyepiece group separately to obtain good starting points due to the complexity of the design. For the initial MLA design, we assume that all the lenslets have the same surface shapes and each lenslet has a square aperture with a lens pitch of 1 mm. Special attention was paid to the marginal rays that were constrained to not surpass the edge of the lenslet to prevent crosstalk among neighboring elemental images. The two surfaces of the lenslet were optimized as aspheric polynomials with coefficients up to 6th order. In the initial design process of the relay and eyepiece group, the design was reversely set up by backward tracing rays from the viewing window toward the eyepiece and relay lenses. Each of the four freeform surfaces of the prism was described by x-plane symmetric XY-polynomials and was optimized with coefficients up to their 10th order. Color aberrations can also be well controlled in this process. On one hand, the optical power of the eyepiece is mainly contributed from $S4$, which is a reflective surface that is naturally color

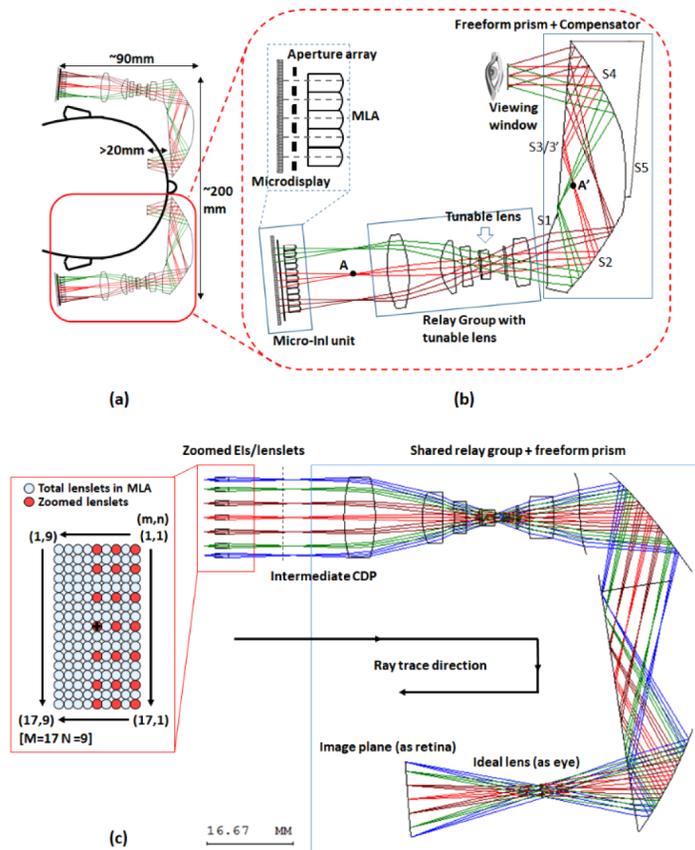


Fig. 6. Real design setup of the integrated system.

aberration free. On the other hand, by carefully picking the glass combination of the relay group, the singlets can function as separated achromats that help to correct the color aberration.

After obtaining the initial designs of both the MLA and the relay-eyepiece group, we integrated the two parts and created an array of 7 by 3 zoom configurations. Figure 6(c) shows the design configuration of the integrated display path in CodeV, plotted with real ray tracing from a fraction of the sampled elemental images and lenslets. The viewing window was placed at the back focal point of the freeform eyepiece. An ideal lens with a focal length equivalent to the eye focal power corresponding to the depth of the virtual CDP was inserted at the viewing window to simulate the retinal images of the elemental images. In the figure only the rays from the center pixel of each elemental image were traced. The MLA consists of 17 by 9 identical lenslets with a lens pitch of 1 mm, and the microdisplay is divided into 17 by 9 elemental images, each with 125 by 125 pixels. Considering the plane symmetry of the freeform prism and convergence speed, totally 7 by 3 sub-systems (elemental images with corresponding lenslets of MLA) for the top half of the total FOV were sampled, and the distribution of the sampled sub-systems is shown in Fig. 6(c) (in red). In each sub-system 9 field points were further sampled covering the whole elemental image. For each of the sub-systems, we optimized on rms spot size and evaluated its performance based on both diffraction and geometrical MTF. Besides, to account for DOF extension as proposed [23] with varying optical power of the tunable lens as well as the depth of virtual CDP, the system was further configured to optimize its performance for the virtual CDP depths of 0, 1, and 3 diopters. The focal length of the ideal lens at the viewing window as well as

the tunable lens is thus adjusted correspondingly to correctly focus the rays on to the image plane. Altogether, combining the zooms of for the 21 sampled MLA-elemental image sub-systems as well as the zooms for different virtual CDP depths, the overall system was modelled with a total 63 zoom configurations, and a total of 567 field points. More field points and sub-systems can be sampled during optimization at the cost of convergence speed of optimization.

Figure 7(a) plots the image contrast of the sampled 7 by 3 sub-systems covering the full field of the display path evaluated at the Nyquist angular frequency of 3 arcmins or 10 cycles/degree (cpd) with the virtual CDP set at 1 diopter away from the viewing window. In each of the sub-systems 5 object fields on their corresponding elemental image were sampled and their contrast values were plotted by different colors. Across the entire 30° by 18° of FOV the image contrast is all well above the threshold of 0.2 at the Nyquist angular frequency with an average of 0.53. Figure 7(b) specifically plots the MTFs of the on-axis field points on three elemental images corresponding to the lenslet centered with optical axis (index (9,5)), the top left corner lenslet (index (1,1)), and the top right corner lenslet (index (17,1)), respectively, covering the whole FOV of the display path. Figure 7(c) further plots the MTFs of the on-axis field points on three elemental images corresponding to the lenslet centered with the optical axis (index (9,5)) but with their virtual CDP adjusted from 3 diopters to 0 diopters away from the viewing window by adjusting the optical power of the tunable lens. It is clear that the optical system demonstrates uniform image contrast and MTF performance across the entire FOV and depth range of over 3 diopters with a degradation of image contrast evaluated at Nyquist angular frequency less than 0.15.

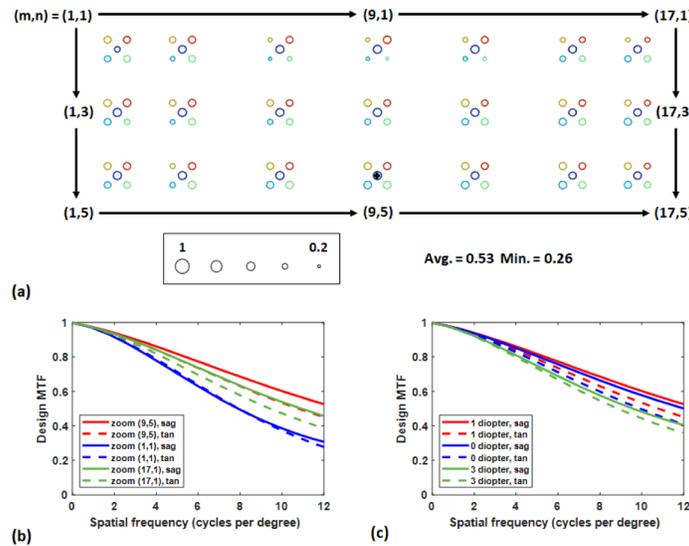


Fig. 7. (a) Full field image contrast plot of the display path. MTFs of (b) the on-axis field points of each elemental images corresponding to three sampled lenslets covering the whole FOV of the display path, and (c) the on-axis field points of each elemental images corresponding to the central lenslets with their virtual CDP sampled from 3 diopters to 0 diopters away from the viewing window.

Figure 8(a) further plots the global distortion grid of the sampled 7 by 3 sub-systems of the display path covering the full display FOV by extracting the chief ray coordinates of the center object field on the corresponding elemental image from each of the sub-systems from real ray tracing of the design example, where the paraxial coordinates of the chief rays are plotted in black solid grid and the actual ray coordinates in blue asterisks. Though the display path suffers a small amount of keystone distortion due to the folded optical path, generally the global distortion for

full display field is relatively small, especially for a design involving freeform optics which easily introduces high order distortion terms. The design target regarding the global distortion GD was set as 0.75° which corresponding to around 2% of the distortion with respect to the full FOV. All of the 7 by 3 sub-systems were optimized within the design target with an average value of GD of 0.22° which corresponds to an average distortion with respect to the full FOV less than 1%.

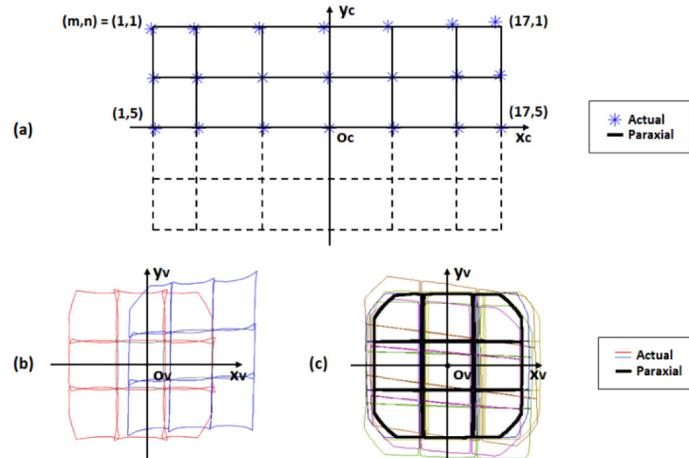


Fig. 8. (a) Distortion grid of display path covering the full field. Footprint diagrams at the viewing window (b) before and (c) after optimization of pupil aberration of the system.

On the other hand, Figs. 8(b) and 8(c) compare the ray footprint diagrams at the viewing window before and after optimization regarding the ray directions of light field function. Figure 8(b) plots the envelopes of the ray footprints on the viewing window for the 9 sampled object fields of the on-axis lenslet (in red, index (9,5)) and the envelopes for the 9 object fields of the edge lenslet located at the top-right corner (in blue, index (17,1)) from the real design setup before constraining the pupil aberration of the system. The ray footprint envelopes for these two lenslets are not only distorted but also severely separated. In comparison, Fig. 8(c) plots the merged envelopes of the footprint diagrams extracted from the real design setup after optimization. In this case, the ray footprints through 9 sampled lenslets of the MLA were plotted in different colors (indexes (1,1), (9,1), (17,1), (1,3), (9,3), (17,3), (1,5), (9,5) and (17,5)). Figure 8(c) also plots the theoretical envelopes (in black) of the ray footprints of the same fields on the lenslets obtained from paraxial calculations which are perfectly aligned with each other across the lenslets and fields as suggested above. The design target PA was set as 0.3 since the human vision system would be less sensitive to the ray directions than positions. 19 of the 7 by 3 sub-systems were optimized within the design target with an average ratio of 0.145 that is well below the design target, which corresponds to an average throughput or size deviation of the real footprints less than 0.27, of which the deviation and deformation of the projected footprints are still acceptable as shown in Fig. 8(c).

As demonstrated, the proposed design setup and optimization method yields a promising design of a high performance InI-HMD. Based on this design, we have built and implemented a prototype of the design example [28]. Figure 9 further shows the test result of the prototype by placing the camera at the viewing window and capturing real images of the displayed scene through the system. As shown in Fig. 9(a), a slanted wall with water drop texture spanning a depth from around 500 mm (2 diopters) to 1600 mm (0.6 diopters) was computationally rendered and to be displayed as the test target. Figure 9(b) shows the central 15 by 7 elemental views rendered on the microdisplay. Figures 9(c)–(e) show the real captured images of the rendered light fields of such a continuous 3D scene by adjusting the focal depth of the camera from the

near side (~ 600 mm), to the middle part (~ 1000 mm), and the far side (~ 1400 mm) of the scene, respectively, which simulates the adjustment of the eye accommodation from near to far distances. The virtual CDP of the prototype was shifted and fixed at depth of 750 mm (1.33 diopters). On each of the three images the depth region corresponding to camera focal depth is marked with a yellow box. It can be observed that the parts of the 3D scene within the same depth of the camera focus (e.g. inside the yellow box) are in sharp focus and of high fidelity compared to the target. In contrast, the other parts of the 3D scene (outside of the yellow box) outside of the camera focal depth are in blur, and the more the depth of the 3D scene is deviated from the camera focus, the blurrier the part of the 3D scene would be, which is similar to what we observe from the real world scene. Such a result clearly demonstrates the ability of the prototype of the design example to render high-quality light field contents, and more importantly, to render correct focus cues to drive the accommodation of the viewer's eye.

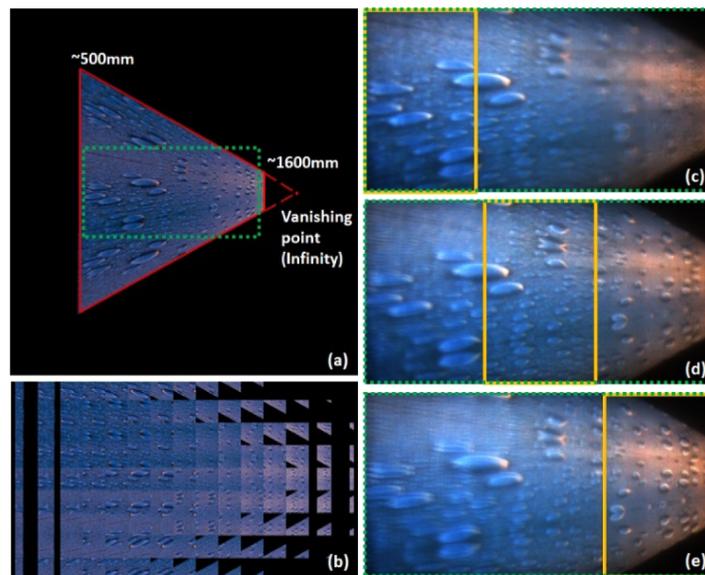


Fig. 9. (a) Computationally rendered 3D scene as test target and (b) rendered array of elemental images of the central part of the 3D scene on microdisplay. Captured images of the test target displayed by the prototype when the camera is focusing on depth of (c) 600 mm, (d) 1000 mm and (e) 1400 mm.

7. Conclusion

In this paper, the optical design methodology including the overall strategy and light field optimization of high-performance InI-HMD system was presented. We analyzed the basic optical principle of such a display system and overcame numerous challenges in optimization to achieve seamless integration of the InI-unit with the eyepiece group and minimal aberration-induced light field rendering taking both the image and pupil aberration into considerations. A design example was further given based on the design methodology which is capable of rendering the light field reconstructed target with high quality. We believe this novel design methodology can be applied as a guidance for designing different types of light field HMDs.

Funding

Directorate for Computer and Information Science and Engineering (14-22653).

Disclosures

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.

References

1. J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," *Vision Res.* **35**(19), 2731–2736 (1995).
2. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.* **8**(3), 33 (2008).
3. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *J. Vis.* **5**(10), 7 (2005).
4. M. Mon-Williams, J. P. Warm, and S. Rushton, "Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display," *Ophthalmic Physiol. Opt.* **13**(4), 387–391 (1993).
5. H. Hua, "Enabling focus cues in head-mounted displays," *Proc. IEEE* **105**(5), 805–824 (2017).
6. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," *Science* **265**(5173), 749–752 (1994).
7. 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).
8. 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).
9. S. Shiwa, K. Omura, and F. Kishino, "Proposal for a 3-D display with accommodative compensation: 3DDAC," *J. Soc. Inf. Disp.* **4**(4), 255–261 (1996).
10. T. Shibata, T. Kawai, K. Ohta, M. Otsuki, N. Miyake, Y. Yoshihara, and T. Iwasaki, "Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence," *J. Soc. Inf. Disp.* **13**(8), 665–671 (2005).
11. S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," *Proceedings of IEEE Int. Symp. Mixed Augmented Reality (ISMAR)* (2008), pp. 32–42.
12. D. Dunn, C. Tippetts, K. Torell, P. Kellnhöfer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs, "Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors," *IEEE Trans. Vis. Comput. Graph.* **23**(4), 1322–1331 (2017).
13. J. P. Rolland, M. W. Krueger, and A. Goon, "Multifocal planes head-mounted displays," *Appl. Opt.* **39**(19), 3209–3215 (2000).
14. 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).
15. 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).
16. 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).
17. 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).
18. A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "An interactive 360° light field display," *ACM Trans. Graph.* **13**, 1–4 (2007).
19. D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.* **32**(6), 1–10 (2013).
20. H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Opt. Express* **22**(11), 13484–13491 (2014).
21. 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).
22. J. Hong, S. W. Min, and B. Lee, "Integral floating display systems for augmented reality," *Appl. Opt.* **51**(18), 4201–4209 (2012).
23. 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).
24. H. Huang and H. Hua, "Systematic characterization and optimization of 3D light field displays," *Opt. Express* **25**(16), 18508–18525 (2017).
25. 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).
26. M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH '96, Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), pp. 31–36.
27. H. Huang and H. Hua, "Effects of ray position sampling on the visual responses of 3D light field displays," *Opt. Express* **27**(7), 9343–9360 (2019).

28. H. Huang and H. Hua, "High-performance integral-imaging-based light field augmented reality display using freeform optics," *Opt. Express* **26**(13), 17578–17590 (2018).
29. J. Schwiegerling, *Field Guide to Visual and Ophthalmic Optics* (SPIE, 2004).