Systematic characterization and optimization of 3D 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 85721, Arizona, USA

*hhua@optics.arizona.edu

Abstract One of the key issues in conventional stereoscopic displays is the well-known vergence-accommodation conflict problem due to the lack of the ability to render correct focus cues for 3D scenes. Recently several light field display methods have been explored to reconstruct a true 3D scene by sampling either the projections of the 3D scene at different depths or the directions of the light rays apparently emitted by the 3D scene and viewed from different eye positions. These methods are potentially capable of rendering correct or nearly correct focus cues and addressing the vergence-accommodation conflict problem. In this paper, we describe a generalized framework to model the image formation process of the existing light-field display methods and present a systematic method to simulate and characterize the retinal image and the accommodation response rendered by a light field display. We further employ this framework to investigate the trade-offs and guidelines for an optimal 3D light field display design. Our method is based on quantitatively evaluating the modulation transfer functions of the perceived retinal image of a light field display by accounting for the ocular factors of the human visual system.

References and links

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), 1–30 (2008).
3. S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," J. Vis. **5**(10), 834–862 (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. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," Science **265**(5173), 749–752 (1994).
6. 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).
7. G. E. Favalora, "Volumetric 3D displays and application infrastructure," Computer **38**(8), 37–44 (2005).
8. 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).
9. 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).
10. 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).
11. Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Opt. Express **18**(9), 8824–8835 (2010).
12. Y. Kajiki, H. Yoshikawa, and T. Honda, "Hologram-like video images by 45-view stereoscopic display," Proc. SPIE **3012**, 154–166 (1997).
13. Y. Takaki, "High-density directional display for generating natural three-dimensional images," Proc. IEEE **94**(3), 654–663 (2006).
14. J. H. Lee, J. Park, D. Nam, S. Y. Choi, D. S. Park, and C. Y. Kim, "Optimal projector configuration design for 300-Mpixel multi-projection 3D display," Opt. Express **21**(22), 26820–26835 (2013).
15. 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).

16. H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," Opt. Lett. **26**(3), 157–159 (2001).
17. H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," Opt. Express **22**(11), 13484–13491 (2014).
18. 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).
19. W. Gordon, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting," ACM Trans. Graph. **31**, 80 (2012).
20. 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), 89 (2014).
21. Y. Kim, J. Kim, K. Hong, H. K. Yang, J. Jung, H. Choi, S. Min, J. Seo, J. Hwang, and B. Lee, "Accommodative Response of Integral Imaging in Near Distance," J. Disp. Technol. **8**(2), 70–78 (2012).
22. S. K. Kim, D. W. Kim, Y. M. Kwon, and J. Y. Son, "Evaluation of the monocular depth cue in 3D displays," Opt. Express **16**(26), 21415–21422 (2008).
23. A. Stern, Y. Yitzhaky, and B. Javidi, "Perceivable light fields: Matching the requirements between the human visual system and autostereoscopic 3D displays," Proc. IEEE **102**(10), 1571–1587 (2014).
24. G. Lippmann, "Epreuves reversibles donnant la sensation du relief," J. Phys. (Paris) **7**, 821–825 (1908).
25. J. Schwiegerling, *Field Guide to Visual and Ophthalmic Optics* (SPIE, 2004).
26. https://optics.synopsys.com/
27. J. E. Greivenkamp, J. Schwiegerling, J. M. Miller, and M. D. Mellinger, "Visual acuity modeling using optical raytracing of schematic eyes," Am. J. Ophthalmol. **120**(2), 227–240 (1995).
28. D. G. Green and F. W. Campbell, "Effect of focus on the visual response to a sinusoidally modulated spatial stimulus," J. Opt. Soc. Am. A **55**(9), 1154–1157 (1965).
29. J. S. Jang and B. Javidi, "Improvement of viewing angle in integral imaging by use of moving lenslet arrays with low fill factor," Appl. Opt. **42**(11), 1996–2002 (2003).
30. C. Jang, K. Hong, J. Yeom, and B. Lee, "See-through integral imaging display using a resolution and fill factor-enhanced lens-array holographic optical element," Opt. Express **22**(23), 27958–27967 (2014).

1. Introduction

Conventional stereoscopic three-dimensional displays (S3D) stimulate the perception of 3D space and shapes from a pair of two-dimensional (2D) perspective images, one for each eye, with binocular disparities and other pictorial depth cues of a 3D scene seen from two slightly different viewing positions. Although they can create compelling depth perceptions, the S3D-type displays are subject to a well-known vergence-accommodation conflict (VAC) problem due to the inability to render correct focus cues, including accommodation and retinal blur effects, for 3D scenes [1,2]. Eye accommodation refers to the focus action of the eye where ciliary muscles contract or relax to change the refractive power of the crystalline lens to obtain and maintain clarity of retinal image for a fixated object of a given depth. The retinal image blur effect refers to the phenomena associated with eye accommodation change in which objects away from the eye's accommodative distance appear blurry in the retinal image due to the limited depth-of-field (DOF) of the eyes. The conventional S3D displays fail to render correct retinal blur effects and stimulate natural eye accommodation response, which causes several cue conflicts and is considered as one of the key contributing factors to various visual artifacts associated with viewing S3D displays, such as distorted depth perception [3] and visual discomfort [4]. In recent years, several display methods that are potentially capable of resolving the VAC problem have been demonstrated, including holographic displays [5], volumetric displays [6,7], multi-focal plane displays [8–10] and light field displays [11–20]. Among these different methods, the light field display method is considered as one of the most promising 3D display techniques.

As illustrated in Fig. 1, a light field 3D (LF-3D) display, which in some cases is referred to as super multi-view (SMV) displays, renders the perception of a 3D object (e.g. the cube) by reproducing the directional samples of the light rays apparently emitted by each point on the object such that multiple ray samples are viewed through each of the eye pupils. The light intensities corresponding to the angular samples of the light rays are anisotropically modulated to reconstruct a 3D scene that approximates the visual effects of viewing a natural 3D scene. Each of the directional samples represents the subtle difference of the object when viewed from slightly different positions and thus is regarded as an elemental view of the

object. To enable the eye to accommodate at the depth of the 3D object rather than the source from which the rays are actually originated, a true LF-3D display requires that multiple different elemental views are seen through each of the eye pupils and they integrally sum up to form the perception of the object. When the eye is accommodated at a given depth (e.g. the blue corner of the cube), the directional ray samples apparently emitted by a point of the same depth naturally form a sharply focused image on the retina, while the rays for points at other depths (e.g. the red and green corners) form retinal blur effects varying with their corresponding depth. Therefore, a LF-3D display has the potential to render correct or nearly correct focus cues for 3D scenes and resolve the VAC problem. Furthermore, it can also potentially render correct motion parallax, which is the change in retinal images caused by the change of eye pupil position, and further improve 3D perception. Conventional two-view S3D displays lack the ability to render motion parallax because the retinal images remain unchanged when the eyes move within each of their own viewing windows.
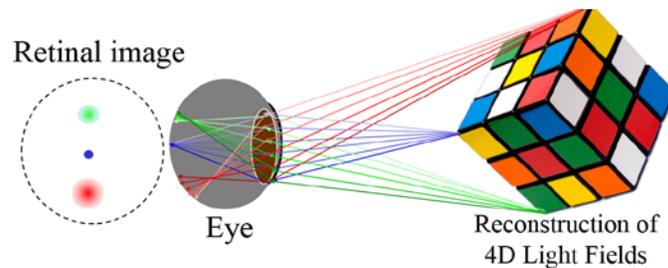

Fig. 1. Schematic illustration of a light field display which reproduces the directional light rays apparently emitted by the object of a given position.

Despite that several research groups have demonstrated the promising potentials of LF-3D displays for addressing the VAC problem, there are very few systematic investigations upon addressing many of the fundamental issues in the process of engineering a LF-3D display, including (1) methods for quantifying the accuracy of the focus cues rendered by a LF-3D display; (2) the threshold requirement for a LF-3D display to render correct or nearly correct focus cues; and (3) the optimal configurations for a LF-3D display that provide both correct accommodative cue and high image quality. Several pioneering works attempted to address some of these issues. Takaki suggested that the angular separation of the directional ray bundles is required to be around 0.2°~0.4° to allow more than two view samples for each eye and to stimulate accommodation response, where the eyes are anticipated to focus at the rendered depth instead of the 2D screens [13]. It remains unclear that how accurate the rendered focus cues would be if such a view density is satisfied. Kim et al. experimentally measured the accommodative responses in viewing a real object and a digital 3D object rendered through integral imaging (InI) and suggested that over 73% of the 71 participants were able to accommodate at the depth of the rendered object instead of a display screen [21]. Kim et al. captured the point spread function (PSF) of different 3D displays and evaluated the DOF and the monocular focus cues of these displays based on their PSF measurements [22]. Stern at al. attempted to combine major perception and human visual system requirements with analytical tools to build an analytical framework for establishing perceivable light fields and determining display device specifications [23]. Overall, however, a critical gap is the lack of a systematic method to quantify the relationships between the accuracy of focus cue rendering and the number of samples per pupil area and a systematic investigation on the trade-off relations between view numbers and retinal image quality.

In this paper, we present a systematic approach to fully address the three aforementioned fundamental issues in designing a LF-3D display. By extracting the common characteristics of the existing LF-3D methods, we present a generalized framework to model their image formation process (Section 2), based on which we present a systematic method to simulate

and characterize the retinal image quality (Section 3) and the accommodative response (Section 4) perceived from a LF-3D display by accounting for both the ocular and display factors. We further employ this framework to investigate the optimal view sampling strategy for LF-3D display designs offering balance between accommodation cue accuracy and retinal image quality (Section 5). Our method is based on quantitative evaluation of the modulation transfer functions (MTF) of the perceived retinal image of light-field displays. Instead of performing an observer-based, objective measurement, we adopt a schematic eye model which takes into account most of the ocular factors such as pupil size, eye aberrations, diffraction and accommodation. Based on the MTF of the light-field images and the DOF of the human visual system, we determine the retinal image qualities and accommodative responses for different viewing conditions and system setups, and further decide the optimal view sampling guidelines for engineering LF-3D displays. It is worth mentioning that the methods and results can be generally applicable to both the emerging head-mounted LF-3D displays and the better-established eyewear-free direct-view LF-3D displays.

2. Generalized model and image formation process of light field displays

Several different methods have been explored to implement LF-3D displays, including SMV displays [11–15], integral-imaging (InI) based displays [16–18], and computational multi-layer light field displays [19,20]. Analogous to the camera array method for light field capture, an SMV display generally employs an array of 2D displays, each of which renders an elemental view of a 3D scene from a given viewpoint, to produce dense samples of the light rays apparently emitted by the scene. For instance, Lee et al. demonstrated the construction of a 100-inch, 300-Mpixel, and horizontal-parallax only light field display system by using 304 projectors arranged in a 19 by 16 array with about 0.17° horizontal parallax interval [14]. Instead of using an array of displays, alternatively, similar light field displays can be implemented in a time-multiplexed fashion. Jones et al. demonstrated a 360° horizontal-parallax light field display system by projecting light-field patterns onto a rapidly spinning anisotropic reflective surface through a single high-speed projector [15]. Most of the existing systems based on SMV-like method, however, only render the horizontal parallax of the light field due to the enormously increased complexity if vertical parallax would be considered. An InI-based display, utilizing the same principle as integral photography technique invented by Lipmamn in 1908 [24], typically consists of a display panel and a 2D array which can be a micro-lens array (MLA) [16,17] or aperture array [18]. While the 2D array angularly samples the directional light rays of a 3D scene, the display renders a set of 2D elemental images, each of which represents a different perspective of the 3D scene through each MLA or aperture. The conical ray bundles emitted by the corresponding pixels in the elemental images intersect and integrally create the perception of a 3D scene that appears to emit light and occupy the 3D space. An InI-based display using 2D arrays allows the reconstruction of a 3D shape with full-parallax information in both horizontal and vertical directions. A computational multi-layer display is a relatively new, emerging class of LF-3D display method that samples the directional rays through multi-layers of pixel arrays. It typically employs a stack of light-attenuating layers illuminated by a backlight [19]. The light field of a 3D scene is computationally decomposed into a number of masks representing the transmittance of each layer of the light attenuators. The intensity value of each light ray entering the eye from the backlight is the product of the pixel values of the attenuation layers at which the ray intersects. The computational multi-layer displays operate in a multiplicative fashion and approximate the light fields of the scene by angularly sampling the directions of the light rays apparently emitted by a 3D scene in a fashion similar to the InI-based display. The pinlight display, consisting of a spatial light modulator (SLM) and an array of point light sources, namely the pinlights, may also be considered as a computational multi-layer display where the back layer of light attenuators next to the backlight is replaced by a pinhole array or an array of point sources [20].

Although the various methods for implementing LF-3D displays summarized above appear to largely differ from each other, they all share the common characteristics of reproducing the directional light rays apparently emitted by a 3D object either in horizontal direction only or in both horizontal and vertical directions. Without loss of generality, we therefore simplify the image formation process of a LF-3D display into a generalized model adapted from the well-known 4-D light field function, *L(s, t, u, v)*, for representing the ray radiance as a function of positions *(s, t)* and directions *(u, v)*. As illustrated in Fig. 2(a), the generalized model can be divided into two subparts: a light field engine (in blue box) and an eye model (in red box).
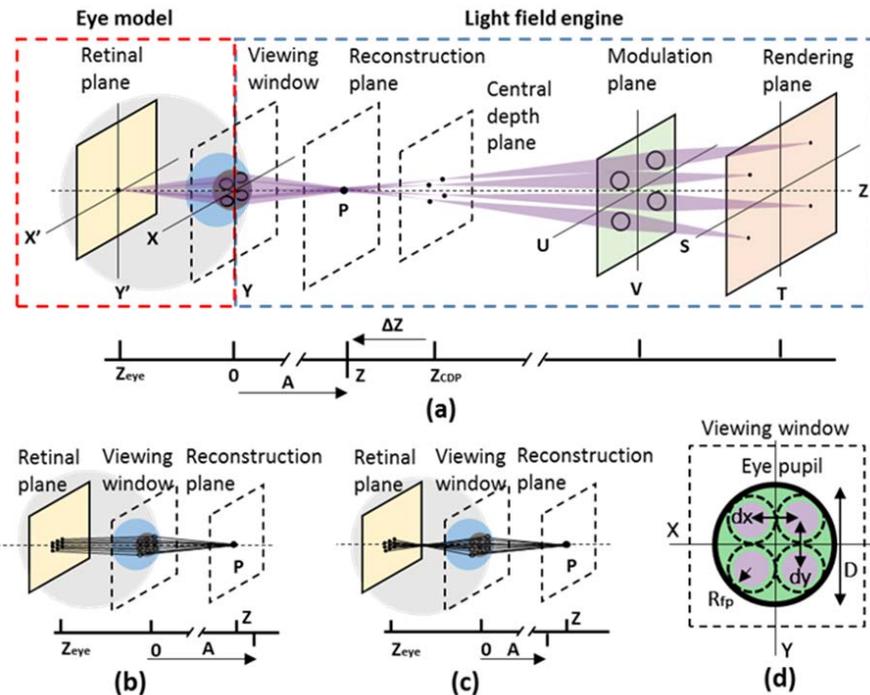

Fig. 2. The generalized schematic model of a light field 3D display, along with the illustration of the retinal response of a reconstructed point when the accommodative distance, *A*, of the eye is (a) the same as, (b) larger than, or (c) smaller than the depth of the reconstructed point. (d) The schematic drawing of the eye model entrance pupil and footprints of elemental views on the viewing window.

The light field engine minimally consists of a rendering plane, a modulation plane, a central depth plane, a reconstruction plane, and a viewing window. The rendering plane is where the elemental images are displayed and may be considered as the source defining the positional information of the light field function. It is an abstract representation of an array of displays in SMV systems [11–15], or a single physical display [16] or its virtual image through an eyepiece [17,18] being spatially divided into multiple regions in the case of InI-based displays, or the modulation layer closest to the backlight in computational multi-layer displays [19], or the point source array in pinlight displays [20]. The elemental images function as spatially-incoherent objects with each representing a unique perspective of a 3D scene. The modulation plane is where the directional samples of the light rays are produced and may be considered as the component defining the directional information of the light field function. It is an abstract representation of an array of optics [11–15], a lenslet array [16,17], an aperture array [18], or SLM(s) [19,20] in the different LF-3D display methods above. In SMV and InI-based LF-3D displays, the modulation plane consists of an array of optical

elements (e.g. lenses or apertures) each of which corresponds to an elemental image on the rendering plane and creates one directional sample of a 3D scene. In a computational multi-layer type displays, each pixel on the modulation plane may be considered as an element and defines one directional ray sample. The central depth plane (CDP) is viewed as a reference plane where the light rays created by a point source in the rendering plane converge after propagating through the modulation plane. It typically refers to the optical conjugate of the rendering plane through the modulation plane. It is where usually the highest spatial resolution of the reconstructed 3D scene can be obtained. Among the aforementioned three methods for LF-3D displays, both the SMV and InI-based displays have a well-defined CDP in the system. The computational multi-layer light field method, however, does not necessarily have a clearly optical conjugate plane in its system construction. To account for the diffraction effects of wave propagation through the layers of pixel structures, we can consider the plane located at the last surface of the SLM stack as its reference plane. The reconstruction plane is a representation of the depth location where a 3D point, P, is to be rendered through a LF-3D display. Its location clearly varies with the 3D point of interest. Unless the depth of the reconstruction scene coincides with that of the CDP, usually the spatial resolution of the reconstructed object will be degraded compared to that on the CDP. The last component of the light field engine is the viewing window, which defines the area with which a viewer observes the reconstructed 3D scene. It is commonly known as the eye box of the system in a head-mounted display or viewing zone in an eyewear-free direct view display. The ray bundle originated from a pixel on the rendering plane propagates through the modulating plane and projects a footprint on the viewing window resembling the geometric arrangement of the corresponding modulation elements on the modulating plane.

On the other hand, the eye model that simulates the optics properties of the human visual system is carefully placed so that its entrance pupil matches the location of the viewing window of the light field engine. We further assume that the entrance pupil area of the eye model is no larger than the viewing window area and thus determines the total number of elemental images or views being imaged at the retina. To better evaluate the eye's response in viewing a LF-3D display, a schematic eye model which is capable of adjusting its optical properties according to the eye's accommodation state is required.

Each pixel on the rendering plane is considered as an elemental point source and emits a conical ray rundle that propagates through an element on the modulation plane. As illustrated in Fig. 2(a), to reconstruct the light field of a 3D point, P, the rays emitted by the selected pixels located on different elemental images are directed by the corresponding elements on the modulation plane such that they intersect at the 3D position of reconstruction, or the reconstruction plane. The eye model then is anticipated to perceive the integral bundles of the rays originated from different pixels as if they were emitted by a source located at the point of reconstruction. The accommodated status of the eye model plays a critical role on the resulted retinal image. When the eye is accommodated at the depth of the reconstruction point, the retinal images of these spatially separated pixels will overlap with each other and form a focused image of the reconstruction point, as illustrated in Fig. 2(a). Otherwise, the retinal images of the individual pixels will be spatially displaced from each other, as illustrated in Figs. 2(b) and 2(C), and integrally create a retinal blur and the level of the retinal blur varies with the difference between the depths of the reconstruction and eye accommodation. In other words, the accommodation state of the eye model will not only change the perceived retinal image of the individual elemental pixel, but also the displacement between them. Under such circumstance, the overall retinal image of the reconstructed point shall be considered as an integral of the retinal images of individual elemental pixels with corresponding displacements. The retinal image quality apparently is expected to vary when the eye is accommodated at different depth and the resulted eye accommodation response to a reconstructed depth depends on the focus cues rendered by the display.

Based on the principle of light field reconstruction described above, we can characterize the retinal image properties of the reconstructed light field by its normalized, accumulated point spread function (PSF), $PSF_{LF\text{-}Accu}$, by integrating the PSFs of the retinal images of all the elemental pixels, PSF_c, assuming incoherent condition which almost all the LF-3D displays satisfy. For the convenience of characterizing the imaging properties of a LF-3D display and eye accommodative response from the point of view of a viewer, the center of the viewing window is set as the origin of the coordinate system OXYZ, the Z-axis is along the viewing direction, and the OXY plane is parallel to the viewing window. The depth of the CDP with respect to the eye is denoted as Z_{CDP}, and the depth of the eye accommodative state is denoted as A. The coordinates of an arbitrary point of construction, P, are defined as $P(x,y,z)$ while its depth is more conveniently defined by its axial displacement from the CDP, Δz. A positive Δz indicates the point of reconstruction is further away from the viewer than the CDP. We further define a reference frame, O'X'Y', for the retinal image plane, with the origin O' located at a distance, Z_{eye}, which is the effective distance from the retinal plane to the eye pupil.

The view sampling property of a LF-3D display is characterized by the view density, σ_{view}, defined as the number of views per unit area on the entrance pupil of the eye. The total number of elemental views entering the eye pupil can be determined by checking the number of elemental views encircled inside the eye pupil and can be explicitly expressed as

$$N_{view} = \frac{\pi D^2 \sigma_{view}}{4}, \tag{1}$$

where D is the entrance pupil diameter (EPD) of the eye. For instance, assuming a 3-mm EPD, a view density of 0.142mm^{-2} corresponds to only one view across the entire eye pupil, or a density of 1.274mm^{-2} corresponds to 9 view samples across the eye pupil.

The footprint of an individual view projected on the viewing window and the distribution of all the elemental views on the viewing window depend on the shape and arrangement of the modulation elements on the modulation plane, as shown in Fig. 2(d). For simplicity, we assume the elemental views are evenly distributed on the eye pupil in a rectangular array symmetric to the optical axis (z-axis), and their footprints are perfectly circular. Under such assumptions, the lateral displacements between two adjacent elemental views on the viewing window, Δd_x and Δd_y, along the x- and y-directions, respectively, are given by

$$\Delta d_x = \Delta d_y = 2\sqrt{\frac{1}{\pi \sigma_{view}}}. \tag{2}$$

The footprint size of an elemental view projected on the viewing window is characterized by its radial distance, R_{fp}, given as

$$R_{fp} = a\sqrt{\frac{1}{\pi \sigma_{view}}}, \tag{3}$$

where a is a scalar factor between 0 and 1 defining the fill factor of each elemental view. A fill factor of 1 suggests the footprint of each elemental view is the same as the pitch of the elemental views, while a fill factor less than 1 suggests that the footprint of each elemental view is smaller than the view pitch.

Based on the model described above, the elemental view of a LF-3D display with a view density greater than 0.142mm^{-2} typically under-fills a 3-mm eye pupil, allowing more than one view to be perceived simultaneously. For convenience, we assume the number of views entering the eye pupil is an integer rounded from the Eq. (1) so that complete sets of elemental views could be perceived. The accumulated PSF of the perceived light field on the retina, $PSF_{LF\text{-}Accu}$, can be therefore expressed as

$$PSF_{LF-Accu}(x',y') = \frac{1}{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K} L(m,n)w(\lambda_k)s(dc_{xm},dc_{ym})}$$

$$\cdot \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K} L(m,n)w(\lambda_k)s(dc_{xm},dc_{yn})\left|PSF_{Cmnk}(x',y')\right|^2, \tag{4}$$

where M and N are the total number of elemental views entering the eye pupil along the X and Y directions, respectively; K is the number of the sampled wavelengths, and PSF_{Cmnk} is the retinal PSF of a given elemental view indexed as (m,n) at a given wavelength, λ_k. In a LF-3D display, the weights of different elemental views to the accumulated retinal image may vary slightly to reproduce the subtle luminance differences of the object when viewed from slightly different positions. We thus introduce L, which may be regarded as the normalized luminance value of an elemental view indexed as (m,n) received by the eye from the point of reconstruction, to account for the weights of different elemental views to the accumulated PSF. w is the weighting function applied to the retinal PSF of an elemental view for the kth sampled wavelengths, λ_k, accounting for the relative luminous response of the human visual system to different illumination sources, and s is another weighting function applied to the PSF of a given elemental view indexed as (m,n) depending on its entry position, and dc_{xm} and dc_{yn}, on the eye pupil, to account for the directional sensitivity of the photoreceptors on the retina, known as the Stiles-Crawford effect [25]. For an elemental view indexed as (m,n), its entry position on the eye pupil, (dc_{xm}, dc_{yn}), is given by

$$dc_{xm} = \frac{(2m-M-1)}{2}\Delta d_x; \qquad dc_{yn} = \frac{(2m-M-1)}{2}\Delta d_y. \tag{5}$$

The exact form of the monochromatic, coherent, retinal PSF of an individual elemental view, PSF_c, through the combined model depends on both the light field engine and eye model adopted and is better computed numerically. Assuming that the light field engine is diffraction-limited (aberration-free) and shift-invariant, the retinal PSF of an individual elemental view can be modeled as

$$PSF_C(x',y') = \frac{1}{\lambda^2 z_{CDP} z_{eye}} \exp\left[j\frac{\pi}{\lambda z_{CDP}}\left(\Delta x_c'^2 + \Delta y_c'^2\right)\right]\exp\left[j\frac{\pi}{\lambda z_{eye}}\left(x'^2 + y'^2\right)\right]$$

$$\cdot \int_{-\infty}^{\infty} P_{eye}(x,y)\exp\left[j\frac{2\pi}{\lambda}W_{eye}\left(\Delta x_c',\Delta y_c';x,y\right)\right]\exp\left[j\frac{\pi}{\lambda}\left(\frac{1}{z_{CDP}}-\frac{1}{A}\right)(x^2+y^2)\right] \tag{6}$$

$$\cdot \exp\left\{-j\frac{2\pi}{\lambda}\left[(\frac{\Delta x_c'}{z_{CDP}}+\frac{x'}{z_{eye}})x + (\frac{\Delta y_c'}{z_{CDP}}+\frac{y'}{z_{eye}})y\right]\right\}dxdy,$$

where $\Delta x_c'$ and $\Delta y_c'$ are the lateral displacements of an elemental view of a given reconstructed point from z-axis on the CDP, P_{eye} is the footprint of each elemental view upon the pupil of the eye model, and W_{eye} is the wavefront function standing for the amount of aberrations in the adopted eye model measured at the pupil, which may vary from model to model, and is influenced by various ocular factors such as accommodation state or EPD. Following the same assumption about view distribution as for Eqs. (2) and (3), $\Delta x_c'$, $\Delta y_c'$ of an on-axis reconstructed target point and P_{eye} for an elemental view indexed as (m,n) can be further expressed by Eqs. (7) and (8), respectively, as,

$$\Delta x_{cm}' = \frac{dc_{xm}\Delta z}{z_{CDP}+\Delta z}; \quad \Delta y_{cn}' = \frac{dy_{xm}\Delta z}{z_{CDP}+\Delta z}, \tag{7}$$

$$P_{eye(m,n)}(x,y) = circ\left(\frac{\sqrt{(x-dc_{xm})^2+(y-dc_{yn})^2}}{R_{fp}}\right), \tag{8}$$

where *circ* is the circular function working as a binary sub-aperture window. P_{eye} helps define the entry position for each of the elemental view at the entrance pupil of the eye model and thus determines which sub-part of the eye model will interact with the corresponding elemental view and changes the form of its corresponding PSF_c calculated from Eq. (6).

Based on the image formation process above, the MTF of the perceived light field on retina, characterized by the ratio of the contrast modulation of the retinal image to that of a sinusoidal object rendered on the display, can be obtained by applying Fourier transform to Eq. (4), which is expressed as

$$MTF_{LF-Accu}(\xi',\eta') = \frac{\int\int\limits_{-\infty}^{\infty} PSF_{LF-Accu}(x',y')\exp\left[-j(\xi'x'+\eta'y')\right]dx'dy'}{\int\int\limits_{-\infty}^{\infty} PSF_{LF-Accu}(x',y')dx'dy'}, \tag{9}$$

where ξ' and η' are the spatial frequencies measured on the retina in the x'- and y'-directions, respectively.

It is worth pointing out that the accumulated PSF and MTF of the perceived light field on retina, $PSF_{LF-Accu}$ and $MTF_{LF-Accu}$, calculated by Eqs. (4) and (9), respectively, only depend on the CDP location (Z_{CDP}), the view density of light field sampling (σ_{view}), the fill factor of each elemental view (a), the axial displacement of the reconstruction depth from the CDP (Δz), the eye pupil diameter (D), the spectral property of the source (λ), the eye accommodation status (A), as well as the wavefront function of the adopted eye model (W_{eye}). The depths of the rendering plane and the modulation plane have no influence upon the retinal image, however, the numerical aperture (NA) of the modulating elements, NA_{mod}, on the modulation plane is required to match the view density σ_{view} or footprint size R_{fp} such that

$$NA_{mod} = \frac{2a}{Z_{CDP}}\sqrt{\frac{1}{\pi\sigma_{view}}} = \frac{2R_{fp}}{Z_{CDP}}. \tag{10}$$

In summary, the generalized model of a LF-3D display described above enables us to quantify the perceived retinal image quality of a LF-3D display by an observer, assess whether the eye would properly accommodate at the depth of a 3D reconstruction, quantify the accuracy of the resulted focus cues when viewing a LF-3D display, and investigate how key system design parameters such as view density and fill factor of each elemental view will influence the retinal response in viewing LF-3D displays.

3. Characterizing the perceived retinal images of a light field display

Equation (4) demonstrates that the overall retinal response of the eye viewing a LF-3D display is actually the integral of the weighted PSFs of all the elemental views reconstructing a 3D point, which suggests that there exist fundamental differences between the reconstructed 3D scene by LF-3D displays and a natural 3D scene, or a 2D scene (a displaying plane) by traditional displays. Assuming that the accommodative depth of the eye, *A*, coincides with the reconstruction depth, this section will focus on characterizing the perceived retinal image quality of a LF-3D display.

To simulate and characterize the perceived retinal image of a LF-3D display based on the generalized model in Fig. 2, we created a multi-configuration zoom system in Code V® [26], in which each configuration represents the path of an elemental view perceived by a shared eye model. For simplicity and without loss of generality, the light field engine was

implemented using the InI-based method similar to the one described in [17]. The rendering plane was modeled by fiducial display with infinite pixel resolution and the modulation plane was assumed to be an aberration-free MLA. A circular aperture was assumed for each element of the MLA. The lenslet pitch and NA of the MLA were set to match the view density and fill factor of interest given by Eqs. (2) and (10), respectively. The microdisplay in the model was set up with five wavelengths, 470nm, 510 nm, 555 nm, 610 nm, 650 nm, respectively, to simulate a full-color LF-3D display, of which the relative weights were set according to human eye's photopic response curve [25] to reflect the polychromatic responses in Eq. (4). With this simplified construction of a light field engine, the CDP was the optical conjugate of the microdisplay and the PSF of each elemental view through the lenslet was simplified to be diffractive-limited. Reconstructed 3D targets for LF-3D displays were also assumed on or near z-axis. In addition, a schematic eye model capable of varying its optical properties according to the eye's accommodation state was required to complete the model for perceived retinal image assessment. Several schematic eye models have been widely used in the fields of visual and ophthalmic optics to predict the performance of an optical system involved with human observers [25, 27]. Among the various eye models, the Arizona eye model was selected, which was designed to match clinical levels of aberrations for both on- and off-axis fields and provides the capability of changing the accommodative distance, A, of the eye by varying the shape and refractive index of the crystalline lens [25]. The parameters and dimensions of the Arizona eye model have been chosen to be consistent with the population-average data so that the conclusions resulted from such a model could be applied to the majority of the potential viewers. The eye model optics was integrated into the multi-configuration zoom system in Code V. The entrance pupil of the eye model was set at 1 diopter (1m) away from the CDP of the light field engine. The accommodative distance of the eye, which determines the lens shape, conic constant and refractive index of the surfaces in the schematic eye, can be varied as needed. We set the entrance pupil diameter of the eye model to be 3mm, which corresponds to the average pupil size when viewing typical displays with luminance around 200 cd/m^2. In order to account for Stiles-Crawford effect [25], a Gaussian apodization filter with an amplitude transmittance coefficient of $\beta = -0.116$mm^{-2} was selected as the weighting function s in Eq. (4). Hence the elemental view that passes through the central part of the eye pupil would have a larger contribution to the accumulated PSF than a view through the edge of the eye pupil.

Based on the setup described above, for a given display configuration, the depth of a reconstruction point, the accommodation state of the eye, and the pixel locations of the elemental views on the rendering plane to reconstruct the point of interest are determined. The retinal PSF of each elemental view given by Eq. (6), can be simulated independently using CODE V, and the accumulated PSF of the retinal image of such a reconstructed point given by Eq. (4) was obtained by integrating the retinal PSFs of all the elemental views passing through the eye pupil, and the corresponding MTF of the perceived LF-3D display was computed using Eq. (9).

We began with investigating how the spatial resolution of a LF-3D display is affected by the view density and the depth of reconstruction. Figure 3(a) plotted the polychromatic MTF of the retinal image as a function of spatial frequencies for reconstructed 3D targets located on the CDP for LF-3D displays with different view densities ranging from 0.142 to 2.26mm^{-2}, corresponding to footprint pitch from 3mm to 0.75mm. In this simulation, the fill factor of the elemental views was set to be 1 (i.e. a = 1) and the accommodative distance, A, of the eye model was assumed to coincide with the depth of the CDP. The MTF plot for the view density of 0.142mm^{-2} is equivalent to the MTF when viewing a traditional 2D display placed at the same depth as the CDP. Apparently, as the view density increases, the MTF of the perceived light field reconstruction at the depth of the CDP decreases rapidly due to the increasing effects of diffraction with small sub-apertures of elemental views. For instance, with a view density of 2.26mm^{-2}, a total of 16 views were integrated to reconstruct a 3D point

located on the CDP and the MTF of the retinal image drops down to zero at approximately 23 cycles/degree mainly due to diffraction effects by the small NA of the modulation elements.
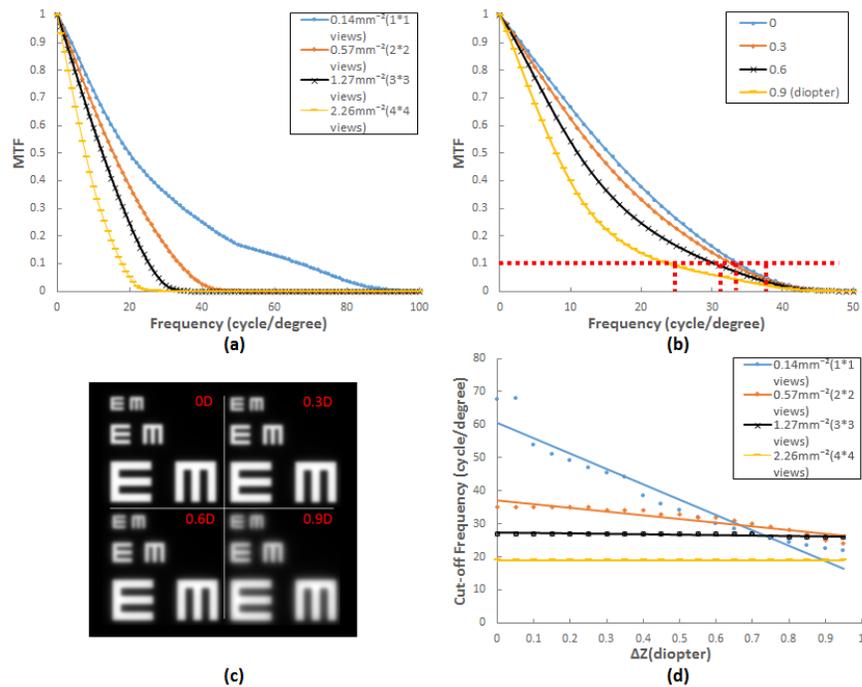

Fig. 3. (a) The MTF plots of the retinal image of light field displays with different view densities ranging from 0.142 to $2.26mm^{-2}$ for reconstructed 3D targets located on the CDP. (b) The MTF plots of the retinal image of a light field display with a view density of $0.57mm^{-2}$ for reconstructed targets located at four different depths from the CDP: 0, 0.3, 0,6, and 0.9 diopter, respectively. (c) Simulation of the perceived retinal images of a series of Snellen "E" s of three different angular resolutions located at four different depths from the CDP. (d) The cut-off frequencies of light field display with different view densities as function of the depth displacement from the CDP.

Figure 3(b) plotted the polychromatic MTF of the retinal image as a function of spatial frequencies for reconstructed 3D targets located at four different depths from the CDP: 0, 0.3, 0.6, and 0.9 diopters, respectively. All the displacements were on the close side of the CDP with respect to the eye. The accommodative distance, A, of the eye model was varied so that it coincided with the depth of the reconstructed targets: 1, 1.3, 1.6, and 1.9 diopters, respectively. The view density was assumed to be $0.57mm^{-2}$ with a fill factor of 1, corresponding to a view pitch of 1.5mm, which yields a total of 4 views over the eye pupil. It is clear that the perceived MTF of such a LF-3D display decreases rapidly as the reconstruction point is displaced away from the CDP, which correlates to a degradation of spatial resolution. By setting a minimal threshold value for the MTF of a LF-3D display, for example, 0.1 as shown by the red dotted line in Fig. 3(b), the cut-off frequencies, or the limiting spatial resolution, for reconstructed 3D points at different depths can be determined. For instance, with the given view density of $0.57mm^{-2}$ and a minimum MTF threshold of 0.1, the cut-off frequencies for the reconstructed scene are approximately 34, 33, 30, and 24 cycles/degree for the depth displacement of 0, 0.3, 0.6, and 0.9 diopters from the CDP, respectively. Figure 3(c) shows the simulation of the perceived retinal images by convolving a series of Snellen 'E's of three different angular resolutions (10, 20 and 30 cycles/degree) with the accumulated PSF corresponding to of the MTF curves as in Fig. 3(b). Retinal image

quality degradation can be clearly observed as the reconstruction depths of the targets are displaced away from the CDP.

Using the same MTF threshold of 0.1, Fig. 3(d) plotted the cut-off frequencies of a LF-3D display as a function of dioptric depth displacement from the CDP for different view densities ranging from 0.142 to 2.26mm^{-2}. For a given view density, the cut-off frequency generally decreases as the depth displacement increases. As the view density increases, the cut-off-frequencies on the CDP, which corresponds to the maximum constructible frequency, decrease rapidly due to the increasing effects of diffraction with small sub-apertures of elemental views. On the other hand, as the view density increases, the spatial resolution degradation, characterized by the slope of the curves, becomes less sensitive to depth displacement from the CDP, which can be explained by the increasing DOF of each elemental view with a decreasing sub-aperture. The cut-off frequencies become nearly independent of the depth displacements when the view density is about 2.26mm^{-2} or greater.

Based on the results shown in Fig. 3, the DOF of a LF-3D display can be defined as the depth range within which a given spatial resolution criteria can be achieved. For instance, the DOF of a LF-3D display constructed with a view density of 0.57mmmm^{-2} is approximately $\pm$ 0.65 diopters to achieve a minimal spatial resolution of 35 cycles/degree, while it is about $\pm$ 1 diopters to achieve a minimal spatial resolution of 25 cycles/degree. Although a larger view density offers a larger DOF for the same resolution criteria, it comes at the cost of lower image resolution on the CDP and also lower image contrast. A lower view density offers higher image resolution on the CDP, but compromises the DOF and potentially larger accommodation error, which to be further discussed.

4. Characterizing the accommodative response of a 3D light field display

The focusing cues rendered by a LF-3D display can be characterized by two components. The first part is the cue rendered by the reconstruction depth at which the rays of the elemental views appear to converge, namely the accommodation cue. The other is the blur cue rendered by the perceived retina image which stimulates the eye to change its accommodation response to maximize retinal image sharpness. The retinal image blur is the sole true stimulus to accommodation and is the visual information used to effect predictable changes in the accommodative response. A light field reconstruction presumably generates the elemental views such that the accommodation cue is correctly rendered. The actual accommodative response, however, is affected by the actual retinal image blur rendered by a LF-3D display which may possibly drive the eye to accommodate on neither the CDP nor the depth of the reconstructed point, but somewhere in between to balance between the two aforementioned components of focusing cues. The mismatch between the actual accommodative response of a viewer to a LF-3D display and the accommodation cue rendered by the images is considered as accommodation cue error which potentially leads to residual VAC problem. One of the main objectives of this section is to characterize the accommodative response and quantify the accommodation cue error rendered by a LF-3D display with respect to key systematic parameters.

Generally, the contrast gradient and contrast magnitude of the retinal image are the key factors that drive and stabilize eye accommodation response. The eye tends to adjust its accommodation to maximize these two factors in the focusing process. Through the MTF in Eq. (9), it becomes relatively straightforward to quantify the contrast magnitude and gradient of the retinal image with respect to eye accommodation status and to characterize the eye accommodative response to a LF-3D display. Based on the same setup described in Sect. 3, the change of retinal image properties and the accommodative response of a LF-3D display can be characterized by varying the accommodation status of the eye model through the depth of interest. The actual accommodative distance, A_m, that offers the maximum image contrast and gradient is located and is considered to be depth at which the eye is likely to

accommodate. The image contrast and gradient resulted from the actually accommodative distance are considered to be focus cues rendered by the LF-3D displays.

To investigate the accommodative response of the eye to a LF-3D display, Fig. 4(a) through 4(c) plotted the MTF of the retinal image of a LF-3D display as a function of eye accommodation shift for reconstructed targets located at three different depths away from the CDP by a dioptric displacement of 0, 0.5, and 1 diopters, respectively. All the displacements of the reconstructed depths were on the same and the close side of the CDP with respect to the eye. Each figure plotted the responses to targets of 5 different spatial frequencies ranging from 5 to 25 cycles per degree (cpd) at a 5 cycles/degree increment. The sampled frequencies were selected based on the guidance provided by Fig. 3(c) for the cut-off frequencies. The CDP of the light field engine was set to be 1 diopter (1m) away from the eye as well as the view density to be 0.57mm^{-2} on the eye pupil, which corresponds to 4 different views passing through the area of a 3-mm eye pupil. The horizontal axis of the plots, denoted as the accommodation shift in diopters, is defined as the dioptric displacement of the eye accommodative distance, A, from the corresponding depth of the reconstruction target. A negative accommodation shift suggests that the eye accommodative distance is shifted closer toward to the CDP. The vertical axis is the polychromatic MTF value of the retinal image for a given eye accommodative distance. The black arrows marked the location where maximum retinal image contrast is achieved for each of the target frequencies. As a comparison, we further examined the MTF curves of the retinal images of real targets placed at 1, 1.5 and 2 diopters away from the eye, as a function of the eye accommodative distance for the same frequency range. The depths of these real retargets correspond to the same depths of the reconstruction planes in Fig. 4(a)-4(c). The same eye model was utilized. The only difference from the Fig. 3(a) is that the target is a real target, rather than being reconstructed by multiple elemental views of small NAs. As an example, Fig. 4(d) plotted the MTF curves of the retinal image of a real target, placed at the same depth as the CDP, as a function of the eye accommodative distance. It is worth pointing out that the accommodative responses of the eye to real targets of other depths are very similar to the one shown in Fig. 4(d) and are omitted without redundancy.

When the reconstructed target is located on the CDP ($\Delta z = 0$ diopter) as shown in Fig. 4(a), it can be observed that the MTF values at different frequencies reach their maximums when the eye is accommodated at the reconstruction depth but gradually decrease as the accommodation shift increases on either side of the reconstruction plane. Across the entire frequency range of simulation, the contrast gradient, which primarily drives the eye to accommodate at the proper depth, of the retinal image of the reconstructed light field on the CDP is lower than that of a natural target shown in Fig. 4(d) but with similar trend. However, at any given eye accommodative distance, the image contrast of the reconstructed light field was noticeably lower than that of a natural target for all the frequencies and the contrast degradation is more prominent in the high frequency range. For instance, the maximum contrast for the frequency of 25 cycles/degree is about 0.2 for the reconstructed target while it is over 0.35 for a real target. Such noticeable difference in image contrast can be explained by the diffraction effects due to the substantially smaller NA of each elemental view than that of a real object to the eye pupil as a whole. Overall, Fig. 4(a) indicates that a reconstructed target on or near the CDP by a LF-3D display can provide correct focus cues, comparable to a natural target.

As the displacement of the reconstruction depth from the CDP significantly increases, as shown in Figs. 4(b) and 4(c), however, the focus cues rendered by a LF-3D display may not be in agreement with the depth of rendering. It can be clearly observed that both the contrast magnitude and gradient of the retinal image are noticeably lower, especially in the 10-20 cycles/degree frequency range. Furthermore, the peak responses are shifted away from depth of the reconstruction. Instead, they are shifted towards the location of CDP, leading to a negative accommodation shift. For instance, as shown in Fig. 4(c) with 1 diopter

displacement from the CDP, the accommodative distance corresponding to the maximum image contrast was shifted away from the depth of reconstruction by as large as about 0.2 diopters for targets of 5 cycles/degree or over 0.05 diopters for targets of 15 cycles/degree. This indicates that the focus cues rendered by a LF-3D display for 3D targets located far from the CDP bear a significant error compared to their rendering depth, which limits the ability to address the VAC problem.
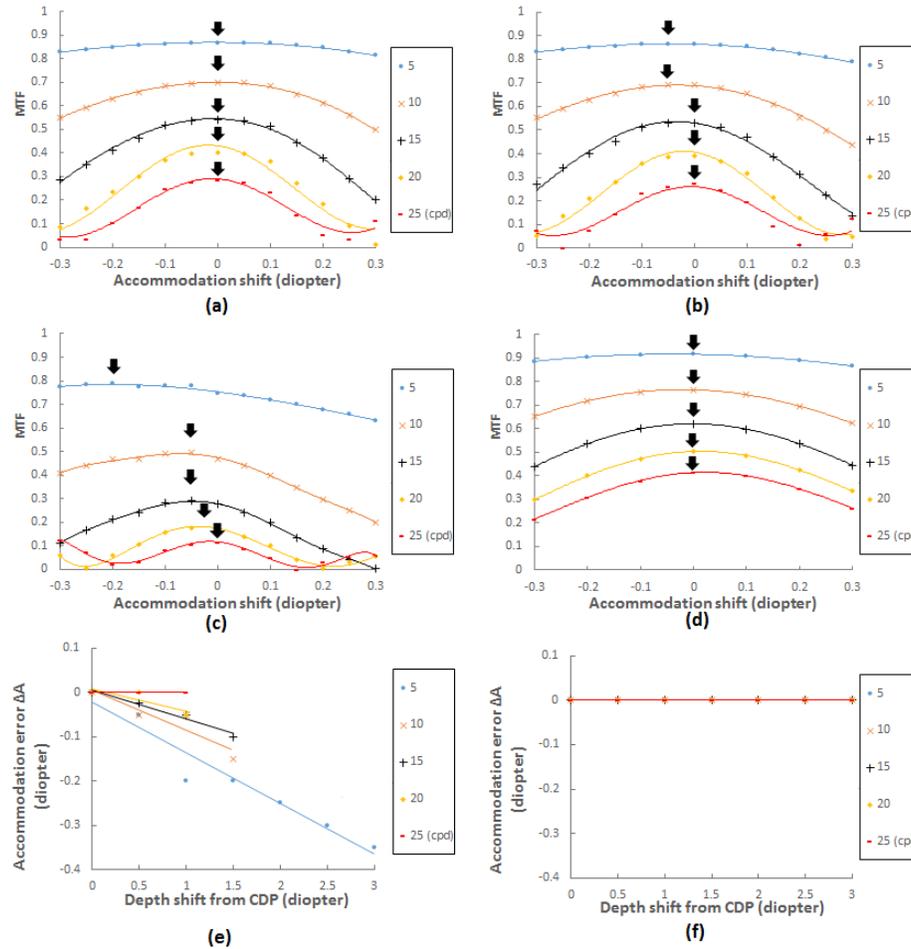

Fig. 4. (a)-(d) The accommodative response curves for different spatial frequencies (cpd stands for cycles per degree) with the reconstruction plane located at (a) $\varDelta z = 0$ diopter, (b) $\varDelta z = 0.5$ diopters, (c) $\varDelta z = 1$ diopter, and (d) with a real target located on the CDP. (e)-(f) Accommodation error for different spatial frequencies as a function of the dioptric displacement of a target depth from the CDP toward the eye for (e) targets reconstructed by a LF-3D display and (f) real targets.

To better illustrate the accuracy of focus cues rendered by of a LF-3D display in relation to the depth of reconstruction, the axial displacement between the actual accommodative distance, A_m, and the reconstruction distance is defined as accommodation error $\varDelta A$. A_m corresponds to the locations marked by the black arrows in Figs. 4(a)-4(d) and is the depth where the eye will practically accommodate, while the reconstruction distance is where the eye theoretically shall accommodate. Along with the contrast magnitude and gradient of the resulted retinal image, $\varDelta A$ provides an objective measurement on whether a given display configuration is able to provide accurate and distinctive focus cues. Figure 4(e) plotted the

accommodation error as a function of the dioptric displacement of a reconstructed target from the CDP for the spatial frequencies ranging from 5 to 25cycles/degree. The depths of the reconstructed targets for simulation were shifted away from the CDP by a magnitude up to 3 diopters at a 0.5 diopter increment toward the eye, corresponding to the reconstruction depth range from 1 to 3.5 diopters away from the eye. As the reconstructed target is displaced away from the CDP by more than 1 diopter, the 3D-LF display configuration used in the simulation is unable to reconstruct targets of high spatial frequencies, which was suggested by the cut-off frequencies in Fig. 3. Therefore Fig. 4(e) only plotted accommodation errors for targets within its corresponding reconstructable depth range for a given frequency. As a comparison, Fig. 4(f) plotted the accommodation error for real targets placed at the same dioptric depths. It is worth noting that the curves for all of the spatial frequencies overlap with each other due to the negligible errors for real targets. The results clearly suggest that the accommodation error increases noticeably as the increase of the dioptric displacement from the CDP for targets rendered by a LF-3D display, while the accommodation errors for real targets are negligible as expected. Although the plots in Fig. 4(e) only considered positive dioptric displacement of a reconstructed target from the CDP (i.e. moving closer to the eye from the CDP), we anticipate nearly symmetric performance for negative displacements from the CDP (i.e. moving further away from the CDP).

Because the relatively low frequency content owns better image contrast in general, there always exists a trade-off between focus cue accuracy and retinal image quality of a light field system. According to the neural transfer function (NSF), which is the contrast of the effective neural image divided by retinal-image contrast as a function of spatial frequency, human eye is most sensitive to the mid-range frequency around 10-15 cycles/degree [28]. This also agrees with the Fig. 4(d), where the MTF curve of the mid-range frequencies has the largest gradient that will effectively drive the eye accommodation response toward the depth which yields maximal image contrast. Therefore, in the following section, we will target on the mid-range frequencies to investigate how we can further reduce the accommodation error by optimizing the system parameters of LF-3D displays.

5. Optimal view sampling of light-field displays

As discussed in Sections 3 and 4, the spatial resolution of a LF-3D display and the accuracy of the focus cues rendered by the display not only vary greatly with the displacement of the reconstruction depth of a 3D scene but also with the view density of the display design. This section aims to investigate the optimal view sampling for LF-3D display designs.

The accommodation cue of a LF-3D display originates from the intersection of multiple elemental views at the depth of reconstruction. Therefore, the number of views that fill the eye pupil or the view density plays a key role in influencing not only the spatial resolution and DOF of a LF-3D display but also the accuracy of focus cues rendered the display to drive proper eye accommodative response. In viewing a natural scene of different depths in the real word, the eye observes infinite number of views entering the eye pupil from the scene, which yields subtle yet accurate focus cues to drive eye accommodative responses. Naturally, in designing a LF-3D displays, the more the number of views or the higher the view density with which we sample the light field to be rendered, the less the accommodation error and thus the more accurate the accommodation cue may be anticipated. Due to the diffraction effect, however, a higher view density may lead to a lower spatial resolution overall. It is of great importance to balance between the accommodation cue and spatial resolution, and therefore the number of views or view density needs to be optimized.

Figures 5(a) and 5(b) plotted the accommodative response of LF-3D displays with different view densities ranging from 0.57 to 2.26mm^{-2}, corresponding to sub-aperture diameter from 1.5mm to 0.75mm (or 4 to 16 views across the eye pupil). The fill factor of the elemental views was assumed to be 1. The CDP of the display was fixed at 1 diopter (1m) and the diameter of the entrance pupil of the eye was set as 3mm. The reconstruction depth of the

target was set on the CDP and 1 diopter away from the CDP for Figs. 5(a) and 5(b), respectively, while both figures plotted the MTF values for the mid-range frequency of 15 cycles/degree.
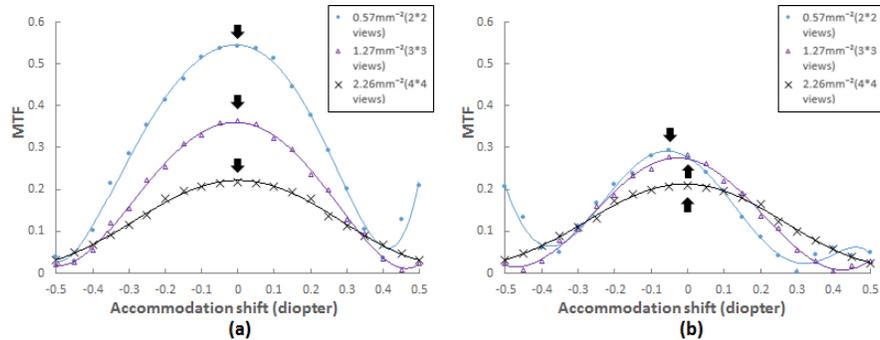

Fig. 5. The accommodative response curves for different view densities with the reconstruction plane at (a) $\Delta z = 0$ diopter and (b) $\Delta z = 1$ diopter. The target frequency was set at 15 cycles/degree.

When the reconstructed target is located on the CDP ($\Delta z = 0$ diopter) as shown in Fig. 5(a), the MTF values in all of the three sampled view densities reach their maximum when the eye is accommodated at the reconstruction depth. As expected, the contrast magnitude and gradient decrease noticeably as the view density increases due to increasing diffraction effects. Overall, all the sampled view densities can provide adequate retinal image contrast and gradient to stimulate correct eye accommodation. When the target reconstruction depth is displaced by 1 diopter from the CDP as shown in Fig. 5(b), however, noticeable accommodation error is observed for displays with low view density or small number of views. As the view density increases, the accommodation error is noticeably reduced. The magnitude and gradient of the image contrast degrade just slightly as the increase of view density. For instance, in a display configuration with a view density of 0.57mm^{-2}, the accommodation error is about 0.05 diopters for targets rendered at 1 diopter away from the CDP. As the displacement from the CDP increases, the magnitude of the accommodation error, the image contrast and contrast gradient degrade dramatically. In a display configuration with a view density of 1.27mm^{-2}, the accommodation error of mid-range frequency content is nearly zero for targets rendered on the CDP and for targets rendered 1 diopter away from the CDP, while the magnitude and gradient of the image contrast are slightly worse than those for a display configuration with a view density of 0.57mm^{-2}. However, the advantage of creating a high view density ceased when diffraction effects dominate. For example, as shown in Fig. 5, a display configuration with a view density of 2.26mm^{-2} lead to noticeably worse image contrast and contrast gradient than a configuration with a view density of 1.27mm^{-2} for targets located on the CDP and 1 diopter away from the CDP. Therefore, the accommodation cue rendered by such a display is not necessarily improved at the cost of spatial resolution.

Figures 6(a)-6(c) plotted the accommodation error, the maximum contrast magnitude, and maximum contrast gradient of the retinal image, respectively, as a function of the axial displacement of the reconstruction depth from the CDP for different view densities from 0.57 to 2.26mm^{-2}. On each figure, the results for targets rendered at four different depths: 0, 033, 0.67, and 1 diopter from the CDP, respectively, were plotted. The red dotted lines in the plots stand for the accommodative curve of viewing an ideal real target. The simulation setup remained the same as those in Fig. 5. The results in this figure can be utilized as a general guideline when selecting the appropriate view density for designing a LF-3D display system. What is desired is a LF-3D display system that generates no accommodation error and also a

high, constant contrast contents along a large DOF. There obviously exists trade-off between these two factors.
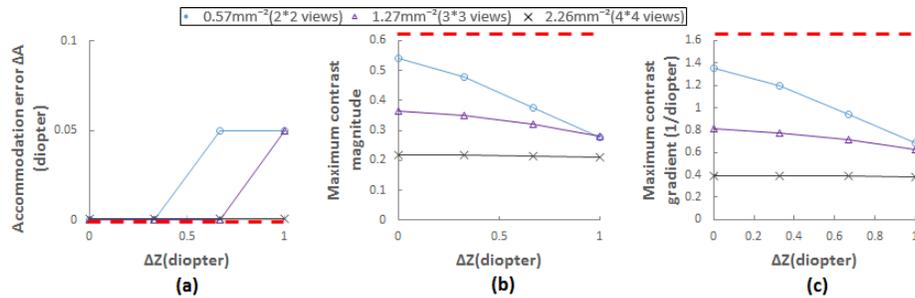

Fig. 6. The plots of (a) accommodation error, (b) the maximum contrast magnitude and (c) maximum contrast gradient for *Δz* from 0 to 1 diopter with varying view densities.

All the simulations above assume a fill factor of 1 for the elemental views. In fact, the fill factor, *a*, of the elemental views defined in Eq. (3) can have significant influence on the perceived light field image. Various methods [29, 30] that involve in adjusting the effective fill factor have been proposed to improve different aspects of the LF-3D display systems. Changing the fill factor will directly redefine the effective working NA of element views (Eq. (10)) and thus dramatically impact the integral light field image. Reducing the fill factor of elemental views is considered as being a simple, straightforward way to improve the accuracy of the accommodation cue, especially for a system with a low view density at the cost of reduced display brightness and potentially compromised spatial resolution.
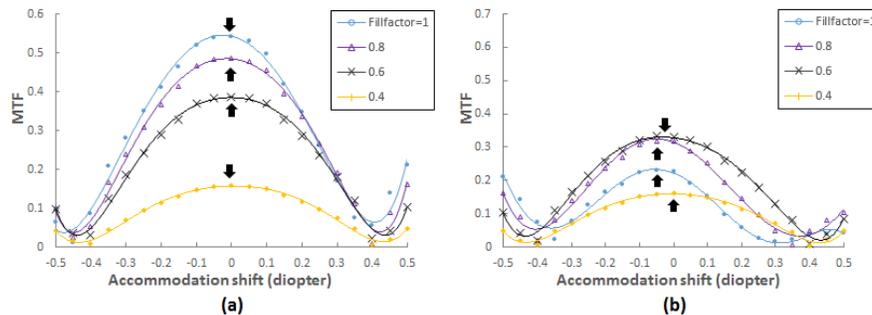

Fig. 7. The accommodative curves for displays with a view density of 0.57mm⁻² and different fill factors ranging from 1 to 0.4 with the reconstruction plane at (a) *Δz* = 0 diopter, (b) *Δz* = 1diopter. The target frequency was set at 15 cycles/degree.

In order to investigate the effects of the fill factor of the modulation element, we slightly modified the model setup in CodeV by reducing the effective aperture size of each lens element of the MLA on the modulation plane while maintaining the same lens pitch by fixing the amount of decenter among the zooms. Figures 7(a) and 7(b) show the result of the accommodative response of display configurations with different fill factors ranging from 1 to 0.4. The CDP of the display was fixed at 1 diopter (1m), the diameter of the entrance pupil of the eye was set as 3mm and the view density was set to be 0.57mm⁻² since under such circumstance, the accommodation error is noticeable as shown in Fig. 6. The reconstruction depth of the target was set on the CDP and 1 diopter away from the CDP, for Figs. 7(a) and 6(b), respectively, while both figures plotted the MTF values for the mid-range frequency of 15 cycles/degree.

Overall, the results in Fig. 7 suggest that reducing the fill factor of elemental views for displays of the same view density can noticeably reduce the accommodation error.

Furthermore, for targets with large depth displacement from the CDP (Fig. 7(b)), properly picking the fill factor can also improve the magnitude and gradient of the retinal image by obtaining good balance between the diffraction and aberration effects. For instance, in a display configuration with a view density of $0.57mm^{-2}$, by reducing the fill factor to 0.6, the accommodation error for targets rendered at 1 diopter away from the CDP can be reduced to nearly zero and the contrast magnitude and gradient of the retinal image are also noticeably improved. In this case, it might be advantageous to adopt a fill factor of 0.6 that offers negligible accommodation error and high image contrast for a depth range of ± 1 diopters from the CDP. For a display configuration with a view density of $1.27mm^{-2}$ or more, reducing the fill factor mainly improves the accommodative error for targets with large displacement from the CDP, but have little improvements or even adverse impacts on the image contrast and contrast gradient since the diffraction effect starts to dominate and thus does not show noticeable benefits.

6. Conclusion

We describe a generalized framework to model the image formation process of light field display methods and present a systematic method to simulate and characterize the retinal image quality and accommodation response rendered by a LF-3D display. We further employ this framework to investigate the trade-offs and guidelines for optimal view sampling in the design of LF-3D displays. By taking both ocular and display factors into account, we determine that increasing the view density generally leads an increase of DOF and a reduction of accommodation error at the cost of spatial resolution and image contrast. The maximally achievable spatial resolution decreases with the increase of view density until the system reaches diffraction domination at a view density of $1.27mm^{-2}$, corresponding to 3 by 3 views across the eye pupil. At a view density of $1.27mm^{-2}$ or higher, the obtainable spatial resolution becomes nearly independent of the depth displacements, but the corresponding image contrast may be significantly compromised compared to a display of lower view density. For displays that can only afford lower view density such as $0.57mm^{-2}$ or lower, the fill factor of the elemental views may be reduced properly to improve the magnitude and gradient of the retinal image and reduce the accommodation error. Overall, the paper provides a framework that can objectively predict the performance and guide the design of LF-3D displays by optimizing the key parameters. In the future, we plan to build a high quality, compact LF-3D display and carry out further experiments to validate our methods.

Disclaimer

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.