INVITED PAPER

Enabling Focus Cues in Head-Mounted Displays

The nature of vergence–accommodation conflict problem in head-mounted displays and the associated visual artifacts are summarized, followed by a comprehensive review of the various technical approaches toward rendering proper focus cues in head-mounted displays, for both virtual reality (VR) and augmented reality (AR) applications.

By Hong Hua

ABSTRACT | Developing head-mounted displays (HMDs) that offer uncompromised optical pathways to both digital and physical worlds without encumbrance and discomfort confronts many grand challenges, both from technological perspectives and human factors. Among the many challenges, minimizing visual discomfort is one of the key obstacles. One of the key contributing factors to visual discomfort is the lack of the ability to render proper focus cues in HMDs to stimulate natural eye accommodation responses, which leads to the well-known problem of vergence–accommodation conflict. This paper provides a comprehensive summary of various technical approaches toward enabling focus cues in HMDs for both virtual reality (VR) and augmented reality (AR) applications.

KEYWORDS | Accommodation and convergence; augmented reality; focus cues; head-mounted displays (HMDs); near-to-eye displays; virtual reality (VR)

I. INTRODUCTION

The human visual system (HVS) exploits a wide range of cues for 3-D percepts and depth estimation. These depth cues can be categorized into psychological and physiological types. Examples of psychological cues include perspective, occlusion, aerial perspective, shadow and shading, apparent size, texture gradient, while examples of physiological cues include binocular disparity, motion parallax, retinal blur effect, accommodation, and vergence [1]. Among these cues, the accommodation cue refers to the focus action of the eye where ciliary muscles change the refractive power of the crystalline lens and minimize the amount of blur for the fixated depth of the scene. Associated with eye accommodation change is the retinal image blur effect in which objects away from the eye's accommodation depth appear blurry in the retinal image due to the limited depth-of-field (DOF) of the eyes. The degree of image blur varies with the distance of the object from the focused distance. The accommodation and retinal image blur effects together are known as the focus cues. The vergence cue refers to the rotation action of the eyes to bring the visual axes inward or outward to intersect at a 3-D object of interest at near or far distances.

This review paper is particularly interested in the well-known vergence–accommodation conflict (VAC) in head-mounted displays (HMDs) and the technical methods that enable correct or nearly correct focus cue rendering to resolve the VAC problem. It is worth noting that the VAC problem is not unique to HMDs, but inherent to all conventional stereoscopic 3-D displays (S3D) which stimulate the perception of 3-D space and shapes from a pair of 2-D 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.

HMDs, also commonly known as near-to-eye displays (NEDs) or head-worn displays (HWDs), have gained significant interests in recent years and stimulated tremendous efforts and resources to push the technology forward for a broad range of consumer applications [2]. For instance, a lightweight optical see-through HMD (OST-HMD), which enables optical superposition of digital information onto a user's direct view of the physical world and maintains

Manuscript received November 1, 2016; revised date December 15, 2016; accepted January 2, 2017. Date of publication January 24, 2017; date of current version April 20, 2017.
The author is with the College of Optical sciences, University of Arizona, Tucson, AZ 85721 USA (e-mail: hhua@optics.arizona.edu).

Digital Object Identifier: 10.1109/JPROC.2017.2648796

see-through vision to the real world, is one of the key enabling technologies to augmented reality (AR) applications. It has been portrayed as a transformative technology in the digital age, redefining the way we access and interact with digital information essential in our daily life [3]. On the other end of the spectrum, a wide field-of-view (FOV), immersive HMD, which immerses a user in computer-generated virtual world or a high-resolution video capture of a remote real world, is a key enabling technology to virtual reality (VR) applications. It finds a myriad of applications in gaming, simulation and training, defense, education, and other fields.

Despite the high promises and the tremendous progress made recently toward the development of both VR and AR displays, minimizing visual discomfort involved in wearing HMDs for an extended period remains to be an unresolved challenge. One of the key contributing factors to visual discomfort is the well-known VAC due to the lack of the ability to render correct focus cues. In recent years, many research and development efforts have been made to address the VAC problem in HMDs and develop display methods that are capable of rendering correct or nearly correct focus cues.

In the rest of the paper, we will summarize the nature of the VAC problem in HMDs and psychophysical findings of the visual artifacts associated with the problem (Section II), present a categorical overview on the various technical methods toward enabling focus cues in HMDs (Section III), and review the recent advancements in implementing the various approaches (Sections IV–VIII).

II. OVERVIEW OF VERGENCE–ACCOMMODATION CONFLICT

In natural vision, the stimulus of convergence (i.e., binocular disparity) and the stimulus of accommodation are highly correlated. Fig. 1 schematically illustrates the naturally coupled actions of eye accommodation and convergence along with retinal blur effect when viewing a natural scene. While the eyes converge to bring the visual axes to intersect at a 3-D object of interest (e.g., the beach ball), the ciliary muscles contract or relax to adjust the shapes of the crystalline lens and bring the object of interest in focus. Objects at other depths than the object of interest (e.g., the cube) are seen blurred as illustrated by the lower right inset. Therefore, the vergence depth of the eyes coincides with its accommodation depth and the retinal image blur of objects in the scene increases with the distances from the eye fixation point to other points at different depths. For a target moving along the cyclopean axis, the functional relationship between accommodative and convergent responses is nearly linear, which is termed as Donder's line [4]. Such highly coupled correlation is exemplified by convergent accommodation (CA) and accommodative convergence (AC). The former refers to the phenomenon where making the eyes converge also causes them to increase optical power, and the latter refers to the situation where making one eye increase its optical power also induces the other eye to converge. These cross-coupled responses are typically measured in ophthalmic exams by the ratio of convergent-accommodation over convergence (CA/C) and the ratio of accommodative convergence over accommodation (AC/A).

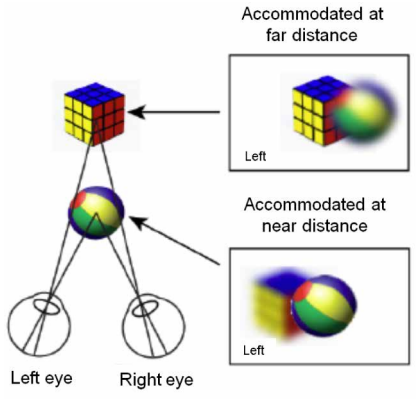

Fig. 1. *Naturally coupled actions of eye accommodation and convergence in real-world view: The vergence depth of the eye where the visual axes intersect coincides with its accommodation depths and the retinal image blur of objects varies as the eye accommodation depth (cube versus beach ball).*

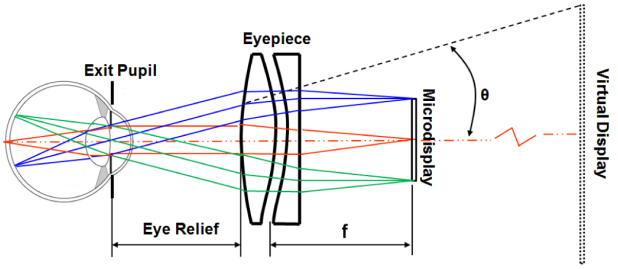

Fig. 2. *Schematic layout of a conventional, monocular HMD: An eyepiece magnifies the image rendered on a microdisplay and forms a virtual display appearing at a fixed, far distance from the eye.*

In viewing conventional HMD images, however, the stimuli for convergence and accommodation may be decoupled and conflicting. The VAC problem in HMDs stems from the fact that the image source mostly is a 2-D flat surface located at a fixed distance from the eye. Fig. 2 shows a schematic layout of a typical monocular HMD, which mainly consists of a 2-D microdisplay as the image source and an eyepiece that magnifies the image rendered on the microdisplay and forms a virtual image appearing at a fixed distance from the eye. An OST-HMD requires an optical combiner (e.g., beamsplitter) placed in front of the eye to combine the optical paths of the virtual display and real scene. The conventional HMDs, regardless of being monocular or binocular, being see-through or immersive, lack the ability to render correct focus cues for the digital information which may appear at other distances than that corresponding to the virtual image plane. As a result, conventional HMDs fail to stimulate natural eye accommodation response and retinal blur effects.

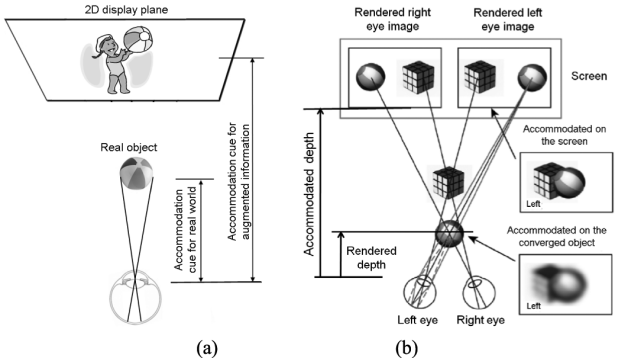

Fig. 3. ***Vergence–accommodation conflicts in HMDs: (a) monocular AR display; (b) stereoscopic display view.***

The problem of lacking correct focus cues in HMDs causes several visual cue conflicts. Marran and Schor [5] provided a thorough review on the various types of accommodative cue conflicts in both VR and AR systems. One of the conflicts in a see-through AR display, regardless of being monocular and binocular, is the mismatch of the accommodation cues between the 2-D virtual image and the real-world scene. As illustrated in Fig. 3(a), the eye is cued to accommodate at the 2-D image plane of a fixed distance for viewing the augmented information while the eye is cued to accommodate and converge at the depth of a real 3-D object onto which the digital information may be overlaid. The distance gap between the display plane and real-world objects can be easily beyond what the human eyes can accommodate simultaneously. A simple example is the use of an AR display for driving assistance where the eyes need to constantly switch attention between the AR display and real-world objects spanning from near (e.g., dashboard) to far (e.g., road signs).

The second conflict is the mismatch of accommodation and convergence between the 2-D image plane and the 3-D virtual world rendered in a stereoscopic display, where pairs of images with binocular disparities are typically presented on 2-D flat screens at a fixed distance from the eye to create the perception of 3-D virtual objects at different distances. As illustrated in Fig. 3(b), when viewing the 3-D virtual objects, the eyes of a user are cued to accommodate at the distance of the fixed 2-D display surface to bring the digital information in focus, while they are forced to converge toward different depths of the rendered 3-D contents dictated by the binocular disparity cues rendered in the images to fuse the stereoscopic pair. The natural coupling of eye accommodation and convergence in viewing a real-world scene (Fig. 1) is thus broken in stereoscopic displays. The conflicting depth cues result in decoupled actions of vergence and accommodation and cause visual fatigue. The decoupled vergence–accommodation depth cues could also have an indirect effect on perceived depth by interacting with binocular disparity and yielding convergence-induced accommodation or accommodation-induced vergence [6], [7]. Usually, convergence-induced accommodation occurs as the eye converges to virtual objects at different depths, but at considerably smaller magnitude than the actual convergence depth variations [6], [8].

The third conflict is the mismatch of the retinal image blur cues between the virtual and real-world scenes. Virtual objects rendered via stereoscopic images, regardless of their rendered depths, are seen all in focus if the viewer accommodates on the image plane, or are seen all blurred if the user accommodate at distances separated from the image plane by distances greater than the DOF of the human eye. The retinal image blur of the virtual scene does not vary with the distances from an eye fixation point to other points at different depths. Fig. 4 demonstrates two photographs captured from a conventional OST-HMD prototype built in our lab, which is schematically the same as the one shown in Fig. 2 except that a 50/50 beamsplitter was inserted between the eyepiece and the exit pupil to enable a see-through view. Three physical targets (black–white gratings) were placed on an optical rail at 1, 3, and 5 diopters, respectively, while three virtual objects textured with color stripes were graphically rendered at the same depths as the physical references by binocular disparities. Just like any other conventional HMDs, the prototype has a fixed focus where the focal distance of the virtual displays was set at approximately 3 diopters. Fig. 4(a) and (b) shows the augmented views captured with a camera focusing at 3 and 5 diopters, respectively. In these pictures, the physical objects show clear focusing cues while the virtual objects are either all in sharp focus in Fig. 4(a) when the camera was focused at the same distance as the virtual display or all blurry in Fig. 4(b) when the camera was focused at the targets at 5 diopters.

Though it is not conclusive, many studies have provided strong supportive evidence that the incorrectly rendered focus cues in conventional HMDs may contribute to various visual artifacts and degraded visual performance [6]–[16]. In a nutshell, incorrect focus cues may contribute to the two commonly recognized issues: distorted depth perception [9]–[13] and visual discomfort. Examples of discomfort include diplopic vision [14], visual fatigue [15], and

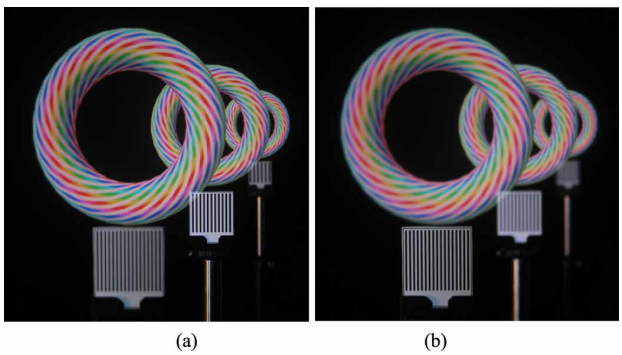

Fig. 4. ***Examples of focus cue mismatch between physical and virtual objects in a conventional OST-HMD: (a) camera focused at 3 diopters; (b) camera focused at 5 diopters.***

degradation in oculomotor responses [6]–[7], especially after viewing such display for an extended period of time. For instance, Watt *et al.* suggested that inappropriate focus cues in stereoscopic displays will adversely affect the depth perception, through both direct and indirect means by image blur and disparity scaling [13]. Inoue and Ohzu suggested that stereoscopic users may perceive a flattened scene rather than the 3-D scene with rendered depths specified by the binocular disparity cues [6]. In immersive virtual environments viewed through stereoscopic displays, people report and act as if they perceive 3-D graphical objects as being closer to them than they really are, in some cases by up to 50% or more [11], [12]. Recent studies by Hoffman *et al.* examined how the focus cues may affect perceived depth directly or indirectly in computer displays [15]. Their results support the argument that inappropriate focus cues in typical stereoscopic displays can contribute to biases in perceived 3-D scene structure under some circumstances.

Symptoms of visual fatigue or asthenopia generally include eyestrain, dried mucus or tears around the eyelids, feeling of pressure in the eyes, aching around the eyes, discomfort when the eyes are open, hot eyes, difficulty in focusing or blurred vision, stiff shoulders, and headaches [18]. Since the emergence of stereoscopic displays, numerous studies have investigated the effect of viewing stereoscopic images on various visual functions. A very common finding is that viewers experience discomfort and visual fatigue [19]. For instance, Mon-Williams *et al.* examined the short-term effects on binocular stability of wearing a commercially available stereoscopic for 10 min [20]. Subjects were examined before and after the exposure to the HMD and there were clear signs of induced binocular stress for a number of the subjects. The primary cause of the discomfort and fatigue in stereoscopic displays has been widely attributed to the forced decoupling of accommodation and convergence cues [19].

The forced decoupling of accommodation and convergence cues in HMDs changes the interactions between the accommodation and convergence system and thus may break the oculomotor balance [21]–[23]. Several studies have measured changes in these oculomotor functions before and after exposure to stereoscopic images for various durations. Hasebe *et al.* have investigated the AC/A ratio, binocular vision functions, convergence width, and refractory changes after using a stereoscopic HMD for 25 min [24]. Slight but significant hyperopic changes were detected in refraction after the task. Amplitude and velocity of convergence varied significantly but inconsistently. Yano *et al.* have measured visual fatigue using a subjective method and compared it to changes in accommodation before and after viewing images by stereoscopic HDTV [25], [26]. They reported evidence supporting the argument that the conflict of convergence eye movement and accommodative function affects visual fatigue. Iwasaki and Tawara have described the various aftereffects of accommodation and pupillary function, which were dependent on the stimulus distance, after 10-min viewing of stereoscopic images at four viewing distances [27]. Significant delay in the accommodation response was shown after the task in all viewing distances, but greater miosis, the condition of excessive pupillary constriction relative to the amount of light the eye receives, was found only after far-distant viewing. Suzuki *et al.* have examined accommodation step response velocity after 30 min of viewing stereoscopic images using the parallax barrier system [28]. A significant slowdown of average accommodation velocity was observed after viewing a stereoscopic display but not after viewing a nonstereoscopic display. Suzuki suggested that viewers should be careful to avoid viewing stereoscopic images for extended durations because visual fatigue might be accumulated. Emoto *et al.* have found more serious subjective visual fatigue and decreased fusional amplitude in stereoscopic TV viewing than in viewing conventional TV [29]. Read and Bohr carried out a lab-based study where 433 viewers aged 4 to 82 years watched a movie either in 2-D or S3D setting. They reported that around 14% viewers experienced adverse effects, mainly headache and eyestrain, due to viewing a S3D movie [30]. A more recent study by Read *et al.* was carried to probe long-term effects of viewing S3D contents and reported that viewing 3-D TV over two months produced no discernible effects on balance, coordination or eyesight [31].

In summary, visual discomfort is a critical concern in applications where users need to work with displays for an extended period of time. Specific applications may minimize the effects of focus cue conflicts by setting the location of the virtual displays according to the tasks to be performed so that the disparities between vergence and accommodation depths presented through the display fall within the "zone of comfort" [32]. For instance, collimated displays are typically preferred in HMDs for far-field applications such as aviation in order to visualize distant information, while the focal distance of HMDs for near-field applications such as surgery is likely set at arm's length. This naturally limits the range of possible depth rendering and thus the presentation of dramatic depth effects. In two common types of viewing conditions the VAC problem may become severe and challenge to manage: virtual objects need to be presented across a wide range of distances to the user, from very close to far away (e.g., driving simulators); or the display is used to augment a relatively close real-world scene with virtual objects and information (e.g., surgical training). It becomes challenging in these applications that involves visualization tasks across a large depth range or near-field tasks where the focus cues are most salient.

III. METHODS FOR ENABLING FOCUS CUES

Several technical methods have been explored to address the VAC problem in conventional HMDs, attempting to approximate the visual effects created by the focus cues in natural vision. Based on the different optical mechanisms for controlling focus cues, Marran and Schor described five different methods that can potentially minimize the visual discomfort associated with conflicting accommodative cues

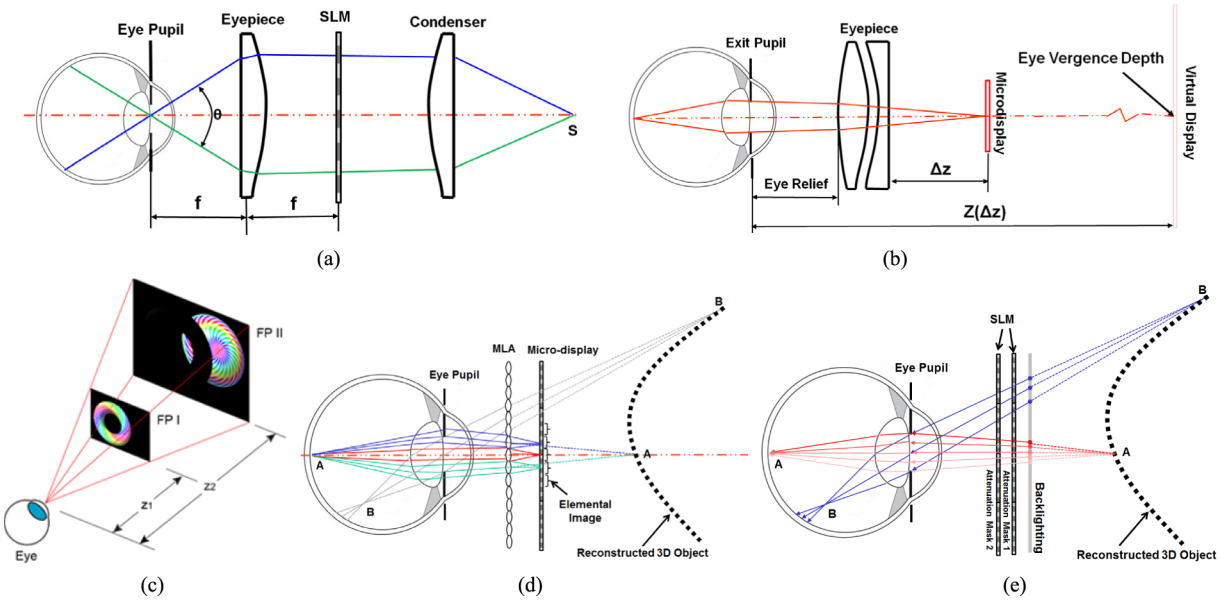

Fig. 5. ***Technical methods for enabling focus cues in HMDs (a): Maxwellian view display; (b) vari-focal approach; (c) multifocal plane displays; (d) integral-imaging-based light-field display; and (e) computational multilayer display. The dotted curves in (d) and (e) illustrate the reconstructed 3-D shapes.***

in HMDs, namely pinhole optics, monocular lens addition, monocular lens addition combined with aniso-accommodation, chromatic bifocal, and bifocal lens systems [5]. A pinhole optics design, which typically is combined with the Maxwellian viewing system [33], limits the ray bundles from the virtual display to the paraxial rays such that the DOF of the virtual image is extended. A monocular lens addition method, also known as a monovision technique, adds a lens to one eye for near viewing and the other for distant viewing, which have been successfully utilized for treating presbyopes [34]. The resulting anisometropic blur from the monovision may affect fine stereopsis and stereoacuity. Aniso-accommodation is the ability of the two eyes to accommodate independently of each other while remaining binocular. Combining the monovision technique with aniso-accommodation may provide a better solution to the VAC problem than the monovision alone. By employing the about 2 diopters of longitudinal chromatic aberration of the eye, a chromatic bifocal method utilizes the lower half of the virtual display for viewing near objects in long-wavelength light (e.g., red) while utilizes the upper half of the display for distant objects in short-wavelength light (e.g., blue). A bifocal lens method is similar to the chromatic bifocal expect removing the limitation of monochromatic virtual image.

Although the classification method by Marran and Schor [5] is still valid to some extent today, to account for the recent advancements in both optical technologies and computational displays, we categorize the existing approaches into five general types: Maxwellian view displays, vari-focal plane displays, multifocal plane (MFP) displays, integral imaging (InI)-based displays, and computational multilayer displays.

A Maxwellian view display, as illustrated in Fig. 5(a), uses the Maxwellian viewing principle in which the diverging rays from a point source S is collimated by a condenser lens and then focused by an eyepiece onto the entrance pupil of the eye instead of being imaged directly on the retina [33]. In a Maxwellian view display, a target or typically a spatial light modulator (SLM), is placed at a focal distance away from the eyepiece such that an image conjugated to the SLM is formed on the retina. Due to the fact that each field on the SLM is imaged through a narrow aperture on the pupil (pinhole), the image is observed with very large DOF without requiring eye accommodation. For this reason, the Maxwellian view display may be characterized as the pinhole optics approach [5].

Vari-focal approach, as illustrated in Fig. 5(b), is a simple remedy to the VAC problem by dynamically compensating the focal distance Z of a single-plane display to match it with the convergence depth of the eye. The focus compensation may be achieved by a mechanical mechanism or an electronic-driven active optical element. The mechanical mechanism may be implemented by mechanically zooming the eyepiece of the display [35] or adjusting the distance between the microdisplay and the eyepiece [36]. Instead of zooming the eyepiece focus through mechanically adjustable parts, a range of electronic-driven active optical elements can be used [37], including a liquid lens [38], liquid crystal lens [39], and deformable mirror [40]. The vari-focal method typically requires dynamically tracking the eye convergence distance in real time. It is able to render correct focus cues for the fixated objects if the focal distance of the virtual display is in synchrony with the distance of eye fixation, but objects at other depths would still have incorrect focus cues.

The vari-focal method may be further improved by applying a depth-dependent blur filter to simulate the retinal image blur and may improve depth perception to some extent.

Multifocal plane (MFP) displays create a stack of discrete focal planes dividing an extended 3-D scene volume into multiple zones along the visual axis. Through these focal planes, they additively reconstruct the light fields of a large scene volume by sampling its projections at different depths. Typically, each of the focal planes samples the projection of the 3-D objects within a depth range centered on it. Fig. 5(c) shows a schematic diagram of a dual-focal plane example in which the projections of a 3-D donut and a sphere are rendered by FPI and FPII, respectively. Additionally, the projections of the objects rendered by the front layer FPI can be rendered as occlusion masks on FPII (e.g., the projection of the donut on FPII) to render correct occlusion relationships among these objects. Multifocal planes may be implemented either by spatially multiplexing a stack of 2-D displays [41]–[43] or by rapidly switching the focal distance of a single 2-D display sequentially by a high-speed vari-focal element (VFE) in synchronization with the frame rendering of multifocal images (i.e., in a time-multiplexed fashion) [45]–[48]. In general, the spatial-multiplexed approach allows for rendering multiple focal planes in parallel and reduces the speed requirements for the display technology. On the other hand, its practical implementation is challenged by the lack of stack display technologies with high transmittance and by the demand for computational power to simultaneously render a stack of 2-D images of a 3-D scene. The time-multiplexed approach at a flicker-free rate demands for high response speed for the active optical element, the display device, and the graphics rendering engine and the response speed is proportional to the number of flicker-free focal planes to be portrayed.

Integral-imaging-based displays reconstruct the light fields of a 3-D scene by angularly sampling the directions of the lightrays apparently emitted by the 3-D scene and viewed from different eye positions [49], [50]. As illustrated in Fig. 5(d), a simple InI-based display typically consists of a display panel and a 2-D array which can be a microlens array (MLA) [51], [52] or pinhole array [53]. The display renders a set of 2-D elemental images, each of which represents a different perspective of a 3-D scene. The conical ray bundles emitted by the corresponding pixels in the elemental images intersect and integrally create the perception of a 3-D scene that appears to emit light and occupy the 3-D space. The InI-based display using 2-D arrays allows the reconstruction of a 3-D shape with full-parallax information in both horizontal and vertical directions, which is its main difference from the conventional autostereoscopic displays with only horizontal parallax using 1-D parallax barriers or cylindrical lenticular lenses [54], [55].

A computational multilayer display, illustrated in Fig. 5(e), typically consists of multiple layers of SLMs illuminated by either a uniform or directional backlight [56], [57]. It is a new, emerging class of display method that shares similarity to the MFP-based and the InI-based display methods. Unlike the additive nature of the MFP approach to light-field rendering, the multiple SLM layers in a computational display method typically operates in a multiplicative fashion. Combining with the backlight, they aim to angularly sample the directions of the lightrays apparently emitted by a 3-D scene in a fashion similar to the InI-based display.

Among the five methods summarized above, the Maxwellian view display attempts to extend the DOF of the virtual display to the extent such that eye accommodation is no longer required to view the virtual scene. This allows the eye to accommodate and converge freely to view real-world scenes, without compromising the image sharpness of the virtual display, which partially resolves the VAC in OST-HMDs. However, it is unable to produce natural retinal blur cues to virtual scenes. The vari-focal approach attempts to adapt the focus of the display in real-time to the vergence depth corresponding to the region of interest. It can be a simple and effective modification to conventional HMDs and is able to address the VAC problems to a great extent, but it is unable to produce natural retinal blur cues. The MFP, integral imaging, and multilayer approaches are commonly referred to be light-field displays, which render a true 3-D scene by sampling either the projections of the scene at different depths or the directions of the lightrays apparently emitted by the scene and viewed from different eye positions. The MFP method is an extension to the bifocal lens method described earlier in [5] by providing a larger number of focal plane sampling and 3-D continuity. Both the InI-based method and the computational multilayer displays share great similarities to the pinhole optics solution.

Besides the methods summarized above, the monovision technique was further experimented in HMDs recently for its utility for mitigating the VAC problem. Konrad *et al.* implemented it in an immersive HMD where lenses with different focal lengths were utilized for the display paths of the left and right eyes [58]. Depending on the eyes' accommodation states, one eye may perceive much sharper image than the other eye. The hypothesis of this technique is that the eye with its focal distance closer to the vergence distance of the fixated object will dictate the viewer's binocular perception and therefore the vergence and accommodation responses will be more similar than they would be in a conventional S3D display. They further compared the effectiveness of the monovision technique against focus-tunable displays for mitigating the VAC problem and suggested that monovision can be a low-cost remedy and potentially provide better user experiences [58]. Johnson *et al.* implemented a similar monovision technique, in which a fixed lens is inserted in front of one eye and conventional stereoscopic contents are presented while no lens is used for the other eye [59]. The authors further experimentally assessed the visual performance and discomfort of the monovision technique in comparison to a vari-focal method and a conventional S3D display. Their experimental

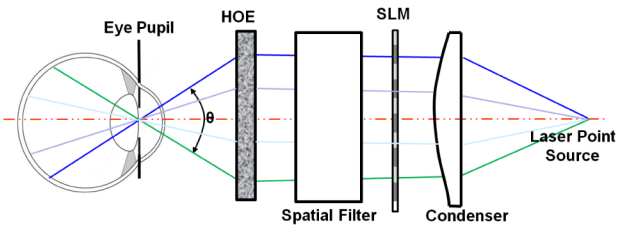

Fig. 6. ***Maxwellian view retinal projection display using a laser source [65].***

results clearly demonstrated the effectiveness of a vari-focal display (VFD) method in improving visual performance and reducing discomfort over the conventional S3D display technique, while the monovision technique demonstrated neither improvements in visual performance nor reductions in discomfort. Given that the monovision technique has not been found effective in addressing the VAC problem and its long-term effects on comfort remain unknown, it is omitted from further discussion in the rest of the paper.

It is worth noting that we mainly concentrate on methods developed for HMDs in this paper. There are a few alternative 3-D display technologies that are able to render nearly correct or correct focus cues similar to those of a real world scene. Examples include volumetric displays [60] that render all voxels within a volumetric space regardless of the viewer's point-of-interest or viewpoint, holographic video displays [61], and super-multiview autostereoscopic displays [62], [63].

IV. MAXWELLIAN VIEW DISPLAYS

One simple way to mitigate the VAC problem in HMDs is to extend the DOF of the virtual display. A Maxwellian view display is one of the well-known solutions, in which a very thin bundle of rays from each field of the display is focused on the retina through a small area on the eye pupil. When the diameter of the ray projections on the pupil is substantially smaller than that of the eye pupil itself, the DOF of the display is extended and the display is considered as being accommodation free.

Several efforts have been made to develop this type of displays aiming to address the VAC problem. A Maxwellian view retinal projection display, as shown schematically in Fig. 6, was proposed by Ando *et al.* for accommodation-free HMDs [64]. A converging laser source is collimated by a collimation lens. The parallel rays from the collimated laser beam pass or reflect through the pixels of a SLM and are focused by a holographic optical element (HOE) at the center of the eye pupil. Each parallel ray from each pixel projects directly onto the retina and stimulates a different point. A spatial filter along with additional focusing and collimation lenses is inserted between the HOE and SLM to mitigate the diffraction effects of the small pixel apertures to the coherent laser rays. As a result, the image pattern shown on the SLM is directly imaged on the retina without eye accommodation. Two optical see-through bench prototypes were demonstrated, one with a reflective DMD and one with a transmissive LCD as the SLM, along with a holographic optical element as the imaging lens [65]. Both prototypes demonstrate a large DOF for the virtual display due to the benefit of the Maxwellian view.

Instead of using coherent laser source [66], Von Waldkirch *et al.* demonstrated a retinal projection display prototype based on Maxwellian view in which a partially coherent light emitting diode (LED) illuminates an LCD through a condenser lens. Before the LCD image is projected onto the retina, the LED source was focused through a sequence of narrow apertures to filter out undesired diffraction patterns of the LCD image and to increase the spatial coherence and thus the DOF [67], [68]. Waldkirch *et al.* further proposed to extend the DOF of a Maxwellian view retinal projection display by applying a multiple imaging technique in which a pupil phase mask is inserted at the aperture stop plane to produce a series of images of the LCD at various focal planes [69]. Waldkirch *et al.* proposed the use of a fast oscillating, variable-focus lens in a partially coherent retinal projection display, in which the lens's dioptric power is oscillating periodically in time instead of being adapted in real time to respond to the current point of interest in a vari-focal approach [70]. As a result, the image pattern rendered on a SLM is projected sequentially at different foci onto the retina and the viewer observes a temporal fusion of multiple in-focus and defocused images, independent of the current eye accommodation.

These conventional Maxwellian view displays have several major limitations, such as reduced spatial resolution. The most significant limitation is its small entrance pupil diameter (EPD) due to the nature of Maxwellian view which requires all the rays converge to a small area on the pupil. The small EPD not only imposes ergonomic challenges for viewers to properly position their eyes, but also poses restrictions on eye rotation and eventually limits the field of view due to vignetting effects caused by natural eye movements [33].

Yuuki *et al.* demonstrated an implementation of a dense Maxwellian view display which is composed of a regular base display such as a LCD panel, a light absorption layer with pinhole patterns, and a plano-convex fly-eye lens sheet [71]. The absorption layer with micropinholes is attached to the back surface of the fly-eye lens sheet and coincides with the back focal point of the lens sheet. There are total 7×7 micropinholes in the area covered by each convex lens of the lens sheet. The rays emitted from the base display pass through the corresponding pinholes of the micro-pinhole groups all over lens sheet intersect with each other and form a 7×7 ray intersection points on the eye pupil. The diameter of the ray intersecting points is smaller than 1 mm on the eye pupil while the 7×7 points are spaced out by a distance about 4 mm. This allows the observer to see all pixels on the base display in clear focus by placing their eye at each of the ray intersecting points.

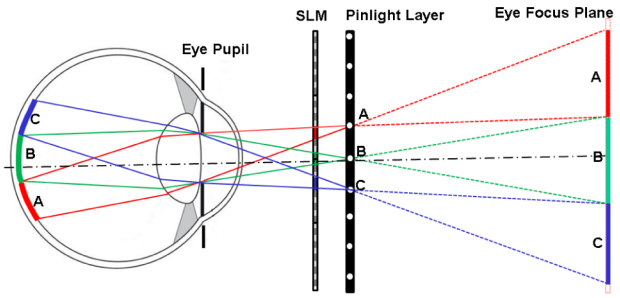

Fig. 7. ***Schematic layout of a Pinlight display [72] consisting of an array of point sources, namely the Pinlight layer, and a SLM. Each Pinlight is modulated by a section of the SLM and form a subimage on the retina.***

Maimone *et al.* demonstrated a compact, wide-FOV OST-HMD design, namely Pinlight displays [72], which is similar to the design by Yuuki *et al.* [71]. Fig. 7 shows the schematic layout of a Pinlight display. It mainly consists of an LCD panel and an array of point light sources placed directly in front of the eye. The array of point sources, namely pinlights, is formed on an edge-lit acrylic sheet into which small cavities are etched. LEDs placed around the edge of the plastic sheet light up the cavities and form the point sources. Lightrays emitted by each pinlight are modulated by a section of the LCD panel which displays subset images spatially decomposed from a target image. The modulated lightrays are refracted by the eye lens and create a "sharp" copy of the modulated image on the retina. The small subset images projected by all the pinlights are tiled together to create a wide FOV image. A prototype display in the form factor of large glasses was demonstrated, offering a 110° diagonal FOV.

Each individual pinlight projector can be thought as a Maxwellian view display in the sense it forms a sharp image of the pattern displayed on the LCD panel, regardless the accommodation state of the eye, as each pixel on the LCD panel only transmits a "single" directional ray, defined by the locations of the pixel and the corresponding pinlight, through the eye pupil. The main difference from a conventional Maxwellian view display is that neither the rays from a pinlight converge at a point on the pupil nor the LCD panel forms a conjugate image on the retina. In the sense of expanding the EPD, the pinlight display share similarity to the design by Yuuki *et al.* [71]. The subset images may be generated in such a way that the rays from multiple pinlight projectors enter the eye pupil at different locations but intersect at the same point on the retina, creating the perception of the same point in the visual space and allowing a large eye box. The pinlight display can also be thought as a simplified implementation of the computational light-field displays in the sense that the directional sampling of the light fields is accomplished by the pinlight projectors.

The Pinlight display suffers from several limitations. The most notable limitation is the diffraction effect due to the small aperture defined by the pixel dimension on the LCD panel, which limits the spatial resolution to both the see-through and virtual display paths. Other limitations include low image contrast and frame rate, low see-through transparency, and toning effects across the tiled FOV.

V. VARI-FOCAL DISPLAY METHOD

There are two key aspects in enabling focus cues in HMDs. One aspect is to create the proper cue to stimulate the eye to accommodate and converge at the same depth, which typically involves controlling the focal depth of the display hardware. The other aspect is to recreate the same retinal blur effects as those observed for a real-world scene.

Simulation of retinal blur effects alone can be easily implemented in the rendering software of a conventional HMD without modifying the hardware, in which a depth-dependent filter technique is applied to simulate the retinal image blur effects based on the distance of a 3-D synthetic object from the eye convergence depth. For instance, a recent system by Hillaire *et al.* [73] was able to track the 3-D convergence point of the viewer and to render corresponding retinal image blur effects in real time. It was reported that video game players indeed preferred the display with the graphically rendered blur cues, as opposed to displays lacking the synthesized blur effects. Mauderer *et al.* also utilized gaze-contingent depth of field to produce realistic 3-D images and investigated its effectiveness for depth perception [74]. They reported that the simulated blur effects through gaze-contingent DOF method increases subjective perceived realism and depth and can contribute to the perception of ordinal depth and distance between objects to some extent. This approach, however, is limited in its accuracy. It is unable to provide correct focus cues and address the VAC problem because the focal plane of the display hardware remains fixed in the visual space, independent of the changing convergence distance of the viewer.

To facilitate the proper cue for eye accommodation response, the vari-focal approach adds the least hardware complexity to a conventional HMD design by directly manipulating the apparent distance of the virtual image plane formed by the HMD eyepiece. Wann *et al.* suggested the method of using an oscillating lens to adjust the focal depth of the virtual image based on the user's gaze point to address the VAC problem in HMDs [8]. Shiwa *et al.* demonstrated the first experimental implementation of a VFD prototype in which a relay lens group inserted between a display device and eyepiece was mechanically moved back and forth along the optical axis by a stepper motor to control the overall optical power of the optics and thus the focal distance of the virtual image [35]. The control of the mechanical zooming was open-loop through a user's manual input (e.g., mouse or keypad). The mechanical zooming took about 0.3 s to change focus from 5 diopters to 0.1 diopters. Shibata *et al.* demonstrated a bench prototype with similar vari-focal function for focus control by axially translating the microdisplay mounted

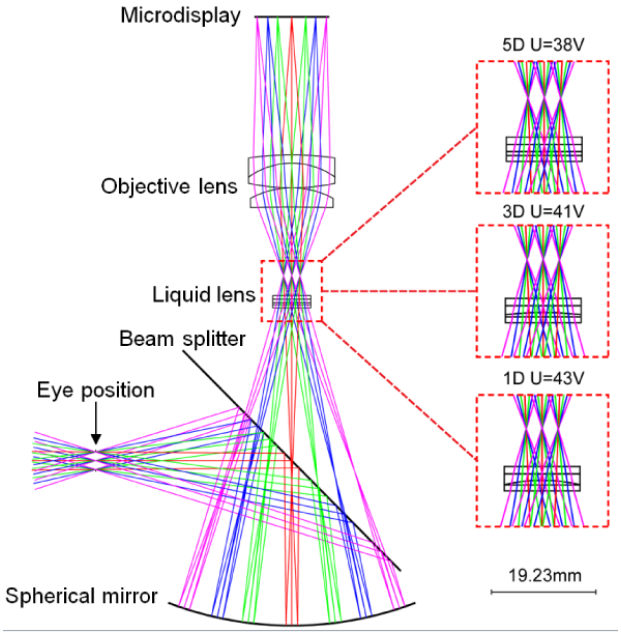

Fig. 8. ***Optical design layout of a VFD using a liquid lens [37].***

on a microcontrolled stage, which allowed the control of focal distance from 3.3 to 0.5 diopters [36].

Liu and Hua demonstrated the first OST-HMD prototype with focus control using an eyepiece design integrated with a liquid lens, for which the optical layout is shown in Fig. 8 [37], [75]. The prototype is capable of dynamically controlling the focal distance of the display from 5 to 0 diopters (the near point of the eye to infinity) by driving the liquid lens from 38 Vrms to 51 Vrms, offering a diagonal FOV of about 28°. As shown on the left side of each subimage in Fig. 9, three bar resolution targets were placed at 16, 33, and 100 cm, respectively, away from the eye position along the visual axis of the HMD. They were used as references to the virtual image with variable accommodation cue. A virtual torus was rendered at approximately 16 cm away from the eye and the voltage applied to the liquid lens was adjusted to match with the depth of the virtual torus. Fig. 9(a)–(c) shows three photos with the camera focused at 6, 3, and 1 diopter, respectively.

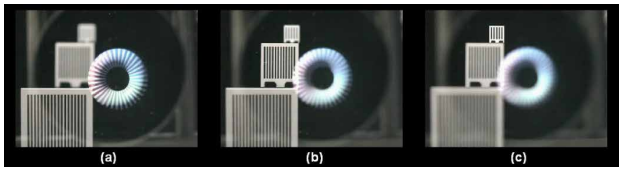

Fig. 9. ***Demonstration of a VFD: the camera focused at (a) 6 diopters, (b) 3 diopters, and (c) 1 diopter [37], [75].***

The prototype design shown in Fig. 8 was further extended by integrating a binocular eyetracking system which enables tracking the eye convergence depth of the viewer, addressing the focal distance of virtual image according to the vergence depth, and rendering gaze-contingent retinal blur cues in real time [76]. Fig. 10(a) shows a photograph of the gaze-contingent vari-focal display (GC-VFD) prototype, in which the binocular eyetrackers are able to detect the eye vergence depth at a speed of approximately 60 Hz. Both the focus and the rendering of the displays are updated at the same frame rate as the tracking speed. Fig. 10(b) shows the result of real-time vergence tracking in responding to three physical resolution targets placed along the visual axis at 3, 2, and 1 diopter, respectively. The eye-tracked convergence distances approximately match with the distances of the real targets, while the slight mismatch may be explained in part by the about 0.6D depth-of-field of the eyes [77]. Fig. 10(c) and (d) compares the addressable focus cues rendered by the GC-VFD prototype against the focus cues of real-world targets. A digital camera with an F/4.8, approximately matching with the F/# of the human eye, was set up at the exit pupil location of the GC-VFD prototype. As shown in Fig. 10(c), when the camera focused at the near distance of 3 diopters, the bunny at the same distance was rendered clearly while depth-dependent blur filters were applied to the bunnies at 1 and 2 diopters. Meanwhile, the focal distance of the GC-VFD was adjusted by the liquid lens device matching with camera's focus depth, and *vice versa* in Fig. 10(d) as the camera focused at 1 diopter. The virtual bunnies located at three discrete depths demonstrate nearly correct focus cues similar to those of the real resolution targets. The captured images demonstrate that the visual effects through a GC-VFD are analogous to the real-world scene with nearly correct focus

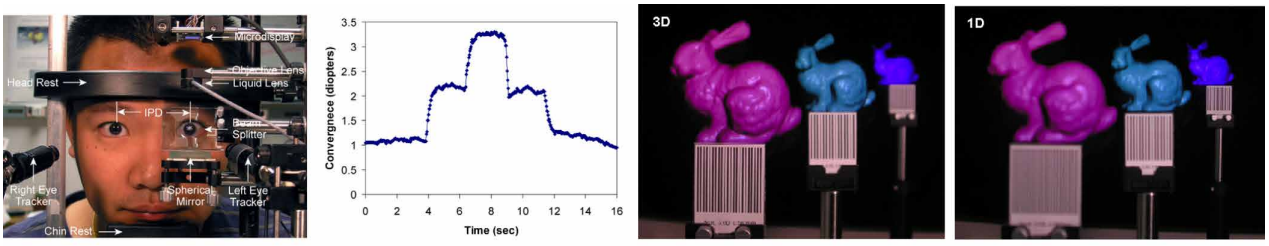

Fig. 10. ***Gaze-contingent VFD mode: (a) prototype; (b) tracking eye convergence depth through eyetrackers; and (c) and (d) camera focused at 3 and 1 diopter, respectively [76].***

cues rendered interactively by the display hardware and software. Overall, the gaze-contingent vari-focus approach provides a good approximation to the viewing condition of the real world. It is able to render nearly correct accommodation cues for 3-D objects close to the fixation point, but objects at other depths would still have incorrect accommodation and retinal blur cues.

Liu *et al.* carried experiments to assess the depth perception and eye accommodative response to a VFD technique [75]. A controlled depth judgment task was utilized to evaluate the perceived depth of a virtual object with the prototype operated in a monocular, vari-focal plane mode in which focus cues were the primary contribution to the perceived depth of the virtual objects. Both subjective and objective measurements of the perceived distance of a virtual object, by verbal report and by positioning a real reference target, respectively, and the eye accommodative responses to virtual objects, were obtained through the experiments. The results suggest that the perceived depths and accommodative responses of the user match with the rendered depths of the virtual object in a VFD when its focal distance is in synchrony with the object depth. This result suggests the potentially improved depth perception induced by appropriately rendered focus cues, in comparison to findings on traditional S3D displays, which suggest distorted and compressed perception of depth.

VI. MULTIFOCAL PLANE APPROACH TOWARD HEAD-WORN LIGHT-FIELD DISPLAYS

An MFP-based HMD can be thought of sampling the light-field projections of a 3-D scene at different depths along the visual axis and the stacks of the projections additively render the light field of the scene seen from a fixed viewpoint. Unlike a vari-focal method, the MFP method does not require a mechanism to determine the point of vergence and is capable of rendering correct focus cues through a 3-D volume.

The pursuit of a MFP-based HMD starts with the spatially multiplexed approach. Marran and Schor suggested the use of bifocal lenses to create different focal depths for different parts of the visual field in HMDs [5]. The experimental implementation of the spatially-multiplexed MFP method can be traced back to the volumetric projection display by Crabtree [78]. A series of 2-D images are projected into a multiplanar optical element (MOE) which acts as a projection screen with a variable depth and the screen depth is synchronized to the frame rendering of the projector. One of their prototypes projected a total of 13.1 million voxels covering a volume of 39.8 cm × 34.3 cm × 23 cm with 50 planes and 512 × 512 pixels per plane [79]. Rolland *et al.* investigated the engineering requirements for adapting the multiplane volumetric display method to HMDs by using a thick stack of planar displays [41]. The analysis concluded that 14 equally spaced (in dioptric spacing) focal planes would be required to accommodate the focusing range from infinity to 0.5 m for a visual acuity of 1 arcminute and a 4-mm pupil diameter. Akeley *et al.* demonstrated the first experimental implementation of a spatially multiplexed three-focal plane display intended for HMD applications with an equal spacing of 0.67 diopters between adjacent focal planes [42]. The prototype covers a fixed depth range from 0.311 to 0.536 m by dividing a flat panel display into three focal planes through 3 beamplitters placed at different distance to the viewer. More recently, Schowengerdt *et al.* suggested a spatial-multiplexing retinal scanning display by utilizing a fiber array to produce multifocal bundle of beams [43]. Cheng *et al.* designed a spatially multiplexed dual-focal plane system by stacking two freeform prisms [44].

In general, the spatially multiplexed MFP approach allows for rendering multiple focal planes in parallel and reduces the speed requirements for the display technology. On the other hand, its practical implementation is challenged by the lack of stack display technologies with high transmittance and by the demand for computational power to simultaneously render a stack of 2-D images of a 3-D scene. It often becomes impractical for implementing a large number of focal planes due to optical path complexity or managing the light efficiency through a stack of SLMs. Another drawback associated with the spatially multiplexed MFP approach is the lack of flexibility to vary the positions and spacing of the focal planes.

An alternative to the spatially multiplexed scheme is a time-multiplexed approach using VFEs, in which 3-D objects of different depths are rendered sequentially on a 2-D focal plane whose focal distance is adjusted in synchronization with the depth of the objects being rendered. If all objects are rendered at a flickering-free rate, they would appear to have correct focus cues simultaneously. Several time-multiplexing MFP prototypes were demonstrated over the last decade. McQuaide *et al.* [45] and Schowengerdt and Seibel [80] demonstrated a dual-focal plane retinal scanning display in which a 2-D image is generated on a pixel-by-pixel basis by raster scanning a modulated laser beam across the visual field and the focus cues of the pixels are rendered by defocusing the laser beam through a deformable membrane mirror device. A full-color pixel can be portrayed by modulating the intensities of three RGB lasers. On the other hand, developing a flickering-free dual-focal plane display with a high resolution on each 2-D image plane requires the speed of the raster scanner to be as high as several megahertz level. Love *et al.* demonstrated a prototype with four fixed focal planes generated through birefringent lenses as the VFEs and high refresh rate CRTs as the image sources [48]. Llull demonstrated a prototype of 6-plane MFP display utilizing a DMD microdisplay [81].

Using a liquid lens as the VFE and an OLED microdisplay as the image source, Liu and Hua demonstrated the first prototype of a dual-focal plane optical see-through AR display,

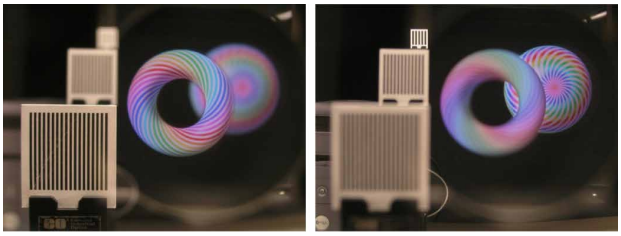

Fig. 11. *View examples captured through a dual-focal plane optical see-through HMD prototype with camera focused at (a) the front resolution target located at 5 diopters and (b) the back resolution target located at 1 diopter distances [47].*

which maintains nonobstructive see-through view to the real world [47]. In the prototype design, the front and back focal planes were placed at a depth of 5 diopters and 1 diopter, respectively, from the eye. Fig. 11 demonstrates two photos captured by a camera from the exit pupil of the display. Three printed resolution targets were placed in the scene at 5, 2, and 1 diopter, respectively, as physical references. A virtual sphere located at a 1-diopter distance was rendered by the back focal plane while a donut located at a 4-diopter distance along with the occlusion mask of the sphere was rendered by the front plane. Due to the large separation of the two virtual objects, no depth blending was applied. Fig. 11(a) and (b) demonstrates the comparison between the camera focus at 5 diopters and 1 diopter, respectively. The virtual objects showed comparable blur cue that varied with the object depth as that of the physical references.

In both spatially multiplexed and time-multiplexed methods, a large number of focal planes and small dioptric spacing are desirable for creating a large scene volume with accurate focus cues and high image quality. It was suggested that the dioptric spacing between adjacent focal planes should be 1/7 diopters to achieve 1-arcminute spatial resolution [41]. At this spacing, 28 focal planes are needed to cover the depth range from infinity to 4 diopters. However, it is practically very challenging to achieve a large number of focal planes with the affordance of current technologies. Reducing the number of focal planes without noticeably compromising the quality of displayed images and the accuracy of rendered focus cues can have significant advantages.

To achieve good image quality with a significantly reduced number of focal planes, the multifocal plane approach may be further improved by incorporating a depth-fused 3-D (DFD) perception [82], where two overlapped images displayed on two transparent screens at two different depths may be perceived as a single-depth image. The perceived depth of the fused image may be manipulated by modulating the luminance ratio between the two images through depth-weighting fusing function to optimize the contrast magnitude and gradient of the fused image which create cues for stimulating and stabilizing the eye accommodation responses. A multifocal plane display incorporating the DFD effect is thus referred to as depth-fused multifocal plane displays, or DFD-MFP displays. This method significantly reduces the number of necessary focal planes to an affordable level to render correct or nearly correct focus cues for 3-D objects spanning in a large depth range to the viewer [83]–[87].

Fig. 12 illustrates the depth fusion concept of two focal planes separated by a dioptric distance of Δz [83], [85]. The dioptric distance from the eye to the front focal plane is z_1 and to the rear plane is z_2. When the images shown on the two-layer displays are aligned such that each pixel on the front and rear planes subtends the same visual angle to the eye, a pixel on the front plane and the overlapping pixel on the back are then fused as a single pixel at the viewpoint. The luminance of the fused pixel (L) is summed from the front and rear pixels (L_1 and L_2, respectively). The two overlapped pixels displayed at two different depths may be perceived as a single-depth pixel, and the luminance ratio between the two pixels may be modulated to change the perceived depth of the fused pixel. These relationships may be expressed as

$$\begin{aligned} L &= L_1(z) + L_2(z) = w_1(z)L + w_2(z)L \\ \hat{z} &= f(w_1, w_2) \end{aligned}$$
and
$$w_1(z) + w_2(z) = 1 \tag{1}$$

where $w_1(z)$ and $w_2(z)$ are the depth-weighted fusion functions modulating the luminance of the front and back pixels, respectively, and z and $\hat{z}$ are the rendered depth and perceived depth of the fused pixel, respectively.

Consider the example of a dual-focal-plane DFD display with the two focal planes placed at 1.2 and 1.8 diopters, respectively. The luminance ratio between the two focal planes was set to be 1:1 (i.e., $w_1 = w_2 = 0.5$), indicating that the fused pixel was being simulated at the dioptric midpoint (i.e., 1.5 diopters) of the front and back focal planes [83], [85]. We further assume a standard observer with normal vision and 3-mm eye pupil which is subject to the average display brightness. Fig. 13(a) shows the contrast gradient of the retinal image of the fused pixel as a function of eye accommodation distance for different spatial frequencies. The peak contrast for each frequency was marked by a black square marker. A contrast-gradient reverse point

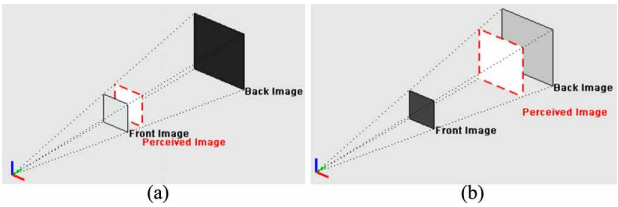

Fig. 12. *Schematic model of a DFD-MFP display: The luminance ratio between the front and back images are modulated to change the perceived depth of the fused image to be (a) near the front image plane and (b) near the back image.*

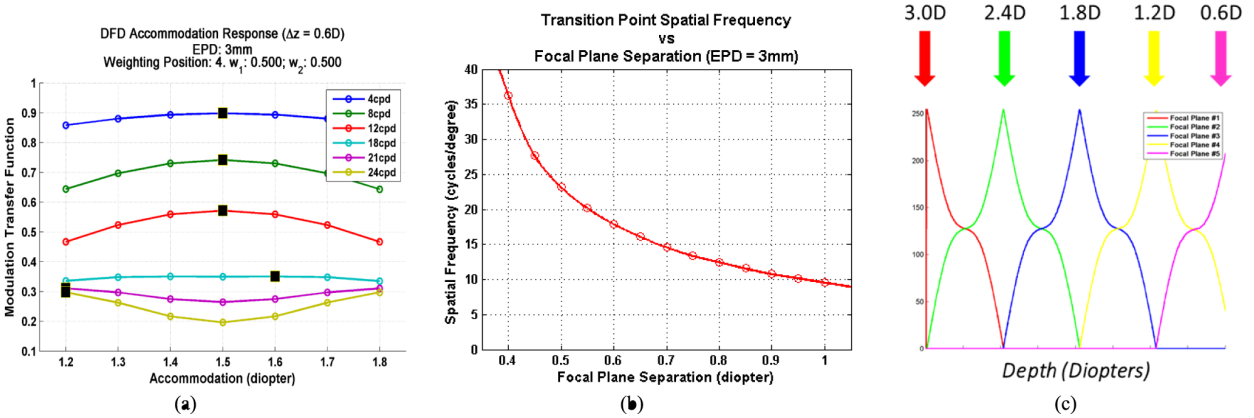

Fig. 13. ***Simulation of a dual-focal-plane DFD display with 0.6-diopter spacing. (a) Contrast gradient of the fused image. (b) The limiting spatial frequency of the fused image to avoid conflicting focus cues as a function of the dioptric separation between adjacent focal planes. (c) Examples of nonlinear depth-fusion functions to achieve ±0.1 diopters of depth difference between the simulated depth and the depth where the maximum retinal image contrast is obtained [85], [86].***

is observed around the frequency of 18 cycles per degree (cpd). Below 18 cpd, the retinal image contrast of the fused pixel is maximized at the dioptric midpoint of 1.5 diopters. As the eye approaches the simulated depth of 1.5 diopters from either the far or near focal planes, the contrast values increase smoothly, providing the appropriate contrast gradient required for driving the eye accommodation response. For frequencies above 18 cpd, however, the contrast of the fused pixel is the highest when the eye is accommodated at the far or near physical focal planes, meaning that the contrast gradient has the tendency to drive the accommodation away from the simulated depth and creates a conflict accommodation cue.

The optimal spacing between adjacent focal planes depends on several factors. The most significant factor is the desired image quality. Fig. 13(b) plots the spatial resolution of the fused image as a function of the dioptric spacing [85], which suggests that a spacing of ~0.6 diopter between two adjacent focal planes is appropriate to render nonconflicting focus cues for MFP displays with a Nyquist frequency of ~18 cpd. A spacing of 0.4 diopters or smaller would be desired for displays affording a resolution of 30 cpd (i.e., 1 arcminute).

In terms of the fusion functions, Akeley *et al.* [42] and more recently Ravikumar *et al.* [88] suggested using a pair of linear depth-blending functions on adjacent focal planes to mitigate image artifacts caused by discrete sampling of focal planes. A linear blending function was demonstrated to be effective in creating continuous 3-D scene through discrete focal planes and can be effective in creating nearly correct focus cues to some extent, it does not necessarily minimize conflicting cues.

To facilitate the rendering of correct or nearly correct focus cues and minimizing conflicting cues, the appropriate fusion functions depend on many factors, such as the tolerance to focus cue error, focal plane spacing, display characteristics, and eye model parameters. Liu and Hua [83], Hu and Hua [85], and Hu [86] presented an optimization framework for DFD-MFP system design, in which the focal

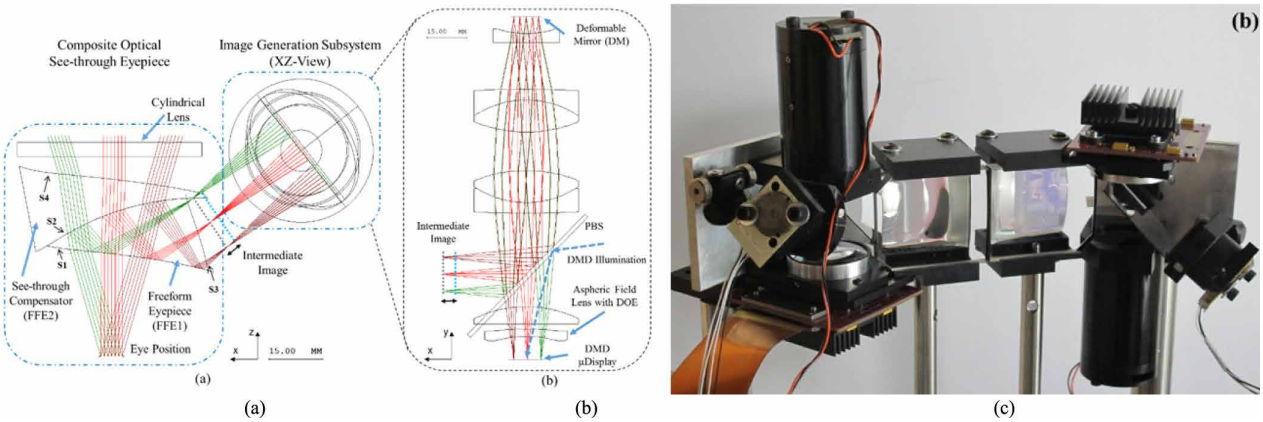

Fig. 14. ***(a) Top-view optical layout of the right-eye module of a DFD-MFP system. (b) Detailed layout of the image generation subsystem. (c) Multifocal plane prototype system with freeform eyepiece [87], [89].***

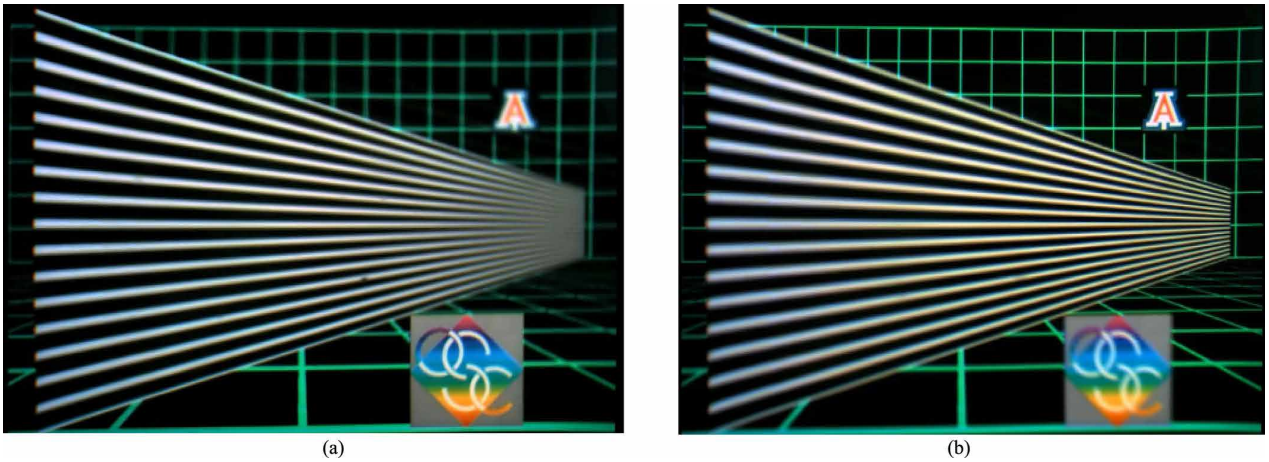

Fig. 15. *Example of 3-D images rendered by the DFD-MFP prototype: the black/white grating target extends from 3 to 0.6 diopters and the green wall grid as well as the UofA logo were located at 0.6 diopters and the OSC logo was placed at 3 diopters. The composed scene was rendered by five focal planes separated by 0.6 diopters apart using the depth-weighted fusing functions shown in Fig. 13(c). The two photos were captured with a F/4.8 camera focused at (a) 3 diopters and (b) 0.6 diopters, respectively [85].*

planes spacing and the fusion functions are optimized to maximize the contrast magnitude and gradient of the retinal image fused by the two images overlapping along the visual axis to avoid conflicting focusing cues and create smooth contrast gradient that helps to stimulate and stabilize the eye accommodation response. Fig. 13(c) shows a set of nonlinear fusion functions applied to five focal planes spaced by 0.6 diopters to ensure less than ±0.1 diopters of depth difference between the simulated depth and the depth where the maximum retinal image contrast is obtained [86]. These nonlinear functions require the pixel intensities distributed on two adjacent focal planes to bias toward the focal plane which is further away from the simulated depth. Further experiment is required to validate the effectiveness of the different types of depth blending functions.

To demonstrate the effectiveness of the DFD-MFP method and the fusion functions, six-focal-plane OST-HMD prototypes were custom designed and built [84], [85], [87], [89]. Fig. 14 shows the optical layout of a monocular OST-HMD system based on the DFD-MFP method [87]. Each monocular setup consists of two parts: the composite optical see-though eyepiece and the image generation subsystem (IGS). The composite eyepiece [Fig. 14(a)], through which a viewer sees the virtual image and the real world, enables the successful integration of MFP display and OST-HMD techniques. It consists of a wedge-shaped freeform eyepiece, a freeform see-through compensator lens, and a cylindrical lens [87], [89]. The IGS [Fig. 14(b)] achieves the core function of generating the multifocal-plane contents. It consists of a 0.7" high-speed digital mirror device (DMD) microdisplay by Texas Instruments, a deformable mirror (DM) device, and relay lens groups. The DMD display, illuminated by an RGB LED light source (not shown), is first magnified by two field lenses and then relayed by a double-pass double-telecentric lens group, forming an intermediate image between the IGS and the eyepiece. By changing the optical power of the DM, the intermediate image shifts axially with respect to the eyepiece without magnification change, while the accommodation cue of the HMD, which is the distance from the eye to the virtual image plane, is varied from far to close. The optical power of the DM can change at a speed as high as 1 kHz, thus virtual images at multiple focal distances can be multiplexed by rapidly changing the optical power of the DM.

Fig. 14(c) shows the actual setup built on an optical bench, while Fig. 15(a) and (b) shows photos of a 3-D scene rendered by the virtual display and captured with a camera placed at the exit pupil position. Six focal planes were dynamically formed at 3.0, 2.4, 1.8, 1.2, 0.6, and 0.0 diopters, respectively, at an overall refresh rate of 60 Hz. The 3-D scene consists of a green floor grid extending from 3.0 to 0.6 diopters, a green wall grid at 0.6 diopters with the University of Arizona (UofA) logo on it and a grating target extending from 3.0 to 0.6 diopters, as well as an OSC logo placed at the 3.0 diopter end. Each focal plane displays a different part of the 3-D scene. By incorporating the depth fusing functions shown in Fig. 13(c), this 3-D scene was rendered continuously by five of the focal planes. Fig.15(a) shows the image with the camera focusing at 3 diopter distance (near), and Fig. 15(b) shows the image focusing at 0.6 diopter distance (far). Natural focus cues are clearly demonstrated and high-contrast continuous targets were correctly fused across five spatially separated focal planes, which visually validate the depth-fusion display method. The optical see-through path of the prototype also achieved superb performance. The optical see-through path of the prototype also achieved superb performance.

The depth-weighted MFP blending method described above concentrates on optimizing retinal image quality and minimizing accommodation cue conflicts when rendering

a continuous 3-D scene volume through a few discrete focal planes with uniform focal-plane spacing. Narain *et al.* recently proposed an alternative depth-blending weighting scheme by computationally optimizing luminance weight to accurately reproduce the defocus behavior of occlusions, reflections, and other nonlocal effects as a function of accommodation [90]. Wu *et al.* explored the benefits of optimizing and adapting the dynamic configuration of a finite number of focal planes based on the characteristics of the 3-D contents to be rendered [91]. They optimized an objective function that characterizes the overall perceptual quality of the rendered 3-D scene, which was referred to as the "multifocal scene defocus quality." They suggested that the perceived visual quality of the content can be significantly improved by explicitly reducing the overall loss of contrast for aggregated accommodate states within the rendered image volume [91].

Overall, the prototypes adequately demonstrated the capability of the DFD-MFP display method for rendering nearly correct focus cues and the potential for addressing the accommodation-convergence conflict. However, the technology suffers from several critical technical obstacles to become a viable solution to truly wearable light-field AR displays. The first major obstacle is the technology miniaturization. Due to the limitations of several enabling technologies, including the high-speed displays and VFE, the current prototype was implemented in the form of a bench prototype, occupying a volume of nearly 500-mm by 300-mm by 100-mm bench space. The second major obstacle is real-time rendering and display. Limited by the availability of high-speed display technology and display-computer interface, the prototype is unable to render and refresh six or more frames of high-resolution full-color images at a speed several times faster than a standard single-frame display. Instead, precomputed images with compromised color depth were preloaded to the memory buffer available through a specialized graphics board. Transforming this display method into a compact wearable solution requires several technical innovations.

Utilizing three focal plane prototype in [42], Hoffman *et al.* conducted experiments to evaluate the influence of focus cues on perceptual distortions, fusion failures, and visual fatigue. They reported findings that MFP displays offering correct or nearly correct focus cues may yield higher depth perception accuracy, faster task performance, higher stereosacuity, and less eye fatigue [15].

VII. INI-BASED HEAD-WORN LIGHT-FIELD DISPLAYS

By angularly sampling the directions of the lightrays apparently emitted by a 3-D scene and viewed from different eye positions, an InI-based display is capable of reproducing the multidimensional light fields appearing to be emitted by a 3-D scene. Since its invention by Lippmann in 1908 [49], the InI-based technique has been widely explored for both capturing the light fields of real scenes [92]–[97] and for its use in eyewear-free autostereoscopic displays [98]–[100]. It has been known for its limitations in low lateral and longitudinal resolutions, narrow DOF, and narrow viewing angle. However, compared with all other nonstereoscopic 3-D display techniques, the simple optical architecture of an InI technique makes it attractive to integrate with HMD optical system and create a wearable light-field display.

Lanman *et al.* demonstrated a non-see-through light-field display by directly placing a microlens array (MLA) and an OLED microdisplay in front of the eyes to render the light field of a 3-D scene for VR applications [51]. The operation principle is similar to the one illustrated in Fig. 5(d). Rays from the pixels corresponding to the same scene point are relayed by the multiple lenslets toward the eye pupil and virtually intersect to form a reconstruction point at the rendered depth. The prototype system, shown in Fig. 16, has an FOV of about 29° × 16° and a spatial resolution of 146 × 78 pixels. The main advantage of this approach is its compactness due to the thin profile of the microdisplay-MLA assembly. Besides its obvious limitation of non-see-through, this approach suffers from the main limitations of low lateral and longitudinal resolutions as well as shallow DOF. Limited by the imaging capability of a stand-alone MLA, the reconstructed scene appears to be only a few inches away from the eye. Hong *et al.* demonstrated a prototype of an integral floating display system [101], but the system FOV is only a few degrees and the system does not offer a useful see-through capability in a wearable device.

Hua and Javidi demonstrated the first practical implementation of an OST-HMD design that integrates a microscopic InI (micro-InI) method for full-parallax 3-D scene visualization with the emerging freeform optical technology for OST-HMD eyepiece optics [52], [102], [103]. This approach enables a compact 3-D integral imaging OST-HMD (InI-OST-HMD) with full-parallax light-field rendering capability. Fig. 17(a) shows the schematic optical layout of an InI-OST-HMD system. The optics consists of a micro-InI unit, a wedge-shaped freeform eyepiece, and a see-through freeform corrector lens. The micro-InI unit, consisting of a high-resolution micro-display and a MLA, reproduces

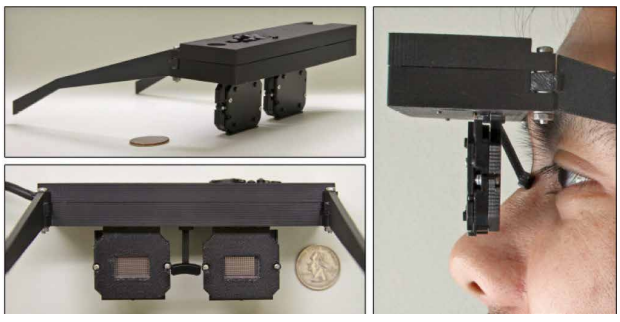

Fig. 16. *Head-mounted lightfield display prototype [51].*

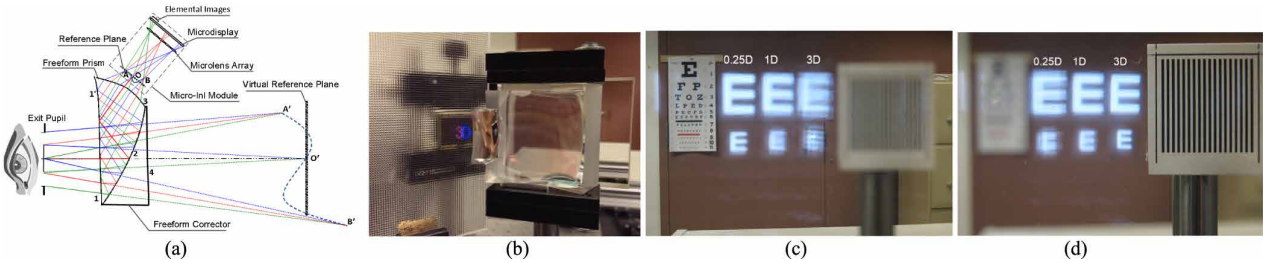

Fig. 17. ***Demonstration of a 3-D integral imaging optical see-through HMD using freeform optical technology. (a) Optical design layout. (b) Prototype of an InI-OST-HMD. (c) and (d) Photograph captured with a camera focusing at 0.25 and 3 diopters, respectively [52].***

the full-parallax light fields of a 3-D scene, which replaces a conventional 2-D microdisplay as the image source. The freeform eyepiece optics directly relays the reconstructed 3-D light fields into a viewer's eye for viewing. In the see-through path, a freeform corrector lens is cemented with the freeform prism and optically enables nonobtrusive view of the real world scene. Song *et al.* demonstrated another OST InI-HMD design using a pinhole array together with a similar freeform eyepiece [104].

Fig. 17(b) shows the photograph of a proof-of-concept monocular prototype using off-the-shelf optical components. The prototype utilized a 0.8" organic light emitting display, offering 1920 × 1200 color pixels with a pixel size of 9.6 μm, an MLA of a 3.3-mm focal length and 0.985-mm pitch, and a wedge-shaped freeform eyepiece with an equivalent focal length of 28 mm, along with a see-through compensator. The gap between the microdisplay and MLA was about 4.3 mm and the reference plane of the micro-InI unit was approximately 10 mm away from the MLA. After being magnified by the eyepiece, the virtual reference plane of the reconstructed 3-D scene in the visual space was approximately 1 diopter away from the exit pupil of the eyepiece. An array of 12 × 11 elemental images were simulated, each of which consists of 102 × 102 color pixels. The FOV of the reconstructed 3-D scene by the eyepiece was about 33.4° diagonally. The viewing angle of the micro-InI unit was about 14°, which yields a crosstalk-free EPD of 5 mm in which the 3-D InI view can be observed. The distance from the system exit pupil to the freeform eyepiece is 19 mm. The spatial resolution of reconstructed 3-D scene by the micro-InI unit was about 22.4 μm on the reference plane, which yields an angular resolution of 2.7 arcminutes on the virtual reference plane in the visual space. Just like other InI-based systems, the spatial resolution degrades as the depth of the reconstructed 3-D object deviates from the virtual reference plane.

Fig. 17(c) and (d) shows two views captured by a camera focusing at 0.25 and 3 diopters, respectively. In the see-through view, two physical references, a Snellen eye chart and a resolution grating, were placed at a distance of 0.25 diopters and 3 diopters, respectively, from the camera. In the virtual view, three columns of letter "E" were rendered at a distance of 0.25, 1, and 3 diopters, respectively. The letters on the bottom row, which were half the size of the letters on the top row, were about the same size as the largest letter on the Snellen chart with similar stroke width or gap. The gap or stroke of the largest letter on the Snellen chart represents an angular resolution of 15 minutes of arc at the distance of 4 m.

The results in Fig. 17 clearly demonstrate that an InI-based HMD method with an eyepiece can produce correct focus cues and true 3-D viewing in a large depth range spanning a few diopters, which is significantly different from the method of using MLA only as in [51]. On the other hand, artifacts admittedly are visible in these images. Like other InI-based display methods, the current InI-HMD method is subject to the limitations of relatively low spatial resolution, narrow viewing angle, as well as crosstalk due to the limited imaging capability and finite aperture of the MLAs, inadequate spatial resolution of displays, and tradeoff relationship between large EPD and high resolution. Further technological innovations are needed to improve the optical performances.

VIII. COMPUTATIONAL MULTILAYER HEAD-WORN LIGHT-FIELD DISPLAYS

Instead of using the InI-based architecture to angularly sample the directional lightrays apparently emitted by a 3-D scene through a pinhole array or MLA, a computational multilayer light-field display samples the directional rays through multilayers of pixel arrays. It typically employs a stack of light-attenuating layers illuminated by either a uniform or directional backlight. As illustrated in Fig. 5(e), the light field of a 3-D scene is computationally decomposed into a number of masks representing the transmittance of each layer of the light attenuators. The intensity value of each lightray entering the eye from the backlight is the product of the pixel values of the attenuation layers which the ray intersects. Unlike the additive nature of the MFP approach to light-field rendering, the computational multilayer displays approximate the light fields of the scene by multiplicatively attenuating the backlight.

It is worth mentioning that a computational multilayer light-field display essentially is an adaptation of the well-known parallax-barrier autostereoscopic displays by

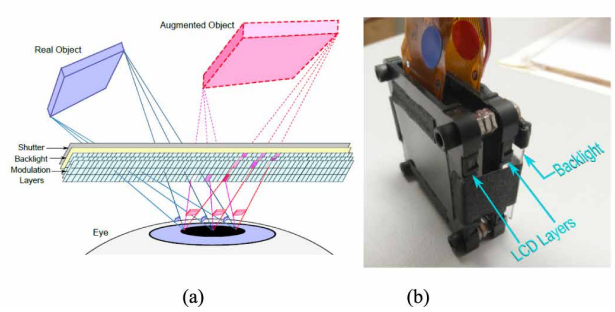

Fig. 18. *A computational multilayer optical see-through light-field display: (a) schematic layout; (vtype implementation [56].*

introducing content-adaptive optimization to compute the proper attenuation values for the mask layers. For instance, Wetzstein *et al.* demonstrated a new tensor display comprised of a stack of time-multiplexed light-attenuating layers illuminated by a directional backlight for autostereoscopic displays [57]. Unlike a conventional parallax barrier display utilizing a fixed barrier to create multiple views, here the modulation pattern of each attenuation layer is optimized to produce images for a desired viewing zone. In this sense, the barrier layer is dynamically controlled based on the 3-D scene. Maimone *et al.* demonstrated the feasibility of stimulating correct eye accommodation by synthesizing dense samples of the light fields over the eye pupil for autostereoscopic displays [106].

Maimone and Fuchs pioneered the work to apply the multilayer computational light-field display technique for its usage in HMDs and demonstrated the first computational multilayer AR display [56]. A stack of transparent SLMs, a thin, transparent backlight, and a high-speed shutter were sandwiched together with a small spacing between the SLM layers. Fig. 18(a) shows a schematic layout and Fig. 18(b) shows a prototype implementation. The sandwiched stack was placed directly in front of the eye, without any other focusing optics placed in between the display stack and the eye. The device operates in two modes: augmented image rendering mode (shutter off) and the occluded real-world image formation mode (shutter on). In the augmented view mode, the real-world view is blocked and a set of optimized patterns are rendered on the SLM layers to attenuate the lightrays from the backlight and produce the final color of the rays entering the eye, which is the product of the attenuation values assigned to each of the intersected pixels across the layers. The multiple SLM layers reproduce a set of lightrays with adequate angular resolution over the eye pupil that appear to be emitted from a virtual object at an apparent location far from the device stack. The eye perceives the sum of these rays integrally which synthesizes a virtual object at a desired depth. Therefore, the image formation process of the multilayer display is similar to that of integral imaging and able to reconstruct the light field of a 3-D virtual object with correct focus cues. In the real-world image formation mode, the backlight is turned off and the shutter is turned on. Occlusion masks can be displayed on the SLM layers to allow selective transmission of the real-world rays, enabling mutual occlusion between virtual and real-world scenes. Due to the close proximity of the SLM stack to the eye, this method could potentially achieve a compact OST-HMD with a wide FOV. Owing to the light-field rendering nature, it can also potentially render correct focus cues and mutual occlusion capabilities. Their early prototype demonstrated these capabilities to some extent. On the other hand, this approach is subject to major limitations. For instance, both the augmented image rendering mode and the occlusion-enabled see-through mode, although recognizable, suffer dramatic resolution loss due to diffraction effects through the SLM stack. The see-through view of the real world is blurry and low contrast due to the diffraction effects of the SLMs as well as the partial transparency.

More recently, Wetzstein *et al.* extended their multilayer factored light-field autostereoscopic display method and demonstrated a light-field stereoscope for immersive VR applications [107]. The schematic layout and the prototype implementation are shown in Fig. 19(a) and (b), respectively. The prototype consists of two stacked liquid crystal display panels with a gap of 53 cm in between and a pair of simple magnifier lenses with a focal length of 5 cm for each eye. Modulation patterns are computed using

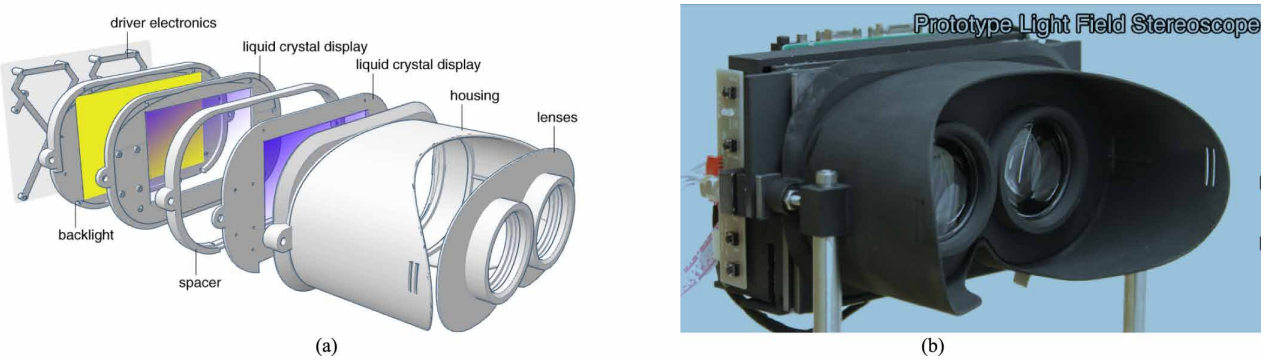

Fig. 19. *A computational multilayer light-field stereoscope: (a) schematic layout; (b) prototype implementation [107].*

a rank-1 light-field factorization process to synthesize and render the light field of a 3-D scene. Although their preliminary demonstration is promising, this approach is subject to a diffraction limit due to the fact that the virtual image pattern on the rear display panel is observed through the front panel.

Overall, the computational multilayer approach to head-worn light-field display is subject to a number of obvious limitations and requires significant innovations to enable improvements. First, it suffers from significantly reduced spatial resolution due to diffraction artifacts caused by the small pixel apertures of high-resolution attenuation layers. The diffraction artifacts can be reduced by utilizing reflective SLMs with high fill factors at the cost of increasing optics complexity and volume [108]. Second, the technique is computational intensive, requiring significant reduction of the optimization time to enable low-latency applications for AR displays. Finally, the prototype also suffers from high light loss due to the low transmittance of the SLM stack.

IX. CONCLUSION

This paper provided a summary of the well-known VAC problem in HMDs and related psychophysical findings of the visual artifacts associated with the problem. In the context of HMD systems for VR and AR applications, we further presented an overview of various technical methods toward resolving the VAC problem and summarized the recent progress. Despite the tremendous progress in the past decades, all of the existing technical approaches are subject to different tradeoffs and limitations. The Maxwellian view display approach offers a relaxation of the accommodation requirement for the virtual display by providing an extended DOF, but it is unable to produce natural retinal blur cues so that certain levels of visual conflicts still exist. The VFD approach provides a low-cost, good approximation to the focus cues produced by natural vision for 3-D scenes near the vergence depth, but discrepancy still exists when the scene volume is large. Moreover, it requires robust tracking of vergence depth through eyetrackers or other mechanisms, which can be a daunting challenge. The MFP display method is capable of rendering correct or nearly correct focus cues for a 3-D scene across a large depth volume at high spatial resolution comparable to conventional non-light-field HMD methods, but it has to overcome several critical obstacles to become a compact, wearable display solution as pointed out in Section VI. The micro-InI based light-field display approach is able to render correct focus cues for a large depth volume, but its optical performances, including spatial resolutions, DOF, and resolution of viewing angles, are relatively low compared to the MFP approach. The computational multilayer approach is still in its preliminary development stage and requires significant innovations to overcome some fundamental limitations such as diffraction artifacts. We still have a long way to go to engineer head-mounted displays with compact and portable form factor and high optical performance. However, we are confident that substantial progress and breakthroughs will be made over the next few years, making possible the commercial launch of HMDs, which are less vulnerable to visual discomfort and visual artifacts, for consumer applications. ■

Acknowledgement

Reviews on the multifocal plane displays are partially based on coauthored papers and the Ph.D. dissertation work by Dr. S. Liu and Dr. X. Hu who were former graduate students affiliated with the 3-D Visualization and Imaging Systems Laboratory at the University of Arizona. The author would like to thank Prof. B. Javidi at the University of Connecticut for the collaborative work on the integral-imaging-based display. She would also like to thank Dr. D. Luebke, Prof. H. Fuchs, and Prof. G. Wetzstein for giving her the permission to use the figures in their publications. Dr. H. 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] S. M. Ebenholtz, *Oculomotor Systems and Perception*. New York, NY, USA: Cambridge Univ. Press, 2001.

[2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov. 2001.

[3] S. K. Feiner, "Augmented reality: A new way of seeing," *Sci. Amer.*, vol. 54, pp. 48–55, Apr. 2002.

[4] J. Semmlow and P. Wetzel, "Dynamic contributions of the components of binocular vergence," *J. Opt. Soc. Amer.*, vol. 69, no. 5, pp. 638–645, 1979.

[5] L. Marran and C. Schor, "Multiaccommodative stimuli in VR systems: Problems & solutions," *Human Factors*, vol. 39, no. 3, pp. 382–388, 1997.

[6] T. Inoue and H. Ohzu, "Accommodative responses to stereoscopic three-dimensional display," *Appl. Opt.*, vol. 36, no. 19, pp. 4509–4515, 1997.

[7] C. Vienne, L. Sorin, L. Blonde, Q. Huynh-Thu, and P. Mamassian, "Effect of the accommodation-vergence conflict on vergence eye movements," *Vis. Res.*, vol. 100, pp. 124–133, Jul. 2014.

[8] T. Takeda, K. Hashimoto, N. Hiruma, and Y. Fukui, "Characteristics of accommodation toward apparent depth," *Vis. Res.*, vol. 39, no. 12, pp. 2087–2097, 1999.

[9] J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural problems for stereoscopic depth perception in virtual environments," *Vis. Res.*, vol. 35, no. 19, pp. 2731–2736, 1995.

[10] J. P. Frisby, D. Buckley, and J. M. Horsman, "Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges," *Perception*, vol. 24, no. 2, pp. 181–198, 1995.

[11] V. Interrante, B. Ries, and L. Anderson, "Distance perception in immersive virtual environments, revisited," in *Proc. IEEE Conf. Virtual Reality*, Mar. 2006, pp. 3–10.

[12] W. B. Thompson, V. Dilda, and S. H. Creem-Regehr, "Absolute distance perception to locations off the ground plane," *Perception*, vol. 36, no. 11, pp. 1559–1571, 2007.

[13] S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *J. Vis.*, vol. 5, no. 10, pp. 834–862, 2005.

[14] P. A. Howarth, "Potential hazards of viewing 3-D stereoscopic television, cinema and computer games: A review," *Ophthalmic Physil. Opt.*, vol. 31, pp. 111–122, Mar. 2011.

[15] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis.*, vol. 8, no. 3, pp. 1–30, 2008.

[16] J. M. Loomis and J. M. Knapp, "Visual perception of egocentric distance in real and virtual environments," *Virtual and Adaptive Environments: Applications, Implications and Human Performance Issues*, L. J. Hettinger and J. W. Haas, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Assoc., 2003, pp. 21–46.

[17] J. Siderov and R. S. Harwerth, "Precision of stereoscopic depth perception from double images," *Vis. Res.*, vol. 33, no. 11, pp. 1553–1560, 1993.

[18] A. Wilkins, *Visual Stress*. Oxford, U.K.: Oxford Univ. Press, 1995.

[19] K. Ukai and P. A. Howarth, "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations," *Displays*, vol. 29, no. 2, pp. 106–116, Mar. 2010.

[20] M. Mon-Williams, J. P. Wann, and S. Rushton, "Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display," *Ophthalmic Physiol. Opt.*, vol. 13, no. 4, pp. 387–391, 1993.

[21] B. B. Rainey, "The effect of prism adaptation on the response AC/A ratio," *Ophthalmic Physiol. Opt.*, vol. 20, no. 3, pp. 199–206, 2000.

[22] C. Schor, "Influence of accommodative and vergence adaptation on binocular motor disorders," *Amer. J. Optometry Physiol. Opt.*, vol. 65, no. 6, pp. 464–475, 1988.

[23] C. Schor and D. Horner, "Adaptive disorders of accommodation and vergence in binocular dysfunction," *Ophthalmic Physiol. Opt.*, vol. 9, no. 3, pp. 264–268, 1989.

[24] H. Hasebe, H. Oyamada, K. Ukai, H. Toda, and T. Bando, "Changes in oculomotor functions before and after loading of a 3-D visually-guided task by using a head-mounted display," *Ergonomics*, vol. 39, no. 11, pp. 1330–1343, 1996.

[25] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV/HDTV images," *Displays*, vol. 23, no. 4, pp. 191–201, Sep. 2002.

[26] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays*, vol. 25, no. 4, pp. 141–150, 2004.

[27] T. Iwasaki and A. Tawara, "Effects of viewing distance on accommodative and pupillary responses following a three-dimensional task," *Ophthalmic Physiol. Opt.*, vol. 22, no. 2, pp. 113–118, 2002.

[28] Y. Suzuki, Y. Onda, S. Katada, S. Ino, and T. Ifukube, "Effects of an eyeglass-free 3-D display on the human visual system," *Jpn. J. Ophthalmol.*, vol. 48, no. 1, pp. 1–6, 2004.

[29] M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *J. Display Technol.*, vol. 1, no. 2, pp. 328–340, Dec. 2005.

[30] J. C. Read and I. Bohr, "User experience while viewing stereoscopic 3D television," *Ergonomics*, vol. 57, no. 8, pp. 1140–1153, 2014.

[31] J. C. Read, A. Godfrey, I. Bohr, J. Simonotto, B. Galna, and T. V. Smulders, "Viewing 3D TV over two months produces no discernible effects on balance, coordination or eyesight," *Ergonomics*, vol. 59, no. 8, pp. 1073–1088, 2016.

[32] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, p. 11, 2011.

[33] G. Westheimer, "Maxwellian viewing system," *Vis. Res.*, vol. 6, pp. 669–682, 1966.

[34] R. A. Koetting, "Stereopsis in presbyopes fitted with single vision contact lenses," *Amer. J. Optometry Arch. Amer. Acad. Optometry*, vol. 47, no. 7, pp. 557–561, 1970.

[35] S. Shiwa, K. Omura, and F. Kishino, "Proposal for a 3-D display with accommodative compensation: 3DDAC," *J. Soc. Inf. Display*, vol. 4, no. 4, pp. 255–261, 1996.

[36] T. Shibata *et al.*, "Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence," *J. Soc. Inf. Display*, vol. 13, no. 8, pp. 665–671, 2005.

[37] S. Liu, D. Cheng, and H. Hua, "An optical see-through head mounted display with addressable focal planes," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2008, pp. 32–42.

[38] S. Kuiper and B. H. W. Hendriks, "Variable-focus liquid lens for miniature cameras," *Appl. Phys. Lett.*, vol. 85, no. 7, pp. 1128–1130, Aug. 2004.

[39] H. Ren, D. W. Fox, B. Wu, and S. T. Wu, "Liquid crystal lens with large focal length tunability and low operating voltage," *Opt. Exp.*, vol. 15, no. 18, pp. 11328–11335, 2007.

[40] E. J. Fernández and P. Artal, "Membrane deformable mirror for adaptive optics: Performance limits in visual optics," *Opt. Exp.*, vol. 11, no. 9, pp. 1056–1069, 2003.

[41] J. P. Rolland, M. W. Kureger, and A. Goon, "Multifocal planes head-mounted displays," *Appl. Opt.*, vol. 39, no. 19, pp. 3209–3215, 2000.

[42] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.*, vol. 23, pp. 804–813, 2004.

[43] B. T. Schowengerdt, M. Murari, and E. J. Seibel, "Volumetric display using scanned fiber array," in *SID Symp. Dig. Tech. Papers*, May 2010, pp. 653–656.

[44] D. Cheng, Q. Wang, Y. Wang, and G. Jin, "Lightweight spatial-multiplexed dual focal-plane head-mounted display using two freeform prisms," *Chin. Opt. Lett.*, vol. 11, no. 3, p. 031201, 2013.

[45] S. C. McQuaide, E. J. Seibel, J. P. Kelly, B. T. Schowengerdt, and T. A. Furness, III, "A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror," *Displays*, vol. 24, no. 2, pp. 65–72, 2003.

[46] S. Suyama, M. Date, and H. Takada, "Three-dimensional display system with dual-frequency liquid-crystal varifocal lens," *Jpn. J. Appl. Phys.*, vol. 39, no. 2A, pp. 480–484, 2000.

[47] S. Liu and H. Hua, "Time-multiplexed dual-focal plane head-mounted display with a fast liquid lens," *Opt. Lett.*, vol. 34, no. 11, pp. 1642–1644, 2009.

[48] G. D. Love, D. M. Hoffman, P. J. W. Hands, J. Gao, A. K. Kirby, and M. S. Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Opt. Exp.*, vol. 17, no. 18, pp. 15716–15725, 2009.

[49] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, pp. 821–825, 1908.

[50] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, and applications," *Appl. Opt.*, vol. 52, no. 4, pp. 546–560, 2013.

[51] D. Lanman and D. Luebke, "Near-eye light field displays," in *Proc. ACM SIGGRAPH*, 2013.

[52] H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Opt. Exp.*, vol. 22, no. 11, pp. 13484–13491, 2014.

[53] 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.*, vol. 12, no. 6, p. 060010, 2014.

[54] T. Peterka, R. L. Kooima, D. J. Sandin, A. Johnson, J. Leigh, and T. A. DeFanti, "Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 3, pp. 487–499, Jun. 2008.

[55] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540–555, Apr. 2011.

[56] A. Malmone and H. Fuchs, "Computational augmented reality eyeglasses," in *Proc. ISMAR*, 2013, pp. 29–38.

[57] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar, "Tensor displays: Compressive light field synthesis using multilayer displays with directional backlighting," in *Proc. ACM SIGGRAPH*, 2012.

[58] R. Konrad, E. A. Cooper, and G. Wetzstein, "Novel optical configurations for virtual reality: Evaluating user preference and performance with focus-tunable and monovision near-eye displays," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, May 2016, pp. 1211–1220.

[59] P. V. Johnson, J. A. Parnell, J. Kim, C. D. Saunter, G. D. Love, and M. S. Banks, "Dynamic lens and monovision 3D displays to improve viewer comfort," *Opt. Exp.*, vol. 24, no. 11, pp. 11808–11827, 2016.

[60] B. G. Blundell and A. J. Schwarz, "The classification of volumetric display systems: Characteristics and predictability of the image space," *IEEE Trans. Vis. Comput. Graphics*, vol. 8, no. 1, pp. 66–75, Jan. 2002.

[61] E. Moon, M. Kim, J. Roh, H. Kim, and J. Hahn, "Holographic head-mounted display with RGB light emitting diode light source," *Opt. Exp.*, vol. 22, no. 6, pp. 6526–6534, 2014.

[62] Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," *Opt. Exp.*, vol. 18, no. 9, pp. 8824–8835, 2010.

[63] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an

interactive 360° light field display," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 40.

[64] T. Ando, K. Yamasaki, M. Okamoto, T. Matsumoto, and E. Shimizu, "Retinal projection display using holographic optical element," *Proc. SPIE*, vol. 3956, p. 211, Mar. 2000.

[65] T. Ando and E. Shimizu, "A head-mounted display using a holographic optical element," in *Three-Dimensional Television, Video, and Display Technologies*, B. Javidi and F. Okano, Eds., New York, NY, USA: Springer-Verlag, 2002, pp. 67–100.

[66] M. von Waldkirch, P. Lukowicz, and G. Tröster, "LCD-based coherent wearable projection display for quasi-accommodation-free imaging," *Opt. Commun.*, vol. 217, nos. 1–6, pp. 133–140, Mar. 2003.

[67] M. von Waldkirch, P. Lukowicz, and G. Tröster, "Effect of light coherence on depth of focus in head-mounted retinal projection displays," *Opt. Eng.*, vol. 43, no. 7, pp. 1552–1560, 2004.

[68] M. von Waldkirch, P. Lukowicz, and G. Tröster, "Defocusing simulations on a retinal scanning display for quasi accommodation-free viewing," *Opt. Exp.*, vol. 11, no. 24, pp. 3220–3233, 2003.

[69] M. von Waldkirch, P. Lukowicz, and G. Tröster, "Multiple imaging technique for extending depth of focus in retinal displays," *Opt. Exp.*, vol. 12, no. 25, pp. 6350–6365, 2004.

[70] M. von Waldkirch, P. Lukowicz, and G. Tröster, "Oscillating fluid lens in coherent retinal projection displays for extending depth of focus," *Opt. Commun.*, vol. 253, pp. 407–418, Sep. 2005.

[71] A. Yuuki, K. Itoga, and T. Satake, "A new Maxwellian view display for trouble-free accommodation," *J. Soc. Inf. Display*, vol. 20, no. 10, pp. 581–588, Oct. 2012.

[72] 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. Graphics*, vol. 33, no. 4, pp. 1–11, 2014.

[73] S. Hillaire, A. Lécuyer, R. Cozot, and G. Casiez, "Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments," in *Proc. IEEE Virtual Reality Conf.* Mar. 2008, pp. 47–50.

[74] M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath, "Depth perception with gaze-contingent depth of field," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2014, pp. 217–226.

[75] S. Liu, H. Hua, and D. Cheng, "A novel prototype for an optical see-through head-mounted display with addressable focus cues," *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 3, pp. 381–393, May 2010.

[76] S. Liu, "Methods for generating addressable focus cues in stereoscopic displays," Ph.D. dissertation, Univ. Arizona, Tucson, AZ, USA, 2010.

[77] F. W. Campbell, "The depth of field of the human eye," *Opt. Acta, Int. J. Opt.*, vol. 4, no. 4, pp. 157–164, 1957.

[78] A. F. Crabtree, "Method and apparatus for manipulating, projecting, and displaying light in a volumetric format," U.S. Patent 55723755, Nov. 5, 1996.

[79] A. Sullivan, "Multiplanar volumetric displays for medical visualization," presented at the Med. Meets Virtual Reality, Adv. Res. Projects Agency Workshop, San Francisco, CA, USA, Jan. 1999.

[80] B. T. Schowengert and E. J. Seibel, "True 3-D scanned voxel displays using single or multiple light sources," *J. Soc. Inf. Display*, vol. 14, no. 2, pp. 135–143, 2006.

[81] P. Llull, N. Bedard, W. Wu, I. Tosic, K. Berkner, and N. Balram, "Design and optimization of a near-eye multifocal display system for augmented reality," *Opt. Soc. Amer. Imag. Appl. Opt. Congr.*, vol. 2015, p. JTH3A, 2015.

[82] S. Suyama, S. Ohtsuka, H. Takada, K. Uehira, and S. Sakai, "Apparent 3-D image perceived from luminance-modulated two 2-D images displayed at different depths," *Vis. Res.*, vol. 44, no. 8, pp. 785–793, 2004.

[83] S. Liu and H. Hua, "A systematic method for designing depth-fused multi-focal plane three-dimensional displays," *Opt. Exp.*, vol. 18, no. 11, pp. 11562–11573, 2010.

[84] X. Hu and H. Hua, "Distinguished student paper: A depth-fused multi-focal-plane display prototype enabling focus cues in stereoscopic displays," in *SID Symp. Dig. Tech. Papers*, vol. 42. 2011, pp. 691–694.

[85] X. Hu and H. Hua, "Design and assessment of a depth-fused multi-focal-plane display prototype," *J. Display Technol.*, vol. 10, no. 4, pp. 308–316, Apr. 2014.

[86] X. Hu, "Development of the depth-fused multi-focal plane display technology," Ph.D. dissertation, College Opt. Sci., Univ. Arizona, Tucson, AZ, USA, 2014.

[87] X. Hu and H. Hua, "High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics," *Opt. Exp.*, vol. 22, no. 11, pp. 13896–13903, 2014.

[88] S. Ravikumar, K. Akeley, and M. S. Banks, "Creating effective focus cues in multi-plane 3D displays," *Opt. Exp.*, vol. 19, no. 21, pp. 20940–20952, 2011.

[89] X. Hu and H. Hua, "Design and tolerance of a free-form optical system for an optical see-through multi-focal-plane display," *Appl. Opt.*, vol. 54, no. 33, pp. 9990–9999, 2015.

[90] R. Narain, R. A. Albert, A. Bulbul, G. J. Ward, M. S. Banks, and J. F. O'Brien, "Optimal presentation of imagery with focus cues on multi-plane displays," *ACM Trans. Graph.*, vol. 34, no. 4, 2015, Art. no. 59.

[91] W. Wu, P. Llull, I. Tosic, N. Bedard, K. Berkner, and N. Balram, "Content-adaptive focus configuration for near-eye multi-focal displays," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2016, pp. 1–6.

[92] B. Javidi, J. Sola-Pikabea, and M. Martinez-Corral, "Breakthroughs in photonics 2014: Recent advances in 3-D integral imaging sensing and display," *IEEE Photon. J.*, vol. 7, no. 3, Jun. 2015, Art. no. 0700907.

[93] J. S. Jang and B. Javidi, "Three-dimensional integral imaging of micro-objects," *Opt. Lett.*, vol. 29, no. 11, pp. 1230–1232, 2004.

[94] B. Javidi, I. Moon, and S. Yeom, "Three-dimensional identification of biological microorganism using integral imaging," *Opt. Exp.*, vol. 14, no. 25, pp. 12095–12107, 2006.

[95] Y.-T. Lim, J.-H. Park, K.-C. Kwon, and N. Kim, "Resolution-enhanced integral imaging microscopy that uses lens array shifting," *Opt. Exp.*, vol. 17, no. 21, pp. 19253–19263, 2009.

[96] M. Levoy, "Light fields and computational imaging," *IEEE Comput.*, vol. 39, no. 8, pp. 46–55, Aug. 2006.

[97] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.

[98] Y. Kim, K. Hong, and B. Lee, "Recent researches based on integral imaging display method," in *3D Research*. New York, NY, USA: Springer-Verlag, 2010.

[99] Q. H. Wang, C. C. Ji, L. Li, and H. Deng, "Dual-view integral imaging 3D display by using orthogonal polarizer array and polarization switcher," *Opt. Exp.*, vol. 24, no. 1, pp. 9–16, 2016.

[100] A. Stern, Y. Yitzhaky, and B. Javidi, "Perceivable light fields: Matching the requirements between the human visual system and autostereoscopic 3-D displays," *Proc. IEEE*, vol. 22, no. 10, pp. 1571–1587, Oct. 2014.

[101] J. Hong, S. Min, and B. Lee, "Integral floating display systems for augmented reality," *Appl. Opt.*, vol. 51, no. 18, pp. 4201–4209, 2012.

[102] H. Hua, X. Hu, and C. Gao, "A high-resolution optical see-through head-mounted display with eyetracking capability," *Opt. Exp.*, vol. 21, no. 25, pp. 30993–30998, 2013.

[103] 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.*, vol. 48, o. 14, pp. 2655–2668, 2009.

[104] 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.*, vol. 12, no. 6, p. 060–010, 2014.

[105] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar, "Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization," *ACM Trans. Graph.*, vol. 29, no. 6, 2010, Art. no. 163.

[106] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs, "Focus 3D: Compressive accommodation display," *ACM Trans. Graph.*, vol. 32, no. 5, p. 153:1, 2013.

[107] F. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM SIGGRAPH*, vol. 33, no. 5, 2015.

[108] M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM SIGGRAPH Emerg. Technol.*, vol. 33, no. 4, 2014.

ABOUT THE AUTHOR

Hong Hua received the B.S.E. and Ph.D. degrees in optical engineering (with honors) from Beijing Institute of Technology, Beijing, China in 1994 and 1999, respectively.

She is currently a Professor in the College of Optical Sciences, University of Arizona, Tucson, AZ, USA. She has over 20 years of experiences in developing and evaluating head-mounted display technologies and developing virtual reality and augmented reality systems and applications. She has been directing the 3-D Visualization and Imaging Systems Laboratory. Her current research interests mainly include various head-mounted displays and light-field displays, optical engineering, virtual and augmented reality, and 3-D human–computer interaction.

Prof. Hua is a Fellow of the International Society for Optics and Photonics (SPIE).