Focus 3D: Compressive Accommodation Display

ANDREW MAIMONE
University of North Carolina at Chapel Hill
GORDON WETZSTEIN and MATTHEW HIRSCH
MIT Media Lab
DOUGLAS LANMAN
MIT Media Lab, NVIDIA Research
RAMESH RASKAR
MIT Media Lab
and
HENRY FUCHS
University of North Carolina at Chapel Hill

We present a glasses-free 3D display design with the potential to provide viewers with nearly correct accommodative depth cues, as well as motion parallax and binocular cues. Building on multilayer attenuator and directional backlight architectures, the proposed design achieves the high angular resolution needed for accommodation by placing spatial light modulators about a large lens: one conjugate to the viewer's eye, and one or more near the plane of the lens. Nonnegative tensor factorization is used to compress a high angular resolution light field into a set of masks that can be displayed on a pair of commodity LCD panels. By constraining the tensor factorization to preserve only those light rays seen by the viewer, we effectively steer narrow high resolution viewing cones into the user's eyes, allowing binocular disparity, motion parallax, and the potential for nearly correct accommodation over a wide field of view. We verify the design experimentally by focusing a camera at different depths about a prototype display, establish formal upper bounds on the design's accommodation range and diffraction-limited performance, and discuss practical limitations that must be overcome to allow the device to be used with human observers.

Categories and Subject Descriptors: I.3.1 [**Computer Graphics**]: Hardware Architecture—*Three-dimensional displays*

General Terms: Algorithms, Design, Experimentation, Human Factors

Additional Key Words and Phrases: eye accommodation, retinal blur, compressive displays, automultiscopic displays, light fields, high angular resolution backlighting, nonnegative tensor factorization

1. INTRODUCTION

Within the last few years, stereoscopic image display has become common in cinemas and in consumer televisions. Professionally produced stereoscopic content is now widely available and the technology for creating stereo images for both synthetically rendered scenes and captured footage has matured to a point where many newly released movies can be enjoyed in 3D. Next-generation displays will remove the need for additional eyewear, while presenting high-quality imagery; emerging computational light field displays show promise to satisfy the high standards set by the consumer market. State-of-the-art light field displays provide thin form factors, high resolution, and light efficient display modes, while supporting binocular disparity and smooth motion parallax over a wide field of view. Unfortunately, most available 3D displays share a common limitation: lack of the focus depth cues, accommodation and retinal blur.

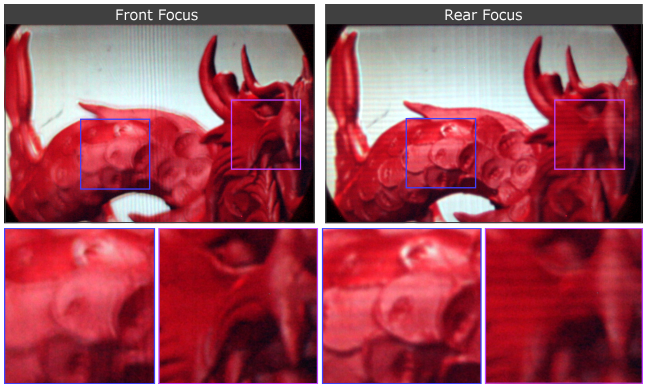

Fig. 1. Photograph of prototype display focused at two different depths. Bottom row shows magnifications of inset regions. The prototype shown was configured with a single LCD layer placed directly in front of a high angular resolution backlight (HARB) and was photographed with a large aperture at a distance of 127 cm. Dragon model courtesy of Stanford Computer Graphics Laboratory.

Accommodation is an important depth cue driven by the focal state of the lens in a human eye; the ciliary muscles contract and relax to change the shape of the lens, causing a change in focus. Takaki [2006] experimentally verified that projecting as few as two different perspectives in one pupil stimulates accommodative responses in a human observer. Retinal blur is a complementary depth cue stimulated by the sensed magnitude of focal blur on the retina; inclusion of this cue has been shown to improve the performance of certain visual tasks [Hoffman and Banks 2010]. When these focus cues are correct or nearly correct (i.e., they closely match the depths of the displayed scene), as in a natural environment, the performance of the visual system is enhanced; however, displays lacking these cues cause significant viewer fatigue, due to a conflict with other cues [Hoffman et al. 2008]. Since retinal blur is preserved by most displays that support accommodation, we concentrate on accommodation in the majority of this paper while also discussing retinal blur in Section 3.5.

With the exception of ultra-high resolution displays, such as holograms, small volumetic displays, and multi-focal devices requiring specialized eye-worn equipment, no existing 3D display simultaneously supports correct accommodation, binocular disparity, and motion parallax over a wide field of view. We propose a new computational display design, dubbed Focus 3D, that has the potential to synthesize light fields with sufficient angular resolution to allow near correct viewer accommodation and retinal blur in addition to smooth motion parallax and binocular disparity. The key innovation is a combination of display optics and compressive light field synthesis through nonnegative light field tensor factorization. Inspired by Tensor Displays [Wetzstein et al. 2012], we explore multilayer display architectures with directional backlighting; however, instead of synthesizing a low angular resolution light field with a predefined field of view, we introduce high angular resolution (HAR) backlighting that allows high-resolution *view cones* to be steered into an observer's eyes. Due to the novel architecture, each view cone has a significantly larger depth of field than previously proposed solutions, offering the potential for the visual system to focus the eyes. We demonstrate the viability of this design through the construction of a prototype display that allows a camera to focus at multiple depths about the display (see Fig. 1).

1.1 Contributions

We explore a computational approach to synthesizing light fields as a set of narrow, but ultra high-resolution view cones that are steered only where required: into the viewer's eyes. Specific contributions are as follows:

(1) We introduce a computational display design that has the potential to support near correct accommodation depth cues in addition to binocular disparity and motion parallax. The proposed design combines two existing optical configurations in a novel way and employs nonnegative tensor factorization (NTF) to decompose a given light field into an optimal set of patterns.

(2) We explore steering of light field view cones to allow an observer to use a display with an increased perceived field of view and depth of field.

(3) We evaluate the proposed design using simulations and demonstrate its feasibility using a hardware prototype driven by an efficient GPU-based NTF solver.

(4) We present a formal analysis of the upper bounds of the accommodation range and diffraction limits for multilayer displays.

1.2 Overview of Benefits and Limitations

Benefits. We describe a new optical display architecture, consisting of stacked display layers and a high resolution directional backlight, that provides a significantly increased depth of field over a small set of view cones steered into the eyes of the viewer, offering the potential to provide binocular disparity, motion parallax, and near correct accommodation. As a compressive light field display, Focus 3D increases the display brightness and field of view while reducing the required number of time-multiplexed frames as compared to conventional displays. As illustrated in Fig. 2, previous displays providing correct accommodation cues require either additional eyewear or a significantly higher optical and computational complexity. To the authors' best knowledge, Focus 3D is the first practical display that has the potential to support near correct accommodative depth cues while allowing the viewer to move around the device from a wide range of viewpoints – including multiple distances from the screen.

Technique	Optical Complexity	Computational Cost	Glasses-free	Binocular Disparity	Motion Parallax	Accommodation
Parallax Barriers & Integral Imaging	low	low	yes	yes	yes	no
Directional Backlights	medium	low	yes	yes	yes	no
Multilayer Displays	medium	medium	yes	yes	yes	no
Volumetric Displays	high	medium	yes	yes	yes	yes*
Supermultiview	high	high	yes	yes	yes	yes
Multi-focus Displays	medium	medium	no	yes	yes	yes
Holograms	high	high	yes	yes	yes	yes
Focus 3D (proposed)	medium	medium	yes	yes	yes	yes

*Within physical display enclosure

Fig. 2. Comparing benefits of a variety of 3D displays. Most existing technologies do not support correct accommodation; those that do require complex optical setups and are computationally expensive or require glasses.

Limitations. As with other multilayer displays, stacking multiple display elements increases moiré and color-channel crosstalk, decreases the overall display brightness, and presents an alignment challenge. Obtaining good performance in the proposed multilayer framework also requires display panels which exceed currently available refresh rates, although upcoming display technologies have been demonstrated with much higher rates [Hagood et al. 2007]. While our current prototype is about 50 cm thick, future generations of the proposed display may benefit from optical folding techniques such as wedge optics [Travis et al. 2013]. We employ an efficient GPU-based implementation of nonnegative tensor factorization to compute content-adaptive light field decompositions. While this approach adds to the computational complexity of the system, no heuristics are known to drive multilayer displays with the proposed type of directional backlighting.

Our prototype display is suitable for testing with a camera; several engineering enhancements would allow the display to be tested with human viewers. Constructing a display with sufficient angular resolution to support multiple depths of focus over a human-sized pupil diameter requires high quality optics. Although we provide simulations with such an aperture, our prototype display is limited to focus over a larger 2 cm camera aperture due to the performance of the inexpensive integrated Fresnel lens, which exhibited poor focus, especially off-axis. Our approach also requires high speed eye tracking; although in this paper we assume the eye positions of the observers are known, we note high speed ($\geq$500 Hz) encumbrance-free commercial trackers are available from such vendors as SensoMotoric Instruments[1]. Finally, the brightness of the display's backlight must be improved to permit human viewing.

2. RELATED WORK

Light Field Displays

Light field displays generally aim to create motion parallax and stereoscopic disparity so that an observer perceives a scene as 3D without having to wear encumbering glasses. Invented more than a century ago, two fundamental techniques enable most light field displays: parallax barriers [Ives 1903] and integral imaging with lenslet arrays [Lippmann 1908]. The former technology has evolved into fully dynamic display systems supporting head tracking and view steering [Perlin et al. 2000; Peterka et al. 2008], as well as high-speed temporal modulation [Lanman et al.

[1] http://www.smivision.com

2010]. Today, lenslet arrays are often used as programmable rear-illumination in combination with a high-speed LCD to steer different views toward tracked observers [Stolle et al. 2008]. Inspired by the evolution of both parallax barrier and lenslet-based displays, we explore computational displays that exploit adaptive view steering to extend the capabilities of current display technology.

Within the 21st century, a variety of glasses-free 3D display technologies have been proposed, including optically switchable diffusers [Sullivan 2003], multilayer architectures [Putilin et al. 2001; Gotoda 2010; Wetzstein et al. 2011; Lanman et al. 2011; Wetzstein et al. 2012] as well as directional backlighting for stereoscopic [Chu et al. 2005; Chien and Shieh 2006; Brott and Schultz 2010] and multiview [Toyooka et al. 2001; Mather et al. 2009; Kwon and Choi 2012] displays. All of these technologies support a low angular resolution that provides only binocular disparity and motion parallax depth cues. We propose a new computational display architecture with the potential to support correct accommodation through the joint design of display optics and compressive light field synthesis.

Travis [1990] proposed mounting a Fresnel lens behind a transparent, light-modulating front panel with an additional high-resolution backlight placed at a distance behind the lens. Subsequently, this backlight was improved by engineering a wedge-shaped light guide that allows for the same functionality to be integrated into a thin, transparent optical element [Travis et al. 2013]. These displays are driven in a time-sequential mode: display optics illuminate a single view direction at each time step while the corresponding view is displayed. While this optical design provides a sufficiently high angular resolution to support correct accommodation, currently available display refresh rates (∼120 Hz) prevent its use beyond a tracked stereoscopic display. We propose a display architecture that combines this high angular resolution backlight with stacks of light-attenuating layers. When driven by compressive light field synthesis, this approach is capable of nearly correct accommodation with currently available display refresh rates.

Displays Supporting Correct Accommodation

Displays supporting correct accommodation are able to create a light field with enough angular resolution to allow subtle, yet crucial, variation over the pupil (see Fig. 3). Such displays utilize three main approaches. Ultra-high angular resolution displays, such as super multiview displays [Takaki 2006; Takaki et al. 2011; Pamplona et al. 2012], take a brute-force approach: all possible views are generated and displayed simultaneously, incurring high hardware costs. In practice, these drawbacks have limited the size, field of view, and spatial resolution of the devices. Multi-focal displays [Akeley et al. 2004; Hoffman and Banks 2009; Shibata et al. 2005], virtually place conventional monitors at different depths via refractive optics. This approach is effective, but requires encumbering glasses. Volumetric displays [Favalora 2005] physically generate light rays at the perceived 3D position, but are limited to small volumes and cannot reproduce occlusion. Closely related light field displays with anisotropic diffusion surfaces [Jones et al. 2007; Cossairt et al. 2007] can reproduce small volumes with occlusion, but accommodation has been demonstrated in the horizontal dimension only within a limited depth range [Jones et al. 2007]. Unlike these approaches, the Focus 3D design is capable of generating near correct accommodation cues with high spatial resolution, in both the horizontal and vertical dimensions, over a wide field of view. The design uses simple hardware (e.g., 2-3 LCD panels and a lens), is scalable to large displays, and does not require glasses.

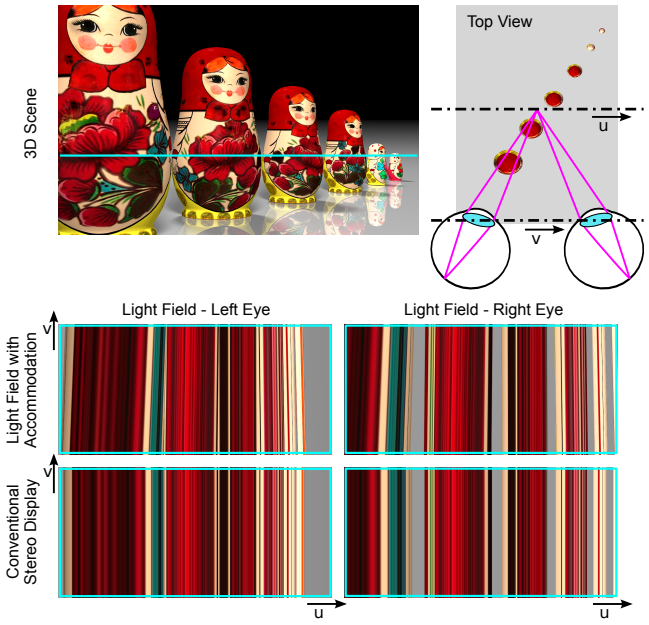

Fig. 3. Natural light fields *(top)* observed by the human visual system (HVS) exhibit stereoscopic disparity (i.e., different left and right eye perspectives) and parallax within the area of each pupil *(center)*. The subtle, but important, variation of the light field over the pupils allows the HVS to accommodate on different depths within the scene. Conventional 3D displays do not provide enough angular resolution to support this important depth cue *(bottom)*. Matryoshka doll model by artist "coboide" of Turbosquid.com.

3. FOCUS 3D ARCHITECTURE

The goal of the Focus 3D architecture is to efficiently provide accommodation, stereo, and motion parallax by steering a set of narrow high resolution light cones directly into the viewer's eyes. Our approach is a hybrid of the view sequential Cambridge display design [Travis 1990] and the multilayer display architecture of Tensor Displays [Wetzstein et al. 2012]; a brief overview follows.

In one variation of the Cambridge display design, an LCD layer is placed against a lens and illuminated by a backlight (refer to Fig. 7). If the backlight and viewer are placed at conjugate distances with respect to the lens, a point light source from the backlight will illuminate the LCD layer and rays will subsequently converge to a point at the viewing plane. Thus an image displayed on the LCD layer will be visible only to an observer in the viewing position corresponding to the illuminated region of the backlight. To create a time-multiplexed multiview display, a set of views are displayed in rapid sequence on the LCD layer, each while the corresponding region of the backlight is illuminated. We observe that it is straightforward to extend this design to support accommodation by incorporating a *high angular resolution (HAR) backlight*; with sufficient backlight resolution, multiple viewpoints can be created within the area of the pupil, providing the focus cues to the eyes. However, such a design would require display rates that far exceed currently technology; for example, a set of 5×5 views over each eye with a 60 Hz refresh rate would require a 3000 Hz display. The result would also be very dim, as each of the M views would be illuminated only a fraction $1/M$ of the time.

Our key innovation is to exploit the correlation between this large set of views using the compressive tensor factorization

framework recently introduced by Wetzstein et al. [2012], enabling eye accommodation with brighter imagery using the refresh rates of current and upcoming displays. In this embodiment, a compressed set of correlated view patterns is displayed in sequence on the LCD layer, each while multiple regions of the backlight (and thus the eyes) are illuminated simultaneously. Furthermore, we can replace the single LCD layer in front of the lens with an N layer *stack* of LCDs, increasing the spatial and angular resolution of the display as well as compression performance.

In the remainder of the this section, we describe the details of this approach and analyze performance and limitations. Section 3.1 establishes how to emit a light field to support correct accommodation using an N-layer, M-frame multilayer display illuminated by a high angular resolution (HAR) backlight. We show that such a display can be optimized using the aforementioned tensor display framework, albeit with a modified backlight illumination model. Section 3.2 assesses the structure of the backlight illumination and layer patterns produced by the decomposition; this analysis reveals the source of enhanced brightness achieved with Focus 3D over prior methods utilizing direct time-multiplexed backlight illumination schemes. Section 3.3 derives upper bounds on the accommodation range for both existing display architectures and Focus 3D. Section 3.4 analyzes how the design is affected by diffraction, and Section 3.5 concludes by showing the influence of diffraction and light field compression on retinal blur quality.

3.1 Displays with HAR Backlighting

As described above and shown in Figs. 4 and 7, Focus 3D consists of an N-layer stack of light-attenuating panels illuminated by a *high angular resolution (HAR) backlight* capable of synthesizing multiple uniform light sources that converge along a closely-spaced set of points spanning the viewer's pupils. Similar to Travis [1990], such a backlight can be fashioned by placing a large lens (e.g., a Fresnel lens or folded waveguide) against the rear layer. If another display is placed at a distance d_b behind the lens, then a virtual layer will be created at a distance $d_v = (fd_b)/(d_b - f)$ in front of the lens. A HAR backlight is obtained when d_b is selected such that d_v equals the distance d_e from the lens to the viewer's pupil.

3.1.1 *Representing Emitted Light Fields.* As shown in Fig. 4, we propose Focus 3D as a generalization of prior displays capable of supporting near correct accommodation through high angular resolution backlighting. Rather than using a single layer placed directly in front of the lens, we propose placing a stack of light-attenuating layers. For greater generality, we further assume that these layers support a higher refresh rate than the human eye, such that the viewer perceives the time average of an M-frame sequence. Such N-layer, M-frame displays have been optimized using the tensor display framework of Wetzstein et al. [2012]. As shown in that work, the emitted light field $\tilde{l}(x, v)$ can be modeled using the following image formation model:

$$\tilde{l}(x,v) = \frac{1}{M}\sum_{m=1}^{M} b_m(x,v) \prod_{n=1}^{N} f_m^{(n)}(x + (d_n/d_r)v), \quad (1)$$

where $b_m(x, v)$ is the light field emitted by the backlight during frame m, $f_m^{(n)}(\xi_n)$ is the transparency of layer n during frame m at position ξ_n, and d_n is the distance of layer n from the lens. In this section we adopt a two-plane light field parameterization, such that a ray (x, v) is defined by its intersection with the u-axis, coincident with the lens, and the v-axis, located a distance d_r from the lens.

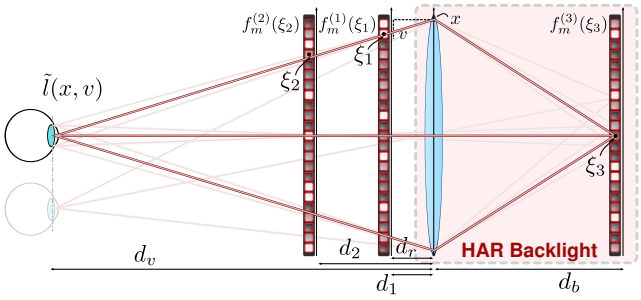

Fig. 4. Focus 3D coordinate system. An N-layer stack of light-attenuating panels is illuminated by a high angular resolution backlight. Each pixel on the backlight layer illuminates a small region of the viewer's pupil. We show a specific embodiment of a HAR backlight, comprising a large lens and a backlight display separated a distance d_b behind the lens, following the design of Travis [1990]. A generalized system is shown in Fig. 7.

The tensor display framework considers two cases: uniform backlighting, such that $b_m(x, v) = 1$, and directional backlighting, such that $b_m(x, v)$ is a low-resolution light field produced by an auxiliary system (e.g., a lenticular display). We observe that Equation 1 can be modified to support high angular resolution backlighting, as depicted in Fig. 4, such that

$$\tilde{l}(x,v) = \frac{1}{M}\sum_{m=1}^{M} f_m^{(N+1)}(\phi(x,v)) \prod_{n=1}^{N} f_m^{(n)}(x + (d_n/d_r)v), \quad (2)$$

where $\phi(x, v)$ defines the point of intersection ξ_{N+1} of ray (x, v) with the backlight layer, $f_m^{(N+1)}(\xi_{N+1})$ denotes the emitted irradiance of the backlight layer during frame m, and $\{f_m^{(n)}(\xi_n)\}$, for $n \in [1, N]$, remain the transparencies of the N layers in front of the lens. We observe that the point of intersection is found by tracing the ray (x, v) backwards through the lens, with focal length f, and propagating a distance d_b to the backlight layer. Using ray transfer matrix analysis [Hecht 2001] with paraxial ray and thin lens approximations, these operations are given by:

$$\begin{pmatrix} \phi(x,v) \\ -\eta/d_r \end{pmatrix} = \begin{pmatrix} 1 & d_b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} x \\ -v/d_r \end{pmatrix}, \quad (3)$$

where η/d_r denotes the slope of the ray leaving the backlight layer. Thus, $\phi(x, v)$ is given by the following expression.

$$\phi(x,v) = \left(1 - \frac{d_b}{f}\right) x - \frac{d_b}{d_r} v \quad (4)$$

3.1.2 *Decomposing Light Fields Using Weighted NTF.* Following the tensor display framework, the light field emitted by a N layer display can be decomposed into a set of M time-multiplexed layer patterns using *nonnegative tensor factorization (NTF)*. Substituting Equation 4 into Equation 2 provides a closed-form expression for the light field emitted by such a display, $\tilde{l}(x, v)$, in terms of the time-multiplexed layer patterns, $\{f_m^{(n)}(\xi_n)\}$. In practice, the decomposition of a target light field, $l(x, v)$, into the layer patterns requires solving the following nonlinear least squares problem:

$$\underset{\{f_m^{(n)}(\xi_n)\}}{\arg\min} \int_{\mathcal{V}}\int_{\mathcal{X}} \left(l(x,v) - \tilde{l}(x,v)\right)^2 dx\,dv, \text{ for } 0 \le f_m^{(n)}(\xi_n) \le 1 \quad (5)$$

To solve this optimization problem, having proven the equivalence of Equations 1 and 2, we refer the reader to Wetzstein et

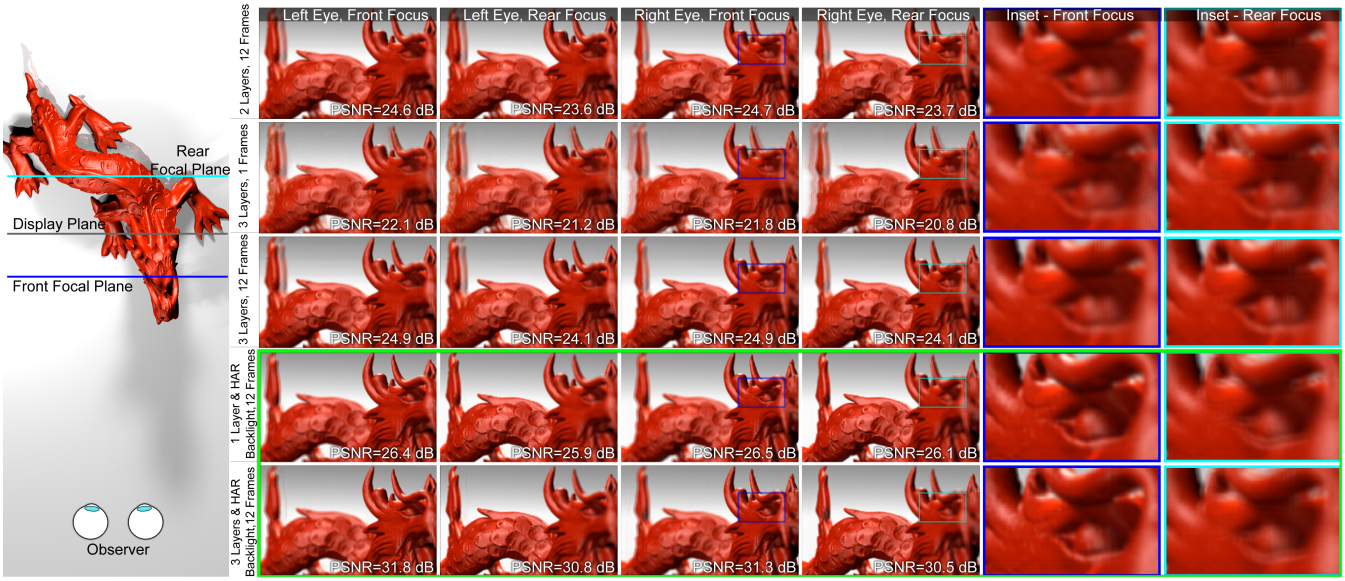

Fig. 5. Performance of computational displays vs. display complexity. We simulate the ability to refocus the light field emitted from various displays, with and without HAR backlighting, following the design proposed in Section 3.1. Up to three layers were placed in front of a Fresnel lens, with focal length $f = 30$ cm, each separated by 0.5 cm. We decompose the target light field to emit 5×5 views spanning each viewer pupil, separated by a distance $d_e = 100$ cm from the lens. *Left:* The light field corresponding to a dragon model is provided as input to the decomposition algorithm. *Right:* The first four columns show the received images for the left and right eye, when focused in front of and behind the lens. The remaining two columns show inset regions centered on the dragon's eye. Five system architectures are compared from top to bottom, with varying numbers of layers and frames. The first three rows evaluate tensor display designs using a uniform backlight ($b_m(x, v) = 1$). The last two rows illustrate the benefits of HAR backlighting, demonstrating that its inclusion enables clear focus cues; note that the dragon's eye can be brought into sharp focus, in contrast to cases without HAR backlighting. Quantitative assessment of focus is provided by the peak signal-to-noise ratio (PSNR) with reference to the original refocused light field, confirming that increasing layers and frames reduces artifacts. Dragon model courtesy of Stanford Computer Graphics Laboratory.

al. [2012] and supplementary Section B; however, for completeness we briefly summarize the tensor-based decomposition framework below.

Assuming discrete pixels, the decomposition into layer patterns, $\{f_m^{(n)}(\xi_n)\}$, is given by the solution of the following optimization problem:

$$\underset{\{\mathbf{F}^{(n)}\}}{\arg\min} \left\| \beta\mathcal{L} - \mathcal{W} \circledast \tilde{\mathcal{T}} \right\|^2, \text{ for } 0 \leq \mathbf{F}^{(n)} \leq 1, \tag{6}$$

where $\circledast$ is the Hadamard (elementwise) product, $\mathcal{L}$ denotes the target light field, $l(x, v)$, represented as a light field tensor, $\mathcal{W}$ is a binary-valued weight tensor (selecting for each ray that passes through a pupil), and $\mathbf{F}^{(n)}$ is a matrix defining the transparency of each pixel in layer n, with the m^{th} column $\mathbf{f}_m^{(n)}$ denoting the values during frame m. Note that β is a brightness scaling factor, controlling the tradeoff between brightness and image fidelity. In this expression, $\tilde{\mathcal{T}}$ is the CP decomposition [Cichocki et al. 2009] of the emitted light field tensor, such that

$$\tilde{\mathcal{T}} = [\![\mathbf{F}^{(1)}, \ldots, \mathbf{F}^{(N+1)}]\!] \equiv \frac{1}{M} \sum_{m=1}^{M} \mathbf{f}_m^{(1)} \circ \cdots \circ \mathbf{f}_m^{(N+1)}, \tag{7}$$

where $\circ$ is the vector outer product. Equation 6 can be solved by applying weighted nonnegative tensor factorization (NTF), with the following update rule:

$$\mathbf{F}^{(n)} \leftarrow \mathbf{F}^{(n)} \circledast \left(\frac{(\mathbf{W}_{(n)} \circledast (\beta \mathbf{L}_{(n)}))\mathbf{F}_{\odot}^{n}}{(\mathbf{W}_{(n)} \circledast (\mathbf{F}^{(n)}(\mathbf{F}_{\odot}^{n})^{\mathsf{T}}))\mathbf{F}_{\odot}^{n}} \right), \tag{8}$$

where $\mathbf{W}_{(n)}$ and $\mathbf{L}_{(n)}$ correspond to the matricization (unfolding) of the weight and target light field tensors along the n^{th} mode, respectively [Kolda and Bader 2009]. In this expression, we define $\mathbf{F}_{\odot}^{n}$ as follows:

$$\mathbf{F}_{\odot}^{n} \equiv \mathbf{F}^{(N+1)} \odot \cdots \odot \mathbf{F}^{(n+1)} \odot \mathbf{F}^{(n-1)} \odot \cdots \odot \mathbf{F}^{(1)}, \tag{9}$$

where $\odot$ is the Khatri-Rao product.

Fig. 5 evaluates the performance of the weighted NTF decomposition for varying display architectures. From these simulations, we conclude that the addition of high angular resolution (HAR) backlighting to the prior tensor display framework is a viable approach to eliminate accommodation-convergence conflicts using current generation and upcoming display technologies. Assuming the viewer's position is known, such a design has the potential to deliver all five "missing" perceptual depth cues for a single user: binocular disparity, convergence, accommodation, retinal blur, and motion parallax.

3.2 Focus 3D Decompositions

While Fig. 5 confirms that the Focus 3D design can successfully synthesize accommodation cues with a sufficient number of layers and frames in simulation, it does not provide intuition into the decomposed patterns. In this section we briefly examine decomposed layer and backlight illumination patterns to understand the expected benefits of our decomposition algorithm over prior direct time-multiplexed backlight illumination schemes. As shown at the top of Fig. 6, direct time-multiplexing requires a single layer placed

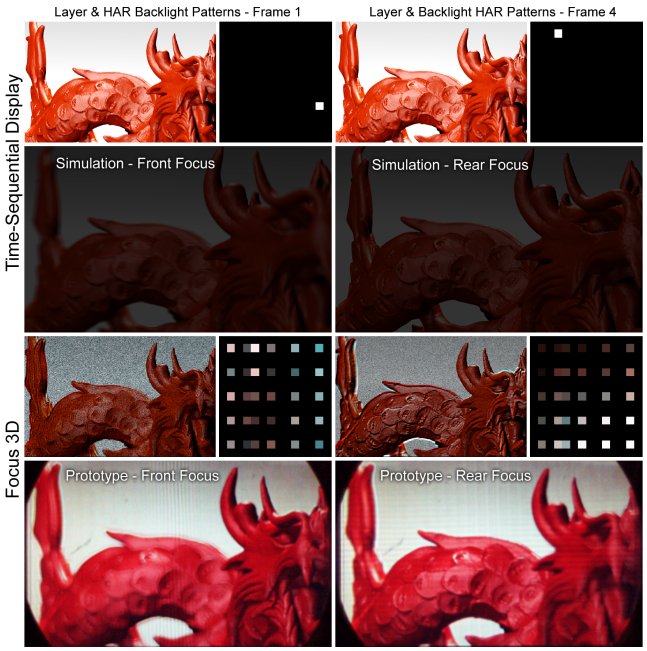

Fig. 6. Comparing direct vs. compressive display modes. A single-layer Focus 3D prototype is considered, comprising one layer in front of a large lens and a backlight conjugate to the viewer's pupil. These examples evaluate a light field with 5×5 views spanning a single pupil located along the optical axis. *Top:* Using direct time-multiplexing, following the approach of Travis [1990], only a single backlight pixel is active in each frame, resulting in a dim image. *Bottom:* Focus 3D exploits correlations between views to illuminate each pupil region for a longer duration, increasing image brightness, as shown in photographs of prototype. Dragon model courtesy of Stanford Computer Graphics Laboratory.

in contact with the lens and a secondary layer placed behind the lens, conjugate to the viewer's pupil. In this mode of operation, each pixel on the backlight that maps to a region of the pupil is sequentially illuminated; simultaneously, the front layer displays the perspective corresponding to a center of projection located in the center of the pupil region. As shown in the refocused images, the depicted light field preserves accommodation cues, but suffers from severe attenuation since each backlight pixel only illuminates the eye for a brief period.

As shown at the bottom of Fig. 6, the decomposition algorithm used with Focus 3D exploits correlations between views to enable each backlight pixel to illuminate the pupil for a longer duration, yielding a brighter image (see supplementary Section A.2 for additional mask decompositions). Yet, similar to prior layered displays described in Section 2, reconstruction artifacts result from the compression process. In summary, Focus 3D opens the door to a new design trade-space between brightness, reconstruction fidelity, and effective frame rate, one that may enable near-term display technologies to resolve the accommodation-convergence conflict.

3.3 Upper Bound on Accommodation Range

In this section we formally assess the benefits of high angular resolution (HAR) backlighting for extending the range over which accommodation cues can be achieved. We adapt the prior frequency-domain analysis of light field displays developed by Zwicker et al. [2006] and Wetzstein et al. [2011; 2012]. While these works derive an upper bound on the depth of field, we perform a similar analysis to reveal an upper bound on the accommodation range for multilayer displays, including those with HAR backlighting.

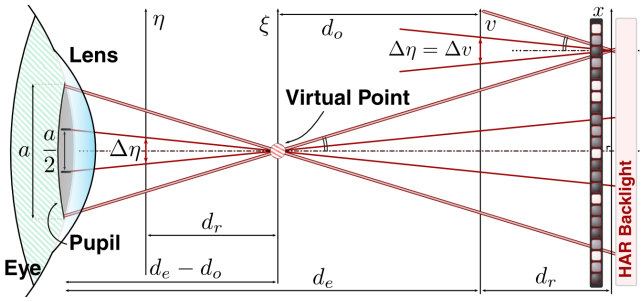

Fig. 7. A virtual point source is created, with accommodation cues. A geometric argument is made for the angular resolution requirements for accommodation in Section 3.3.1. Here, we observe that the angular sampling frequency at the eye, $\Delta\eta$, equals Δv, the angular sampling frequency of the HAR backlight, enforcing a lower bound such that at least two rays enter the pupil of the eye.

3.3.1 *Accommodation Threshold.* Consider the arrangement depicted in Fig. 7, in which a virtual point light source is located a distance $d_e - d_o$ in front of the viewer's pupil, where d_e and d_o are the distance from the eye to the display and from the virtual point to the display, respectively. Following Takaki et al. [2006; 2011], we assume that a minimum of two rays must enter the pupil from this point to support correct accommodation. Let each ray (ξ, η) passing through the virtual point be defined using a two-plane parameterization, where the ξ-axis is coincident with the point and the η-axis is located a distance d_r in front. Under this parameterization, the maximum angular sampling rate $\Delta\eta_{\max}(d_o)$ supporting accommodation is:

$$\frac{\Delta\eta_{\max}(d_o)}{d_r} = \frac{a}{2(d_e - d_o)}. \tag{10}$$

As proven below, the angular sampling rate for a light field display is invariant to the depth of a virtual point. In other words, the maximum angular sampling rate $\Delta v_{\max}(d_o)$ equals $\Delta\eta_{\max}(d_o)$, as defined in the two-plane parameterization of the emitted light field (see Fig. 4). As a result, the angular sampling rate required for accommodation (ω_v) must satisfy the following expression:

$$\omega_v(d_o) \geq \frac{1}{2\Delta v_{\max}(d_o)} = \frac{d_e - d_o}{d_r a} \tag{11}$$

As shown at the bottom of Fig. 8, the supported accommodation range for a given light field display can be estimated by determining the point of intersection of the maximum angular frequency, $\omega_{v\max}$, supported by the display architecture with the *accommodation threshold* given by Equation 11. Points closer to the eye than this point of intersection (i.e., $0 < d_e - d_o \leq d_r a \omega_{v_{\max}}$) will emit a minimum of two rays into the viewer's pupil, whereas points further away will not.

3.3.2 *Maximum Angular Frequency for a Multilayer Display.* To estimate the accommodation range, the maximum angular frequency $\omega_{v\max}$ is required for a given light field display. A direct analysis for conventional architectures, including parallax barriers and integral imaging displays, is possible. Yet, for multilayer displays, it is not clear how to estimate the maximum angular frequency. We propose an upper bound on the maximum angular fre-

quency, based on frequency-domain analyses previously applied to characterize the depth of field of such displays.

Equation 2 can be transformed into the following simplified form:

$$\tilde{l}(x,v) = \frac{1}{M}\sum_{m=1}^{M}\left[\prod_{n=1}^{N+1} f_m^{(n)}(x + (d_n/d_r)v)\right], \quad (12)$$

where $f_m^{(N+1)}(\xi_{N+1})$ now denotes the effective transparency of the virtual layer corresponding to the image of the backlight layer formed by the lens. In this interpretation, the virtual layer is located a distance $d_{N+1} = d_v = (fd_b)/(d_b - f)$ in front of the lens. Taking the two-dimensional Fourier transform of this expression yields an estimate of the emitted light field spectrum in terms of angular frequency ω_v and spatial frequency ω_x for a display with HAR backlighting:

$$\hat{l}(\omega_x,\omega_v) = \frac{1}{M}\sum_{m=1}^{M}\left[\mathop{\ast}_{n=1}^{N+1} \hat{f}_m^{(n)}(\omega_x)\delta(\omega_v - (d_n/d_r)\omega_x)\right], \quad (13)$$

where $*$ denotes convolution and the repeated convolution operator is defined such that

$$\mathop{\ast}_{n=1}^{N+1} \hat{f}_m^{(n)}(\omega_x,\omega_v) \equiv \hat{f}_m^{(1)}(\omega_x,\omega_v) * \cdots * \hat{f}_m^{(N+1)}(\omega_x,\omega_v). \quad (14)$$

Following the procedure outlined by Zwicker et al. [2006] and Wetzstein et al. [2011; 2012], the spatio-angular bandwidth of a multilayer display is determined by the region of non-zero support in the emitted light field spectrum $\hat{l}(\omega_x,\omega_v)$. Intersecting the line $\omega_v = (d_o/d_r)\omega_x$ with the spectral support provides a geometric construction for the upper bound on the depth of field. Fig. 8 compares the upper bound on the depth for field for two competing display architectures: a two-layer display with uniform backlighting and a single-layer display with HAR backlighting. However, this upper bound does not account for limitations of our decomposition algorithm; in practice, the number of ray constraints (i.e. non-zero values of tensor $\mathcal{W}$ in Equation 6) and the compressibility of the input light field determine actual performance. Section 5 provides a performance evaluation in simulation and on a prototype device.

A similar upper bound on the maximum angular frequency $\omega_{v\max}$ can be derived by analyzing the spatio-angular bandwidth of a given multilayer display. Depth-of-field analysis is facilitated by considering the frequency-domain properties of a Lambertian surface located a distance d_o in front of the display. For such a surface, the emitted light field, $\tilde{l}(x,v)$, equals $f(x + (d_o/d_r)v)$, corresponding to the line $\omega_v = (d_o/d_r)\omega_x$ in the frequency domain (see Fig. 8, top and center row). Similarly, a uniform directional area source emits a light field $l(x,v)$ such that

$$\tilde{l}(x,v) = f(v). \quad (15)$$

Taking the two-dimensional Fourier transform of this expression yields an estimate for the corresponding light field spectrum:

$$\hat{l}(\omega_x,\omega_v) = \hat{f}(\omega_v)\,\delta(\omega_x), \quad (16)$$

where $\delta(\xi)$ is the Dirac delta function. Thus, the spectrum of a directional source located *any* distance d_o from a light field display is approximated by a vertical line in the emitted light field spectrum. As a result, the maximum angular frequency $\omega_{v\max}$ supported by any light field display is provided by the intersection of the spatio-angular bandwidth with a vertical line, evaluated along the ω_v-axis.

In summary, we present a new connection linking depth-of-field analysis to bounds on the accommodation range of a light field display. As shown in Fig. 8, the accommodation range is found by intersecting the maximum angular frequency $\omega_{v\max}$ with the accommodation threshold given by Equation 11. In this example, we find that HAR backlighting is necessary to support accommodation within the depth of field centered near the display surface.

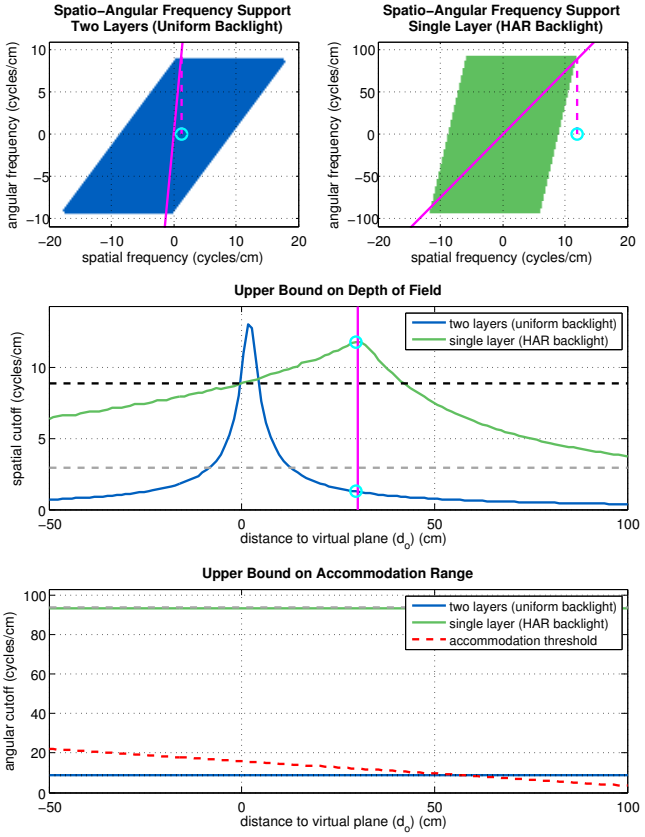

Fig. 8. HAR backlighting is required to support accommodation within the depth of field of a multilayer display. We compare a two layer display with uniform backlighting and a single layer display with HAR backlighting. *Top:* The spatio-angular bandwidths, evaluated following Section 3.3.2. Note that HAR backlighting significantly increases the maximum angular frequency. *Middle:* Upper bounds of depth of field. The dashed black line denotes the maximum spatial frequency corresponding to the physical pixel pitch. The dashed gray line denotes the maximum spatial frequency supported by the virtual panel, given by the magnified image of the backlight layer. The magenta lines illustrate the relationship between the spatio-angular bandwidth and depth of field plots for a reference plane at $d_o = 25$ cm. *Bottom:* Accommodation is supported for virtual plane distances d_o where the display's angular cutoff frequency (blue and green lines) is above the accommodation threshold (Equation 11, red dotted line). Note that without HAR backlighting a two-layer display only supports accommodation when the virtual layer is separated by $d_o \gtrsim 65$ cm from the display (well outside the depth of field). With HAR backlighting, accommodation is predicted throughout the depth of field, as reflected in experiments. The plots reflect our prototype testing configuration: a pupil diameter $a = 2.0$ cm, an eye to display distance $d_e = 127$ cm, and an $f = 31.8$ cm focal length lens. The two layer display used a layer separation of 4.0 cm.

3.4 Diffraction

Light passing through an aperture spreads out angularly (diffracts) to a degree inversely related to the aperture size. For a multiview display, this relationship enforces a limit on the maximum angular resolution that can be achieved for a given spatial resolution; for a given display pixel aperture size, views can be spaced no more closely than the corresponding angular spread of diffraction without overlapping. For multiview displays supporting correct accommodation, diffraction is an important consideration as ultra-high angular resolution is required. In this section, we analyze the relationship between the maximum spatial and angular resolution that can be attained by a diffraction-limited multiview display utilizing spatial light modulators.

In the following analysis we assume that each pixel consists of a round aperture, and light is focused on the plane of the eye. Views are evenly spaced over the observer's pupil, where the outermost views are centered on the pupil edges. Diffraction causes light to spread out to form an Airy disk; as an approximation, we consider the central element of the disk, which is bound by the first minimum at the following angle:

$$\theta_d \approx \arcsin \frac{1.22\lambda}{p}, \tag{17}$$

where λ is the wavelength of the light and p is the diameter of the pixel aperture. Using this model, adjacent views will not overlap due to diffraction at viewing distance d_e if the diameter of the central element of the Airy disk is less than or equal to the view spacing over the pupil, i.e.:

$$2d_e \tan\theta_d \leq \frac{a}{n-1}, \tag{18}$$

where a is the pupil diameter and n is the number of views spaced over the pupil. If the diameter of the central element of the Airy disk exceeds this value, adjacent views will begin to overlap and degrade. By the Rayleigh criterion, two point-light sources are considered "just resolved" when the central element of the Airy disk of one source coincides with the minimum of the other. By this definition, when the diameters of the Airy disk center elements exceed $4d_e \tan\theta_d$, the maximum of the disk corresponding to each view will extend beyond the first minimum of the neighboring views, and adjacent views are no longer resolvable.

Fig. 9 shows the diffraction-limited spatial and angular resolution configuration space for multiview displays that support multiple focal depths. The analysis assumes a human-sized pupil diameter, $a = 5$ mm, and optimal viewing distance of our prototype display, $d_e = 127$ cm. The figure shows that reasonable configurations (spatial resolution of 20-30 cycles/degree, angular resolution of 2-3 views over pupil) are attainable, but lie close to the diffraction limits. Section 3.5 provides simulations to show how diffraction affects the focus quality of a light field.

3.5 Retinal Blur

Along with the ability to focus at different depths about a display, it is also important that the blur of out-of-focus imagery is accurate; retinal blur has been shown to help the human visual system solve the binocular correspondence problem and interpret monocular occlusions [Hoffman and Banks 2010]. From a theoretical standpoint, the quality of retinal blur in our proposed display design is influenced by two primary factors: the light field compression performance of the tensor factorization algorithm and diffraction. (In practice, the blur quality will also be affected by the performance of the optical components).

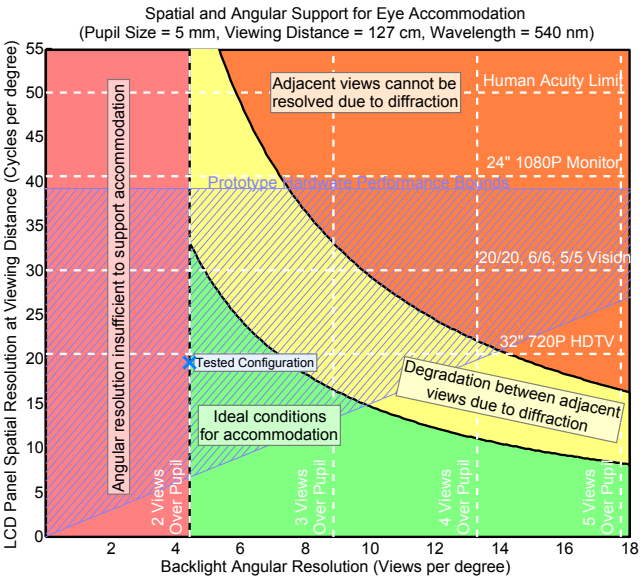

Fig. 9. Diffraction limits on spatial and angular resolution. *Red area:* configurations with insufficient angular resolution to support correct eye accommodation (i.e. fewer than 2 views over the pupil). *Shaded blue area:* theoretical performance bounds of our prototype display (described in Section 4) with regards to the LCD panel resolution and backlight angular resolution. Plot assumes that the backlight LCD panel resolution matches the resolution of the front LCD layers; hence backlight angular resolution increases as LCD panel resolution increases. The tested configuration (see Section 5) is marked with a blue cross. *Green area:* configurations that support correct eye accommodation and have no overlap between views due to diffraction. *Yellow area:* configurations that support correct eye accommodation, but diffraction causes some crosstalk between adjacent views. *Orange area:* configurations that may support eye accommodation, but diffraction is so severe that adjacent views can no longer be resolved. Diffraction is approximated as in Section 3.4 for pupil size $a = 5$ mm, viewing distance $d_e = 127$ cm, and wavelength $\lambda = 540$ nm.

To analyze the effect of diffraction on retinal blur, we simulate the effect of diffraction on a light field. Following the assumptions of Section 3.4, we approximate the diffracted energy distribution using the following Gaussian function [Zhang et al. 2007]:

$$I(v) = e^{\frac{-v^2}{2w^2}}; w = 0.42\lambda\frac{d_e}{p}, \tag{19}$$

where v is the radial distance from the view center on the pupil, λ is the wavelength of the light, d_e is the viewing distance, and p is the diameter of the pixel aperture. A diffracted light field is created according to:

$$\tilde{l}_d(x,v) = I(v) * \tilde{l}(x,v) \tag{20}$$

where $\tilde{l}$ is the emitted light field without diffraction, such that the v axis is coincident with the viewer's pupil (d_r equals d_e in Fig. 4), and $*$ denotes convolution. This process is performed separately on each color channel using the appropriate wavelength, λ, in I.

Fig. 10 provides a comparison of retinal blur between a ground truth light field and light fields compressed through tensor factorization while simulating diffraction as described above. Note that most of the test cases fall above the diffraction limits shown in Fig. 9 in order to provide an estimate of maximum performance in

a diffraction-limited system. We make the following observations from the results:

(1) In the nominal compressed case, the average PSNR is 31 dB for the in-focus images and 37 dB for the out-of-focus images. It is clear that diffraction and compression limit the performance of our approach, but performance on the order of lossy video compression ($\geq$ 30 dB) can still be achieved.

(2) As expected, the in-focus performance decreases as the number of views and time-multiplexed frames are reduced. With too many constraints for the available degrees of freedom (e.g. 5×5 views, 2 frames), focus performance is poor.

(3) High PSNR is not indicative of qualitative blur performance. The most numerically accurate out-of-focus blur occurred in the 2x2 view case, in which the radius appears most accurate. However, the blur accuracy is low as compared to the nominal 5x5 view case – two distinct out-of-focus images can be seen. This issue can be resolved in future work by employing error metrics inspired by the human visual system.

From these observations we conclude that the proposed design theoretically supports focus at multiple depths over a human sized pupil with high quality retinal blur. We also note that the most accurate blur required many views over the pupil, an approach that is only practical with a compressive framework. In Section 5, we describe the actual performance of a prototype display.

4. IMPLEMENTATION

4.1 Hardware

As shown in Fig. 11, our Focus 3D prototype is constructed using off-the-shelf components – three spatial light modulating layers and a large Fresnel lens. The entire optical train is suspended from rails, enabling the placement of the lens and spatial light modulating layers at various distances from the viewer to support the experiments detailed in Section 5. The light modulating layers and backlight consist of modified Viewsonic VX2268wm 120 Hz LCD panels. The diffusing front polarizers were removed from the two front panels and replaced with clear polarizers, enabling image formation through the panels. The lens element is a Fresnel Technologies Inc. #32, 254 mm diameter Fresnel lens with $f = 318$ mm, optimized for conjugates at 424 mm and 1270 mm. We address the impact of the low optical quality of Fresnel lenses in Section 5.

Both simulation and driver software for our prototype run on an Intel Core i7 workstation with 6GB RAM and an external Nvidia QuadroPlex 7000 unit containing two Quadro 7000 GPUs and a G-Sync card. This configuration enables us to drive all three LCDs synchronously at 120 Hz over standard dual-link DVI connections.

4.2 Software

All light fields displayed on the prototype were generated by rendering multiple views of a 3D scene in OpenGL or POV-Ray. A total of 5×5 views were generated at a spatial resolution of 840×525 per eye. For stereoscopic image display, two sets of views were generated at an interocular distance of 64 mm. We note natural light fields can be captured efficiently using compressive techniques [Marwah et al. 2013].

Following Wetzstein et al. [2012], we implement tensor factorization (NTF) on the GPU. This solver implements the multiplicative update rules outlined in Section 3.1 using OpenGL and Cg. These operations are computationally and memory efficient; the full light field matrix or tensor is never stored in memory – only the target views, 32-bit off-screen buffers for the decompositions, and intermediate buffers. The key insight allowing efficient computation is that the mathematically abstract matrix and tensor update rules applied to compressive light field synthesis directly map to hardware-accelerated operations such as perspective rendering and projective texture mapping. Solver runtimes for the above light field resolution are typically a few minutes for 100-200 iterative multiplicative updates. We note that light field rendering and factorization runtimes can be reduced by computing these stages jointly with adaptive sampling [Heide et al. 2013].

5. EXPERIMENTAL ASSESSMENT

We first compare Focus 3D to conventional, time-sequential displays – highlighting the increased display brightness and lower required display framerates. We then evaluate the display system with respect to the supported depth cues, optical design variations, and viewer position. All photographs of the prototype were taken as long exposures on a camera with a 2 cm lens aperture. The minimum aperture size is limited by the focal spot size of the low quality lens used in our prototype – Fig. 10 and 16 provide simulations for human-sized (5 mm) pupils.

5.1 Focus 3D Architecture

Focus 3D fundamentally differs from conventional, time-sequential displays in its compressive approach to light field synthesis. As described in Section 3.2, the computational framework utilized in this paper allows a target light field with an arbitrary number of views to be compressed, in a numerically optimal manner, into the available display refresh rate. This approach enables both practical display architectures and brighter images. This is demonstrated in Fig. 6 – the target light field, containing 25 views over the pupil size of a camera, is compressed into only six frames. An overall brightness gain factor of five was achieved by setting the brightness scaling factor to $\beta = 0.2$ during tensor factorization (see Equation 6).

5.2 Accommodation and Binocular Disparity

Near correct accommodation and binocular disparity are naturally supported by the proposed tensor framework. For this application, two light fields – each with a narrow angular baseline corresponding to one pupil – are rendered and decomposed with the mathematical framework introduced in Section 3. Fig. 12 demonstrates the display prototype supporting both binocular disparity and multiple focal depths. The matryoshka doll images were photographed from two different positions spaced 64 mm apart and were optically focused on three different depths at each viewpoint. Note the focus/defocus effect in the closeups. This scene contains 5×5 viewpoints for each eye – 50 views total – and was successfully decomposed into 12 available frames displayed on a single LCD in front of a Fresnel lens and another LCD at the conjugate distance to the pupil plane behind the lens.

Fig. 13 evaluates the image quality, using peak signal-to-noise ratio (PSNR) as a metric, for a varying number of time-multiplexed frames and light-attenuating layers placed within close proximity in front of the lens. Compared to the previous state of the art multilayer displays [Wetzstein et al. 2012], the proposed work shows improved performance under the same conditions and a greater ability to scale with the number of time-multiplexed frames – even with a single layer. While the PSNR is theoretically improved for multiple stacked layers, designing such systems in practice is challenging due to the necessity of precise layer alignment. Experiments with our prototype show the difficulty of achieving the necessary preci-

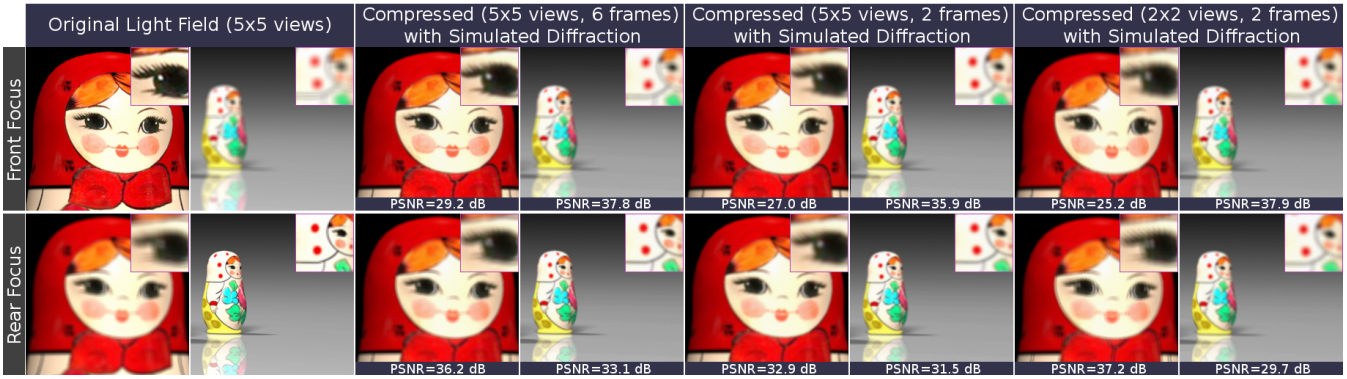

Fig. 10. Simulated retinal blur and diffraction. Images show closeups of close and far matryoshka dolls from the light field shown in Fig. 3, and inset images show further magnification. The larger doll is virtually positioned at 17 cm *in front of* the display and the rear doll at 18 cm *into* the display. The views of the light field are evenly spaced over a pupil of $a = 5$ mm for a single eye, with the outermost views centered at the pupil edges. Compressed images reflect the configuration of our prototype display: 1 LCD layer in front of a HARB, a $f = 31.8$ cm focal length lens, a viewing distance of $d_e = 127$ cm, and native panel resolution of 39.1 cycles/degree at this distance. Diffraction is approximated using the method described in Section 3.5 using the wavelengths $\lambda_{red} = 700$ nm, $\lambda_{green} = 546$ nm and $\lambda_{blue} = 435$ nm. *Rows:* Synthetically refocused images of front doll *(first rows)* and rear doll *(second rows)*. *First column:* Source light field. *Following three columns:* Compressed version of the source light field using the decomposition algorithm described in Section 3.1.2 in the noted configurations. High quality retinal blur can be achieved in the presence of diffraction (*second column*), but quality suffers if compression is too high (*third column*) or angular resolution is too low (*fourth column*). Matryoshka doll model by artist "coboide" of Turbosquid.com.

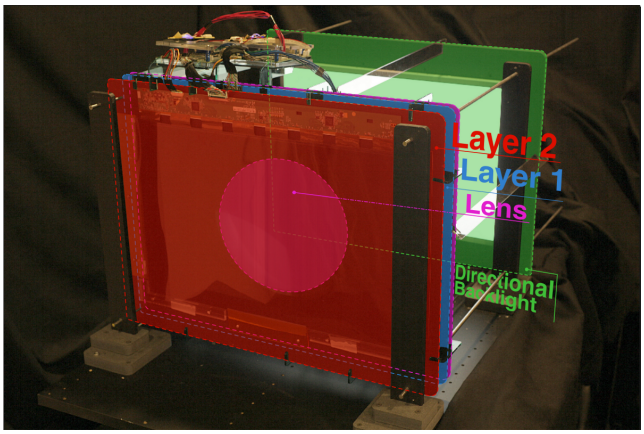

Fig. 11. Focus 3D prototype. A stack of two transparent LCDs is mounted on rails in front of a Fresnel lens with an additional LCD monitor behind the lens. The rear monitor and the lens form a HAR backlight.

sion in practice (see Fig. 14); hence, all photographs of the prototype display utilize only a single LCD and the directional backlight.

5.3 Motion Parallax and View Steering

Motion parallax and view steering are evaluated in Fig. 15. We capture three different viewpoints, centered around the display normal, within a lateral range of 30 cm at a viewing distance of 127 cm. The display optically steers a small light cone into the direction of each view without consideration of any other view. Motion parallax is clearly visible in the three rows of Fig. 15. Additionally, two different focal settings show, for each viewpoint, the front and rear of the shark in focus, respectively. The lateral range of supported viewpoints is practically limited by the quality of the refractive display element – the inexpensive Fresnel lens used in our prototype exhibits significant radial image distortion, coma, and dispersion for off-axis viewpoints at steeper angles. To show the theoretical performance of our system with higher quality optics, we simulate a tracked observer moving around the display at a much wider range in supplementary Section A.1.

5.4 Moving Away from the Conjugate Plane

Moving away from the conjugate plane results in an optical configuration in which the pupil plane does not correspond to the conjugate plane of the backlight. If the observer moves far enough from the display, this optical arrangement practically results in a multilayer display – the backlight is a virtual layer placed at the conjugate plane in front of the physical display enclosure. This approach is similar to that of Gotoda [2011], who noted that placing a lens over an LCD in a multilayer display changes its apparent position. Fig. 16 simulates this case for an observer at a distance of 127 cm, while the conjugate plane of the backlight is located 57 cm in front of the screen. The decompositions use six time-multiplexed frames and the target light field has 5×5 viewpoints over an eye aperture of 5 mm. As shown in the top row, multiple focal depths are still supported. The decompositions (see Fig. 16, bottom row), however, differ from the case where the conjugate plane is in the pupil plane (see Fig. 6) – they show a flipped version of the mask patterns that appear on the virtual layer floating in front of the other layers.

6. CONCLUSION AND FUTURE WORK

Future Work. The incorporation of tracking, a brighter backlight, and a higher quality lens would improve the practicality of our display and allow evaluation on human subjects. Investigation of optical folding techniques, such as Wedge optics [Travis et al. 2013] may also allow the form factor of our current prototype to be significantly reduced – resulting in a thin, light-efficient glasses-free 3D display that overcomes many of the limitations of current-generation devices. Optimization criteria can also be adjusted to compensate for diffraction through the display layers, as well as optical aberrations in the human eye [Huang et al. 2012]. Furthermore, exploring the combination of multiple stacked layers and

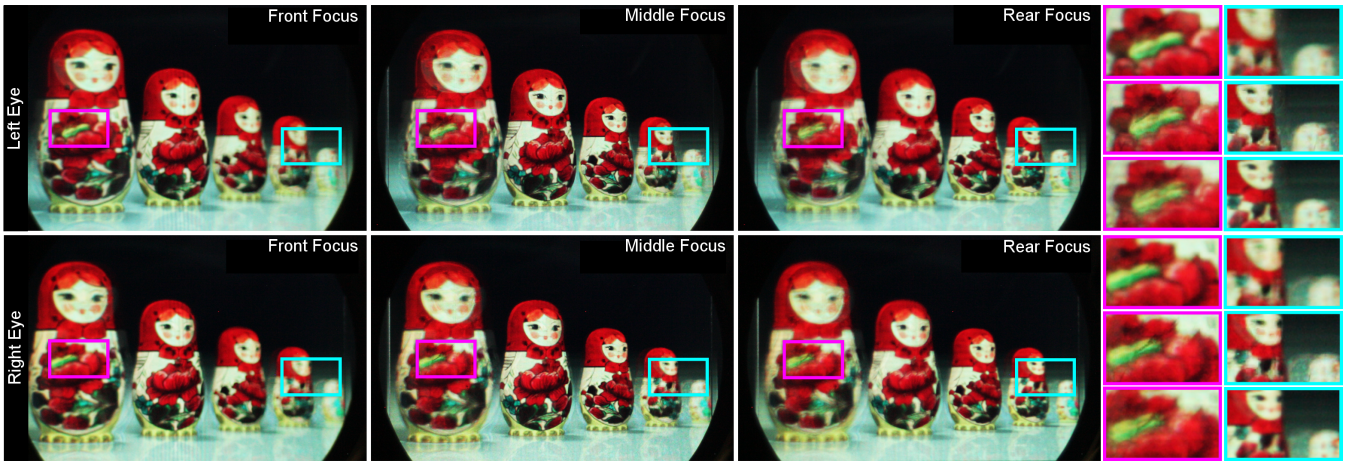

Fig. 12. Photographs of prototype demonstrating binocular disparity (*rows*) and multiple depths of focus (*columns*). Rightmost columns show magnified inset regions. The prototype was configured with a single LCD layer placed directly in front of the lens and was photographed at a viewing distance of 127 cm. Matryoshka doll model by artist "coboide" of Turbosquid.com.

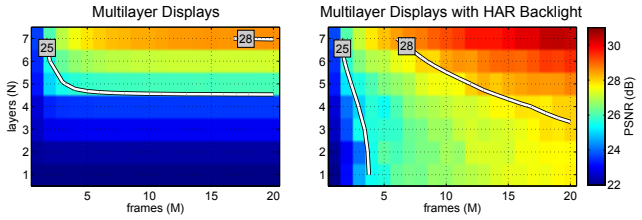

Fig. 13. PSNR scaling with the number of attenuation layers and time-multiplexed frames for the dragon light field shown in Fig. 5. *Left:* Multilayer displays. *Right:* Multilayer Display with HAR Backlight.

Fig. 14. Failure case photographs for dual layer architectures. In practice imprecise alignment of prototype layers creates artifacts. Artifacts are also observed in simulation, as narrow view spacing poses a challenge for large LCD pixel sizes. Matryoshka doll model by artist "coboide" of Turbosquid.com.

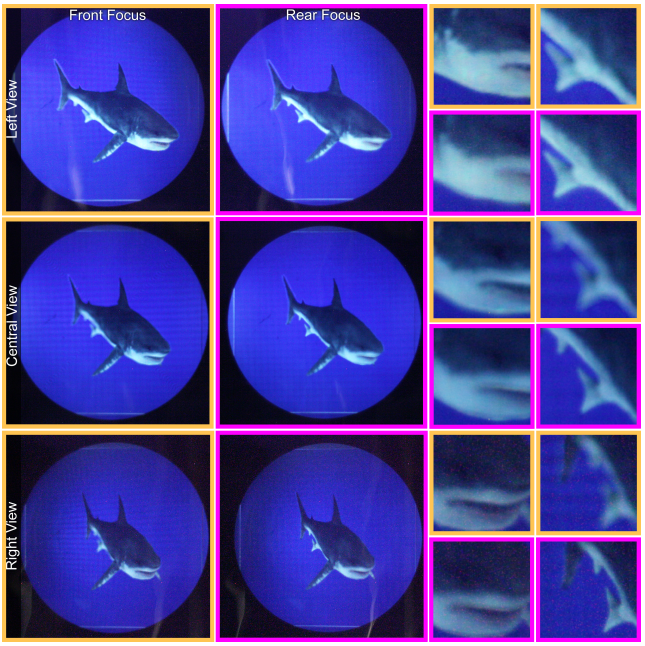

Fig. 15. Photographs of prototype demonstrating motion parallax and multiple depths of focus. The prototype was configured with a single LCD layer placed directly in front of the lens and was photographed at a viewing distance of 127 cm. Three viewpoints, laterally shifted parallel to the display, are shown in the rows while the left and center columns show the front and rear of the shark in focus, respectively. Shark model by artist "wibarra88" of Turbosquid.com.

arbitrary refractive optical elements holds much promise. A single, planar backlight, for instance, will be reimaged onto the focal surface of an arbitrary refractive element; more sophisticated elements, such as parabolic lenses, or catadioptric systems will be fruitful to explore.

Summary. Within the past few years, stereoscopic 3D displays have become commonplace in the home and in theaters – yet many users lament the visual discomfort induced by the accommodation-vergence conflict. By harnessing the emerging field of computational displays – joint designs of display optics and compressive light field synthesis – we have developed a display design with the potential to support the widely sought accommodative depth cue while avoiding the impractical resolution and bandwidth requirements of existing designs. We have demonstrated that computational displays' flexible architectures can further augment the visual experience by supporting nearly correct accommodation cues over a wide field of view and without the need for glasses. We are inspired by the promise of future generations of displays that ap-

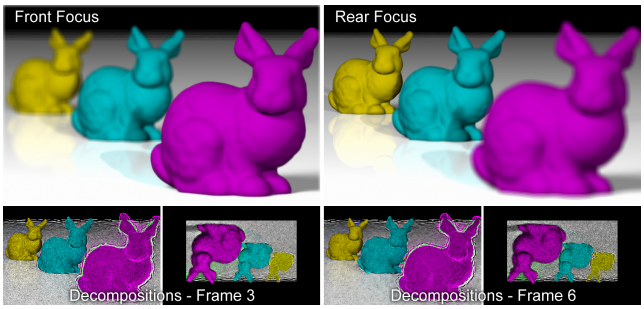

Fig. 16. Multiple focal depths are also supported when the observer moves away from the conjugate plane of the backlight. *Top row:* Simulation shows two differently focused views. *Bottom row:* Two frames of the decomposed patterns for front layer and backlight. Bunny model courtesy of Stanford Computer Graphics Laboratory.

proach the realism of the physical world, and offer comfortable, natural viewing for a wider audience.

ACKNOWLEDGMENTS

We thank the reviewers for their valuable feedback and recognize the support of the Camera Culture group and the MIT Media Lab sponsors. Andrew Maimone was supported by the BeingThere Centre, a collaboration of UNC Chapel Hill, ETH Zurich, NTU Singapore, and the Media Development Authority of Singapore. Gordon Wetzstein was supported by an NSERC PDF, the DARPA SCENICC program, and the NSF through grants 1116452 and 1218411. Douglas Lanman was supported by NSF Grant IIS-1116452 and by the DARPA MOSAIC program. Ramesh Raskar was supported by an Alfred P. Sloan Research Fellowship and a DARPA Young Faculty Award.

REFERENCES

AKELEY, K., WATT, S. J., GIRSHICK, A. R., AND BANKS, M. S. 2004. A stereo display prototype with multiple focal distances. *ACM Trans. Graph. (SIGGRAPH) 23*, 804–813.

BROTT, R. AND SCHULTZ, J. 2010. Directional backlight lightguide considerations for full resolution autostereoscopic 3D displays. *SID Digest*, 218–221.

CHIEN, K.-W. AND SHIEH, H.-P. D. 2006. Time-multiplexed three-dimensional displays based on directional backlights with fast-switching liquid-crystal displays. *Applied Optics 45,* 13, 3106–3110.

CHU, Y. M., CHIEN, K. W., SHIEH, H. P. D., CHANG, J. M., A. HU, Y. C. S., AND YANG, V. 2005. 3D mobile display based on dual-directional light guides with a fast-switching liquid-crystal panel. *J. Soc. Inf. Display 13,* 10, 875–879.

CICHOCKI, A., ZDUNEK, R., PHAN, A. H., AND ICHI AMARI, S. 2009. *Nonnegative Matrix and Tensor Factorizations*. Wiley.

COSSAIRT, O. S., NAPOLI, J., HILL, S. L., DORVAL, R. K., AND FAVALORA, G. E. 2007. Occlusion-capable multiview volumetric three-dimensional display. *Applied Optics 46,* 8, 1244–1250.

FAVALORA, G. E. 2005. Volumetric 3D displays and application infrastructure. *IEEE Computer 38*, 37–44.

GOTODA, H. 2010. A multilayer liquid crystal display for autostereoscopic 3D viewing. In *SPIE Stereoscopic Displays and Applications XXI*. Vol. 7524. 1–8.

GOTODA, H. 2011. Reduction of image blurring in an autostereoscopic multilayer liquid crystal display. In *SPIE Stereoscopic Displays and Applications XXII*. Vol. 7863. 1–7.

HAGOOD, N., BARTON, R., BROSNIHAN, T., FIJOL, J., GANDHI, J., HALFMAN, M., PAYNE, R., AND STEYN, J. L. 2007. 35.5l: Late-news paper: A direct-view mems display for mobile applications. *SID Symposium Digest of Technical Papers 38,* 1, 1278–1281.

HECHT, E. 2001. *Optics (Fourth Edition)*. Addison-Wesley.

HEIDE, F., WETZSTEIN, G., RASKAR, R., AND HEIDRICH, W. 2013. Adaptive Image Synthesis for Compressive Displays. *ACM Trans. Graph. (Proc. SIGGRAPH)*.

HOFFMAN, D., GIRSHICK, A., AKELEY, K., AND BANKS, M. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8,* 3 (March), 33.

HOFFMAN, D. M. AND BANKS, M. S. 2009. Stereo display with time-multiplexed focal adjustment. In *SPIE Stereoscopic Displays and Applications XX*. Vol. 7237. 1–8.

HOFFMAN, D. M. AND BANKS, M. S. 2010. Focus information is used to interpret binocular images. *Journal of Vision 10,* 5, 13.

HUANG, F.-C., LANMAN, D., BARSKY, B. A., AND RASKAR, R. 2012. Correcting for optical aberrations using multilayer displays. *ACM Transaction on Graphics 31*.

IVES, F. E. 1903. Parallax stereogram and process of making same. U.S. Patent 725,567.

JONES, A., MCDOWALL, I., YAMADA, H., BOLAS, M., AND DEBEVEC, P. 2007. Rendering for an interactive 360° light field display. *ACM Trans. Graph. (SIGGRAPH) 26*, 40:1–40:10.

KOLDA, T. G. AND BADER, B. W. 2009. Tensor decompositions and applications. *SIAM Review 51,* 3, 455–500.

KWON, H. AND CHOI, H.-J. 2012. A time-sequential multiview autostereoscopic display without resolution loss using a multi-directional backlight unit and an LCD panel. In *SPIE Stereoscopic Displays and Applications XXIII*. Vol. 8288. 1–6.

LANMAN, D., HIRSCH, M., KIM, Y., AND RASKAR, R. 2010. Content-adaptive parallax barriers: Optimizing dual-layer 3D displays using low-rank light field factorization. *ACM Trans. Graph. (SIGGRAPH Asia) 29*, 163:1–163:10.

LANMAN, D., WETZSTEIN, G., HIRSCH, M., HEIDRICH, W., AND RASKAR, R. 2011. Polarization fields: Dynamic light field display using multi-layer LCDs. *ACM Trans. Graph. (SIGGRAPH Asia) 30*, 1–9.

LIPPMANN, G. 1908. Épreuves réversibles donnant la sensation du relief. *Journal of Physics 7,* 4, 821–825.

MARWAH, K., WETZSTEIN, G., BANDO, Y., AND RASKAR, R. 2013. Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections. *ACM Trans. Graph. (Proc. SIGGRAPH)*.

MATHER, J., BARRATT, N., KEAN, D. U., WALTON, E. J., AND BOURHILL, G. 2009. Directional backlight, a multiple view display and a multi-direction display. U.S. Patent Application 11/814,383.

PAMPLONA, V., OLIVEIRA, M., ALIAGA, D., AND RASKAR, R. 2012. Tailored displays to compensate for visual aberrations. *ACM Trans. Graph. (SIGGRAPH) 31*, 1–11.

PERLIN, K., PAXIA, S., AND KOLLIN, J. S. 2000. An autostereoscopic display. In *ACM SIGGRAPH*. 319–326.

PETERKA, T., KOOIMA, R. L., SANDIN, D. J., JOHNSON, A., LEIGH, J., AND DEFANTI, T. A. 2008. Advances in the Dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE TVCG 14,* 3, 487–499.

PUTILIN, A. N., LUKIANITSA, A. A., AND KANASHIN, K. 2001. Stereodisplay with neural network image processing. In *SPIE Advanced Display Technologies*. Vol. 4511. 245–250.

SHIBATA, T., KAWAI, T., OHTA, K., OTSUKI, M., MIYAKE, N., YOSHIHARA, Y., AND IWASAKI, T. 2005. Stereoscopic 3-D display with optical correction for the reduction of the discrepancy between accommodation and convergence. *SID 13,* 8, 665–671.

STOLLE, H., OLAYA, J.-C., BUSCHBECK, S., SAHM, H., AND SCHWERDTNER, A. 2008. Technical solutions for a full-resolution autostereoscopic 2D/3D display technology. In *Proc. SPIE*. 1–12.

SULLIVAN, A. 2003. A solid-state multi-planar volumetric display. In *SID Digest*. Vol. 32. 207–211.

TAKAKI, Y. 2006. High-Density Directional Display for Generating Natural Three-Dimensional Images. *Proc. IEEE 94,* 3.

TAKAKI, Y., TANAKA, K., AND NAKAMURA, J. 2011. Super multi-view display with a lower resolution flat-panel display. *Opt. Express 19,* 5, 4129–4139.

TOYOOKA, K., MIYASHITA, T., AND UCHIDA, T. 2001. The 3D display using field-sequential LCD with light direction controlling backlight. *SID Digest*, 177–180.

TRAVIS, A., LARGE, T., EMERTON, N., AND BATHICHE, S. 2013. Wedge optics in flat panel displays. *Proceedings of the IEEE 101,* 1, 45–60.

TRAVIS, A. R. L. 1990. Autostereoscopic 3-D display. *Applied Optics 29,* 29, 4341–4342.

WETZSTEIN, G., LANMAN, D., HEIDRICH, W., AND RASKAR, R. 2011. Layered 3D: Tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph. (SIGGRAPH) 30*, 1–11.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH) 31*, 1–11.

ZHANG, W., YE, Z., ZHAO, T., CHEN, Y., AND YU, F. 2007. Point spread function characteristics analysis of the wavefront coding system. *Opt. Express 15,* 4 (Feb), 1543–1552.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3D displays. In *EGSR*.