

# The Light Field Stereoscope

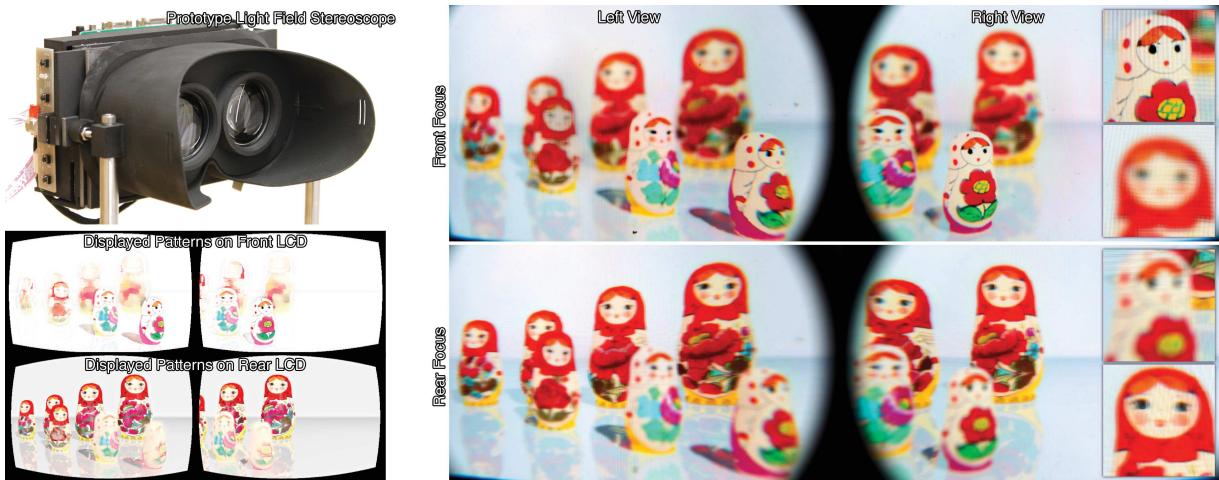
## Immersive Computer Graphics via Factored Near-Eye Light Field Displays with Focus Cues

Fu-Chung Huang

Kevin Chen

Gordon Wetzstein

Stanford University



**Figure 1:** The light field stereoscope is a near-eye display (top left) that facilitates immersive computer graphics via stereoscopic image synthesis with correct or nearly correct focus cues. As opposed to presenting conventional 2D images, the display shows a 4D light field to each eye, allowing the observer to focus within the scene (center and right). The display comprises two stacked liquid crystal displays (LCDs) driven by nonnegative light field factorization. We implement these factorizations in real-time on the GPU; resulting patterns for front and rear LCDs, including the views for both eyes and inverse lens distortion, are shown (bottom left).

## Abstract

Over the last few years, virtual reality (VR) has re-emerged as a technology that is now feasible at low cost via inexpensive cell-phone components. In particular, advances of high-resolution micro displays, low-latency orientation trackers, and modern GPUs facilitate immersive experiences at low cost. One of the remaining challenges to further improve visual comfort in VR experiences is the vergence-accommodation conflict inherent to all stereoscopic displays. Accurate reproduction of all depth cues is crucial for visual comfort. By combining well-known stereoscopic display principles with emerging factored light field technology, we present the first wearable VR display supporting high image resolution as well as focus cues. A light field is presented to each eye, which provides more natural viewing experiences than conventional near-eye displays. Since the eye box is just slightly larger than the pupil size, rank-1 light field factorizations are sufficient to produce correct or nearly-correct focus cues; *no time-multiplexed image display or gaze tracking is required*. We analyze lens distortions in 4D light field space and correct them using the afforded high-dimensional image formation. We also demonstrate significant improvements in resolution and retinal blur quality over related near-eye displays. Finally, we analyze diffraction limits of these types of displays.

**CR Categories:** I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism —Virtual Reality

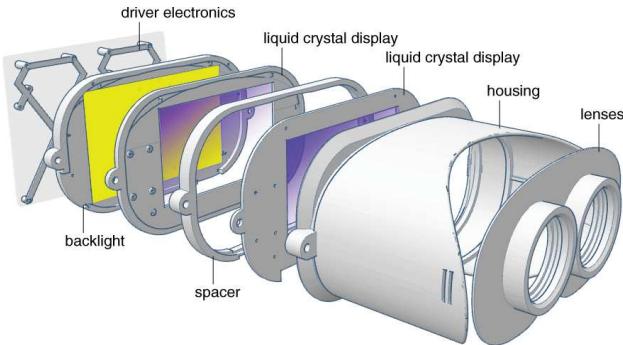
**Keywords:** computational displays, focus cues, light fields

## 1 Introduction

Virtual reality has gained significant traction in the last few years. Although most emerging consumer products are being advertised for gaming and entertainment applications, near-eye display technology provides benefits for society at large by providing a next-generation platform for education, collaborative work, teleconferencing, scientific visualization, remote-controlled vehicles, training and simulation, basic vision research, phobia treatment, and surgical training (e.g., [Hale and Stanney 2014]). For example, immersive VR has been demonstrated to be effective at treating post-traumatic stress disorder [Rothbaum et al. 2001] and it is an integral component of modern, minimally invasive surgery systems, such as the da Vinci surgical system<sup>1</sup>.

To realize these applications and make VR practical for everyday and long-term use, it is crucial to create visually comfortable experiences. Current-generation VR displays support many depth cues of human vision: motion parallax, binocular disparity, binocular occlusions, and vergence. However, focus cues are usually not supported by stereoscopic displays, including head mounted displays (HMDs). Focus cues and vergence are artificially decoupled, forcing observers to maintain a fixed focal distance (on the display screen or its virtual image) while varying the vergence angle of their eyes. The resulting vergence-accommodation conflict

<sup>1</sup>[www.intuitivesurgical.com/](http://www.intuitivesurgical.com/)



**Figure 2:** Schematic of factored near eye light field display. Two stacked, transparent liquid crystal displays modulate the uniform backlight in a multiplicative fashion. When observed through a pair of lenses, the display provides focus cues in addition to binocular disparity afforded by conventional VR displays.

may lead to visual discomfort and fatigue, eyestrain, diplopia, vision, headaches, nausea, compromised image quality, and it may even lead to pathologies in the developing visual system of children [Rushton and Riddell 1999]. Recent studies in the vision literature show that focus cues play a critical role in depth perception (e.g. [Marshall et al. 1996]). In particular, correct or nearly correct focus cues significantly improve stereoscopic correspondence matching [Hoffman and Banks 2010], 3D shape perception becomes more veridical [Watt et al. 2005], and people can discriminate different depths better [Held et al. 2012]. Vergence and accommodation cues are neurally coupled in the human brain; it seems intuitive that displays supporting all depth cues improve visual comfort and performance in long-term experiences.

Our goal is to provide a practical, inexpensive display technology that supports focus cues in a wearable form factor. Inspired by recent advances in computational multilayer displays (e.g. [Seetzen et al. 2004]) and compressive light field displays [Lanman et al. 2010; Wetzstein et al. 2011; Wetzstein et al. 2012] in particular, we design and implement a near-eye stereoscopic light field display, dubbed *The Light Field Stereoscope*. The Light Field Stereoscope combines design aspects of conventional stereoscopic displays (e.g., [Wheatstone 1838]) with modern approaches to factored light field synthesis. Instead of a magnified 2D image, The Light Field Stereoscope presents a different 4D light field to each eye of the viewer. Observing these light fields with a finite pupil size will result in correct or nearly correct retinal blur and the ability to freely focus the eyes within the scene.

One of the most important insights of this work is that benefits and challenges of factored near-eye light field displays are very different from other types of light field displays. Loss in brightness by stacking LCDs is not as critical in immersive environments, because the human visual system quickly adapts to that intensity level. Increased computation is manageable, because VR displays are often driven by personal computers rather than wearable electronics. Finally, the eye box size of the target light field for a single eye is significantly smaller than the field of view required for television-type displays; this allows light field stereoscopes to provide high image quality without time multiplexing (rank-1 light field factorization) using only two stacked LCD panels.

In particular, we make the following contributions:

- We introduce The Light Field Stereoscope as a near-eye display technology with support for all binocular and monocular depth cues, including accommodation and retinal blur.

	multi-focal plane displays	microlens-based near-eye displays	factored near-eye light field displays
resolution	high	low	high*
hardware complexity	high	low	low-medium
form factor	large	small	small-medium
brightness	high	high	medium
computational cost	low	medium	medium
accommodation range	high	low	high
retinal blur quality	medium	low	high

**Figure 3:** Overview of closely-related, focus-supporting display technologies. The light field stereoscope is the first device offering high image resolution and focus cues in a small form factor. The asterisk indicates diffraction limits.

- We design and implement prototype display systems from off-the-shelf consumer electronics and drive them in real-time using efficient GPU implementations of light field factorization.
- We analyze fundamental limits of the proposed technology and show that light field stereoscopes achieve higher resolutions and better focus cues than existing near-eye displays.
- We demonstrate that stereo light fields captured with a light field camera (i.e. Lytro Illum) can be used to generate live action content for these types of displays.

## 2 Related Work

**Head-mounted Displays** Since the first electronic head mounted displays were demonstrated in the 1960s [Sutherland 1968], a lot of research and commercial development has focused on making this technology practical. State-of-the-art reviews of head-mounted displays were recently published by Cakmakci and Rolland [2006] and also by Kress and Starner [2013]. The advent of developer kits such as the Oculus Rift<sup>2</sup> and Google Glass<sup>3</sup> promises consumer HMDs to be widely available in the near future. However, there is much potential to improve the visual comfort of these devices; mitigating the vergence-accommodation conflict inherent to most HMDs is one opportunity to do so.

**Displays supporting Focus Cues** A number of display technologies support focus cues, but none has been demonstrated to be practical for wearable displays. Holography [Benton and Bove 2006], for example, has the promise to synthesize physically-accurate wavefronts that support all depth cues of human vision. Unfortunately, the extreme requirements on feature size, algorithmic computation, and also on the coherence of the light source make dynamic holography an ill-suited technology for HMDs in the near future. Volumetric displays overcome many of these limitations using mechanically spinning [Favalora 2005; Jones et al. 2007] or electronically switchable diffusers [Sullivan 2003] to create an emissive display volume within which a user can accommodate. However, the form factor of these devices is usually prohibitive for wearable applications.

Closely related to volumetric displays are multi-focal-plane displays. Beam splitters [Akeley et al. 2004], switchable lenses [Liu et al. 2008; Love et al. 2009], or vibrating membranes [Schowengerdt and Seibel 2006] can be used to approximate a continuous volume with a few additively superimposed planes. In near-eye display applications, a multi-plane volume would be synthesized for each eye separately, providing correct stereoscopic cues and an approximation of the monocular focus cues. For most of the visible depth range, our ability to discriminate different depths (“axial resolution”) using only monocular cues is relatively

<sup>2</sup><http://oculusvr.com>

<sup>3</sup><https://www.google.com/glass>

low compared to our visual acuity (“lateral resolution”). Hence, multi-focal-plane displays with only a few planes seem to make tradeoffs that are well aligned with the limitations of human vision. MacKenzie et al. [2010] reported about five focal planes spaced at one diopter to be sufficient for supporting accommodation within a range of 25 cm to optical infinity. However, even such a moderate number of depth planes requires the employed displays to provide refresh rates of at least 300Hz such that each depth plane is updated with 60 Hz (approx. the critical flicker fusion of human vision). Unfortunately, only specialized display technologies provide refresh rates in the hundreds of Hz, and none are currently used in wearable displays. Time-multiplexed image presentation may also lead to perceived flicker, which is usually undesirable.

We advocate for near-eye light field displays, implemented by stacked liquid crystal displays. The main difference between multi-focal-plane displays and multi-layer LCD displays is that the image formation is *additive* in the former technology and *multiplicative* in the latter. We demonstrate that this subtle difference allows for correct or nearly-correct focus cues to be supported over larger depth ranges or it alternatively reduces the number of required display planes. Since the amount of parallax observed over the eye box is relatively small, we demonstrate that no time-multiplexing is required for the proposed technology, reducing perceived flicker, making mechanically-moving optical elements in the system obsolete, and significantly relaxing required display refresh rates. Note that our approach to light field display accounts for the entire eye box at once and, therefore, does not require eye tracking. Existing multi-focal-plane displays usually restrict the pupil to a fixed position.

**Light Field Displays supporting Focus Cues** Recently, light field displays have been shown to allow for very small form factors of near-eye displays [Lanman and Luebke 2013], even in see-through display modes [Hua and Javidi 2014]. By placing a microlens array on a small screen close to the eye, near-eye light field displays allow for images to be synthesized that appear to be floating outside the physical device enclosure, such that the observer can accommodate within a limited range. A similar idea has recently also been exploited to correct visual aberrations of observers, including defocus and higher-order aberrations, in the display [Huang et al. 2014]. The microlenses used in most previous work [Lanman and Luebke 2013; Hua and Javidi 2014], however, impose a direct tradeoff between achieved spatial resolution and the supported depth range. A similar tradeoff exists in near-eye pinlight displays [Maimone et al. 2014]. Super-multi-view displays [Takaki et al. 2011] have the ability to overcome this tradeoff but require multiple tiled screens, which is impractical for most near-eye displays.

Our work builds on recent advances in factored multi-layer light field displays [Lanman et al. 2010; Wetstein et al. 2011; Wetstein et al. 2012; Maimone et al. 2013], but we are the first to explore this approach for near-eye VR displays. Although Maimone et al. [2013] demonstrate a television-type light field display supporting stereoscopic disparity, motion parallax, and focus cues via head tracking, time-multiplexed display, and a special directional backlight, their system is not suitable for wearable applications. The main differentiating characteristic between the above-listed technologies and near-eye light field displays is that the latter usually *only* have to support focus cues but *not any of the other depth cues*, because separate microdisplays are used for each eye. This simplifies the display design dramatically. The work closest to ours is the multi-layer see-through AR display proposed by Maimone and Fuchs [2013]. Unfortunately, reduced peak brightness and strong diffraction blur imposed by see-through LCDs on physical objects at larger distances make these types of displays impractical for (outdoor) AR applications today. As a hybrid between light field displays and multi-focal-plane displays for VR applications, our configuration achieves image resolutions that far exceed those of microlens-based near-eye VR displays with display refresh rates easily provided by commercially-available screens. Reduced peak brightness is not as critical in immersive VR applications and our system does not rely on eye tracking. Overall, the benefits of factored light field display seem ideal for immersive VR applications.

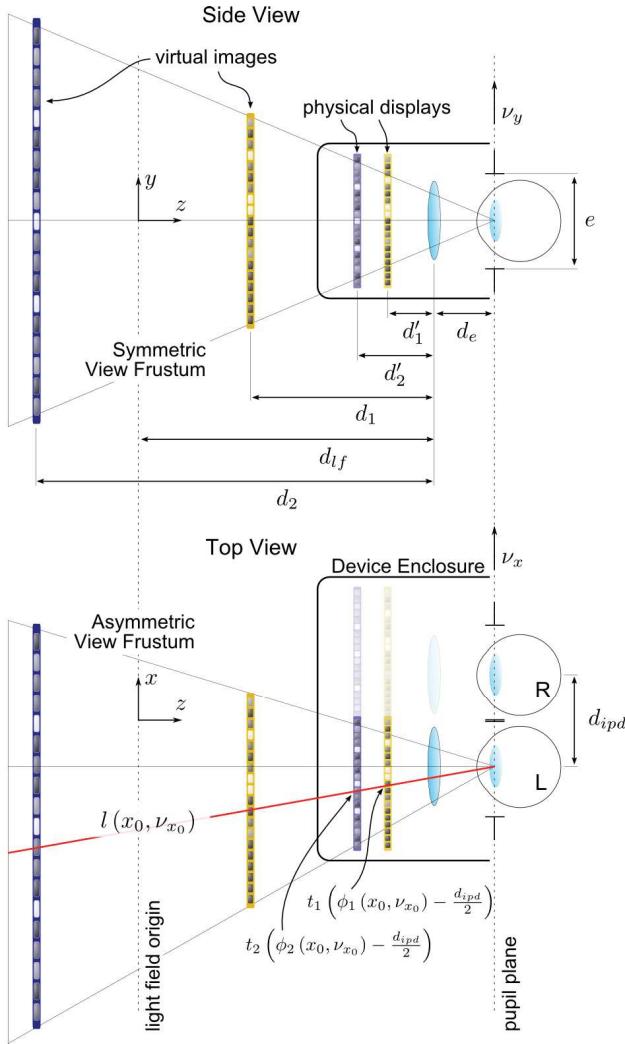
cal for (outdoor) AR applications today. As a hybrid between light field displays and multi-focal-plane displays for VR applications, our configuration achieves image resolutions that far exceed those of microlens-based near-eye VR displays with display refresh rates easily provided by commercially-available screens. Reduced peak brightness is not as critical in immersive VR applications and our system does not rely on eye tracking. Overall, the benefits of factored light field display seem ideal for immersive VR applications.

### 3 Requirements on Focus-Supporting HMDs

We briefly outline the target specifications for immersive near eye displays, as reported in the literature. An overview of some of these parameters can be found in Kress and Starner [2013]. For example, they list the required eye box size, i.e. the area within which the pupil moves (see Fig. 4), as  $10 \times 10$  mm for consumer applications. The instantaneous pupil diameter is, on average, about 3-4 mm, but displaying a light field for the entire eye box accounts for different pupil positions within and also for a range of pupil diameters. The visual field (field of view) of the human visual system (HVS) spans approx.  $190^\circ$  horizontally, of which only about  $120^\circ$  are used for binocular vision, and  $135^\circ$  vertically [Ruch and Fulton 1960]. “Normal” or corrected lateral resolution of human vision is usually specified as 20/20, which corresponds to an angular resolution of 1 minute of arc. The temporal resolution of the HVS is often cited as the critical flicker fusion threshold, which can be approximated as 60 Hz but depends on the adaptation luminance and spatial frequency [Kelly 1979]. Using this as a guideline, the latency requirement on a head mounted display is at least 17 ms without taking head motion into consideration. Egocentric motions, however, place a significantly higher demand on display refresh rates and latency in practice, requiring ultra-low latency for truly immersive experiences.

Focus cues are not supported by conventional near-eye displays, so rigorous specifications are difficult to find in the literature. For super-multi-view displays, Takaki [2006] lists a minimum of two light field rays (or views) per 2D pupil as the minimum for triggering accommodation. This seems to be a very optimistic bound; more recent studies have shown that about  $3 \times 3$  rays entering the pupil can achieve a plausible approximation of retinal blur [Huang et al. 2014]. Microlens-based near-eye light field displays [Lanman and Luebke 2013] build on similar ideas as super-multi-view displays, but the achieved accommodation range is directly traded for spatial resolution [Zwicker et al. 2006]. Therefore, either high resolution can be achieved or a wide focus range, but never both simultaneously. Multi-focal-plane displays [Akeley et al. 2004; Love et al. 2009] can be interpreted as “3D light field displays” with sparse additive depth layers, which suggests that the full 4D light field may not be necessary to create *plausible focus cues*. Since *physically accurate focus* cannot be easily created, the question at hand is the following: *If practical displays can only approximate correct focus cues, how accurate do they have to be to create plausible and comfortable viewing conditions?*

For multi focal plane displays, one would expect a focal range of about 25 cm to optical infinity to be supported. Monocular depth discriminability via retinal blur and accommodation is reported [MacKenzie et al. 2010] as  $\pm 0.25 - 0.3$  diopters (1 / distance in meters) and accommodation is driven mostly by medium spatial frequencies [Love et al. 2009] in the range of 4-10 cycles per degree (cpd). These numbers provide a guideline for any emerging display technology. With the prototype developed for this project, we achieve a field of view of  $87^\circ \times 91^\circ$ , an accommodation range from approx. 0.2 m to 1.2 m (covering a large portion of the depth range where accommodative depth cues matter [Cutting and Vishwanath 1995]), and an end-to-end system latency of about 65 ms. Al-



**Figure 4:** Technical illustration of the proposed near-eye light field display system. Two attenuating spatial light modulators (SLMs), such as liquid crystal displays, are mounted inside the physical device (right, black outline). The cascaded SLMs modulate the backlight (not shown) in a multiplicative manner. When observed through a magnifying lens, virtual and magnified images of the SLMs are created – one close to optical infinity and the other close to the observer. The viewing frusta used for rendering computer-generated content are vertically symmetric but horizontally asymmetric, as illustrated in the bottom.

though we demonstrate high-quality focus cues, other system parameters, most notably latency and field of view, need to be further improved to compete with commercial solution. Please see Section 5 for more implementation details.

## 4 Factored Light Field Synthesis

This section briefly reviews the image formation of factored near eye displays employing simple magnifying lenses. We include aberrations of the lenses in the image formation and derive formulations for the inverse problem of light field synthesis using a dual-layer rank-1 light field display. The device and its most important design parameters are illustrated in Figure 4.

### 4.1 Single Magnifier

Consider an object at a distance  $d'$  to a lens with focal length  $f$ . The lens will create a real or a virtual image at a distance  $d$ . The relationship between these distances can be approximated by the Gaussian thin lens formula:

$$\frac{1}{d'} + \frac{1}{d} = \frac{1}{f} \Leftrightarrow d = \frac{1}{\frac{1}{f} - \frac{1}{d'}}. \quad (1)$$

For near-eye displays, we are particularly interested in the case where micro displays are located at distances close to the lens, i.e.  $d' \leq f$ , such that a virtual and magnified image of the display is created on the same side of the lens (see Fig. 4). The magnification is given by

$$M = \frac{f}{f - d'}. \quad (2)$$

For example, the virtual image of a micro display with a physical width  $w'$  will be perceived to have a width of  $w = Mw'$ . The 2D “flatland” light field synthesized by two stacked attenuating micro displays  $t_1$  and  $t_2$  is then

$$\begin{aligned} \tilde{l}(y, \nu_y) &= \prod_{k=1}^2 t_k(\phi_k(y, \nu_y)) \\ &= \prod_{k=1}^2 t_k\left(\frac{1}{M}\left(y + \frac{(\nu_y - y)(d_{lf} - d_k)}{d_{lf} + d_e}\right)\right), \end{aligned} \quad (3)$$

where  $d_e$  is the eye relief,  $d_{lf}$  is the user-defined distance between lens and light field origin, and  $d_k$  is the distance between the virtual image of a micro display and the lens. We employ a two-plane parametrization of the light field [Levoy and Hanrahan 1996], where  $y$  denotes the coordinate on the light field plane and  $\nu_y$  is an absolute coordinate on the pupil plane. The mapping function  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  computes the intersection of a light ray  $(y, \nu_y)$  with a plane at some distance.

The pupil diameter is not restricted in this formulation and the eyeball can freely rotate within a limited range called eye box  $e$ , such that  $\nu_y \in [-e/2, e/2]$ . In practice, the lens diameter and eye relief may restrict the observable part of the emitted light field. Without loss of generality, we neglect this restriction in our notation for simplicity. The above image formation applies when the lens is aligned with the center of the micro display, as illustrated in the side view of the proposed near-eye light field display (Fig. 4, top).

### 4.2 Stereo Magnifier

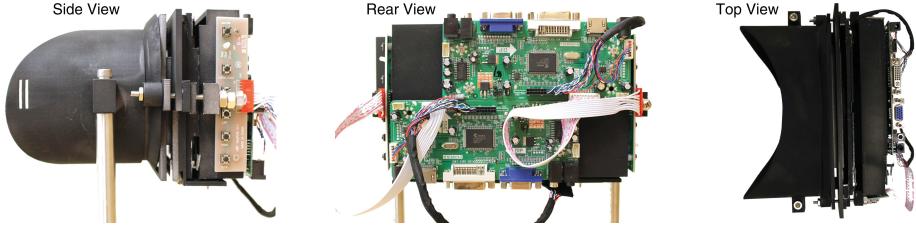
The image formation in a stereo setup is very similar to Equation 3. However, the inter-pupillary distance  $d_{ipd}$  needs to be taken into account, resulting in a slightly different formulation for the light field of the left (L) and right (R) eyes:

$$\tilde{l}_{R/L}(x, \nu_x) = \prod_{k=1}^2 t_k\left(\phi_k(x, \nu_x) \pm \frac{d_{ipd}}{2}\right). \quad (4)$$

Note that the visual field of the eyes is not symmetric in the horizontal dimension. For wide field of view near-eye displays, this creates an asymmetric viewing frustum, as illustrated in Figure 4 (bottom).

### 4.3 Lens Distortions

In practice, a magnifying lens will optically distort the image or light field emitted by any display. Lens distortion is often dominated by a radial term [Brown 2000] which accurately models field



**Figure 5:** Different perspectives of a prototype light field stereoscope.

curvature, but may only approximate other types of aberrations. In commercial VR displays, radial distortions can be partially compensated by applying a lateral pre-distortion to the image coordinates. Common analytical forms of this distortion are usually of the form  $x_d = x(1 + k_1 r^2 + k_2 r^4)$ , where  $r^2 = x^2 + y^2$  and  $k_1, k_2$  are lens-specific distortion coefficients. However, in practice the distortion is not accurately modeled by lateral distortion alone. Field curvature, for example, results in a 3D plane to be re-imaged to a curved surface that also affects the focus cues and perceived image blur. With a light field display, one has the unique opportunity to partially compensate for such high-dimensional distortions, which can be modeled as

$$l_\varphi(x, \nu_x) = l(x + \varphi_x(x, \nu_x), \nu_x + \varphi_\nu(x, \nu_x)). \quad (5)$$

To be consistent with radial distortions used for conventional near-eye displays and also in the computer vision literature (e.g. [Zhang 2000]), we chose to model the distortion functions as a radially-symmetric polynomial  $\varphi(r) = ar^2 + br^4$ , such that the lateral distortions become

$$\varphi_x(x, \nu_x) = \varphi(r)\varphi'(r) \approx x(1 + k_1 r^2 + k_2 r^4), \quad (6)$$

with  $k_1 = 2a^2$  and  $k_2 = 6ab$ . The corresponding angular light field distortions are

$$\varphi_\nu(x, \nu_x) = \tan^{-1}(\varphi'(r)) = \tan^{-1}(2ar + 4br^3). \quad (7)$$

The coefficients  $a$  and  $b$  can be chosen in a wavelength-dependent manner to also account for chromatic lens aberrations. Partial compensation of all of these aberrations with the proposed display is achieved by pre-distorting the target light field with the inverse distortion functions  $\varphi_x^{-1}$  and  $\varphi_\nu^{-1}$  (see Fig. 6) before solving the factorization problem outlined in the following.

#### 4.4 Light Field Factorization

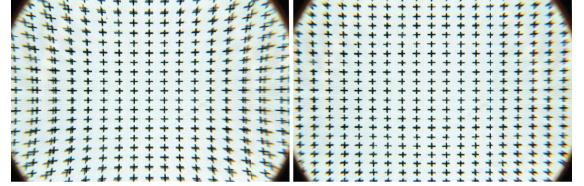
We formulate the full 4D image formation by combining and discretizing Equations 3 and 4 as

$$\tilde{\mathbf{I}} = \Phi_1 \mathbf{t}_1 \circ \Phi_2 \mathbf{t}_2. \quad (8)$$

Here,  $\tilde{\mathbf{I}} \in \mathbb{R}^L$ ,  $\mathbf{t}_{1/2} \in \mathbb{R}^N$  are vectorized forms of the target light field and the display patterns, respectively, and  $\circ$  is the Hadamard or element-wise product. The mapping functions  $\phi_k$  are encoded in the matrices  $\Phi_k \in \mathbb{R}^{L \times N}$ . Each of the matrices is sparse and has a well-defined structure:  $\Phi_k = \left[ (\Phi_k^{(1)})^T (\Phi_k^{(2)})^T \dots (\Phi_k^{(V)})^T \right]^T$ . Discretizing the number of light field positions on the pupil plane to  $V$  locations, each of the submatrices  $(\Phi_k^{(j)})^T$  models how every pixel on the display maps to that particular view of the light field.

An objective function can be formulated as

$$\underset{\{\mathbf{t}_1, \mathbf{t}_2\}}{\text{minimize}} \|\beta \mathbf{I} - (\Phi_1 \mathbf{t}_1) \circ (\Phi_2 \mathbf{t}_2)\|_2^2, \quad \text{s.t. } 0 \leq \mathbf{t}_{1/2} \leq 1 \quad (9)$$



**Figure 6:** Without correction, we observe strong pincushion distortion (left). The distortion also prevents crosses displayed on the two displays to line up properly. With the appropriate inverse 4D distortion, this can be corrected (right).

Model	Geometry #v / #f	Draw mono	Solve mono	Draw stereo	Solve stereo	FPS stereo
lion	5K / 10K	4.5	57.5	9	54.5	15.7
hands	10K / 20K	7.5	55.0	14	49.5	15.7
bench	30K / 50K	15.5	49.5	31	37.5	14.6
armadillo	25K / 50K	15.5	50.5	29.5	41.0	14.2

**Table 1:** Runtime for 5 NMF iterations per frame in ms.

and solved in an iterative fashion using update rules that are similar to those derived in recent work on factored displays [Lanman et al. 2010; Wetzstein et al. 2012; Hirsch et al. 2014; Heide et al. 2014]. The update rules for display 1 are

$$\mathbf{t}_1 \leftarrow \mathbf{t}_1 \circ \frac{\Phi_1^T (\beta \mathbf{I} \circ (\Phi_2 \mathbf{t}_2))}{\Phi_1^T (\tilde{\mathbf{I}} \circ (\Phi_2 \mathbf{t}_2)) + \epsilon} \quad (10)$$

and a similar rule updates the pattern for display 2 but with swapped subscripts. Following a general alternating least-squares (ALS) strategy, each update is carried out in an alternating fashion. We note that for the specific case of rank-1 factorization, as considered in this paper, Equation 10 is numerically equivalent to the rank-1 residue update used in recent work on factored super-resolution displays [Heide et al. 2014]. Real-time implementations on the GPU have been demonstrated [Wetzstein et al. 2012; Heide et al. 2014]. Without considering a time-multiplexed display system, the question at hand is whether or not a rank-1 light field factorization with the proposed displays provides sufficient degrees of freedom for high-quality light field synthesis. We argue that this is, in fact, the case and provide experimental validation in the following.

## 5 Implementation and Assessment

**Hardware** We built several prototypes using different liquid crystal panels, but found the Chimei Innolux N070ICG-LD1 panels to be some of the lowest cost and most widely-available (e.g., on ebay). However, resolution is limited to  $1280 \times 800$  pixels. Magnifying lenses of all prototypes are aspheres available on ebay with



**Figure 7:** We show a variety of computer-generated scenes photographed from one of our prototype light field stereoscopes. All results are generated with the proposed rank-1 light field factorization — factored patterns are shown in the right column. Retinal blur is clearly visible in all presented light fields. Slight moiré effects are barely visible to a human observer, but could be further mitigated using diffusers specifically engineered for this display configuration. The diffraction blur of the front panel limits the resolution observed on rear objects.

a focal length of  $f = 5$  cm and a diameter of 2". The lenses are separated by 6.4 cm. The virtual images of the LCDs are located at 19 cm and 123 cm (5.26 D and 0.81 D) whereas the physical panels are at 3.8 cm and 4.8 cm, respectively, from the lenses. The display housing is modified from Adafruit's 3D printed wearable video goggles<sup>4</sup> and fabricated with a Makerbot Replicator 2. The software runs on a Dell XPS 8700 workstation with an Intel i7 CPU, 8 GB RAM, and an NVIDIA GTX 970 GPU.

<sup>4</sup><https://learn.adafruit.com/3d-printed-wearable-video-goggles/>

**Software** All software is implemented in C++, OpenGL, and CUDA. Rendering is done with OpenGL and the factorization routines outlined in Section 4 are implemented in CUDA. This allows for real-time framerates; for example, we render, factorize, and displays stereo light fields of simple scenes with a resolution of  $640 \times 800$  pixels and  $5 \times 5$  views with about 20 frames per second (fps) on an NVIDIA GTX 970 and with about 35 fps on an NVIDIA Quadro K6000. We list the performance for several complex scenes on the NVIDIA GTX 970 in Table 1. The factorization uses a single

NMF iteration in each frame and its result is used as the initial guess for the next frame. Therefore, static objects are progressively refined over multiple frames whereas quickly moving objects exhibit a slight amount of motion blur (see supplemental video). The motion blur created by the factorization could be reduced by increasing the number of NMF iterations in each frame, but that would decrease framerates. Overall, latency and motion blur are remaining challenges that may counteract benefits provided by the focus cues in some applications. Improving end-to-end latency is a crucial part of future work on factored near-eye light field displays.

Stereo light field rendering time is roughly twice that of the mono light field rendering mode. The solve time for stereo mode is, however, slightly less than that in the mono mode: this is due to the off-axis frustum rendering where the overlapping pixels between panels are fewer. There is a fixed cost of resources-binding which takes 4 ms and 8 ms for mono and stereo modes respectively. The runtime is measured for 5 iterations, and each iteration of solving the panels takes less than 10 ms. Finally, we observe that the solve time is mostly independent on the model’s geometric complexity; in fact, it only depends on the resolution of the panel.

**Calibration** We measure gamma curves of the displays as  $\gamma = 2.4$  and calibrate their blacklevels. The latter is accounted for in the factorization, whereas gamma correction is done on the factorized patterns before they are being displayed. Lateral displacement between panels is mechanically corrected as best as possible. Slight misalignment is further corrected in software. Lens distortion is an issue (see Fig. 6, left) and not only results in distorted images but also in images of the two stacked panels not lining up properly. These distortions are corrected by interactively adjusting the distortion parameters and performing an inverse distortion of the light field as discussed in Section 4.3. The diffraction blur observed on the rear display panel is accounted for and partially corrected in the solver (see [Hirsch et al. 2014] supplemental material for more details). However, image degradation for farther objects is inevitable due to diffraction blur.

**Results** Figure 7 shows a variety of results captured in monoscopic mode. Please find the corresponding stereo results in the supplemental material. We also evaluate result with mid-range focus in Figure 8 and the supplemental video. As discussed above, objects located far away from the front panel are not perfectly sharp due to diffraction blur. In addition, the rank-1 constraint may in some cases limit the image sharpness of 3D objects located in between the virtual images of the physical display panels.

## 6 Analysis and Evaluation

### 6.1 Quality of Retinal Blur

Focus cues comprise two different characteristics: retinal blur and accommodation. Retinal blur triggers accommodative responses of the eye; hence, we evaluate its quality, as produced by a range of different near-eye displays in Figure 9. In particular, we evaluate the microlens-based near eye light field display recently proposed by Lanman and Luebke [2013], additive multi-focal-plane displays (e.g., [Akeley et al. 2004; Schowengerdt and Seibel 2006; Liu et al. 2008; Love et al. 2009]), and the factored, multiplicative multi-layer display proposed in this paper. All of these near-eye displays provide an approximation of the physically realistic retinal blur (Fig. 9, center left column). We simulate the light field in all cases with  $15 \times 15$  views and an eye box size of 8 mm. Although the instantaneous pupil size is significantly smaller than the simulated eye box, the chosen parameters help us visualize the effect for any



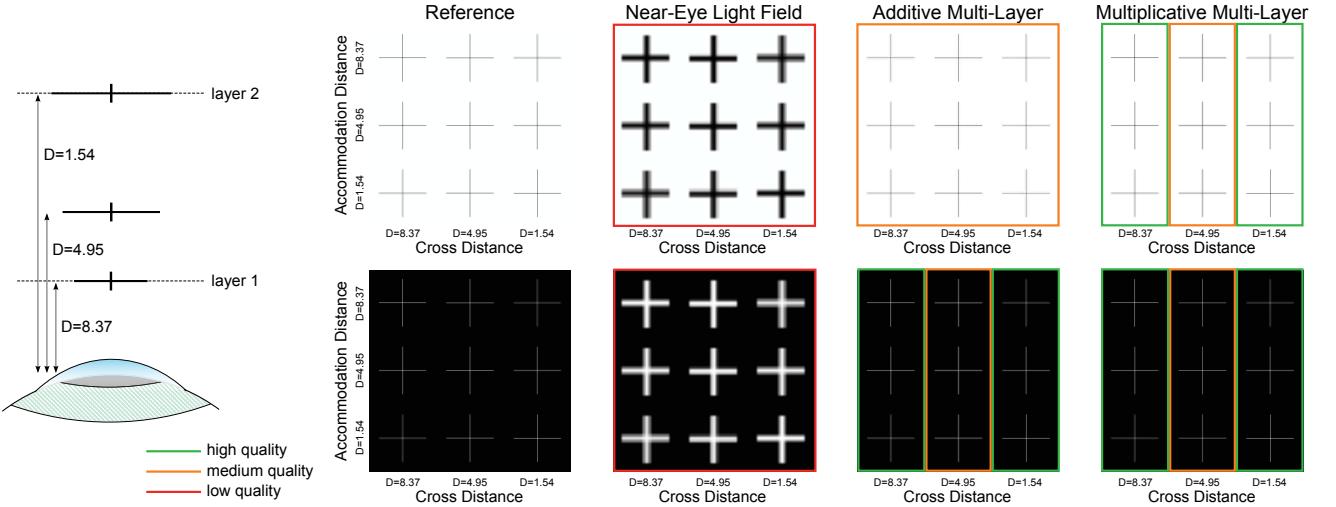
**Figure 8:** Photographs of the prototype when focusing on virtual objects located on the front and rear panels as well as in between. The proposed rank-1 factorization provides high-quality imagery even for virtual objects that are not located on the physical panels.

pupil size in an intuitive manner. All simulated device parameters are those of the prototype (see Sec. 5), except for the layer spacing which is simulated to be that of one of our early prototypes (8.37 D and 1.54 D). The trends observed in this experiment generalize to other hardware configurations, including the particular layer spacing of the physical prototype. Additive layers are optimized with the tomographic solver proposed by Wetzstein et al. [2011], which provides the optimal solution in an  $\ell_2$ -error sense.

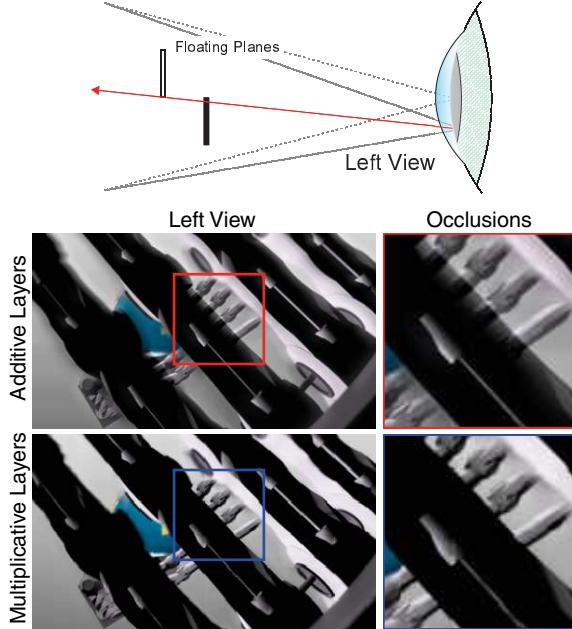
Microlens-based near-eye displays may provide sufficiently high-quality retinal blur, but the inherent loss of spatial resolution seems unacceptable for most applications (Fig. 9, center column). All multi-plane displays provide physically accurate retinal blur for a white object on a dark background (bottom row) when the object coincides with the physical display layers. When located in between physical layers, retinal blur is approximated. The retinal blur approximations of additive and multiplicative layers seem comparable in this case (bottom row, center right and right columns). However, additive displays cannot reproduce accurate retinal blur for dark objects in front of bright backgrounds (top row). This is due to the fact that light can only be emitted, but never be blocked. Hence, even a dark cross on a physical display layer appears semi-transparent because the background illumination is shining through it; retinal blur is not physically accurate in this case. Multiplicative layers produce better retinal blur in this case.

### 6.2 Monocular Occlusions

In the previous subsection, retinal blur is evaluated for a single object. Most interesting scenes, however, contain many different objects that may partially occlude each other. In this case, the quality of reproduced monocular occlusion (over the pupil size) is a critical factor that determines the quality of produced retinal blur. We evaluate this for additive and multiplicative multi-plane displays in Figure 10. Similar to the experiment presented in Figure 9, additive displays cannot produce accurate retinal blur when bright objects are occluded by dark objects. Because light cannot be blocked, dark objects always appear semi-transparent. Multiplicative multi-plane



**Figure 9:** Retinal blur evaluated for three different near-eye displays. We present a cross at different distances to the observer’s eye, simulate the output of the different display types, and evaluate retinal blur for a range of accommodation distances. For intuition, the front and rear distances are chosen to coincide with the two physical layers of the simulated multi-layer displays. For the center cross, some type of interpolation has to be performed for multi-layer displays (left). The light field is simulated with  $15 \times 15$  views, which results in a significant resolution loss (red box) for near-eye light field displays based on integral imaging (center left column). Additive and multiplicative multi-layer displays achieve approx. the same quality of retinal blur for a white cross with a dark background (bottom row). However, additive layers can never accurately reproduce retinal blur for a dark cross with a white background because the ambient light always shines through the objects, making them appear semi-transparent (top row). Multiplicative layers produce physically accurate retinal blur in this case when the object is directly on the physical layers (green box) and an approximation otherwise (orange box). Color around boxes indicates quality.



**Figure 10:** Left view of reconstructed light fields showing a scene with bright and dark objects in background and foreground, respectively. Additive multi-focal-plane displays cannot accurately reproduce monocular occlusions in this case (top row), because the bright light emitted by farther objects always shines through the darker ones in the foreground. Multiplicative layers accurately reproduce occlusions over the eye box (bottom row), thereby providing a better approximation of retinal blur for complex scenes.

displays produce accurate occlusions over the pupil size, therefore a better approximation of physically realistic retinal blur.

### 6.3 Diffraction Limits

Any optical system is fundamentally limited by diffraction. For the stacked display architecture discussed in this paper, the virtual image of the rear display panel is always observed through the front panel. For a high-enough pixel density, one would expect diffraction to degrade the image quality of the rear panel. This is illustrated in Figure 11.

A small aperture that approaches the size of the wavelength of light will diffract it. Depending on its size and shape, a single pixel may create an angular diffraction pattern that is similar to an Airy pattern. The diffraction angle  $\theta$  that models the first minimum of the Airy pattern can be used to estimate the diffraction-limited spot size. This angle can be approximated as

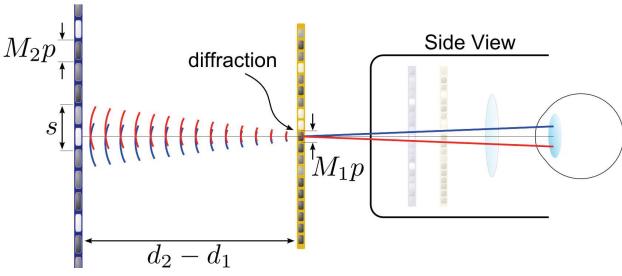
$$M_1 p \sin(\theta) = 1.22\lambda, \quad (11)$$

where  $\lambda$  is the wavelength of light and  $p$  is the pixel size on the physical display panel. Propagating the diffracted pattern through free space to the virtual image of the rear panel creates a diffracted spot size  $s$  of

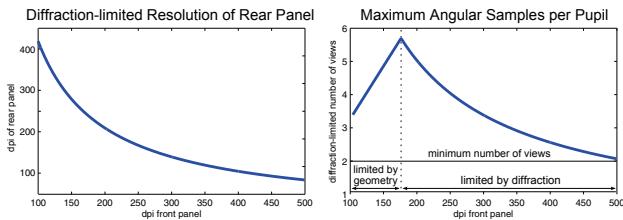
$$s = 2d \tan\left(\sin^{-1}\left(\frac{1.22\lambda}{M_1 p}\right)\right) = \frac{2.44d\lambda}{M_1 p \sqrt{1 - \left(\frac{1.22\lambda}{M_1 p}\right)^2}}, \quad (12)$$

where  $d = d_2 - d_1$  is the distance between layers. If  $s > M_2 p$ , the observed resolution of the rear panel is not limited by its pixel density but by diffraction of the first panel.

In addition to possible resolution loss, diffraction also places an upper limit on how many different angular samples  $\tilde{V}$  can possibly



**Figure 11:** Diffraction-limits. The pixel size of the physical panels  $p$  appears to be magnified on the two virtual images, which are separated by  $d$ . Diffraction of the front panel creates a blurred spot size  $s$  that may degrade perceived resolution. The angular sampling rate over the observer’s pupil is determined by the diffraction-limited resolution of the rear panel; at least two different views must enter the same pupil to allow for focus cues to be supported.



**Figure 12:** Diffraction limits of resolution (left) and angular sampling rate (right). The higher the resolution of the front panel, the more blur is created on the rear panel via diffraction. Assuming that, for the most part, observers will focus on the virtual image of the rear panel (here placed at 1.23 m), high-resolution viewing experiences will only be possible using a low pixel density on the front panel (left). For a fixed resolution of the front panel, we plot the maximum number of light field views entering a 3 mm wide pupil (right). Until about 175 dpi, the angular sampling rate is limited by geometric constraints but for a higher resolutions the angular sampling rate becomes diffraction-limited. Nevertheless, even for 500 dpi panels accommodation could theoretically be achieved.

enter a pupil of size  $e$ :

$$\tilde{V} = \frac{ed}{\max(s, M_2 p)(d_e + d_1)} \quad (13)$$

As long as  $s < M_2 p$ , the sampling rate will be determined by geometric constraints. Increasing diffraction blur, however, reduces the angular sampling rate and thereby limits the ability to produce high-quality focus cues.

Figure 12 analyzes the diffraction limits of spatial resolution on the rear panel and also angular sampling rate over the observer’s pupil. We simulate a 1D display with a wavelength of  $\lambda = 555$  nm (the peak of the luminosity function), a magnifying lens with  $f = 5$  cm, an eye relief distance of  $d_e = 1$  cm, and the virtual images of front and rear panel placed at 19 cm and 123 cm (close to optical infinity), respectively. For an increasing resolution of the front panel, the diffraction-limited resolution of the rear panel quickly drops below any acceptable range (Fig. 12, left). To preserve high-resolution viewing experiences, a relatively low-resolution front panel should be employed. We also plot the number of angular light field samples entering a pupil with a diameter of 3 mm (Fig. 12, right). For lower display resolutions, the angular sampling rate is limited by geometric constraints (i.e. the feature size of the virtual image of the rear panel). For higher-resolution panels, the angular sampling

rate becomes diffraction limit. As a minimum distance between two diffracted rays intersecting the rear panel, we use a conservative distance of twice the Rayleigh criterion. Although the *one-dimensional* angular sampling rate is always above the absolute minimum of two views [Takaki 2006], it is mostly above the recently employed heuristic of three views [Huang et al. 2014], but it also never exceeds six views for the simulated conditions. For this experiment, we simulate front and rear panel to have the same resolution. Possible RGB subpixel structures of the displays are ignored. These may slightly lower predicted bounds for specific orientations (e.g., more diffraction blur horizontally than vertically).

#### 6.4 How Many Display Planes are Necessary?

Additive multi-plane displays have been thoroughly evaluated with layer spacings up to 1D [MacKenzie et al. 2010]. It is argued that about five display layers spaced anywhere between 0.6 to 1 D are required to achieve natural accommodation of observers in a reasonably large depth range [MacKenzie et al. 2010; Ryana et al. 2012]. Our display prototype uses only two layers at a more extreme distance of 4.45 D, which is significantly larger than any layer spacing used by existing additive displays. We argue that the improved quality of retinal blur and especially the support for better monocular (intra-pupillary) occlusion in our multiplicative display results in an overall improvement of retinal blur in complex scenes. This improvement affords increased layer spacing for comparable approximation of retinal blur. A detailed evaluation and analysis of the optimal layer spacing in light field stereoscopes, however, is beyond the scope of this paper and left for future work.

## 7 Discussion

In this paper, we propose a near-eye display technology that supports correct or nearly correct focus cues. The hardware design is a combination of stereoscopic display principles and factored light field display using two stacked liquid crystal panels driven by non-negative, rank-1 light field factorization. The light field stereoscope is the first factored or compressive near-eye display; its characteristics are quite unique compared to other compressive displays. As opposed to television-type light field displays, focus-supporting near-eye displays only require very small viewing areas: the size of the pupil or the eye box at most. The small baseline makes individual light field views very similar, hence compressible to the point where a rank-1 factorization is sufficient. Whereas previous factored light field displays [Lanman et al. 2010; Wetzstein et al. 2012; Maimone et al. 2014] require time-multiplexing via high-speed display panels, the light field stereoscope is afforded by conventional panels as it does not affect the refresh rate of the display itself. We demonstrate that the multiplicative image formation offered by stacked LCDs provides a better approximation of retinal blur cues and occlusions compared with previously-proposed additive multi-plane displays. Finally, we analyze fundamental bounds for resolution achieved with the proposed setup and show results captured with a prototype device.

#### Cinematic Content Creation for Focus-Supporting Displays

We believe that focus cues are a critical feature that will contribute to the success of near-eye displays in the consumer market. Most immersive VR displays today are being marketed for gaming applications. Hence, most of the content shown in our paper is synthetically generated in real time using OpenGL or offline via raytracing. Nevertheless, as near-eye displays continue to become more popular, capturing cinematic content will become a major challenge. Experimental platforms for capturing stereo panoramas are currently being developed by Jaunt VR and Samsung recently an-



**Figure 13:** Stereo light field camera. A stereo pair is captured sequentially by moving the light field camera on a translation stage.

nounced Project Beyond. However, these imaging systems only capture stereo cues and do not contain focus cues that are provided by the proposed and also other near-eye display. Eventually, a need for capturing stereo light field panoramas is unavoidable.

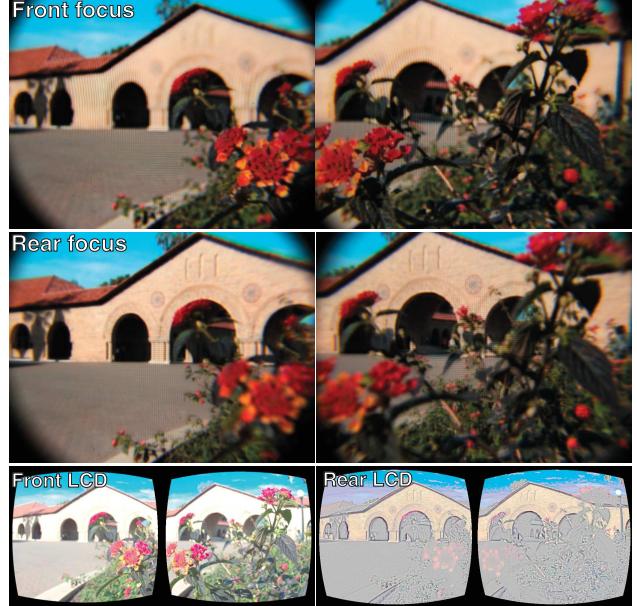
We take a first step towards capturing photographic content for focus-supporting displays using a simple stereo light field camera setup. For this purpose, we mount a Lytro Illum on a manual translation stage (see Fig. 13) that allows us to record two light fields with a baseline corresponding to the inter-pupillary distance of approx. 6.4 cm. The Illum operates at f/2 with a variable focal length between 9.5 and 77.8 mm. We shoot stereo light fields with the shortest focal length, which corresponds to a pupil / eye box diameter of 4.75 mm. Figure 14 shows an example of a resulting factored light field displayed on our prototype near-eye light field display.

**Limitations** Stacking multiple LCD panels reduces the light throughput of the display. Although this is a challenge of all LCD-based multi-layer displays, the quality of an immersive VR experience using this technology is not decreased much. Due to the fact that VR displays cover large parts of the visual field, the human visual system quickly adapts to the brightness conditions. In general, the peak intensity for VR displays is not critical, but it certainly is for AR applications where the display brightness is directly comparable to bright background illumination, for example outdoors. We believe the light field stereoscope is most suitable for immersive VR applications.

As an example of a computational display, the light field stereoscope is driven by advanced nonnegative factorization algorithms. We implement the proposed rank-1 factorization in CUDA and demonstrate real-time framerates. Again, VR applications are unique in that the displays are often connected to a workstation rather than being a wearable display. We can directly leverage high-performance GPUs to provide interactive experiences. Nevertheless, additional computation increases end-to-end latency, which is especially crucial for egocentric motions in VR.

Currently, light field rendering is the bottleneck to the computational performance. In each frame,  $2 \times 5 \times 5$  images, which are the light fields serving as input to the factorization, are rendered. With an increasing number of rendered light field views, which is desirable, rendering times grow quadratically. We expect future graphics pipelines with shading reuse (e.g., [Clarberg et al. 2014]) to significantly accelerate light field rendering.

In addition to slightly increased latency, the main limitation of the light field stereoscope is diffraction. As discussed in Section 6, increased display resolution will increase observed diffraction blur of the rear panel. Ideally, light field stereoscopes should be built with lower-resolution front panels (possible grayscale) and high-



**Figure 14:** Front and rear focus of a stereo light field photographed with a Lytro Illum and displayed on a light field stereoscope. Additional results showing captured, factored, and displayed stereo light fields can be found in the supplement.

resolution rear panels to optimize the viewing experience. Alternatively, the proposed display could be build with reflective micro displays, such as liquid crystal on silicon. If pixel fill factors approaching 100% were achieved, diffraction artifacts could be mitigated. Unfortunately, we are not aware of any available micro display with such a fill factor. Using off-the-shelf components, it may be difficult to achieve ultra-high-resolution image display and high-quality focus cues simultaneously. Although with our display design, we overcome the inherent tradeoff between spatial and angular resolution inherent to integral imaging, we now operate at the diffraction limit. Spatial and angular super-resolution display techniques may be possible but are outside of the scope of this paper. The pixel pitch of consumer displays today is in the order of tens to hundreds of microns; as we move closer to pixel sizes that are in the order of the wavelength of light, holographic display technologies may also become feasible options for focus-supporting near-eye displays. However, speckle and coherent illumination as well as the “big data” problem of high-resolution, wide field of view displays with ultra small pixels will present new challenges.

**Future Work** Exploring the possibility of eye tracking would be an interesting avenue of future work. Foveated light field rendering would be feasible to increase the framerates of the display system. Further, knowing the exact inter-pupillary distance as well as pupil diameter and location would allow for adaptive light field rendering and also relax requirements on eye box size. The rank of a light field is directly coupled to the amount of parallax observed over the eye box; eye tracking would decrease the eye box and improve the quality of rank-1 light field factorizations.

Evaluating the effectiveness of the proposed display with a user study is important. Although we capture many different scenes with an SLR camera simulating a realistic pupil diameter (approx. 4–5 mm) and demonstrate convincing retinal blur quality, the ability for humans to accommodate should be experimentally verified. A commercial autorefractor is probably the easiest way to verify ac-

commodative responses. Whereas experiments with these devices are easily done with benchtop displays, the small form factor of our device did not allow us to evaluate it with the Grand Seiko autorefractor that was at our disposal. The device is too big and optical paths cannot be matched. We believe that other types of user evaluations in future work are crucial to fully evaluate the effectiveness of the proposed technology.

On the capture side, opto-computational imaging systems that allow for the recording of 360° panoramic stereo light fields are highly desirable. The brute-force approach would be to use arrays of light field cameras, but more elegant solutions, for example employing compressive light field photography [Marwah et al. 2013] may be feasible as well.

## 8 Conclusion

The light field stereoscope is a crucial step towards highly-immersive but also comfortable experiences. Along with content, latency, field of view, and resolution, visual comfort may be one of the most important factors determining the eventual success of emerging near-eye displays in the consumer market. With applications in education, collaborative work, teleconferencing, scientific visualization, remote-controlled vehicles, training and simulation, and surgical training, immersive computer graphics provide societal benefits beyond entertainment. Although the benefits of focus cues may be less pronounced for older adults or for virtual scenes that are far away, many applications require interaction with objects that are within reach. With the proposed technology, we hope to contribute a practical technology for improving visual comfort. We strongly believe our technology to be a crucial advance towards more immersive and comfortable VR experiences.

## Acknowledgements

The authors would like to thank Marty Banks, Chia-Kai Liang, and the anonymous reviewers for valuable feedback. Gordon Wetzstein was supported by a Terman Faculty Fellowship.

## 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.
- BENTON, S., AND BOVE, V. 2006. *Holographic Imaging*. John Wiley and Sons.
- BROWN, D. 2000. Decentering distortion of lenses. *Photogrammetric Engineering* 32, 3, 444–462.
- CAKMAKCI, O., AND ROLLAND, J. 2006. Head-worn displays: a review. *Journal of Display Technology* 2, 3, 199–216.
- CLARBERG, P., TOTH, R., HASSELGREN, J., NILSSON, J., AND AKENINE-MÖLLER, T. 2014. Amfs: Adaptive multi-frequency shading for future graphics processors. *ACM Trans. Graph.* 33, 4 (July), 141:1–141:12.
- CUTTING, J. E., AND VISHTON, P. M. 1995. Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. *W. Epstein and S. Rogers (Eds.), Perception of space and motion*, 69–117.
- FATALORA, G. E. 2005. Volumetric 3D displays and application infrastructure. *IEEE Computer* 38, 37–44.
- HALE, K., AND STANNEY, K. 2014. *Handbook of Virtual Environments*. CRC Press.
- HEIDE, F., LANMAN, D., REDDY, D., KAUTZ, J., PULLI, K., AND LUEBKE, D. 2014. Cascaded displays: Spatiotemporal superresolution using offset pixel layers. *ACM Trans. Graph. (SIGGRAPH)* 33, 4, 60:1–60:11.
- HELD, R., COOPER, E., AND BANKS, M. 2012. Blur and Disparity Are Complementary Cues to Depth. *Current Biology* 22, R163–R165.
- HIRSCH, M., WETZSTEIN, G., AND RASKAR, R. 2014. A Compressive Light Field Projection System. *ACM Trans. Graph. (Proc. SIGGRAPH)* 33, 4, 1–12.
- HOFFMAN, D. M., AND BANKS, M. S. 2010. Focus information is used to interpret binocular images. *Journal of Vision* 10, 5, 13.
- HUA, H., AND JAVIDI, B. 2014. A 3d integral imaging optical see-through head-mounted display. *OSA Opt. Exp.* 22, 11, 13484–13491.
- HUANG, F.-C., WETZSTEIN, G., BARSKY, B. A., AND RASKAR, R. 2014. Eyeglasses-free display: Towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph. (SIGGRAPH)* 33, 4, 59:1–59:12.
- 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.
- KELLY, D. H. 1979. Motion and vision. ii. stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America* 69, 1340–1349.
- KRESS, B., AND STARNER, T. 2013. A review of head-mounted displays (hmd) technologies and applications for consumer electronics. In *Proc. SPIE*, vol. 8720, 87200A–87200A–13.
- LANMAN, D., AND LUEBKE, D. 2013. Near-eye light field displays. *ACM Trans. Graph. (SIGGRAPH Asia)* 32, 6, 220:1–220:10.
- 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.
- LEVOY, M., AND HANRAHAN, P. 1996. Light Field Rendering. In *ACM SIGGRAPH*, 31–42.
- LIU, S., CHENG, D., AND HUA, H. 2008. An optical see-through head mounted display with addressable focal planes. In *Proc. ISMAR*, 33–42.
- LOVE, G. D., HOFFMAN, D. M., HANDS, P. J., GAO, J., KIRBY, A. K., AND BANKS, M. S. 2009. High-speed switchable lens enables the development of a volumetric stereoscopic display. *OSA Optics Express* 17, 18, 15716–15725.
- MACKENZIE, K. J., HOFFMAN, D. M., AND WATT, S. J. 2010. Accommodation to multiplefocalplane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision* 10, 8.
- MAIMONE, A., AND FUCHS, H. 2013. Computational augmented reality eyeglasses. In *Proc. ISMAR*, 29–38.
- MAIMONE, A., WETZSTEIN, G., HIRSCH, M., LANMAN, D., RASKAR, R., AND FUCHS, H. 2013. Focus 3d: Compressive accommodation display. *ACM Trans. Graph.* 32, 5, 153:1–153:13.

- MAIMONE, A., LANMAN, D., RATHINAVEL, K., KELLER, K., LUEBKE, D., AND FUCHS, H. 2014. Pinlight displays: Wide field of view augmented reality eyeglasses using defocused point light sources. *ACM Trans. Graph. (SIGGRAPH)* 33, 4.
- MARSHALL, J. A., MARSHALL, J. A., ARIELY, D., BURBECK, C. A., ARICLY, T. D., ROLLAND, J. P., AND MARTIN, K. E. 1996. Occlusion edge blur: A cue to relative visual depth. *OSA JOSAA* 13, 681–688.
- 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)*.
- ROTHBAUM, B., HODGES, L., READY, D., GRAAP, K., AND ALARCON, R. 2001. Virtual reality exposure therapy for vietnam veterans with posttraumatic stress disorder. *Ann Surg* 62, 8, 617–22.
- RUCH, T., AND FULTON, J. F. 1960. *Medical physiology and biophysics*. W. B. Saunders Company.
- RUSHTON, S. K., AND RIDDELL, P. M. 1999. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Appl Ergonomics* 30, 69–78.
- RYANA, L., MACKENZIEA, K., AND WATTA, S. 2012. Multiple-focal-planes 3d displays: A practical solution to the vergence-accommodation conflict? In *Proc. 3D Imaging (IC3D)*, 1–6.
- SCHOWENGERDT, B. T., AND SEIBEL, E. J. 2006. True 3-d scanned voxel displays using single or multiple light sources. *Journal of the Society for Information Display* 14, 2, 135–143.
- SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROCOV, A. 2004. High dynamic range display systems. *ACM Trans. Graph. (SIGGRAPH)* 23, 3, 760–768.
- SULLIVAN, A. 2003. A solid-state multi-planar volumetric display. In *SID Digest*, vol. 32, 207–211.
- SUTHERLAND, I. 1968. A head-mounted three dimensional display. In *Proc. AFIPS Fall Joint Computer Conference*.
- 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.
- TAKAKI, Y. 2006. High-Density Directional Display for Generating Natural Three-Dimensional Images. *Proc. IEEE* 94, 3.
- WATT, S., AKELEY, K., ERNST, M., AND BANKS, M. 2005. Focus cues affect perceived depth. *Journal of Vision* 5, 10, 834–862.
- 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.
- WHEATSTONE, C. 1838. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London* 128, 371–394.
- ZHANG, Z. 2000. A flexible new technique for camera calibration. *IEEE Trans. PAMI* 22, 11, 1330–1334.
- ZWICKER, M., MATSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultoscopic 3D displays. In *EGSR*.