Wide Field of View Compressive Light Field Display using a Multilayer Architecture and Tracked Viewers

Andrew Maimone[*], *Renjie Chen*[*], *Henry Fuchs*[*], *Ramesh Raskar*[**], *Gordon Wetzstein*[**]

[*]**University of North Carolina at Chapel Hill, USA**
[**]**MIT Media Lab, USA**

Abstract

In this paper, we discuss a simple extension to existing compressive multilayer light field displays that greatly extends their field of view and depth of field. Rather than optimizing these displays to create a moderately narrow field of view at the center of the display, we constrain optimization to create narrow view cones that are directed to the viewer's eyes, allowing the available display bandwidth to be utilized more efficiently. These narrow view cones follow the viewer, creating a wide apparent field of view. Imagery is also recalculated for the viewer's exact position, creating a greater depth of field. The view cones can be scaled to match the positional error and latency of the tracking system. Using more efficient optimization and commodity tracking hardware and software, we demonstrate a real-time, glasses-free 3D display that offers a 110x45 degree field of view.

1. Introduction

Since the invention of the stereoscope by Sir Charles Wheatstone in 1838, the public has been fascinated by the idea of viewing photographs, cartoons, or movies in 3D. Although a vast range of three-dimensional displays have been invented in the last 150 years [10], 3D technologies have had limited commercial success thus far despite the ubiquity of 2D displays among smartphones, tablets, computer monitors, and televisions. Part of the reason for this failure may be that most existing 3D displays require additional eyewear, such as polarizing glasses, which is impractical for everyday use. Light field or glasses-free 3D displays are an alternative, but usually sacrifice image resolution for 3D capabilities [1]. Unfortunately, the human visual system is extremely sensitive to resolution. Degrading overall image quality to support 3D presentation does not increase the overall viewing experience – rather it appears to decrease it. With a critical flicker fusion threshold of about 60 Hz, however, the human visual system is rather insensitive to fast motions. Time-multiplexed 3D displays offering high image resolution and 3D capabilities simultaneously therefore seem like the most promising avenue of practical 3D displays (e.g., [3,9]). One of the major challenges of time-multiplexed displays, however, is that most readily available display hardware, such as liquid crystal displays (LCDs), support only limited refresh rates of 60-240 Hz. The number of time-multiplexed images is therefore limited to only a few sub-frames, making these technologies suitable for stereoscopic but not multiview or light field displays.

Over the last few years, a new generation of displays has started to emerge: compressive light field displays. By combining unconventional optical setups, such as multilayer LCDs [2,4,5,8,11] or directional backlights [6,12], with compressive computation, these types of displays support unprecedented image resolution and 3D capabilities using commodity hardware. The key idea behind all of these displays is to directly exploit the compressibility of the presented light field image content. However, as opposed to conventional 2D image compression, compressive light field displays employ a joint optical and computational approach to presenting compressed content that allows the human visual system to act as a decoder. Despite the efficiency gains made by compressive displays, however, these displays still have limited display bandwidth: they are unable to present glasses-free imagery over a wide field of view.

In this paper, we explore the combination of head tracking and compressive light field displays. Whereas previously-described light field displays support limited fields of view of about 10 to 20 degrees, we demonstrate that head tracking can significantly increase the field of view of a compressive light field display as well as its depth of field.

2. Related Work

Stereoscopic and multiview 3D displays have been an active area of research for more than a century [10]. With the emergence of high-speed displays, time-multiplexed 3D image presentation has become one of the most promising directions [3,9]. These technologies can be made more efficient by exploiting the compressibility of the displayed multiview image content. In particular, compressive light field displays [4,12] use modern matrix and tensor factorization algorithms to enhance image brightness and reduce the requirements on display refresh rates as compared to conventional time-multiplexing techniques. Whereas most existing 3D displays are limited to binocular depth cues, compressive light field displays have also demonstrated support for eye accommodation [6].

One of the primary challenges for all existing light field display technologies is support for a wide field of view. Current display designs dictate a tradeoff on achievable field of view: image resolution must be degraded [1] or display refresh rates must be increased [3,9] to widen the field of view. Whereas most compressive light field display architectures have thus far only demonstrated limited fields of view [2,4,5,11,12], head and eye tracking has been shown to allow for more flexible display modes in some configurations [6,7]. However, these existing tracked works [6,7] have not demonstrated a real-time, wide field of view display system that operates in high resolution.

In this paper, we explore the promise of head-tracked compressive light field displays. By flexibly steering a "view cone" with a small instantaneous field of view towards the direction of the observer, we demonstrate that the observed field of view of a dual-LCD light field display can be significantly increased. This methodology applies to a variety of other light field displays as well.
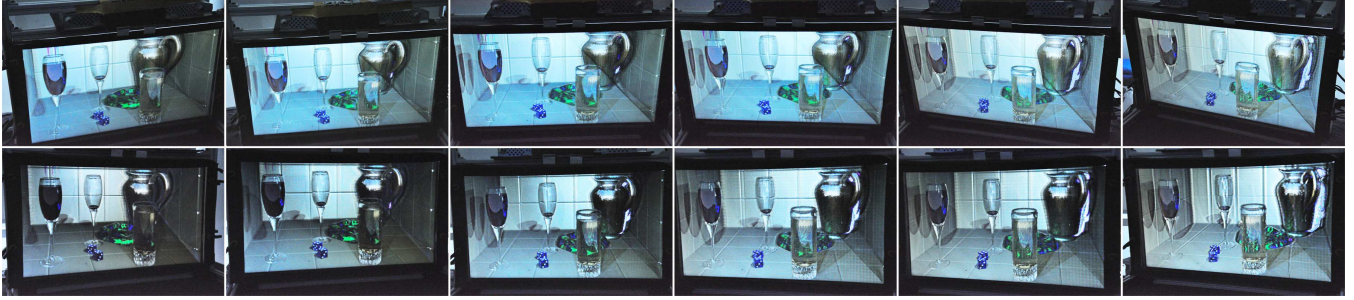
Figure 1. Wide field of view compressive light field display. We extend the instantaneous field of view of a dual-layer light field display using head tracking. Whereas the instantaneous field of view provided by the dual-layer display is only about 10x10 degrees, this can be flexibly steered into the direction of the observer without any hardware modifications. For this experiment, we achieve a perceived field of view of 110x45 degrees using the proposed techniques. 3D Model credit: Gilles Tran

3. Compressive Image Synthesis

This section briefly reviews image formation, rank-deficiencies of multi-layer light field displays, and strategies for efficient image synthesis. Our basic description follows Lanman et al. [4] and Wetzstein et al [12]; please see these works for full details.

3.1 Rank Deficiency of Parallax Barriers

We employ an *absolute two plane* parameterization for the light field emitted by a dual-layer display. The discrete light field **L** produced by two stacked attenuating layers *f[i]* and *g[k]*, for instance liquid crystal displays (LCDs), can be represented as the outer product:

$$L[i,k] = f[i] \otimes g[k] = f[i]g^T[k]$$

In this analysis we show 1D masks and a 2D light field for simplicity, but this generalizes to 4D by reordering the 4D light field into a 2D matrix.

From the above equation, it is clear that any pair of 2D masks can only produce a rank-1 approximation of a 4D light field. Here, rank can be understood to indicate the *degrees of freedom* available, thus a rank-1 approximation of a light field will be of poor quality: the field of view is small and the depth of field is shallow. Experiments shown in past work [4] indicate that a rank-3 reconstruction should be sufficient to obtain a PSNR of 30 dB.

To achieve a higher rank reconstruction, we take advantage of a perceptual effect in the human visual system: flicker fusion. A rapid sequence of images will be perceived as their temporal average. Thus, a series of rank-1 matrices can be integrated to achieve a higher rank matrix. Any sequence of *T* 1D mask pairs creates (at most) a rank-*T* decomposition of a 2D light field matrix such that

$$L[i,k] = \frac{1}{T}\sum_{t=1}^{T} f_t[i] \otimes g_t[k] = \frac{1}{T}\sum_{t=1}^{T} f_t[i]g_t^T[k],$$

where $f_t[i]$ and $g_t[k]$ denote the rear and front masks for frame *t*, respectively. The light field matrix must be decomposed as the matrix product

$$L = FG^T,$$

where **F** and **G** are $N_i \times T$ and $N_k \times T$ matrices, respectively. Column *t* of **F** and column *t* of **G** are the masks displayed on the rear and front LCD panels during frame *t*.

3.2 Optimizing Pixel States

Mask pairs $\{f_t[i,j],\ g_t[k,l]\}$ must be non-negative since they represent the pixel states of the LCDs. We aim for a light field factorization $L' = FG^T$, that minimizes the weighted Euclidean distance to the target light field **L**, under the necessary non-negativity constraints, such that

$$\underset{F,G}{\operatorname{argmin}} \frac{1}{2}\left\| \beta L - FG^T \right\|_W^2, for\ 0 \le F, G \le 1$$

The field of view can be adapted to multiple viewers by specifying elements of the weight matrix **W**, a binary valued matrix which indicates which rays are constrained. The weight matrix ensures that a low-rank approximation obtains higher accuracy by artificially reducing the rank of the target light field. Brightness scaling factor *β* is used to control the tradeoff between image quality and brightness.

This expression can be solved using non-negative matrix factorization (NMF). We use the weighted multiplicative update rule, as described in past work [4]. Figure 3 represents typical mask pairs found by the NMF approach outlined here. Note that these masks are significantly more transmissive than pinhole arrays. In fact, although we can predict their structure, we cannot obtain them analytically. The masks create localized 1D parallax barriers, which follow the contours of the angular gradient of the light field. This observation gives us intuition about the maximum performance of our algorithm. The rotated 1D barrier intuition tells us that we can never increase the brightness of a scene more than that achieved by a horizontal-parallax only 3D display.

3.3 Extending Field of View through Head Tracking

The main challenge of dual-layer compressive light field displays is that only a small field of view and shallow depth of field can be achieved. Although these shortcomings can be alleviated using multiple stacked layers or directional backlighting [6,12], this is either optically less efficient (stacks of LCDs block a lot of light) or more challenging to engineer. In this paper, we propose a simple technique to significantly increase the perceived field of view of a dual-layer display for a single observer by flexibly steering a small field of view directly towards the observer.

In other words, we constrain the optimization only for rays that reach a small area around the eyes of the viewer; all other rays are
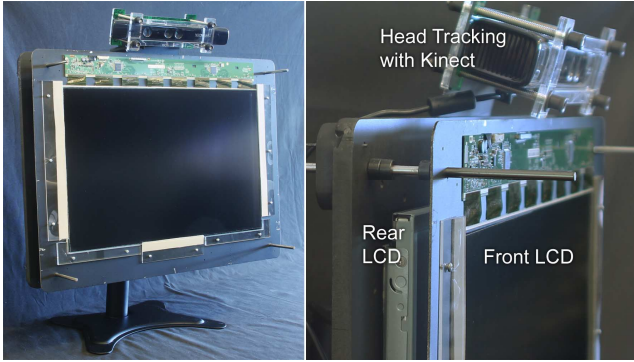
Figure 2. Prototype compressive light field display. The device comprises two stacked, off-the-shelf liquid crystal displays (LCDs) and a Microsoft Kinect sensor that is used to estimate the position of the observer's eyes.

unconstrained. This is implemented by setting the weight matrix W (see Section 3.2) to indicate only this set of rays.

As shown in Figure 1, this allows us to increase the observed field of view from about 10×10 degrees to 110x45 degrees for this experiment. The same technique is applicable to other multi-layer light field displays and is indifferent to the underlying head tracking mechanism. A Microsoft Kinect and vision-based tracking is used to find the 3D positions of the eyes.

4. Implementation
In this section, we describe software and hardware implementation details of the proposed compressive display.

4.1 Hardware
Our prototype display consists of two stacked spatial light modulator (SLM) layers and a tracking camera mounted on a metal frame. The SLM layers are LCD panels obtained from 27" Asus VG278H monitors which operate at 144 Hz with a resolution of 1920x1080. The LCD panels were removed from their housing and mounted on a frame with a spacing of 3.3 cm. The rear polarizer of the front LCD was removed, and the front polarizer of the front LCD was replaced with a clear polarizer so that there was an alternating set of crossed polarizers about the LCD stack. A mildly diffusing film was placed in front of the rear LCD panel to reduce the effects of moiré between the two panels. A Microsoft Kinect color plus depth camera was mounted above the LCD layers to provide 3D eye tracking of the viewer. The prototype display is illustrated in Figure 2.

4.2 Software
Our prototype system is capable of real-time multilayer optimization that allows the display to be observed over a wide field of view by an eye-tracked user. The input light field for the optimization process can be pre-rendered or rendered in real-time on the GPU; Figure 1 shows a pre-rendered scene raytraced using POV-Ray while Figure 3 shows a scene rendered in real-time using OpenGL. At each frame, the initial state of the layer masks is seeded with the state of the previous frame before performing optimization.
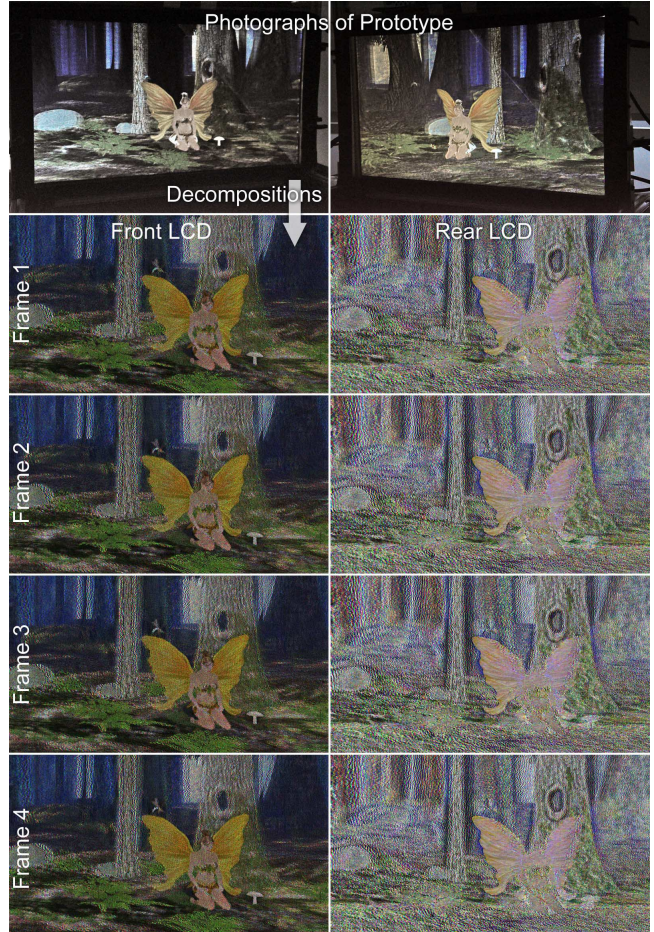

Figure 3. Photographs of two perspectives of the "fairy" scene (top). Each of the perspectives is dynamically decomposed into a set of four patterns for each LCD. These patterns are displayed at high speed on the respective LCD layers.

Because of the small change in the scene between frames, this allows the optimization to converge with few iterations. Optimization runtimes were an average of 4.44 ms per iteration, enabling real-time operation. Optimization was performed with four time-multiplexed frames and a brightness scaling factor of β =0.5 (see Section 3.2) while the weight matrix W was set to consider only the viewpoints around the viewer's eyes during optimization. To obtain the user's 3D eye positions, eyes were tracked in 2D using the FaceTracker library[1], and the depth coordinate was obtained using the depth map from the Kinect depth camera. For the results shown in this paper, we used a printed face – recognized as a real face by the tracking system – to obtain the 3D positions at several locations in front of the display where a camera was placed to photograph the display. For the photographs, we constrained the optimization to only the exact position of a camera. However, for moving tracked user, the optimization can be constrained to a small view *cone* around the user's eyes, allowing the correct image to be seen in the presence of tracker error and latency. Using this technique, we were able to use the display with human viewers during informal testing.

---

[1] https://github.com/kylemcdonald/FaceTracker

5. Results

Figure 1 shows photographs of our prototype with the "glasses" scene. A small instantaneous field of view showing the scene from the observer's perspective is rendered and pixel states of the display optimized accordingly. The total observed field of fiew of this scene is 110×45 degrees – a significant improvement over previously reported fields of view for light field displays.

Figure 3 shows photographs of the prototype for an additional scene. We also show the decomposed patterns displayed on the front and rear LCD for one of the perspectives. Whereas these patterns almost look like noise, when optically overlaid on the sceen and displayed at a high refresh rate, the viewer perceives a consistent, high-resolution 3D image of the target scene.

6. Discussion

In summary, we explore the combination of compressive light field displays and head tracking. By steering a small instantaneous field of view dynamically into the direction of a single, tracked observer, we demonstrate how the perceived field of view of the display is significantly improved.

Our current prototype exhibits artifacts, which can mostly be attributed to limitations in the precision of LCD panel alignment. The depth of field of each perspective is limited by the panel refresh rate. Artifacts observed for objects that extend from the physical display plane, such as the jug on the right-hand side of the "glasses" scene in Figure 1, could be mitigated by using faster LCD panels or other high-speed light modulators.

In the future, we would like to explore display settings that facilitate wide fields of view for multiple observers. Currently, the most practical solution for this problem would be high-speed light modulators.

7. Acknowledgements

8. References

[1] Fattal, D., Peng, Z., Tran, T., Vo, S., Fiorentino, M., Brug, J., Beausoleil, R. 2013. A multi-directional backlight for a wide-angle, glasses-free three-dimensional display. Nature 495, 348-351

[2] Gotoda, H. 2013. Design of time-multiplexed auto-stereoscopic displays based on virtual stacking of multi-layer panels. Proc. SPIE 8648

[3] Kim, Y., Kim, J., Kang, J.-M., Jung, J.-H., Choi, H., Lee, B. 2007. Point light source integral imaging with improved resolution and viewing angle by the use of electrically movable pinhole array. Optics Express 15, 26, 18253–18267

[4] 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.

[5] 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.

[6] Maimone, A., Wetzstein, G., Lanman, D., Hirsch, M., Raskar, R., Fuchs, H. 2013. Focus 3D: Compressive Accommodation Display. ACM Trans. Graph. 32(5)

[7] Nashel, A., Fuchs, H. 2009. Random Hole Display: A non-uniform barrier autostereoscopic display. 3DTV Conference.

[8] Ranieri, N., Heinzle, S., Barnum, P, Matusik, W., Gross, M. 2013. Light-Field Approximation using Basic Display Layer Primitives. Proc. SID, vol. 44, no. 1, 408-411.

[9] Travis, A. 1990. Autostereoscopic 3-D display. Appl. Opt. 29, 4341-4342.

[10] Urey, H., Chellappan, K.V., Erden, E., Surman, P. 2011. State of the Art in Stereoscopic and Autostereoscopic Displays. *Proc. IEEE*, vol.99, no.4, 540-555.

[11] 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.

[12] 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.