Dynamic Field of View in a Tomographic Light Field Display

Jason Ginsberg
Stanford University
jasong2@stanford.edu

Neil Movva
Stanford University
nmovva@stanford.edu

Figure 1. Two-Layer Stacked Attenuation Display

Abstract

In this paper, we extend the field of view of a glasses-free multi-layer light field display, which we have assembled. By integrating a camera into the rendering and display pipeline, we constrain our light-field display to a narrow field of view at the center of the display. As the viewing cone follows the viewer's gaze, we optimize our iterative tomographic reconstruction to be utilized more efficiently and extend the apparent field of view from 10° to 100°. Finally, we demonstrate real-time generation of dynamic light field scenes on the GPU.

1. Introduction

Traditional 2D displays often depend on monocular cues to replicate 3D effects. Illusions of perspective, occlusion, and shading give a simulated effect of depth. These effects, however, are limited in driving true oculomotor responses like vergence and accommodation. 3D displays extend the class of replicated depth cues to those which are binocular, including: motion parallax, occlusion, translucency, and specularity. Though engineers have built designs for 3D displays as early as the turn of the 19th century, modern displays have enabled the accurate, high-resolution depiction of these effects in a compact, cost-effective form factor.

In particular, this paper considers an automultiscopic display composed of compact volumes of light-attenuating material (stacked LCDs). Compared to other 3D display techniques, this sub-class of display requires no encumbering hardware (e.g. special glasses) and enables objects to appear beyond the display.

In this paper, We have implemented iterative tomographic reconstruction for image synthesis on a stack of spatial light modulators (multiple low-cost iPad LCDs) using Simultaneous Algebraic Reconstruction Techniques. We illuminate these volumetric attenuators with a backlight to recreate a 4D target light field. Although five-layer decomposition generates the optimal tomographic reconstruction, our two-layer display costs less than 100 dollars and incurs less computational cost.

We also explore the combination of head tracking and attenuation-based light field displays. Whereas previously described light field displays support limited field of view of about 10° to 20°, we demonstrate that head tracking can virtually extend the field of view of a light field display and optimize rendering and display. Finally, we demonstrate the generation of these light fields in real-time, taking advantage of the GPU.

1.1. Goals

Our minimal requirements for building an effective 3D display were as follows:

Binocular Depth Cues. Binocular cues are necessary in driving the human visual system to interpret a scene as three-dimensionally realistic. The display must also be able to present content beyond the bounds of the LCDs, so that it may replicate deep real-world scenes. A display which

elides this requirement would not generate convincing or immersive content.

Non-Encumbering. Requiring the user to adopt additional hardware accessories would restrict collaborative use. Since displays are ubiquitous in workplace environments and the public sphere, requiring the user to don additional hardware would significantly restrict adoption. Head-mounted hardware would similarly be a non-starter for many users, unwilling to obstruct their vision or significantly alter their daily sartorial habits.

No Moving Parts. Moving parts greatly increase the cost and complexity of manufacturing displays. Great assurance would be needed to protect and maintain high-speed components. For a display to achieve wide adoption, we contend that fully solid-state operation is a requirement.

Dynamic Real-Time Scenes. Much viewable content is dynamic or interactive. To serve these purposes, a display must be able to generate and update content at or above 60Hz, with minimal input latency (roughly 10 to 100ms, depending on user and use case).

Wide Field of View. We expect users of a 3D display to be placed spread throughout a space, or alternatively move across the display. To maintain a realistic and immersive 3D viewing-experience for large audiences, the display must be able to maintain its 3D effect across a wide field of view and for multiple viewers.

1.2. Constraints

Given our desire to build an affordable, reproducible display, our design was informed by the following constraints:

- Less than 100 dollars for parts and assembly (excluding rendering computer)
- Min. 60Hz refresh rate
- Min. Two display layers
- Max. 100ms input latency
- Solid-state operation

1.3. Contributions

We present a low-cost light field display with real-time dynamic rendering. We also establish a technique to effectively extend the field of view. Specific contributions are as follows:

- Prove the effectiveness of a low-cost two-layered attenuation-based light field display.
- Show how SART allows light fields to be optimally displayed.
- Demonstrate that a dynamic viewing cone can virtually extend the field of view.

2. Related Work

Designs for 3D displays have been suggested as early as the turn of the 19th century. Frederic E. Ives was the first to demonstrate a glasses-free autostereoscopic 3D display with his invention of the Parallax Barrier in 1901; though the design significantly limited resolution and attenuated light. [2] Ives' work would be improved upon by Gabriel Lippman just five years later with his invention of Integral Imaging; by adding lenslets to the barrier apertures, Lippman was able produce a brighter albeit still low-resolution display compared to Ives'. [6]

More recent glasses-less 3D display technologies can be divided among Additive Volumetric, Holographic, and Multiplicative Volumetric. Given our aforementioned goals, this paper focuses on multiplicative volumetric displays. Though all three display types can produce binocular depth cues without encumbering hardware, Multiplicative Volumetric displays are the only type which can present dynamic scenes without moving parts and beyond the bounds of the display.

Lanman et al. first presented content-adaptive parallax barriers in 2010; a dual-layer display, which was optimized by temporally-varying attenuation found from non-negative matrix factorization. [3] In comparison, our iterative tomographic reconstruction provides the optimal rank-1 decomposition, in the two layer case, guaranteed to converge to the global, rather than a local, minimum. [4]

Wetzstein et. al. [2011] introduced Layered 3D, which developed tomographic techniques for image synthesis on compact light-attenuation based displays. [9] Layered3D the primary motivation for our display prototype. This paper however did not focus on real-time rendering and projection, field of view extension, or limited cost and compute considerations.

The Tensor Display is another compressive light field display developed by Wetzstein et al. [2012]. [10] The Tensor Display employs time-multiplexing, N-Layers with M-Frames, and a directional backlight to project order N, Rank-M light field scenes. Wetzstein's implementation even utilized the GPU to render dynamic and interactive applications. Though Tensor Displays emit high-resolution 3D scenes, the display framework requires more than two high refresh rate displays ($\geq$120Hz), a large form-factor,

optical elements, and a directional backlight; these components place the Tensor Display out of the scope and budget of our constraints.

In Real-time Image Generation for Compressive Light Field Displays, Wetzstein et. al [2013] delve deeper into mapping computed tomographic light field synthesis onto the standard graphics pipeline. [8] Additionally, the paper discusses efficient GPU-based implementations of light field rendering with real-time frame rates. The algorithms described, in particular SART, greatly informed our strategies in synthesizing and displaying light fields.

Wide Field of View Compressive Light Field Display using a Multilayer Architecture and Tracked Viewers by Maimone et al. [2014] provides the primary motivation for our research. [7] Using commodity tracking hardware and software, Maimone et al. were able to greatly extend field of view and depth of field in compressive multilayer light field displays.

Polarization Fields by Lanman et. al [2011] improves upon the attenuation-based light field display in terms of resolution, brightness, and depth of field. [5] The display further adopts a GPU-based SART algorithm to render interactive dynamic scenes. Though a Polarization Field-based display would be of higher quality than our two-layer display, it would not adhere to our constraints as it increases cost, complexity, and thickness significantly.

In The Light Field Stereoscope, Huang et al. [2015] use a near-eye light field display to produce correct focal cues in a virtual reality headset; these cues help to resolve visual discomfort arising from the vergence-accommodation conflict. [1] Though we focus on non-encumbering displays, the Light Field Stereoscope proves an alternative use case with similar constraints as it also relies on a non-time multiplexed, 2-layer light-field display.

It is important to note that our paper does not present any single new method to rendering or synthesizing light fields. Rather our objective was to combine multiple approaches into a compact and reproducible form-factor in order to produce interactive binocular depth cue-driven 3D scenes. We see this as necessary for large-scale adoption of light field displays, which has not occurred at the rates many of these aforementioned papers predicted.

3. Method

Overview of the stages in our rendering/display pipeline [tracked face–position–light field–to light field projection iterative approximation, show matrix image for solution–display]

3.1. Computational Equipment

Our rendering workloads target an Nvidia Pascal GPU (GP104), programmed through OpenGL in a Linux environment. We run the face tracking algorithm on an Intel RealSense R200, an intelligent camera platform that uses structured light to provide scene depth annotation, enabling better viewer pose estimation.

3.2. Display Construction

We first procured two 9.7" diagonal, 2048x1536 60Hz IPS LCDs, commonly found as the displays used in the Apple iPad 3. We chose this part for its mature, highly available supply which has kept the price low, and critically, the use of a native 'embedded DisplayPort' interface that enables direct, passive connection to any DisplayPort source. Since our goal is to use the LCDs as spatial light modulators, we remove the backlight assembly and extract the IPS pixel layer.

The enclosure was then designed such that the stack of LCDs would be well-fastened and closely aligned. The enclosure also leaves space on the back panel to mount and conceal driver circuitry. To fabricate the enclosure, we laser cut our design on 1/8" sheets of Duron (similar to plywood).

In this initial prototype, we screwed the back LCD into place and then attached the front LCD with adhesive so that we could manually adjust the display for approximate pixel alignment.

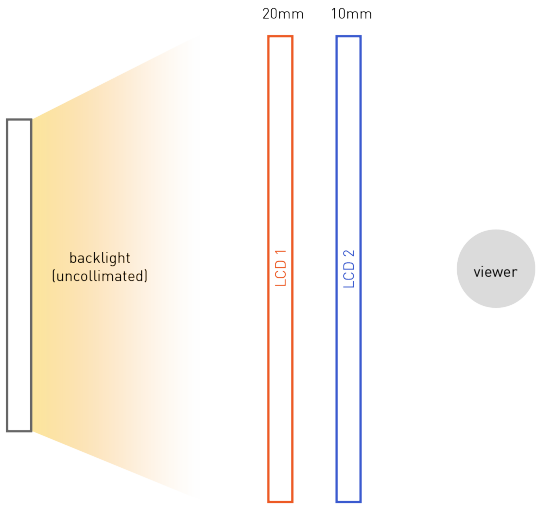

Figure 2. LCD stack diagram, two layers. Note that the backlight produces uncollimated, multidirectional light, and may be somewhat nonuniform over the LCD area.

Figure 3. Final LCD stack prototype. RealSense R200 camera is mounted to bottom edge of the LCD stack. Notice the teapot image faintly visible in the center of the display - bright lighting conditions very quickly overwhelm our prototype's brightness capability.

3.3. Face Tracking

Our face tracking algorithm makes use of a Haar feature-based cascade classifier [Rapid Object Detection using a Boosted Cascade of Simple Features, 2001] and relies on the OpenCV library for a performant, tuned implementation. Our face tracking software provides estimated coordinates for viewer positions relative to the center of the LCD stack, which we then pass as camera coordinates to our light field rendering and display function.

3.4. Tomographic Light Field Synthesis

The 2D volumetric attenuator is defined as a continuously varying attenuation map $\mu(x, y)$. In this model, $\mu(x, y)$ is computed to obey Beer-Lambert's law so that $I = I_0 * e^{-\int_c(\mu(r)dr)}$

By the Weber-Fechner law, however, the human visual system recognizes logarithmic changes in illumination as nearly linear, so the illumination is re-computed as $I^- = ln(I/I_0)$.

In the forward model, the Radon transform $p(u, a)$can take the attenuation map μ and the width and height of the layered slabs to encode all possible line integrals through the attenuation map along each ray (u, a). Here, the orientation of ray (u, a) is defined by the slope $a = s - u = d_r * tan(\theta)$ where d_r is the distance of the s-axis from the u-axis. The oblique light field can then be described as $l^-(u, a) = p(u, a)$for linear angle a.

With parallel beam tomography, an estimate of the attenuation map, $\mu_{approx} * (x, y)$ is recovered from the projections $p(u, a)$ using the inverse Radon transform. In turn, the filtered back-projection algorithm estimates a volumetric attenuator capable of emitting the target light field. The projection matrix P_{ij}^k corresponds to line integrals through every basis function k along each ray (i, j) can be expressed as a linear combination of N_b non-negative basis functions with coefficients α. This system, which models the attenuation as a linear system of equations when considering a discrete light field, is expressed in matrix-vector form as $P * \alpha = -l^- + e^-$, where e^- is the approximation error. As a result, we cast attenuation map synthesis as the following non-negative linear least-squares problem:

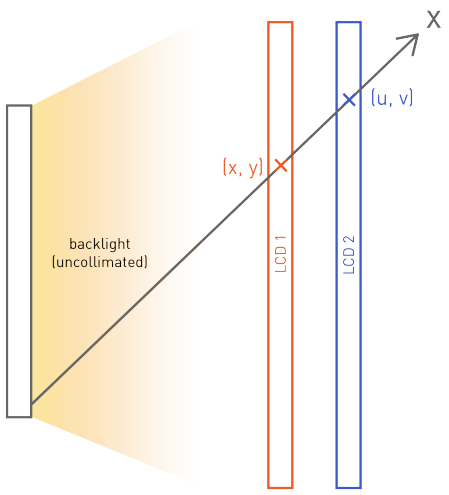

Figure 4. Illustration of light field tomography. Light rays from the backlight are defined by straight-line paths through coordinates on each LCD masking plane, resulting in per-ray multiplicative attenuation that can be controlled through rendering of the tomographic mask.

$$\arg\min \alpha \text{of } ||l^- + P * \alpha||^2 \text{for } \alpha >= 0$$

For multi-layered attenuators, the form of the projection matrix P is modified, now encoding the intersection of every ray with each mask. Thus, a similar optimization solves the inverse problem of constructing an optimal multi-layered attenuator. Practically, however, layers have a finite contrast and the approximation is solved as a constrained least-squares problem.

3.5. SART

Once a light field representation of the scene is rendered, we are presented with an optimization problem: solving for the set of tomographic masks that best reconstructs the rendered light field. The Simultaneous Algebraic Reconstruc-

tion Technique (SART) algorithm is well-suited to this task, and is described in more detail by Wetzstein et. al in [9].

We exploit the high temporal correlation of video by starting each frame's guess with the solution from the prior frame. Furthermore, the SART algorithm is largely composed of matrix multiplications and independent element-wise operations, which are generally amenable to parallel execution. Using compute shaders implemented in the Cg language by Wetzstein et. al, we can afford to run multiple SART iterations within our 100ms latency budget when executing on the Nvidia GP104 GPU of our rendering hardware.

4. Assessment

Figure 5. Demonstration setup of light field display. The LCD stack is in the center right of the frame, immediately to the left of the observer's head. In the upper left quadrant of the image, a large conventional display serves as debug output by showing the 2 computed tomographic masks currently being sent to the LCD stack, as well as showing the face-tracking software's estimate of the current viewer position.

4.1. Optical Efficiency

Using uncalibrated light metering (provided by a commodity smartphone ambient light sensor), we coarsely estimate each of our LCDs to achieve no more than 5% transmissivity, meaning our two-layer stack would suffer a light attenuation factor of roughly 400x. This severely limits the current practical use of our display, especially in nominal daylight conditions. We considered circumventing this limitation by significantly increasing backlight strength, but were quickly limited by thermals - at low transmissivity, much of the backlight energy instead contributes to radiant heating of the LCD layers, necessitating significant forced air cooling to maintain safe operating temperatures. Such measures were not feasible in our demonstration environment, and so we were limited to a 5-10W LED source (an original iPad 3 display backlight).

4.2. Real Time Rendering

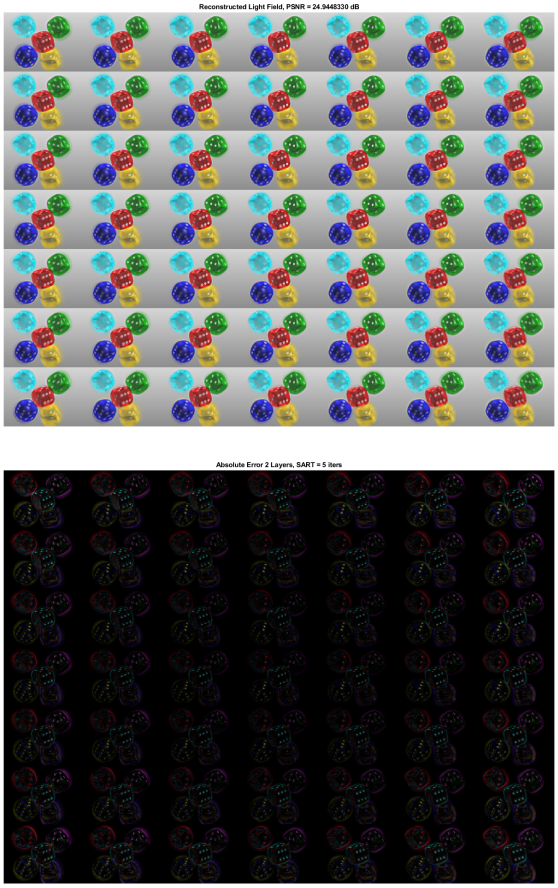

Figure 6. (Upper Frame) Computed layers of reconstructed 384x512x7x7 light field covering a field of view of 10°. PSNR vs. ground truth light field is 24.94 dB. (Lower Frame) Visualization of absolute error between the ground truth and reconstructed light field. The layers are equally separated with a depth range of 1/4", which corresponds to two 1/8" Duron sheets.

Guided by fine-grained performance profiling, we are able to tune our rendering pipeline to achieve approximately 80% utilization of our GPU (an Nvidia GP104) execution resources, with the remaining inefficiency explained mostly by OpenGL draw call overhead. Given a global latency maximum of 100ms (which includes input latency and LCD response time), we target a total budget of 50 ms to render each frame (i.e. set of 2 tomographic masks, to be sent to the LCD stack). That time is split approximately equally between light field rendering and iterative SART optimization, and we dynamically reduce SART iterations in cases

where we would exceed the target 50 ms budget. On average, this allows about 4 SART iterations per frame.

4.3. Dynamic Field of View

Empirically, we observe that the error of our tomographic solution becomes unacceptably high when attempting to reconstruct a light field with a static field of view greater than 10°. We fix our rendering pipeline (i.e. light field synthesis, then tomographic reconstruction) to this 10° FoV, and dynamically retarget according to viewer position updates. Since our face-tracking system can run at 60 updates/sec, in parallel with the rendering pipeline, retargeting can occur within our 50ms render time budget. Qualitatively, we find the system to feel quite responsive to user movement. Thus, we extend our apparent field of view (in the single viewer case) to that of the face tracking camera, which we were able to validate out to 100°.

For a demonstration of dynamic retargeting, we've recorded a short video of our system's debug views, available here. In the video, the tomographic mask views are shown such that the left frame is sent to the front (closest to viewer) LCD, and the right frame is sent to the rear LCD.

5. Discussion

5.1. Benefits and Limitations

We find modern LCDs to be highly cost-effective when used as spatial light modulators. Increasing spatial resolution or stacking more LC layers both appear to be promising, low-cost routes to high-fidelity light field reconstruction, and the recent history of LCD development suggest continued progress on both these axes. However, the stacked LCD architecture in general pays a steep price in light efficiency, and this tradeoff is only exacerbated by the same trends we seek to exploit (higher resolution reduces transmissivity, stacking layers compounds attenuation). Thus, alternative engineering solutions will be required, potentially negating many of the cost savings and imposing additional fundamental restrictions in terms of power and thermal design.

5.2. Future Work

Brightness. The most apparent usability issue in our current prototype is the low brightness of the displayed scene. This is a problem inherent to the stacked LCD approach, as each LCD only transmits a fraction of the incident backlighting. As discussed in our analysis, we experience light attenuation of 2-3 orders of magnitude . There exist some LCD panels that achieve higher fill factors and correspondingly higher transmissivity, but for the future of our prototype, we think it simplest to pay off this heavy transmission loss by engineering a high intensity backlight, managing thermals as needed. From our current prototype, we envision this to be possible, but will require even further compromises in power consumption and footprint, essentially binding us to immobile desktop use cases.

Low Refresh Rate. The original stacked LCD design from Wetzstein et. al called for high refresh rate (120+ Hz) LCDs, such that successive frames could multiplex many light rays in time. Unfortunately, fast refresh LCDs remain somewhat niche and thus expensive; additionally, computing high-resolution, low error tomographic mask solutions within a tighter latency bound will demand significantly more computing resources. We want to first characterize exactly how much bearing this multiplexing has on the perceptual quality of the scene, and if not absolutely critical, explore partial solutions while avoiding perceptual flicker at just 60 Hz.

Pixel Alignment. Our future prototypes will be designed around a more rigid frame, with better facilities for fine positional adjustments. We think the current prototype suffers from significant misalignment of the two LCDs, resulting in what appears to be a 'smeared, blurry' scene.

Color Crosstalk. Adding more layers to our stack will afford us finer control over the aggregate tomographic mask, helping reduce the crosstalk we currently observe from our lack of an ideally directional backlight.

6. Conclusion

We propose an inexpensive, compact two-layer light field display that produces binocular 3D depth cues. Though we trade-off brightness and resolution compared to polarization fields, we maintain significantly lower cost and complexity. Notably, our display demonstrates 3D effects across a wide field of view for real-time dynamic, interactive scenes. We hope that our work can prove the practicality of 3D displays without the need for encumbering head mounted hardware or glasses. We strongly believe that glasses-less 3D displays are necessary for enabling a new paradigm in human-computer interaction that does not sacrifice collaboration or shared viewing experiences.

7. Acknowledgements

We thank Professor Gordon Wetzstein for mentoring our project and the Computational Imaging Group for lending equipment. We also thank the Stanford student organization Rabbit Hole VR for funding the project hardware.

References

[1] F. Huang, K. Chen, and G. Wetzstein. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-

Eye Light Field Displays with Focus Cues. *ACM Trans. Graph. (SIGGRAPH)*, (4), 2015.

[2] F. E. Ives. Parallax stereogram and process of making same. 1903. U.S. Patent 725,567.

[3] D. Lanman, M. Hirsch, Y. Kim, and R. Raskar. Content-adaptive parallax barriers: Optimizing dual-layer 3d displays using low-rank light field factorization. *ACM Trans. Graph. (SIGGRAPH Asia)*, (29):163:1163:10., 2010.

[4] D. Lanman, G. Wetzstein, W. Heidrich, and R. Raskar. Beyond parallax barriers: Applying formal optimization methods to multi-layer automultiscopic displays. *Proceedings of SPIE - The International Society for Optical Engineering*, 8288(02):7, 2012.

[5] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar. Polarization fields: Dynamic light field display using multi-layer LCDs. *ACM Trans. Graph.*, 30(6), 2011.

[6] G. Lippman. Epreuves rversibles donnant la sensation du relief. *Journal of Physics*, 7(4):821825., 1908.

[7] A. Maimone, R. Chen, H. Fuchs, R. Raskar, and G. Wetzstein. Wide field of view compressive light field display using a multilayer architecture and tracked viewers. *SID Symposium Digest of Technical Papers*, 45(05), 2014.

[8] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. Layered 3d: tomographic image synthesis for attenuation-based light field and high dynamic range displays. *ACM Trans. Graph.*, 30:95:1–95:12, 2011.

[9] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. Real-time image generation for compressive light field displays. *Journal of Physics Conference Series*, 415(012045), 2013.

[10] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):1–11, 2012.