Real-Time Analysis and Synthesis of Imagery for Light-Field Displays

Tianyu Wu*, Hee-jin Choi**, Benjamin Watson*

***Visual Experience Lab, North Carolina State University, Raleigh NC, USA**
****Sejong University, Seoul, Korea**

Abstract

Current light-field displays increase resolution and reduce cross-talk with head tracking, despite using simple lens models. With a more complete model, our real-time technique uses GPUs to analyze the current frame's light flow at subpixel precision, and to render a matching image that further improves resolution and cross-talk.

Author Keywords

Light-field display; integral imaging; eye-tracking; real-time ray tracing.

1. Introduction

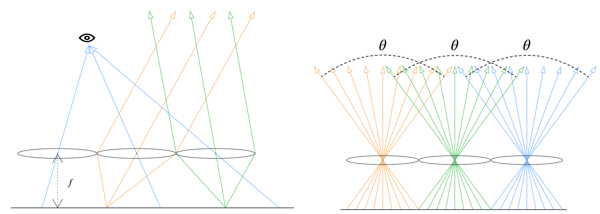
The integral imaging light-field display (LFD) provides stereoscopic imagery to its viewers, delivering images through an array of lenses. It achieves both binocular parallax and motion parallax, allowing displayed objects to be perceived in 3D from various viewing positions, without requiring the user to wear any specialized equipment (e.g. glasses). However, conventional LFDs (without head-tracking) are far from ideal: they suffer from low resolution, limited viewing angle, and cross-talk, in which the user sees multiple images, including some not matching the current view. With head-tracking, LFDs can mitigate some of these issues, especially cross-talk, by identifying and addressing display areas containing it [5, 6]. Yet because they use simple optical models and rendering methods, existing head-tracked LFDs cannot eliminate cross-talk. With more accurate, real-time optical modeling and rendering, our technique significantly reduces cross-talk, improves resolution.

2. Ray Tracing and LFDs

Our technique relies on close integration of rendering with display. The two primary rendering methods are rasterization [1] and ray tracing [2,3]. Until recently, rasterization has dominated because of its better fit to hardware. But use of ray tracing is increasing in games and elsewhere, thanks to recent hardware support [4]. Because rasterizers build imagery triangle by triangle, with projected triangles spanning many display pixels, rasterizers must assume that pixels are viewed from a single point. With LFD lens arrays altering light paths, this single viewpoint assumption is at best an approximation. For this reason, we use ray tracing, which builds imagery pixel by pixel and allows a different view projection per pixel, supporting accurate modeling of LFD light paths. Lee et al. also used ray tracing in their LFD research [5], though rather than using it to generate and display imagery, they used ray tracing to make ground truth comparative images to evaluate their work.

3. Conventional Optical Modeling for LFDs

In conventional LFDs, lenslets are placed so that the distance from their centers to the display panel is equal to their focal lengths. With this placement, light coming from any point on the display will become parallel after traveling through a lenslet; and when the user looks at a lenslet, they see the display panel locations that intersect with the rays from their eyes through the lenslet center, as Figure 1 (a) shows. Hence conventional LFDs segment viewing zones pixel by pixel, defined by the light beam originating from the center of each pixel through the lenslet directly above it, as is shown in Figure 1 (b). Thus, when making imagery for an LFD, one need only render the colors of the views along the per-pixel beams.



(a) (b)

Figure 1. (a) parallel light beams through lenses, where f is the focal length of lenses; (b) the views defined by each pixel beneath a lens, defining the viewing angle θ.

While this optical model works adequately, users often notice cross-talk. Most commonly, this happens when the user's eye is inside overlapping viewing zones, shown in Figure 2. In effect, viewers see not an infinitely small point below each lenslet, but something more like a line segment.

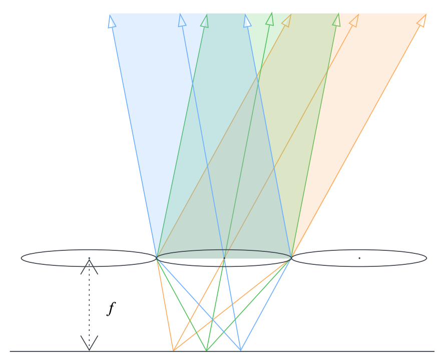


Figure 2. Overlapping viewing zones indicate cross-talk: multiple pixels visible under the same lenslet.

This cross-talk has several sources, including the angular range of the light moving from each lenslet to the eye, the use of single-sided lenlets, and vertical refraction.

The angular range of the light entering the eye from the lenslet is quite different from the single light ray modeled by the conventional optical model, particularly when the user is close to the display. This difference is enough to make several pixels visible under the lenslet, rather than just one, contributing to cross-talk. Furthermore, often only parts of some pixels are visible, distorting the colors users see.

Another source of cross-talk is single-sided lenses. Frequently the lenslet sheet is made with a uniform material, meaning only the side facing the user has small curved lenslets, while the other side is flat and pressed onto the display panel with little to no space in between. This breaks the conventional model's focal length lens assumption: without double-sided lenses, focal length and lens-to-display distance are not the same, and light traveling from a pixel through a lenslet may not be parallel when it leaves the lenslet. Lens sheets with double-sided lenslets are very uncommon and difficult to make, because the refractive index of the substrate material under the lens is not the same as the air above the lens sheet [7]. Single-sided lenslets increase the number pixels seen through them still further, increasing cross-talk.

Finally, the conventional optical model assumes that each pixel row in a display is viewed identically and perpendicularly, meaning that the model ignores vertical refraction. Yet of course, the user's eye can only be perpendicular to at most one row, while typically all rows are visible, and vertical refraction affects the light path of the majority of pixels. Neglect of vertical refraction creates significant sub-pixel-level misalignment, particularly at the display periphery.

4. Real-Time, Complex Optical Modeling

To address all of these sources of cross-talk in conventional LFD optical modeling, our technique combines a more accurate optical model with head-tracking and parallel GPU computation, enabling accurate calculation of the display areas visible through the lens sheet, in real-time. As shown in Figure 3, for each pixel row, we model light paths from the eye location to lenslet edges, refract the light paths on the lenslet's surface, and find the intersections between the paths and the boundaries between pixel rows at the bottom of the lens sheet, i.e. the display panel. This defines the visible strip, that is the portion of the pixel row visible under a lenslet. Note that the visible strip varies in width both across each lenslet, and across each pixel row. Thus this model is 2D and must be calculated across the entire display. Finding the two light paths that move from a pixel row boundary through a lenslet's edges to the eye is non-trivial. The calculation involves several unknown variables, most importantly the refraction point, and requires solving a set of equations.

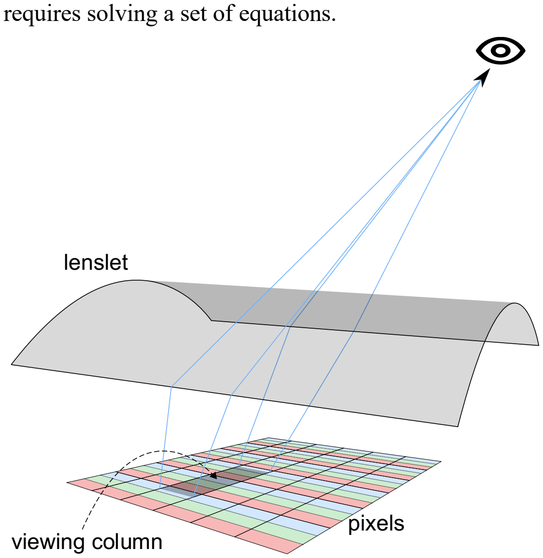


Figure 3. Rays entering the eye originating from the four corners of a viewing strip, refracted through the lenslet.

Solving these equations requires several display parameters, including the horizontal positions of each lenslet's edges, the vertical position of each display pixel row, the tangent angle of each lenslet's curve at its edges, and the lenslet's refractive index. We solve for the two vertical positions on each edge of the lenslet through which light traveling from the pixel row boundaries through the lenslet would reach the eye. We also solve for the horizontal positions on the pixel row boundaries that the user sees through each lenslet edge. The results determine the areas of the display that the ray tracer samples, defined with precision exceeding sub-pixel resolution.

5. Sub-Pixel Color Display

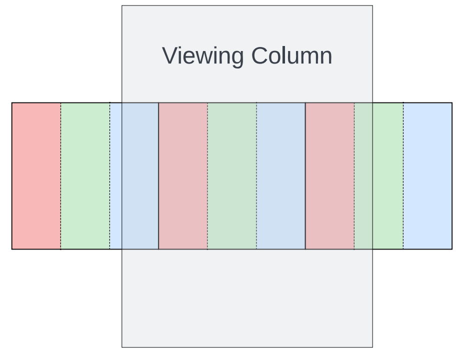


Figure 4. Part of a viewing column crossing one pixel row.

We call all the strips visible through a single lenslet a viewing column. Each column's shape is almost rectangular, however, its vertical edges are slightly curved. Viewing column edges do not align with subpixel edges, as is shown in Figure 4. Given this, sub-pixel color adjustments are needed for the best viewing experience.

To improve the color displayed in each viewing column, we work one visible strip at a time. First, we adjust the brightness of the leftmost sub-pixel in the strip. If the target luminance of the sub-pixel is L and the proportion of the sub-pixel inside the strip is p, then we set its brightness to $\min\left(\frac{L}{p}, 1\right)$. Next, if the leftmost sub-pixel is green or blue, we shift pixel boundaries to the right by 1 or 2 sub-pixels. For example, if the intended pixel colors in a viewing strip are (r_0, g_0, b_0), (r_1, g_1, b_1), (r_2, g_2, b_2), (r_3, g_3, b_3), $\cdots$ and the leftmost sub-pixel is green, we set the actual pixel values to $\left(0, \min\left(\frac{g_0}{p}, 1\right), b_0\right)$, (r_0, g_1, b_1), (r_1, g_2, b_2), (r_2, g_3, b_3), $\cdots$. When the number of sub-pixels is not a multiple of 3, we combine the 1 or 2 extra sub-pixels at the right edge of the visible strip with the rightmost full pixel. For example, if the leftmost sub-pixel of a visible strip is red, and the color in the rightmost full pixel is (r, g, b), the extra blue sub-pixel is outside the strip, and the proportion of the extra green sub-pixel inside the strip is p, we set the color of full pixel to $\left(\frac{r}{2}, \frac{g}{1+p}, b\right)$, and the color of the display pixel on the right strip boundary to $\left(\frac{r}{2}, \frac{g}{1+p}, 0\right)$. Thus the combined luminance of the two pixels is (r, g, b).

6. Preliminary Evaluation

To verify the correctness of our complex optical model, we display one of two patterns on the LFD: a visible columns (VC) image and a not visible columns (NVC) image. The VC image displays maximum brightness on sub-pixels that should be seen from a viewing position and turns off the sub-pixels that cannot be seen. Through the refraction of the lens sheet, the VC image should appear completely white. The complement of this VC image is the NVC image, it turns off visible sub-pixels and displays maximum brightness on those that cannot be seen, so the perceived image should appear completely black. To capture the appearance of these images through the lens array, we used a DSLR camera in manual mode. We kept all camera parameters unchanged throughout the evaluation, excepting focal length and focus so that we could capture the best view of the LFD.

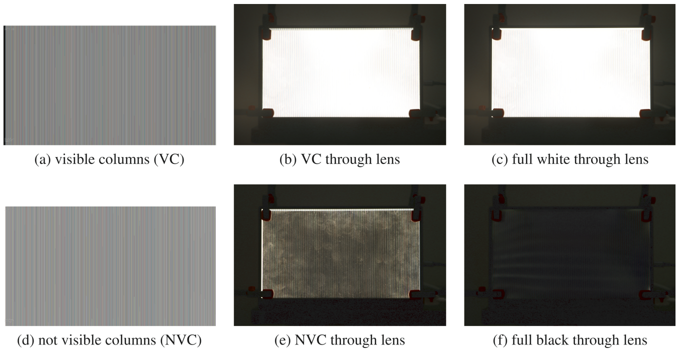


Figure 5. Testing optical model accuracy from a view centered on the display. (a) is the visible columns (VC) image seen without a lens, with white in visible pixels. (d) is the complementary not visible columns (NVC) image, with white in pixels not visible. (b) is VC viewed through the lens array, while (e) is NVC viewed through the array. For comparison, (c) is a fully white image viewed through the array; while (f) is a fully black image.

Figure 5 shows the VC and NVC images without lenses, corresponding photos taken with a camera centered in front of the LFD and viewing images through the lens sheet, and additional photos of the display fully turned on or off, again viewed through the lens array. The VC image (a) appears completely white when perceived through the lens sheet in (b), nearly the same as viewing a completely white image through the lens array in (c). On the other hand, the NVC image (d) appears largely black when perceived through the lens sheet (e), but it is brighter than the completely black image viewed through the lens array (f). Nevertheless, the NVC image (e) is significantly darker than VC image (b), despite inevitable leaked light through subsurface scattering.

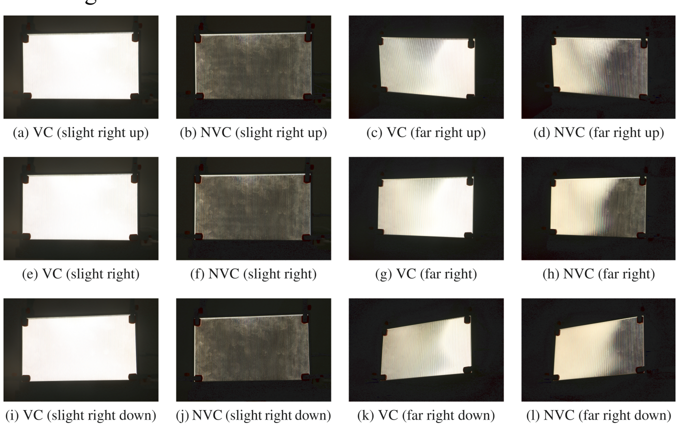


Figure 6. Testing optical model accuracy from off-center views. Accuracy looks good in slight right views, but less so in the far right views.

We further tested our model from more viewing positions (7° rightward and 3° up and down, and 23° rightward and 15° up and down), with the results shown in Figure 6. Accuracy holds up well from slightly angled perspectives, but less so from very angled perspectives, most noticeably in the NVC photos (d), (h), and (l), where the distant side of the display becomes white instead of black. We believe this is because of imprecision in the current test process, as the display and camera were placed loosely by hand, making the parameters used in the model calculation inaccurate. We plan to reduce these inaccuracies but are nevertheless encouraged by the preliminary results for the centered and slightly angled views, which indicate accurate optical modeling.

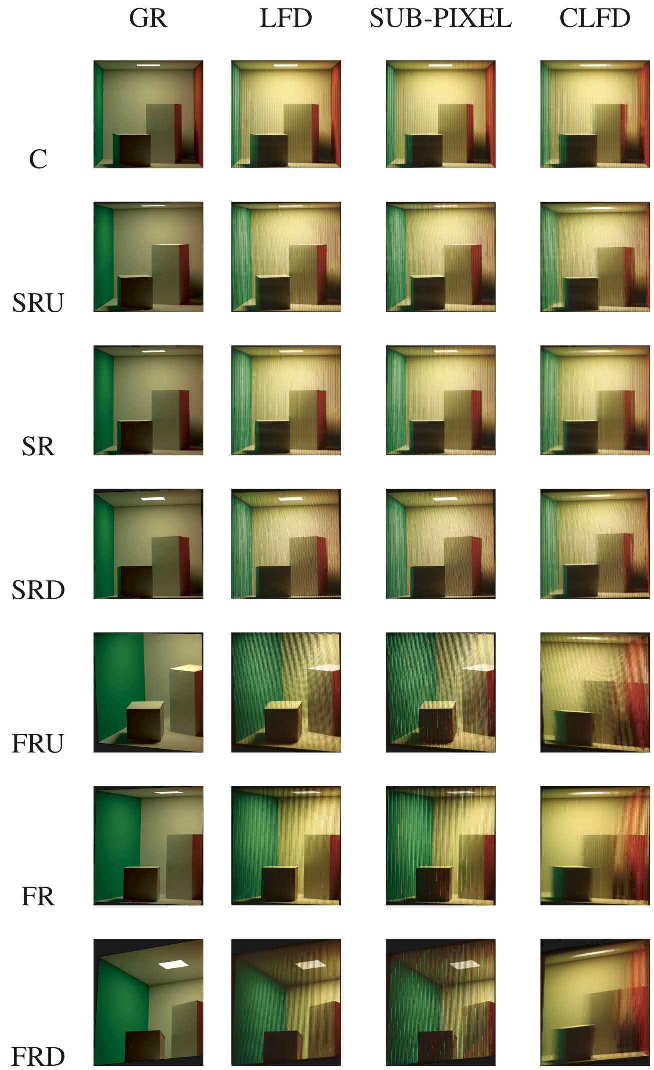


Figure 7. Cornell Box images rendered by different LFD techniques (horizontal), across different views (vertical). Renderings include GR (ground truth), LFD (our LFD technique without sub-pixel adjustment), SUB-PIXEL (our technique sub-pixel adjusted), and CLFD (conventional LFD rendering). Views include C (viewed from the center), SRU (slightly right & up), SR (slightly right), SRD (slightly right & down), FRU (far right & up), FR (far right), and FRD (far right & down).

Next we compare images generated with our LFD technique to images displayed on a conventional LFD, and a lensless, very high-quality "ground truth". All images show the well-known Cornell Box scene from the same set of views used in Figure 6. Our technique's images contain only one view, while the conventional LFD's images contain many. The ground truth imagery is also single-view, rendered at a much higher resolution using all display pixels, captured without the lens sheet.

Figure 7 compares these LFD rendering techniques. From left to right, each column shows ground truth, our LFD technique without any sub-pixel adjustment, our technique with sub-pixel adjustment, and conventional LFD rendering. From top to bottom, the figure shows the center view; the slightly right up, middle, and down views; and the far right up, middle, and down views. Across all these views, the images built by our head-tracked LFD techniques are very close to the ground truth, while the conventional LFD, which only contains a fixed set of views, differs greatly in views from above and below. Echoing our model accuracy results in Figure 6, our sub-pixel-adjusted LFD techniques show artifacts in the far right views. Note that in this preliminary evaluation, our LFD only provides a single view and does not support stereoscopic viewing; we plan to implement stereoscopic 3D soon.

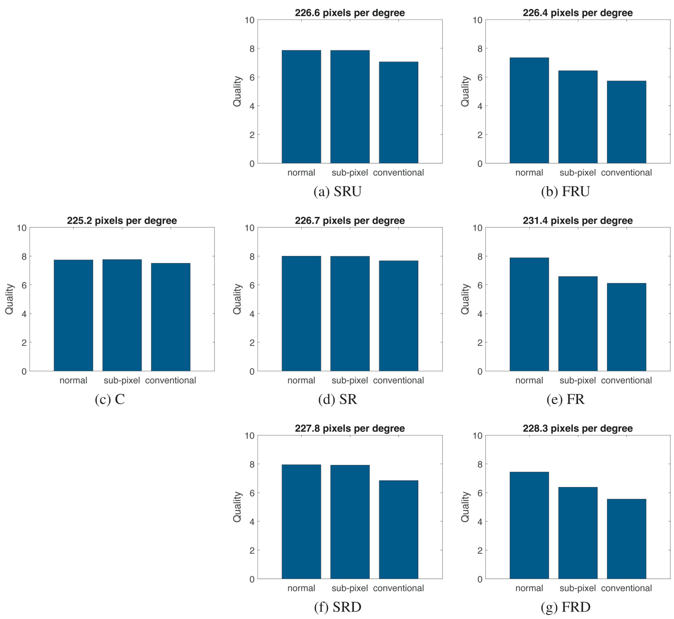


Figure 8. Image quality comparisons of conventional LFD and our techniques with and without sub-pixel adjustment against ground truth using HDR-VDP, with the same abbreviations as Figure 7.

To objectively quantify the accuracy of these images, we used HDR-VDP [8,9], an image quality assessment tool based on a human visual model. We compared the photos of our and conventional LFD techniques against the ground truth imagery. We graph the results in Figure 8, with higher values indicating more similarity to ground truth. In all of the views, our LFD technique without sub-pixel adjustment performs better than the conventional LFD due to significantly reduced cross-talk, and sub-pixel-adjusted imagery perform as well or slightly better than not adjusted in the center and slightly right views; in the far right views, due to the artifacts caused by inaccuracy of the experimental setup, our LFD technique with sub-pixel adjustment performs noticeably worse than without, but still better than conventional LFD. It is also worth noting that in the centered view, which is less prone to parameter errors, the sub-pixel adjusted technique yields slightly better image quality than unadjusted.

7. Conclusions & Future Work

Conventional and even head-tracked LFDs use simple optical models that result in significant cross-talk. By using a more accurate model that we update pixel by pixel and frame by frame, we are able to significantly reduce cross-talk, while also improving resolution.

Our work has several limitations, each of which we plan to address in the near future. First, our research to date has focused on LFDs with vertical lens sheets. Current commercial LFDs use slanted lenses, improving the resolution aspect ratio. We are currently modeling optical flow for slanted lenses, increasing real-world relevance. Second, our LFD technique currently renders several times per second. Solving the equations in our optical model requires use of a high precision math library. Unfortunately the only available high precision GPU libraries no longer function on current GPUs. We are now updating those GPU libraries to achieve real time speeds. Lastly, our evaluation, while promising, is incomplete. We are preparing evaluations that include stereoscopic imagery, interactive human viewing, and more precise experimental camera positioning.

8. Acknowledgements

We thank Turner Whitted for providing valuable inspiration and advice. This work was supported in part by NSF award IIS-2008590.

9. References

1. Wylie C, Romney G, Evans D, Erdahl A. Half-tone perspective drawings by computer. In: Proceedings of the November 14-16, 1967, fall joint computer conference. 1967. p. 49–58.
2. Appel A. Some techniques for shading machine renderings of solids. In: Proceedings of the April 30--May 2, 1968, spring joint computer conference. 1968. p. 37–45.
3. Whitted T. An improved illumination model for shaded display. In: ACM Siggraph 2005 Courses. 2005. p. 4-es.
4. Burgess J. Rtx on—the nvidia turing gpu. IEEE Micro. 2020;40(2):36–44.
5. Lee J-H, Yanusik I, Choi Y, Kang B, Hwang C, Park J, et al. Automotive augmented reality 3D head-up display based on light-field rendering with eye-tracking. Optics Express. 2020;28(20):29788–804.
6. Aoyama K, Yokoyama K, Yano T, Nakahata Y. 48-5: Eye-sensing Light Field Display for Spatial Reality Reproduction. In: SID Symposium Digest of Technical Papers. 2021. p. 669–72.
7. Ramakrishna SA. Physics of negative refractive index materials. Reports on progress in physics. 2005;68(2):449.
8. Mantiuk R, Kim KJ, Rempel AG, Heidrich W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Transactions on graphics (TOG). 2011;30(4):1–14.
9. Narwaria M, Mantiuk RK, Da Silva MP, Le Callet P. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. Journal of Electronic Imaging. 2015;24(1):010501–010501.