Perceptually-Guided Foveation for Light Field Displays

QI SUN, Stony Brook University and NVIDIA Research
FU-CHUNG HUANG and JOOHWAN KIM, NVIDIA Research
LI-YI WEI, University of Hong Kong
DAVID LUEBKE, NVIDIA Research
ARIE KAUFMAN, Stony Brook University

(a) foveation left, focus far (b) foveation right, focus far (c) foveation right, focus near

Fig. 1. *Foveated light field display and rendering.* (a), (b), (c) are our simulated retinal images under foveation with different tracked eye gazes (shown in green circles) and different focus planes. Specifically, (b) has the same gaze position but different focus plane from (c), and the same focus plane but different gaze position from (a). Our method traces only 25% of the light field rays while preserving perceptual quality.

A variety of applications such as virtual reality and immersive cinema require high image quality, low rendering latency, and consistent depth cues. 4D light field displays support focus accommodation, but are more costly to render than 2D images, resulting in higher latency.

The human visual system can resolve higher spatial frequencies in the fovea than in the periphery. This property has been harnessed by recent 2D foveated rendering methods to reduce computation cost while maintaining perceptual quality. Inspired by this, we present foveated 4D light fields by investigating their effects on 3D depth perception. Based on our psychophysical experiments and theoretical analysis on visual and display bandwidths, we formulate a content-adaptive importance model in the 4D ray space. We verify our method by building a prototype light field display that can render only 16% – 30% rays without compromising perceptual quality.

CCS Concepts: • **Computing methodologies** → **Perception**;

Additional Key Words and Phrases: light field, computational display, foveation, sampling

ACM Reference format:
Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-Guided Foveation for Light Field Displays. *ACM Trans. Graph.* 36, 6, Article 192 (November 2017), 13 pages.
https://doi.org/10.1145/3130800.3130807

0730-0301/2017/11-ART192 $15.00
https://doi.org/10.1145/3130800.3130807

1 INTRODUCTION

Advances in graphics algorithms and hardware have enabled high quality and low latency for traditional 2D displays. However, consistent 3D depth perception, which is important for perceptual comfort, remains out of reach for many users.

Light field displays support focal cues [Huang et al. 2015; Lanman and Luebke 2013; Maimone and Fuchs 2013; Maimone et al. 2013; Narain et al. 2015], but current rendering techniques cannot generate high quality content in real time. With gaze tracking, foveated rendering reduces computational costs while maintaining perceptual quality [Guenter et al. 2012; Patney et al. 2016]. However, existing methods are designed for 2D images; foveating 4D light field displays remains a challenging open problem. The human visual system automatically reconstructs 2D retinal images from 4D light fields. However, light field foveation cannot be simply reduced to image foveation due to the lack of reliable technology for tracking accommodation, a major factor of monocular depth perception.

Inspired by prior work on 4D light field display and 2D foveated image rendering, we present the first foveated light field rendering and display system that supports low latency and high quality, as well as focus accommodation to improve depth perception and reduce vergence-accommodation conflicts. Based on our psychophysical studies, our main idea is to derive an importance sampling model in the 4D light field ray space based on both foveation and accommodation. Conceptually, this can be achieved by tracing rays from retina cells back through the eye and into the scene, and varying the focal length of the eye to sweep the ray space.

We derive the spectral bounds of the light field imaging pipeline, including the display, the eye lens, and the retina. Based on these

bandwidths, we propose a sampling and reconstruction method for real-time rendering of foveated 4D light fields.

Our study also addresses a long-standing argument among the display and vision communities [Huang et al. 2015, 2014; Maimone et al. 2013; Narain et al. 2015; Pamplona et al. 2012; Takaki 2006; Takaki et al. 2011] on the number of rays necessary to support focal cues. Our spectral analysis shows that the number depends on several factors including the display/eye optics, the retinal eccentricity, and the scene content. The analysis allows us to significantly reduce the rendering cost while preserving perceptual quality.

We evaluate our method by conducting psychophysical studies through our hardware prototype running a variety of scenes with different characteristics. Our system is shown to render up to 3× faster than prior work and trace only 16% ∼ 30% of all rays of the light field display while maintaining similar visual quality.

The main contributions of this paper include:

- We analyze the bandwidth bounds for perceiving 4D light fields based on the display property, the eye lens, and the retinal distribution, and derive a minimum sampling rate to answer the argument among the display, graphics, and vision communities.
- Based on the spectral bounds and the depth perception measurements, we propose a 4D light field rendering method with importance sampling and a sparse reconstruction scheme, with reduced computation cost. The minimum 4D rendering supports both foveation and accommodation.
- We have built a hardware prototype for foveated light field display from commodity components including a gaze tracker, and a GPU-based light field rendering engine that runs in real time . Our prototype hardware + software system achieves better performance and quality than alternative methods, as verified through different scenes and user studies with multiple participants.

2 PREVIOUS WORK

A comfortable and immersive 3D experience requires displays with high quality, low latency, and consistent depth cues.

Depth perception and light field display. Understanding and navigating 3D environments require accurate depth cues, which arise from multiple mechanisms including motion parallax, binocular vergence, and focus accommodation [Patney et al. 2017]. Conventional 2D desktop and stereoscopic displays lack proper focus cues and can cause vergence-accommodation conflict [Akeley et al. 2004]. Although light field displays can support proper focal cue by 4D light rays [Huang et al. 2015; Lanman and Luebke 2013; Wetzstein et al. 2011, 2012], they are considerably more costly to render or acquire than 2D images. Thus they often lack sufficient speed or resolution for fully immersive VR applications which are sensitive to simulator sickness. Despite prior physiological studies in retinal blur and cell distributions [Watson 2014; Watson and Ahumada 2011], it remains an open problem to build a perceptually accurate and quantitative model for fast content synthesis for light field displays. This project aims to address this challenge and answer the fundamental question: how should we sample a 4D light field to support focal cues with minimum cost and maximum quality?

Foveated rendering. The human visual system has much denser receptors (cones) and neurons (midget ganglion cells) near the fovea than the periphery. Foveated rendering harnesses this property to reduce computation cost without perceptual quality degradation in desktop displays [Guenter et al. 2012] and VR HMDs [Patney et al. 2016]. The potential benefits of foveation for path tracing is surveyed in [Koskela et al. 2016]. However, foveation has not been explored in higher dimensional displays, such as for 4D light fields.

This paper explores sampling/reconstruction and hardware requirements to foveate 4D displays with perceptual preservation.

Light-field sampling. Light field analysis in the spectral [Chai et al. 2000; Levin et al. 2009; Ng 2005; Ramachandra et al. 2011] or ray-space [Gortler et al. 1996; Levoy and Hanrahan 1996] domain improves quality and performance of rendering [Egan et al. 2011a,b, 2009; Hachisuka et al. 2008; Lehtinen et al. 2011; Yan et al. 2015] and acquisition [Dansereau et al. 2017; Iseringhausen et al. 2017; Ng 2005; Wei et al. 2015; Wender et al. 2015].

Prior work on light field rendering and reconstruction [Hachisuka et al. 2008; Lehtinen et al. 2011, 2012] focuses on the projected 2D images with distributed effects, e.g., depth of field [Yan et al. 2015], motion blur [Egan et al. 2009], and soft shadows [Egan et al. 2011b; Yan et al. 2015]. However, foveating light field displays needs sparsely sampled 4D rays with sufficient fidelity for the observer to accommodate the scene content and integrate the retinal image.

Using gaze tracking, we augment traditional 4D light field sampling and rendering with two main components: visual foveation and accommodation. The former guides sampling to the retinal cells distribution; the latter allows adaptation to the scene content.

3 OVERVIEW

To understand the visual factors, we perform perceptual studies with both optical blur and our light field display prototype [Kim et al. 2017]. Driven by the study discoveries, we further analyze the whole light field system, including the display, the eye lens, and the eye retina, in both the primary and frequency domains in Section 4. Based on this perceptual model, we describe our 4D sampling and reconstruction methodology for foveated light field rendering in Section 5, and implementation details including hardware prototype and software system in Section 6. We validate our system via psychophysical studies and performance analysis in Section 7.

4 ANALYSIS: FREQUENCY BOUNDS

Light field displays require dense sampling from multiple viewpoints, which are orders of magnitude more expensive to render than traditional displays. Sheared filters with spatial-angular frequency bounds save samples for global illumination [Egan et al. 2011a,b, 2009; Yan et al. 2015]. However, image reconstruction from a 4D light field display is automatic through and further bounded by human eyes. Thus, we derive spatial-angular frequency bounds in the realms of display, lens, and retina. The outcome of this analysis and the subsequent sampling strategy (Section 5.1) also answer the long standing question on the minimum number of rays required to support accommodation with a light field display.

In the ray space, we model the perceived retinal image $I(\mathbf{x})$ (Figure 2a) as an angular integration of the retinal light field $L(\mathbf{x}, \mathbf{u})$

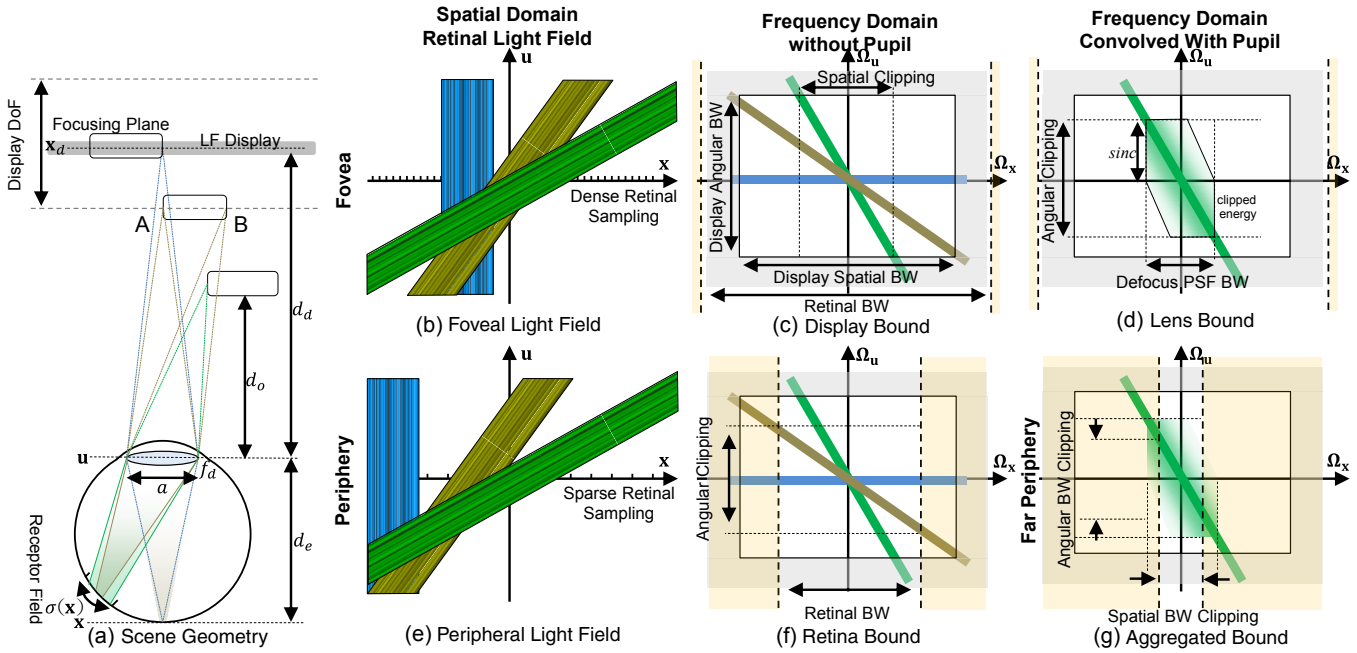


Fig. 2. *Light-field analysis in ray space and frequency domain.* The setup (a) of the eye focusing on the display has a foveal and a peripheral light fields shown in (b) and (e), and their frequency domain spectrum in (c) and (f) respectively. The perceivable light field is subject to spatial clipping due to the display bound (c) shown in retinal coordinates, angular clipping due to the lens bound (d), and spatial and angular clipping due to the retina bound (f). The final perceivable spectrum is obtained by aggregating all bounds (g): the narrower spatial retinal bound not only reduces the spatial bandwidth, but it also further lower the angular bandwidth from (d).

(Figure 2b) across the pupil $\sqcap(\mathbf{u}/a)$. The corresponding frequency spectrum (Figure 2c, colored lines) is then obtained through Fourier slice theorem:

$$\begin{aligned} I(\mathbf{x}) &= \int L(\mathbf{x},\mathbf{u}) \sqcap (\mathbf{u}/a) \mathrm{d}\mathbf{u} \\ \hat{I}(\omega_{\mathbf{x}}) &= \left(\hat{L} \star \hat{\sqcap}\right)(\omega_{\mathbf{x}}, \omega_{\mathbf{u}} = 0) \end{aligned}, \tag{1}$$

where $\hat{\cdot}$ denotes Fourier transform and $\star$ denotes convolution. When the eye has focal length f and diameter d_e, the frequency domain slope of any out-of-focus object at depth d_o is

$$\frac{\omega_{\mathbf{u}}}{\omega_{\mathbf{x}}} \triangleq \hat{k}(d_o, f) = -d_e \left(\frac{1}{d_e} + \frac{1}{d_o} - \frac{1}{f}\right). \tag{2}$$

We approximate the spherical eyeball via a 2-plane parameterization, which suffices in many cases as the fovea is only within 5 degree and the periphery is blurred. A spherical parameterization [Dansereau et al. 2017] will be more accurate to model the retinal geometry and other phenomena, e.g. Stiles-Crawford effect. Detailed derivations of Equations (1) and (2) and ray space analysis are shown in [Huang et al. 2014] and Appendix A. Note that the slope $\hat{k}$ is linearly proportional to objects' diopter depths because both are inverses of metric depths.

Retina bound. The spatial resolution of retina decreases with larger eccentricity primarily because the midget Retinal Ganglion Cell receptor field (mRGCf) increases dendritic field size [Curcio and Allen 1990] while maintaining a constant area sampling rate [Drasdo et al. 2007]. This inspires recent work [Guenter et al. 2012; Patney et al. 2016] in reducing the rendering cost via foveation. The visual acuity falls monotonically as the visual eccentricity grows, and the fall-off is known to follow the density of ganglion cells [Thibos et al. 1987]. Watson [2014] combined results from several studies to construct a model that predicts the receptive field density of midget ganglion cells as a function of retinal eccentricity $r = \sqrt{x^2 + y^2}$, for $(x, y) \in \mathbf{x}$ and the meridian type m:

$$\begin{aligned} \rho(r, m) = 2 \times \rho_{cone} &\left(1 + \frac{r}{41.03}\right)^{-1} \\ &\times \left[a_m \left(1 + \frac{r}{r_{2,m}}\right)^{-2} + (1 - a_m) \exp\left(-\frac{r}{r_{e,m}}\right)\right], \end{aligned} \tag{3}$$

where $\rho_{cone} = 14,804.6 \deg^{-2}$ is the density of cone cell at fovea and $a_m, r_{2,m}, r_{e,m}$ are all fitting constants along the four meridians of the visual field; details can be found in [Watson 2014]. Figures 5a and 5b visualize the densities. In practice, we use the spacing

$$\sigma(\mathbf{x}) = \sigma(x, y) = \frac{1}{r}\sqrt{\frac{2}{\sqrt{3}}\left(\frac{x^2}{\rho(r,1)} + \frac{y^2}{\rho(r,2)}\right)} \tag{4}$$

to derive the retinal spatial bandwidth:

$$B^{retina}_{\omega_{\mathbf{x}}}(\mathbf{x}) = 1/(2\sigma(\mathbf{x})). \tag{5}$$

Figures 5c and 5d show corresponding sampling based on this bandwidth bound only. The corresponding angular bandwidth is obtained from the definition of $\hat{k}$ in Equation (2):

$$B^{retina}_{\omega_{\mathbf{u}}}(\mathbf{x}) = \hat{k}(d_o, f) B^{retina}_{\omega_{\mathbf{x}}}(\mathbf{x}). \tag{6}$$

The angular bound depends on both content depth and gaze eccentricity. The example in Figure 2f shows different angular bounds for objects at the same eccentricity.

Lens bound. For an out-of-focus object, its perceivable frequency spectrum is governed by the energy contributed to the slicing axis $\omega_{\mathbf{u}} = 0$ in Equation (1) through convolution with the Fourier transformed pupil function $\hat{\sqcap}(\mathbf{u}/a) = \text{sinc}(a\omega_{\mathbf{u}})$. The bounds are primarily limited by the pupil aperture a, and because sinc$(\cdot)$ degrades rapidly after its first half cycle π, as shown in Figure 2d, we can derive the angular bandwidth $B^{lens}_{\omega_{\mathbf{u}}} = \pi/a$, and the corresponding spatial bandwidth is given by:

$$B^{lens}_{\omega_{\mathbf{x}}} = \begin{cases} \dfrac{\pi}{a\hat{k}(d_o, f)}, & \text{if } a > \dfrac{2\pi d_e \Delta\mathbf{x}_d}{\hat{k}(d_o, f) d_d} \\ \\ \dfrac{d_d}{2d_e\Delta\mathbf{x}_d}, & \text{otherwise,} \end{cases} \tag{7}$$

where $\frac{d_e}{d_d}\Delta\mathbf{x}_d$ is the spatial sampling period of the light field display projected onto the retina, and it caps the spatial bandwidth by $1/\left(2\frac{d_e}{d_d}\Delta\mathbf{x}_d\right) = \frac{d_d}{2d_e\Delta\mathbf{x}_d}$ (the *otherwise* clause). The *if* clause has further reduced bound due to the object slope $\hat{k}(d_o, f)$.

Display bound. Let $\Delta\mathbf{x}_d$ and $\Delta\mathbf{u}_d$ be the spatial and angular sampling periods of the display. With its angular bound $B^{display}_{\omega_{\mathbf{u}}} = 1/(2\Delta\mathbf{u}_d)$, Zwicker et al. [2006] have shown a spatial bound $B^{display}_{\omega_{\mathbf{x}}}$ when an object's depth extends outside the depth of field of the display (Figure 2c); details are described in Appendix B.

Overall bound. The aforementioned bounds are aggregated into the smallest bandwidth among them:

$$B^{all}_{\{\omega_{\mathbf{x}},\omega_{\mathbf{u}}\}}(\mathbf{x}) = \min\left(B^{retina}_{\{\omega_{\mathbf{x}},\omega_{\mathbf{u}}\}}, B^{lens}_{\{\omega_{\mathbf{x}},\omega_{\mathbf{u}}\}}, B^{display}_{\{\omega_{\mathbf{x}},\omega_{\mathbf{u}}\}}\right)(\mathbf{x}), \tag{8}$$

An example is shown in Figures 2a and 2g.

How many rays do we need? It has been asked for a decade that how many rays entering the pupil, i.e. the angular sampling rate, are needed for a light field display to support proper focus cue. As we have studied and derived, the display, the optics of the eye, and the anatomy of the retina all affect the final perceivable image. Based on the discoveries, we present a closed-form and spatially-varying ray sampling strategy in Section 5.

5 METHOD: SAMPLING AND RENDERING

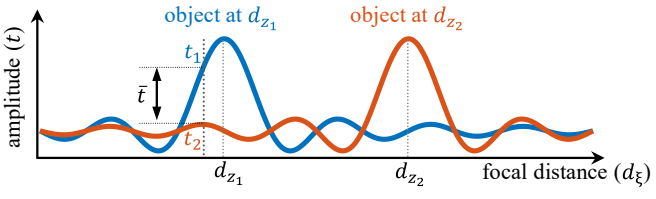


Fig. 3. *Sampling strategies illustration.* X-axis represents the accommodative depth d_ζ. Y-axis shows the amplitude t from Equation (10). Varying objects depths demonstrate different amplitude distribution w.r.t d_ζ. The differential amplitude $\bar{t}$ in Equation (11) is the distance between intersections.

The bandwidth bounds in Section 4 include optical and retinal components. However, variations in scene depth content [Kim et al. 2017], the eye's focus and movement ([Charman and Tucker 1978; Watson and Ahumada 2011]), and occlusions [Zannoli et al. 2016] also decide our depth perception. Considering those additional factors, we extend the bounds in Equation (8) for an importance-based model for sampling and rendering. As illustrated in Figure 3, we consider the perceived amplitude difference among objects ($\bar{t}$) as the depth stimulus strength. Based on this, we derive an importance value W for each light ray $(\mathbf{x}, \mathbf{u})$ with regard to the static range and dynamic movements of accommodative depth d_ζ. This importance distributes the ray budget for the final shading and filtering.

5.1 Content-Adaptive Light Field Sampling

To formally analyze the increased importance due to occlusion, consider two objects at distances d_{z_1} and d_{z_2} to the eye and are visible within a small window centered on a light ray $(\mathbf{x}, \mathbf{u})$. In the frequency domain, their retinal light field spectra have slopes $\hat{k}(d_{z_1}, f_\zeta)$ and $\hat{k}(d_{z_2}, f_\zeta)$ (Equation (2)) with a time-varying focal length of the eye f_ζ. When they are out-of-focus, their perceivable bandwidth with respect to the focus distance[1]

$$d_\zeta = \left(\frac{1}{f_\zeta} - \frac{1}{d_e}\right)^{-1} = \frac{f_\zeta d_e}{d_e - f_\zeta} \tag{9}$$

to the eye is equal to the contribution of amplitude spreading toward the slicing axis $\omega_{\mathbf{u}} = 0$, and is given by

$$t(d_{z_i}, d_\zeta, \omega_{\mathbf{x}}) = \left\|\hat{s}_i\left(-\frac{d_e}{d_{z_i}}\omega_{\mathbf{x}}\right)\right\| \text{sinc}\left(a\omega_{\mathbf{x}}\hat{k}\left(d_{z_i}, f_\zeta\right)\right), \tag{10}$$

where $\|\hat{s}\|$ is the amplitude of the surface texture in the frequency domain. Please refer to [Huang et al. 2014] and Appendix F for detailed derivations. In monocular vision, the eye perceives depths through the differences in the defocus blur. Thus, given the constant focusing distance d_ζ, we consider their differences in the perceivable signal amplitudes:

$$\bar{t}(d_{z_1}, d_{z_2}, d_\zeta, \omega_{\mathbf{x}}) = \left\|t(d_{z_1}, d_\zeta, \omega_{\mathbf{x}}) - t(d_{z_2}, d_\zeta, \omega_{\mathbf{x}})\right\|. \tag{11}$$

Static sampling. Following our blur and depth perception studies [Kim et al. 2017], and the display-eye bandwidth discussions (Section 4), Equation (11) presents an analytical modelling for defocus blur with a constant focusing distance and two objects, as visualized in Figure 3. We consider all the visible objects within a ray and compute the corresponding importance indicator for sampling:

$$\begin{aligned} w_s(d_\zeta) &= \sum_{\forall i,j \in objects}^{i\neq j} \int_{\Omega_x} \bar{t}\left(d_{z_i}, d_{z_j}, d_\zeta, \omega_{\mathbf{x}}\right) \mathrm{d}\omega_{\mathbf{x}} \\ &\propto \int_{\Omega_x} \bar{t}\left(d_z^-, d_z^+, d_\zeta, \omega_{\mathbf{x}}\right) \mathrm{d}\omega_{\mathbf{x}}, \end{aligned} \tag{12}$$

where $[d_z^- = \min_{\forall i} d_{z_i}, d_z^+ = \max_{\forall i} d_{z_i}]$ is the scene's local depth range around the ray. The above formulation requires the knowledge of focal distance d_ζ, which is not directly available due to lack of accommodation tracking technologies. We address this limitation by integrating d_ζ over the estimated accommodation range $[d_\zeta^-, d_\zeta^+]$

[1] d_ζ is focal distance, f_ζ is focal length, as illustrated in Figure 4.

for the final importance estimation in Equation (14). The real-time acquisition of $d_\zeta^\pm$ and $d_z^\pm$ are described in Section 6.

Dynamic sampling. The static weighting above considers a fixed d_ζ. However, accommodation can also be guided by the modulation of retinal images as the eye changes its focal distance (e.g. through micro fluctuation [Charman and Tucker 1978]). These motivate us to consider a dynamic factor that reflects a changing d_ζ:

$$w_d(d_\zeta) = \int_{\Omega_x} \frac{\partial \bar{t}\left(d_z^-, d_z^+, d_\zeta, \omega_\mathbf{x}\right)}{\partial d_\zeta} \mathrm{d}\omega_\mathbf{x}. \tag{13}$$

Figure 4 shows the matching trend between normalized $w_d(d_\zeta)$ and prior vision science discovery from Watson and Ahumada [2011] that the strongest blur discrimination occurs when the accommodation depth (d_ζ) lies slightly off-center to object depths ($d_z^\pm$).

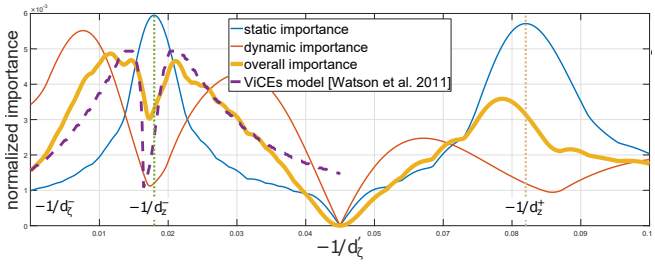


Fig. 4. *Importance values and the model from [Watson and Ahumada 2011].* The three solid curves plot normalized values of Equations (12) to (14) in transformed coordinate (Appendix C). The dashed curve shows the trend of depth perception of the object at depth $d_z^- = 4D$ from ViCEs prediction model [Watson and Ahumada 2011] by assuming its inversed detectable threshold to be the importance. The x-axis represents different accommodation d'_ζ within the range of d_ζ^- and object at depth d_z^+. Because the ViCEs model considers only one of those two objects due to symmetry, its plot has the x-axis range between d_ζ^- and $\frac{d_z^- + d_z^+}{2}$. Coordinates of d'_ζ are transformed as $\frac{-1}{d'_\zeta}$ for easier visualization. Symbols are illustrated in Figure 6.

Overall sampling. Combining the above stimuli strengths modeled with scene content and accommodation preference, we have the importance $w_d(d_\zeta)w_s(d_\zeta)$ for a specific focal distance d_ζ. To fully construct the importance for a light ray $(\mathbf{x}, \mathbf{u})$, we consider its effective local amplitude differences by integrating over the focal distance range $[d_\zeta^-, d_\zeta^+]$. We estimate this range as the min-max depths in fovea since people usually observe and focus on objects within this area. To further accelerate the calculation, we transform each integration to a uniform coordinate frame (via the operator η below):

$$\begin{aligned} W(\mathbf{x}, \mathbf{u}) &= \int_{d_\zeta^-}^{d_\zeta^+} w_d(d_\zeta) w_s(d_\zeta) \mathrm{d}d_\zeta \\ &\overset{\eta}{=} \int\int w'_d\left(\frac{\omega'_\mathbf{u}}{\omega'_\mathbf{x}}\right) w'_s\left(\omega'_\mathbf{x}, \omega'_\mathbf{u}\right) \mathrm{d}\omega'_\mathbf{x} \mathrm{d}\omega'_\mathbf{u}, \end{aligned} \tag{14}$$

where $(\omega'_\mathbf{x}, \omega'_\mathbf{u}) = \eta(d_\zeta, \omega_\mathbf{x}, \omega_\mathbf{u})$ is the transformed frequency coordinate, and $\{w'_s, w'_d\}$ are the pointwise importance functions in the new frame; details are derived and discussed in Appendix C. The

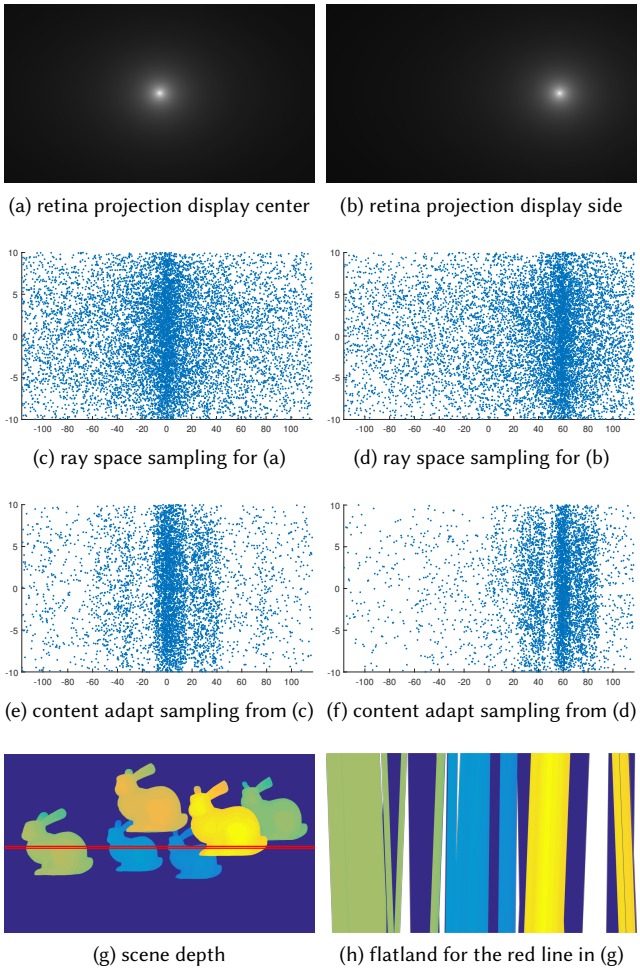


Fig. 5. *Spatial-angular content adaptive sampling.* (a) and (b) show the retinal ganglion density (Equation (3)) projected on the display when the gaze is at the center or side of the display. (c) and (d) show the corresponding ray space sampling for (a) and (b). Based on (c) and (d), (e) and (f) further adapt to the content shown in (g) and (h). The flatland visualizations in (c), (d), (e), (f), and (h) are in the display space with *mm* as units in both axes.

integrating ranges in Equation (14) are bounded by the frequency bandwidth $B^{all}_{\{\omega_\mathbf{x}, \omega_\mathbf{u}\}}$ in Equation (8), and the range of focal length and distance:

$$\begin{aligned} (\omega_\mathbf{x}, \omega_\mathbf{u}) &\in [-B^{all}_{\omega_\mathbf{x}}(\mathbf{x}), B^{all}_{\omega_\mathbf{x}}(\mathbf{x})] \times [-B^{all}_{\omega_\mathbf{u}}(\mathbf{x}), B^{all}_{\omega_\mathbf{u}}(\mathbf{x})] \\ \frac{\omega_\mathbf{u}}{\omega_\mathbf{x}} &\in [\hat{k}(d_\zeta^-, f_\zeta^-), \hat{k}(d_\zeta^+, f_\zeta^-)]. \end{aligned} \tag{15}$$

This analytical importance function can be computed in closed form to allow real-time performance, as is shown in Appendix F. It guides spatially-varying and perceptually-matching ray allocations given a specified rendering budget. As visualized in Figures 3 and 4, our min-max estimation will only increase the numbers of samples, thus being more conservative. In Appendix D, we also present the minimum budget required given a display-viewer setup.

5.2 Sparse Sampling and Filtering for Rendering

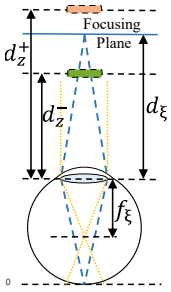


Fig. 6. *Symbols for Figure 4.*

We perform a two-stage GPU-based **sampling** to realize the importance model above, as visualized in Figure 5. To compute preliminary saving (Figures 5c and 5d) without expensive global Fourier transform, we first estimate each local ray region's maximum sample number s_{el} (Appendix D) by distributing the total budget with retina bounds $B^{retina}_{\{\omega_\mathbf{x},\omega_\mathbf{u}\}}(\mathbf{x})$ to consider eccentricity effect. We then compute, for each ray, its aggregate bounds $B^{all}_{\{\omega_\mathbf{x},\omega_\mathbf{u}\}}$ (Equation (8)) to delineate the domain (Equation (15)) for the importance value $W(\mathbf{x},\mathbf{u})$ in Equation (14). We multiply s_{el} with W/ξ to finalize the sample count for each ray (Figures 5e and 5f). ξ is a global ratio to rescale W into $[0,1]$, with $\xi = 320$ based on our specific hardware setup and experiments to balance between performance and perceptual quality. ξ can be further increased for stronger savings, but more thorough evaluation may be needed. To avoid zero samples for flat regions, we clamp the ratio W/ξ to be within $[0.3,1]$. The min clamping value 0.3 can be further reduced with higher resolution displays (e.g., 4K instead of 2K).

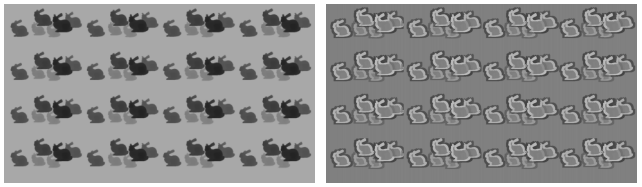
The sparsely sampled ray set is **filtered** for rendering a light field display with uniformly spaced pixels. We implement a separable 4D Gaussian radial basis function for the sparse reconstruction and handle occlusions using the coarse depth map (Figure 7); details are shown in Appendix E. Finally, similar to [Patney et al. 2016], a contrast-preserving filter is applied to improve quality.

6 IMPLEMENTATION

Depth disparity estimation. In each frame we render a multi-view low spatial resolution (500×300) depth mipmap, as shown in Figure 7a, to estimate the local depth variations. Specifically, depending on the specific scene complexity, we render no more than 4×4 depth maps using simultaneous multi-viewport projection supported by modern GPUs. From this multi-view depth mipmap, we find the local minimum and maximum depth for each coarse pixel by performing a mix-max comparison around the local neighborhood and pyramid layers, as show in Figure 7b. Combining the two maps using bilinear interpolation, we obtain the values of $d_\zeta^\pm$ and $d_z^\pm$ to compute Equation (14) for any ray $(\mathbf{x},\mathbf{u})$.

Ray-tracing. We implement our system using the NVIDIA OptiX ray tracer. For comparison, we also implement two full-resolution light field rendering techniques by ray tracing [Lanman and Luebke 2013] and rasterization [Huang et al. 2015].

The foveated rendering pipeline requires asynchronous computation of importance sampling. So, we separate the rendering into two stages similar to the decoupled shading [Ragan-Kelley et al. 2011]: we first create a queue of rays to be shaded, and then use the scheduler to processes the shading. Similar to the foveated rasterization [Patney et al. 2016], we also suffer performance penalty without dedicated hardware scheduler which supports coarse pixel shading.



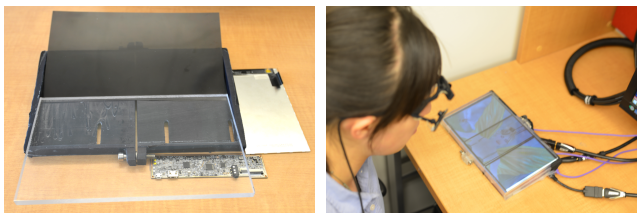
(a) low res zbuffer (b) low res analysis

Fig. 7. *Depth disparity estimation of local regions.* (a): Depth buffer from multiview projection. (b): Real-time depth disparity analysis of local regions; with brighter colors representing larger disparities.

However, our method still shows performance gains in both frame rates and number of shaded rays; see Figure 11.

Hardware. To validate the foveated light field rendering, the prototype hardware needs to offer a high spatial/angular resolution, a wide depth of field , and a wide field of view to separate foveal and peripheral regions. We build a parallax-barrier based light field display by tiling three 5.98-inch 2560×1440 panels (part number TF60006A) from Topfoison. The parallax-barrier at $9.5mm$ from the panels is printed with $300\mu m$ pitch size using a laser photoplotter; its pinhole aperture is $120\mu m$ to avoid diffraction. The final light field display has 579×333 hardware spatial resolution at 10-inch diagonal size and 8×8 views angular resolution (3.2 views/degree), larger than the 5×5 angular resolution in [Huang et al. 2015] which can already support proper accommodation. The components and the interfaces are shown in Figure 8. Assuming an eye with $6mm$ pupil aperture viewing the display from $30cm$ away, we ensure 10 rays/pixel entering the eye to support accommodation. The renderer is driven by a PC with an 2.0GHz 8-core CPU with 56GB of RAM, and an NVIDIA GTX 1080 graphics card. Example elemental image using our foveated ray tracing (Section 5) and the display hardware can be found in Figure 9.

We augment the light field display with a PupilLab [Kassner et al. 2014] eye tracker. The head-mounted tracker offers real-time streaming of gaze positions in the display space. We drive the tracker with a laptop. The foveal accommodation range $[d_\zeta^-, d_\zeta^+]$ in Equation (15) are obtained by combining the eye-tracked gaze position and a ray propagation from the eye to the gaze.



(a) components (b) system setup

Fig. 8. *Our hardware design and system setup.* (a) shows components to build our light field display in Section 6. (b) shows our system setup: a user wearing glass-style eye tracker watches the display.

Fig. 9. *A foveated light field elemental image from the framebuffer of our prototype display.* The 4D light field is generated through propagating rays from each pixel to its corresponding pinhole. The tracked gaze position is at the face of the fairy. Please zoom-in for details.

7 EVALUATION

For perceptual and performance evaluation, we choose 11 scenes with different depth distribution, geometric complexity, and field of view. Figures 1 and 11 show simulated renderings while Figure 10 shows captured results; detailed scene information is in Appendix G.

7.1 Perceptual quality

We conducted a user study to evaluate the effectiveness and quality of our method, by comparing with full-resolution and uniformly down-sampled light fields with the same number of rays as our method. Our goal is to show that foveated light fields achieve the quality of former with the performance of the latter.

Setup. The experimental setup consisted of our prototype light field display, a head-mounted eye tracker [Kassner et al. 2014], and machines (Section 6) that rendered and drove the system. We used a $12mm \times 12mm$ eye box at $0.3m$ from the display.

Stimulus. The stimulus was the fairy scene. Objects contain both high and low spatial frequency textures. The light field of the stimulus was generated using one of the three methods: full resolution, foveated, and uniformly downsampled. The full resolution condition sampled all the rays represented by the hardware (579×333 spatial and 6×6 angular given the eyebox size). Foveated condition used our framework in Section 5, resulting in 24.8% samples (Table 2) compared with full resolution. Uniformly downsampled condition had the same number of rays as the foveated one but uniformly distributed the samples across retina.

Task. Subjects examined and memorized details of the full resolution stimulus before the beginning of the experiment. During each trial, the display presented a stimulus rendered using one of the three methods for 4 seconds. Subjects were instructed to gaze at the fairy's head to avoid big saccades (fast and ballistic eye movements) and choose on keyboard about whether the stimulus looked the same as the examined full resolution stimulus. The entire experiment consisted of 42 trials, 14 per each rendering method. The order of all trials was randomized. Similar to previous studies on foveated effects ([Patney et al. 2016; Wallis et al. 2016]), we inserted blank frames between trials. 14 subjects participated in the experiment (4 females and 10 males, aged 27 to 35). All subjects had normal or corrected-to-normal visual acuity. None of the subjects were aware of the experimental hypothesis or number of rendering methods.

case \ user	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
full resolution	12	14	6	12	14	13	14	12	14	13	7	14	14	13
foveated	14	14	6	7	13	13	14	7	10	14	9	14	14	12
uniform	4	5	0	0	2	0	0	0	0	8	0	0	0	4

Table 1. *User study results.* The values are number of trials (out of 14) where subjects did not notice artifacts. Some subjects reported visible artifacts even in full-resolution condition, reflecting individual differences in criteria. The difference in perceived image quality was significant between full-resolution vs. uniform and foveated vs. uniform ($p < 0.0001$), but not significant between full-resolution vs. foveated ($p = 0.67$).

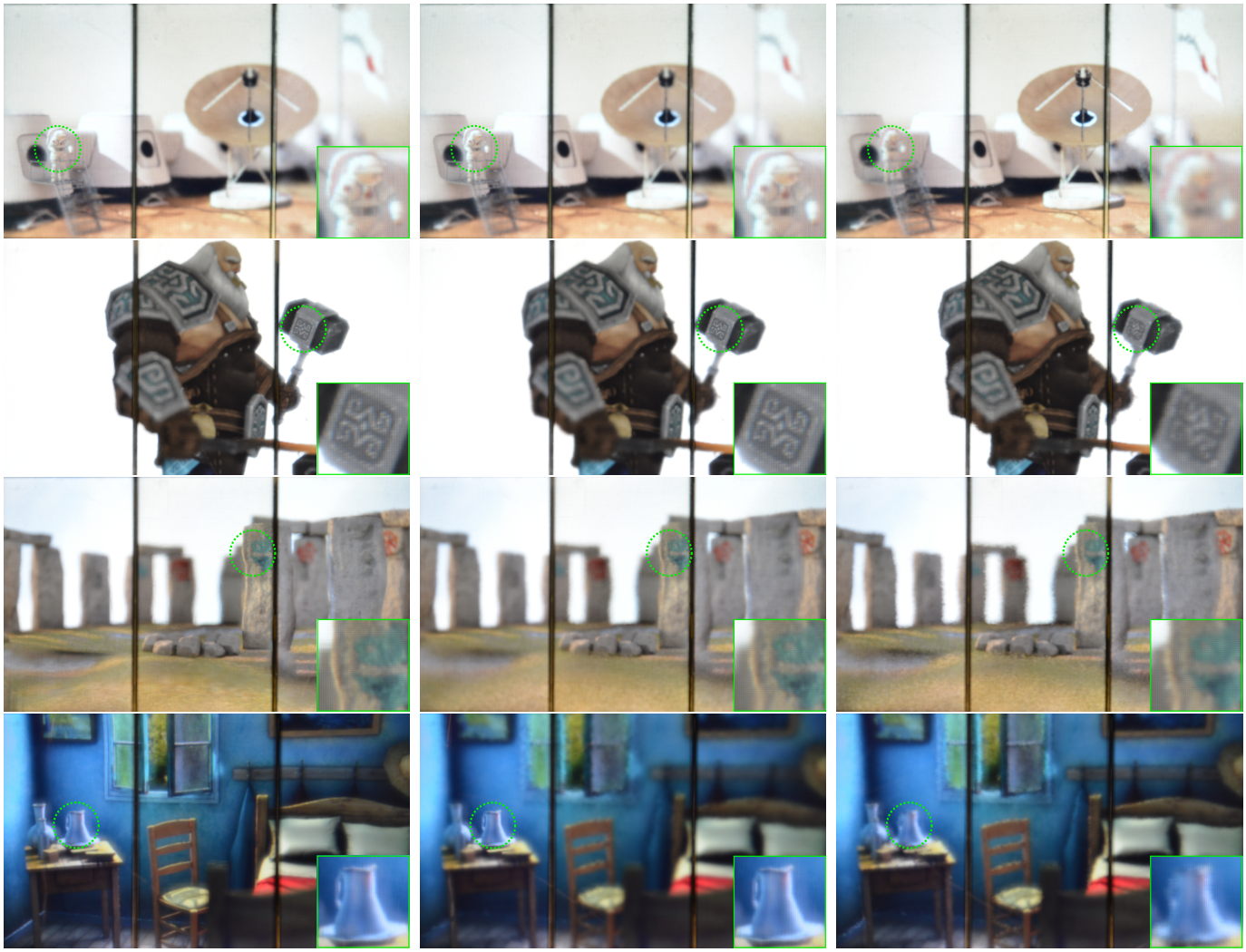
Result. Table 1 shows the number of trials where subjects reported that the stimulus looked the same as full resolution. A one-way within-subjects ANOVA showed that the effect of rendering method is significant ($F_{(2,26)} = 121.1, p < 0.0001$). Note that the difference in perceived image quality was significant between full-resolution vs. uniform and foveated vs. uniform ($p < 0.0001$, paired t-test with Bonferroni correction), but not foveated vs. full-resolution ($p = 0.67$). The experimental results demonstrate that our framework lowers sampling rate without degrading perceived image quality. Figures 1 and 10 show more quality comparisons. Please refer to our supplementary video for live capture of a user interacting with our prototype display.

7.2 Performance

plane 16.42%	fairy 24.80%	Mars 27.20%	Sponza 29.38%	toaster 25.78%	farm 24.69%
craftsman 24.67%	marbles 28.6%	Stonehenge 21.59%	van Gogh 18.57%	Viking 24.59%	chess 26.96%

Table 2. *Ratio of number of traced rays in foveaton relative to full resolution.*

Table 2 shows the ratio of the minimal number of traced light field rays with foveation (as computed in Appendix D) compared with full resolution rendering. Since our method is content-adaptive, the saving in sampling and ray tracing is related to the scene complexities. One extreme scene is a flat plane, in which the ratio is only 16.42%. Our most challenging case is Crytek Sponza containing large variation in depth along the wall; the ratio increases to 29.38%, but the overall time performance is still 2× faster than that in [Huang et al. 2015], as shown in Figure 11. Compared to the recent 2D foveated rendering method [Patney et al. 2016], our 4D light field foveation saves more pixel computation (up to 80%+ vs. up to 70%). Note that the method in [Patney et al. 2016] is constrained by GPU design thus only offer theoretical saving rather than actual performance (frame rates) benefit. Our system demonstrates actual performance gain with modern GPUs.



(a) full resolution (b) our foveated display (c) uniform down-sampling

Fig. 10. *Photograph results from our prototype tiled display with 3 panels.* Our foveated results in (b) have similar quality to full-resolution rendering in (a), and higher quality than uniform sampling with the same number of rays in (c). Because uniform sampling does not consider either retinal receptor distribution or scene content, it introduces blur in fovea and aliasing near occlusion boundaries. The tracked gaze positions are marked in green circles with insets for zoom-in. All captured results are from our prototype (gamma correction enabled) in Figure 8 by a Nikon D800 DSLR camera with a 16-35mm f/4G lens. Corresponding retinal image simulations are available in the supplementary material. From top to bottom: Mars, craftsman, Stonehenge, van Gogh.

8 LIMITATIONS AND FUTURE WORK

Real-time foveated light fields involve multiple disciplines: display, rendering, content analysis, and human perception. Each component contains challenging open problems. We have proposed a starting point for this broad topic in which industry and consumers are gaining significant interests. Our current method and implementation still depend on the perceptual diversities of the observers [Kim et al. 2017], the precisions of trackers, and the capabilities of the GPUs.

Perception. Our psychophysical data and perceptual model can benefit general foveated rendering goals focusing on accommodative depth perception, but other individual factors, including stereoscopic depth [Siderov and Harwerth 1995], high-order refractive aberrations, pupil size, eye dominance, prismatic deficiencies, contrast/color sensitivities, etc., may also influence light field perception. Thus, the saving can be conservative by using the bounds from the anatomical structure. Fully immersive VR/AR applications may require identification of thresholds at eccentricities wider than the 15 deg in our perceptual experiments. These factors are worth study

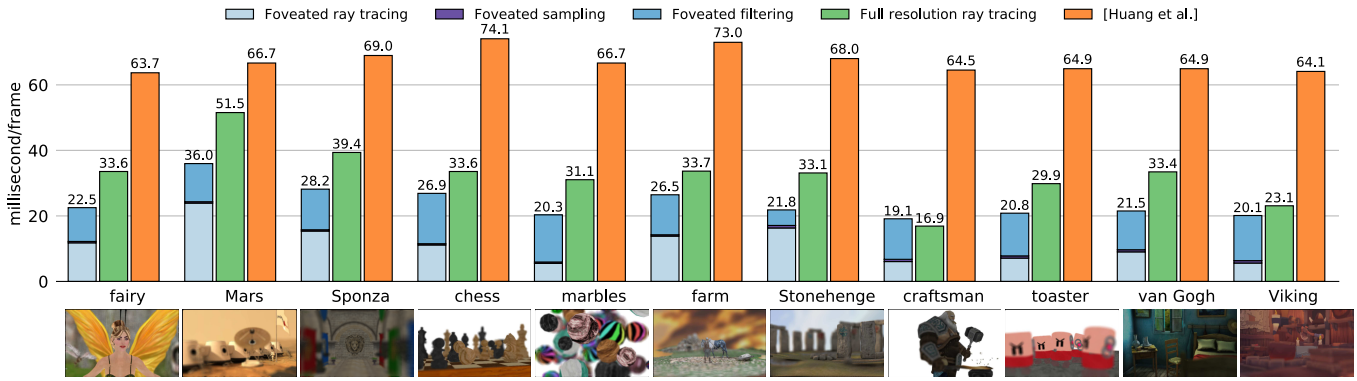


Fig. 11. *Performance comparison and breakdown.* Performance comparison with full resolution ray tracing [Lanman and Luebke 2013] and rasterization [Huang et al. 2015]. Y-axis is the time consumption per frame measured in million-seconds. We also break down the timing for our method into the main components: sampling, ray tracing, and post-filtering. By sampling much less rays (Table 2), our method demonstrates lower overall computation costs, in particular the ray tracing part compared with full resolution ray tracing. Scene courtesies of Ingo Wald, admone, Crytek, Olexandr Zymohliad, Andrew Kensler, Raúl Balsera Moraño, ruslans3d, olmopotums, Andrew Kensler, rusland3d and nigelgoh respectively.

as potential future works but beyond a single paper which first explores foveated light fields.

Tracking. In [Kim et al. 2017], we discouraged users from making big saccades, but saccadic movement is known to help improve depth perception. While our entire system latency (tracker-renderer-display) is shorter than the accommodative reaction time, it is still longer than saccade-proof ($< 60ms$ [Loschky and Wolverton 2007]). Enlarging foveal area balances the system latency, but it affects the accuracy of the psychophysical data which derives and validates our methods. However, we believe the development of fast eye tracking and rendering hardware can help future foveated displays.

GPUs. Rendering light field using ray-tracing might not be the optimal because modern GPUs are originally designed for rasterization. For the latter, further performance improvement can be achieved with future hardware supporting content adaptive shading [Vaidyanathan et al. 2014]. Our current implementation adds overhead on the post-filtering process (Figure 11), but similar to [Heide et al. 2013], integrating the rendering to a compressive display hardware could deliver better performance and image quality.

Scene. Although we have analyzed the bandwidth bounds for Lambertian objects, highly specular 4D surfaces, (semi)transparent objects and high-frequent objects, need further examination on the extended area by the BRDF/BTDF bounds. The occlusion effect is not analyzed in our frequency analysis, so we can only address them in the spatial domain through importance sampling; insight from sheared filter in light transport [Mehta et al. 2012] may contribute to this area. Our analysis and implementation do not consider the temporal dimension: sampling for temporal anti-aliasing across the retina ([Tyler 1987]) is a potential future direction.

9 CONCLUSION

Light field displays resolve the vergence-accommodation conflict that causes eye-strain and double vision, and improve 3D perception even for monocular vision. However, 4D light fields incur heavier rendering workload than 2D images. Inspired by the vision of Egan [1994], we address this challenge by conducting content-aware physiological studies, deriving a perceptual model, and designing a real-time foveated 4D light field rendering and display system. Our prototype system offers both theoretical and actual performance gain with current GPUs (Section 7.2) and preserves perceptual quality when the visual system automatically reconstructs retinal images (Section 7.1).

Across the retinal eccentricity, going from the anatomical receptor distribution, spatial acuity, blur sensitivity, to the depth perception, is a long path. Each individual connection is a long standing research topic in the community. By analyzing the entire optical process from display to retina, our method guides an optimized allocation strategy given hardware budget and user input. It also suggests the minimum sampling required to provide proper accommodation.

For the future, we envision 3D display technologies such as digital hologram for near eye display or vari-/multi-focal display can also benefit from foveated light fields.

ACKNOWLEDGMENTS

We would like to thank Ia-Ju Chiang and Suwen Zhu for helping us conducting the experiments; Anjul Patney, Kaan Akşit, Piotr Didyk, Chris Wyman, and the anonymous reviewers for their valuable suggestions. This work has been partially supported by National Science Foundation grants IIP1069147, CNS1302246, NRT1633299, CNS1650499, and Hong Kong RGC general research fund 17202415.

REFERENCES

Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. *ACM Trans. Graph.* 23, 3 (2004), 804–813.

Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. 2000. Plenoptic Sampling. In *SIGGRAPH '00*. 307–318.

WN Charman and J Tucker. 1978. Accommodation as a function of object form. *Optometry & Vision Science* 55, 2 (1978), 84–92.

Christine A. Curcio and Kimberly A. Allen. 1990. Topography of ganglion cells in human retina. *The Journal of Comparative Neurology* 300, 1 (1990), 5–25.

Donald G Dansereau, Glenn Schuster, Joseph Ford, and Gordon Wetzstein. 2017. A Wide-Field-of-View Monocentric Light Field Camera. In *CVPR '17*.
Neville Drasdo, C. Leigh Millican, Charles R. Katholi, and Christine A. Curcio. 2007. The length of Henle fibers in the human retina and a model of ganglion receptive field density in the visual field. *Vision Research* 47, 22 (2007), 2901 – 2911.
Greg Egan. 1994. *Permutation City.* Millennium Orion Publishing Group.
Kevin Egan, Frédo Durand, and Ravi Ramamoorthi. 2011a. Practical Filtering for Efficient Ray-traced Directional Occlusion. *ACM Trans. Graph.* 30, 6, Article 180 (2011), 10 pages.
Kevin Egan, Florian Hecht, Frédo Durand, and Ravi Ramamoorthi. 2011b. Frequency Analysis and Sheared Filtering for Shadow Light Fields of Complex Occluders. *ACM Trans. Graph.* 30, 2, Article 9 (2011), 13 pages.
Kevin Egan, Yu-Ting Tseng, Nicolas Holzschuch, Frédo Durand, and Ravi Ramamoorthi. 2009. Frequency Analysis and Sheared Reconstruction for Rendering Motion Blur. *ACM Trans. Graph.* 28, 3, Article 93 (2009), 13 pages.
Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. 1996. The Lumigraph. In *SIGGRAPH '96*. 43–54.
Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. *ACM Trans. Graph.* 31, 6, Article 164 (2012), 10 pages.
Toshiya Hachisuka, Wojciech Jarosz, Richard Peter Weistroffer, Kevin Dale, Greg Humphreys, Matthias Zwicker, and Henrik Wann Jensen. 2008. Multidimensional Adaptive Sampling and Reconstruction for Ray Tracing. *ACM Trans. Graph.* 27, 3, Article 33 (2008), 10 pages.
Felix Heide, Gordon Wetzstein, Ramesh Raskar, and Wolfgang Heidrich. 2013. Adaptive Image Synthesis for Compressive Displays. *ACM Trans. Graph.* 32, 4, Article 132 (2013), 12 pages.
Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. 2015. The Light Field Stereoscope: Immersive Computer Graphics via Factored Near-eye Light Field Displays with Focus Cues. *ACM Trans. Graph.* 34, 4, Article 60 (2015), 12 pages.
Fu-Chung Huang, Gordon Wetzstein, Brian A. Barsky, and Ramesh Raskar. 2014. Eyeglasses-free Display: Towards Correcting Visual Aberrations with Computational Light Field Displays. *ACM Trans. Graph.* 33, 4, Article 59 (2014), 12 pages.
Julian Iseringhausen, Bastian Goldlücke, Nina Pesheva, Stanimir Iliev, Alexander Wender, Martin Fuchs, and Matthias B. Hullin. 2017. 4D Imaging Through Spray-on Optics. *ACM Trans. Graph.* 36, 4, Article 35 (2017), 11 pages.
Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *UbiComp '14 Adjunct*. 1151–1160.
Joohwan Kim, Qi Sun, Fu-Chung Huang, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptual Studies for Foveated Light Field Displays. *CoRR* abs/1708.06034 (2017).
Matias Koskela, Timo Viitanen, Pekka Jääskeläinen, and Jarmo Takala. 2016. Foveated Path Tracing. In *ISVC '16*. 723–732.
Douglas Lanman and David Luebke. 2013. Near-eye Light Field Displays. *ACM Trans. Graph.* 32, 6, Article 220 (2013), 10 pages.
Jaakko Lehtinen, Timo Aila, Jiawen Chen, Samuli Laine, and Frédo Durand. 2011. Temporal Light Field Reconstruction for Rendering Distribution Effects. *ACM Trans. Graph.* 30, 4, Article 55 (2011), 12 pages.
Jaakko Lehtinen, Timo Aila, Samuli Laine, and Frédo Durand. 2012. Reconstructing the Indirect Light Field for Global Illumination. *ACM Trans. Graph.* 31, 4, Article 51 (2012), 10 pages.
Anat Levin, Samuel W. Hasinoff, Paul Green, Frédo Durand, and William T. Freeman. 2009. 4D Frequency Analysis of Computational Cameras for Depth of Field Extension. *ACM Trans. Graph.* 28, 3, Article 97 (2009), 14 pages.
Marc Levoy and Pat Hanrahan. 1996. Light Field Rendering. In *SIGGRAPH '96*. 31–42.
Lester C. Loschky and Gary S. Wolverton. 2007. How Late Can You Update Gaze-contingent Multiresolutional Displays Without Detection? *ACM Trans. Multimedia Comput. Commun. Appl.* 3, 4, Article 7 (2007), 10 pages.
Andrew Maimone and Henry Fuchs. 2013. Computational augmented reality eyeglasses. In *ISMAR '13*. 29–38.
Andrew Maimone, Gordon Wetzstein, Matthew Hirsch, Douglas Lanman, Ramesh Raskar, and Henry Fuchs. 2013. Focus 3D: Compressive Accommodation Display. *ACM Trans. Graph.* 32, 5, Article 153 (2013), 13 pages.
Soham Mehta, Brandon Wang, and Ravi Ramamoorthi. 2012. Axis-Aligned Filtering for Interactive Sampled Soft Shadows. *ACM Trans. Graph.* 31, 6 (2012), 163:1–163:10.
Rahul Narain, Rachel A. Albert, Abdullah Bulbul, Gregory J. Ward, Martin S. Banks, and James F. O'Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.* 34, 4, Article 59 (2015), 12 pages.
Ren Ng. 2005. Fourier Slice Photography. *ACM Trans. Graph.* 24, 3 (2005), 735–744.
Vitor F. Pamplona, Manuel M. Oliveira, Daniel G. Aliaga, and Ramesh Raskar. 2012. Tailored Displays to Compensate for Visual Aberrations. *ACM Trans. Graph.* 31, 4, Article 81 (2012), 12 pages.
Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards Foveated Rendering for Gaze-tracked Virtual Reality. *ACM Trans. Graph.* 35, 6, Article 179 (2016), 12 pages.
Anjul Patney, Marina Zannoli, George-Alex Koulieris, Joohwan Kim, Gordon Wetzstein, and Frank Steinicke. 2017. Applications of Visual Perception to Virtual Reality Rendering. In *SIGGRAPH '17 Courses*. Article 1, 38 pages.
Yuyang Qiu and Ling Zhu. 2010. The best approximation of the sinc function by a polynomial of degree with the square norm. *Journal of Inequalities and Applications* 2010, 1 (2010), 1–12.
Jonathan Ragan-Kelley, Jaakko Lehtinen, Jiawen Chen, Michael Doggett, and Frédo Durand. 2011. Decoupled Sampling for Graphics Pipelines. *ACM Trans. Graph.* 30, 3, Article 17 (2011), 17 pages.
V. Ramachandra, K. Hirakawa, M. Zwicker, and T. Nguyen. 2011. Spatioangular Prefiltering for Multiview 3D Displays. *IEEE Transactions on Visualization and Computer Graphics* 17, 5 (2011), 642–654.
John Siderov and Ronald S Harwerth. 1995. Stereopsis, spatial frequency and retinal eccentricity. *Vision research* 35, 16 (1995), 2329–2337.
Y. Takaki. 2006. High-Density Directional Display for Generating Natural Three-Dimensional Images. *Proc. IEEE* 94, 3 (2006), 654–663.
Yasuhiro Takaki, Kosuke Tanaka, and Junya Nakamura. 2011. Super multi-view display with a lower resolution flat-panel display. *Opt. Express* 19, 5 (2011), 4129–4139.
LN Thibos, FE Cheney, and DJ Walsh. 1987. Retinal limits to the detection and resolution of gratings. *JOSA A* 4, 8 (1987), 1524–1529.
Christopher W Tyler. 1987. Analysis of visual modulation sensitivity. III. Meridional variations in peripheral flicker sensitivity. *JOSA A* 4, 8 (1987), 1612–1619.
K. Vaidyanathan, M. Salvi, R. Toth, T. Foley, T. Akenine-Möller, J. Nilsson, J. Munkberg, J. Hasselgren, M. Sugihara, P. Clarberg, T. Janczak, and A. Lefohn. 2014. Coarse Pixel Shading. In *HPG '14*. 9–18.
Thomas S. A. Wallis, Matthias Bethge, and Felix A. Wichmann. 2016. Testing models of peripheral encoding using metamerism in an oddity paradigm. *Journal of Vision* 16, 2 (2016), 4.
Andrew B. Watson. 2014. A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision* 14, 7 (2014), 15.
Andrew B. Watson and Albert J. Ahumada. 2011. Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision* 11, 5 (2011), 10.
Li-Yi Wei, Chia-Kai Liang, Graham Myhre, Colvin Pitts, and Kurt Akeley. 2015. Improving Light Field Camera Sample Design with Irregularity and Aberration. *ACM Trans. Graph.* 34, 4, Article 152 (2015), 11 pages.
Alexander Wender, Julian Iseringhausen, Bastian Goldluecke, Martin Fuchs, and Matthias B. Hullin. 2015. Light Field Imaging through Household Optics. In *Vision, Modeling & Visualization*.
Gordon Wetzstein, Douglas Lanman, Wolfgang Heidrich, and Ramesh Raskar. 2011. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. *ACM Trans. Graph.* 30, 4, Article 95 (2011), 12 pages.
Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. 2012. Tensor Displays: Compressive Light Field Synthesis Using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph.* 31, 4, Article 80 (2012), 11 pages.
Ling-Qi Yan, Soham Uday Mehta, Ravi Ramamoorthi, and Fredo Durand. 2015. Fast 4D Sheared Filtering for Interactive Rendering of Distribution Effects. *ACM Trans. Graph.* 35, 1, Article 7 (2015), 13 pages.
Marina Zannoli, Gordon D. Love, Rahul Narain, and Martin S. Banks. 2016. Blur and the perception of depth at occlusions. *Journal of Vision* 16, 6 (2016), 17.
Matthias Zwicker, Wojciech Matusik, Frédo Durand, Hanspeter Pfister, and Clifton Forlines. 2006. Antialiasing for Automultiscopic 3D Displays. In *SIGGRAPH '06 Sketches*. Article 107.

A RAY SPACE ANALYSIS

We first consider an observer focusing on a light field display at a distance $d_d = (d_e f_d)/(d_e - f_d)$ where f_d is the focal length of the eye when focusing on the display and d_e is the diameter of the eyeball, as shown in Figure 2a. The display light field L_d propagates along the free space and is refracted by the eye lens, and the retina receives an image I by integrating the retinal light field L along the angular dimension $\mathbf{u}$ parameterized at the pupil:

$$\begin{aligned} I(\mathbf{x}) &= \int L(\mathbf{x}, \mathbf{u}) \sqcap (\mathbf{u}/a) \mathrm{d}\mathbf{u} \\ &= \int L_d(\phi(\mathbf{x}, \mathbf{u}), \mathbf{u}) \sqcap (\mathbf{u}/a) \mathrm{d}\mathbf{u}, \end{aligned} \tag{16}$$

where a is the pupil aperture, $\sqcap(\cdot)$ is the rectangular function, and ϕ maps the intersection of a retinal light ray $(\mathbf{x}, \mathbf{u})$ with the display

spatial point $\mathbf{x}_d$:

$$\mathbf{x}_d = \phi(\mathbf{x}, \mathbf{u}) = -\frac{d_d}{d_e}\mathbf{x} + d_d\kappa(d_d, f_d)\mathbf{u},$$
$$\kappa(d, f) = \left(\frac{1}{d_e} - \frac{1}{f} + \frac{1}{d}\right). \tag{17}$$

For an out-of-focus virtual object being presented at a distance $d_o \neq d_d$ to the eye, we can obtain its corresponding retinal light field through the inverse mapping of Equation (17), with slope

$$k(d_o, f_d) = (d_e\kappa(d_o, f_d))^{-1} \tag{18}$$

in the flatland diagram, as shown in Figure 2b. Since we integrate all rays over the pupil to obtain the retinal image in Equation (16), the image is blurred by a retinal Circle-of-Confusion (CoC) of diameter

$$CoC = \frac{a}{k(d_o, f_d)} = ad_e\kappa(d_o, f_d). \tag{19}$$

In the case of an out-of-focus object, intuitively we can sample it at frequency inversely proportional to the circle-of-confusion size. Similarly, inspired by recent work on foveated rendering where peripheral vision has lower retinal resolution, rendering cost can be dramatically reduced as well at large eccentricity. However, there is no theoretical guideline on the savings, and prior techniques do not apply to light field sampling. We show that, through Fourier analysis, more theoretical bounds for saving can be revealed in both spatial and angular dimensions.

B ANALYSIS OF FREQUENCY BOUND DUE TO DISPLAY

Zwicker et al. [2006] have shown that when object extends beyond the depth of field (DoF) of the light field display, the spatial domain is subject to frequency clipping and thus low-pass filtered.

$$B^{display}_{\omega_\mathbf{x}} = \begin{cases} \frac{1}{2\Delta\mathbf{u}_d\hat{k}(d_o,f)}, & \text{if } \hat{k}(d_o, f) \geq \frac{\frac{d_e}{d_d}\Delta\mathbf{x}_d}{\Delta\mathbf{u}_d} \\ \frac{d_d}{2d_e\Delta\mathbf{x}_d}, & \text{otherwise,} \end{cases} \tag{20}$$

These bounds are illustrated in Figure 2c.

C SAMPLING TRANSFORMATION

In Section 5.1, each d_ζ from

$$W(\mathbf{x}, \mathbf{u}) = \int_{d_\zeta^-}^{d_\zeta^+} w_d(d_\zeta)w_s(d_\zeta)\mathrm{d}d_\zeta, \tag{21}$$

defines an independent coordinate system $(\omega_\mathbf{x}, \omega_\mathbf{u})$ with the slope $\hat{k}\left(d_\zeta, f_\zeta\right) = 0$. For fast and closed form computation of the integration, we transform them, through operator η, into one uniform coordinate frame such that $\hat{k}\left(d_\zeta^-, f_\zeta^-\right) = 0$ (i.e., relative to the coordinate frame when the eye is focusing at d_ζ^- with focal length f_ζ^-). The transformed d_ζ and $(\mathbf{x}, \mathbf{u})$ are defined as d'_ζ and $(\mathbf{x}', \mathbf{u}')$.

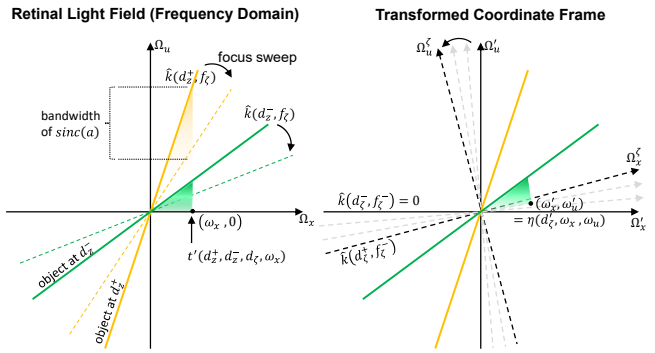


Fig. 12. *Illustration of importance function and coordinate transformation.* The left figure shows original coordinate system for a given d_ζ before transformation: The (sync-smeared) yellow and green lines represent two object points at different depths $d_z^\pm$. Their perceptual bandwidths $t(d_z^+, d_\zeta, \omega_\mathbf{x})$ and $t(d_z^-, d_\zeta, \omega_\mathbf{x})$ are evaluated at $(\omega_\mathbf{x}, 0)$, and their difference represents $\bar{t}(d_z^+, d_z^-, d_\zeta, \omega_\mathbf{x})$, whose integration (along the Ω_x axis) yields the static weight, $w_s(d_\zeta)$. The dynamic weight $w_d(d_\zeta)$ is similarly integrated but from the rate of change of $\bar{t}$ with respect to d_ζ, i.e. the two lines rotate with varying d_ζ. The right figure shows the transformed system: all coordinates are transformed to the one (Ω'_x, Ω'_u) respect to d_ζ^-. Correspondingly, all the importance evaluations of d_ζ (transformed as d'_ζ) are performed at Ω_x^ζ axis.

In the transformed frequency frame, a point $(\omega'_\mathbf{x}, \omega'_\mathbf{u})$ can be computed as:

$$\begin{bmatrix}\omega'_\mathbf{x} \\ \omega'_\mathbf{u}\end{bmatrix} = \left(1 + \hat{k}\left(d_\zeta^-, f_\zeta\right)^2\right)^{-\frac{1}{2}} \begin{bmatrix} 1 & \hat{k}(d_\zeta^-, f_\zeta) \\ -\hat{k}(d_\zeta^-, f_\zeta) & 1 \end{bmatrix} \begin{bmatrix}\omega_\mathbf{x} \\ \omega_\mathbf{u}\end{bmatrix} \tag{22}$$
$$\triangleq \eta(d_\zeta, \omega_\mathbf{x}, \omega_\mathbf{u}).$$

We define its slope as

$$\hat{k} \triangleq \frac{\omega'_\mathbf{u}}{\omega'_\mathbf{x}}. \tag{23}$$

Then its corresponding transformed signal amplitude as

$$t'(z_i, \omega'_\mathbf{x}, \omega'_\mathbf{u}) = \left\|\hat{s_i}\left(-\frac{d_e}{d_{z_i}}\omega'_\mathbf{x}\right)\right\| \times \mathrm{sinc}\left(a\omega'_\mathbf{x}\left\|\frac{\hat{k} - \hat{k}(d_{z_i}, f_\zeta^-)}{\sqrt{1 + \hat{k}^2(d_{z_i}, f_\zeta^-)}}\right\|\right). \tag{24}$$

With the formulation above, the static importance defined on the point $(\omega'_\mathbf{x}, \omega'_\mathbf{u})$ is

$$w'_s(\omega'_\mathbf{x}, \omega'_\mathbf{u}) = \left\|t'(d_z^+, \omega'_\mathbf{x}, \omega'_\mathbf{u}) - t'(d_z^-, \omega'_\mathbf{x}, \omega'_\mathbf{u})\right\|, \tag{25}$$

and the dynamic importance defined on the line with slope $\hat{k}$ becomes

$$w'_d(\hat{k}) = \int_0^{B_s} \frac{w'_s(\omega'_\mathbf{x}, \omega'_\mathbf{u})}{\partial\hat{k}}\mathrm{d}\omega'_\mathbf{x}. \tag{26}$$

Now Equation (21) can be recomputed as:

$$\int\int w'_d(\hat{k})w'_s\left(\omega'_\mathbf{x}, \omega'_\mathbf{u}\right)\mathrm{d}\omega'_\mathbf{x}\mathrm{d}\omega'_\mathbf{u}. \tag{27}$$

This closed form integration is derived in Appendix F.

Note that the display ($B^{display}_{\omega_x}$) and lens ($B^{lens}_{\omega_x}$) spatial bounds may also transform along with η. However, the actual range of $\hat{k}$ under a common light field display is small ($\approx \pm 0.037$ with our prototype), and the major influence in periphery is from the untransformed $B^{retina}_{\omega_x}$, so we keep those two bounds invariant when computing Equation (27).

D MINIMUM DISPLAY SAMPLING

To reach a high perceptual threshold, we allow more rays to be sampled than the minimum number required at locations in the adaptive light field sampling Equation (14). Specifically, we guarantee full sampling in the foveal area (within 5 deg eccentricity). For periphery, according to our bandwidth guideline, we compute the local budget s_{el} for minimum sampling of the display proportional to the density function of the local retinal bandwidth σ^{-1}(Equation (4)):

$$s_{el}(\mathbf{x}_d) = s_e \frac{\sigma^{-1}\left(\phi^{-1}(\mathbf{x}_d, \mathbf{u}_d)\right)}{\int \sigma^{-1}(\mathbf{x})\, d\mathbf{x}}, \tag{28}$$

where s_e is the total peripheral sampling budget, $(\mathbf{x}_d, \mathbf{u}_d)$ is a ray passing the center of eyebox, and ϕ^{-1} maps the display coordinate to retina space (Equation (17)).

To guarantee perception preservation, we also ensure the number of rays to satisfy the condition where the footprint of a ray (e_b/s_{el}) over the eyebox weighted by the spatial retinal bandwidth is smaller than the smallest solid angle of the hardware ray $\Delta\mathbf{u}_d$ on the pupil plane:

$$\frac{e_b/s_{el}(\mathbf{x}_d)}{\sigma(\phi^{-1}(\mathbf{x}_d, \mathbf{u}_d))} \leq \epsilon \frac{\Delta\mathbf{u}_d}{\sigma(\mathbf{x}_{fovea})}, \tag{29}$$

where $\mathbf{x}_{fovea}$ represents the foveal boundary, $\epsilon > 1$ tolerates strong sampling reduction from content-adaptive importance Equation (21). We set $\epsilon = 1.2$ in our experiments.

E OCCLUSION AWARE POST-FILTERING

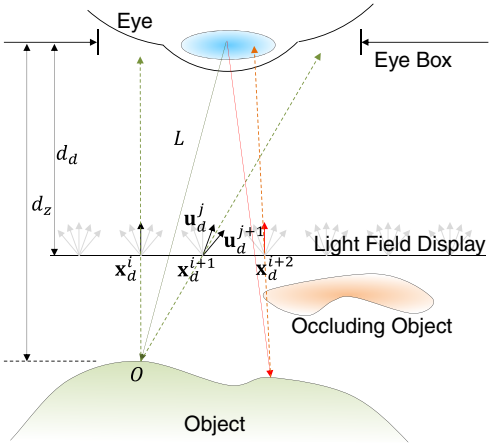


Fig. 13. *Reconstructing rays for light field display.* The display rays $L_d(\mathbf{x}_d, \mathbf{u}_d)$ can be reconstructed from the sparsely sampled rays L (solid lines) through 4D Gaussain radial basis function by intersecting the reflected rays (dashed lines) to the display pixels.

The sparsely sampled set of rays is then filtered to be shown on a light field display of rays with uniform spacing. We implement a separable 4D Gaussian radial basis function for the sparse reconstruction. We first trace the ray $L(\mathbf{x}, \mathbf{u})$ to the scene and intersect it with the point O, and then splat the reflected rays to the the light field display rays $L_d(\mathbf{x}_d^i, \mathbf{u}_d^j)$, as shown in Figure 13, such that their extensions are within the eyebox e_b:

$$\begin{aligned} L_d(\mathbf{x}_d^i, \mathbf{u}_d^j) =& L_d(\mathbf{x}_d^i, \mathbf{u}_d^j) + \mathcal{N}\left(\mathbf{x}_d^i - \phi(\mathbf{x}, \mathbf{u}), \frac{1}{B^{all}_{\omega_x}(\mathbf{x})}\right) \\ & \times \mathcal{N}\left(\mathbf{u}_d^j - \frac{d_d(\mathbf{x}_d^i - O)}{d_z - d_d}, \frac{1}{B^{all}_{\omega_u}(\mathbf{x})}\right) \times L(\mathbf{x}, \mathbf{u}) \\ & \forall_{(i,j)} \text{ such that } (\mathbf{x}_d^i + d_d \mathbf{u}_d^j) < \frac{e_b}{2}. \end{aligned} \tag{30}$$

Proper occlusion handling is crucial in the post-filtering that we use the depth map obtained in the first stage of sparse sampling to cull out rays blocked by the occluder, as shown in Figure 13. Finally, similar to Patney et al. [2016], a contrast-preserving filter is applied to the rendering.

F CLOSED FORM IMPORTANCE SAMPLING

To calculate Equation (14) from the transformed frame in Appendix C, we first simplify Equation (10). Because of the small range of $\hat{k}^2(d_{z_i}, f_{\zeta}^-)$ described in Appendix C, Equation (24) can be approximated and simplified as

$$t'(z_i, \omega'_{\mathbf{x}}, \omega'_{\mathbf{u}}) \approx \left\| \hat{s}_i \left(-\frac{d_e}{d_{z_i}} \omega'_{\mathbf{x}} \right) \right\| \operatorname{sinc}\left(a\omega'_{\mathbf{x}} \left\| \hat{k} - \hat{k}(d_{z_i}, f_{\zeta}^-) \right\| \right). \tag{31}$$

Note that we have applied contrast preserving step in the post filtering Section 5.2, during sampling stage, we can make an conservative estimation by assuming high frequency amplitude over all surfaces, thus Equation (31) can be further simplified as

$$\begin{aligned} t'(z_i, \omega'_{\mathbf{x}}, \omega'_{\mathbf{u}}) &= s_h \operatorname{sinc}\left(a\omega'_{\mathbf{x}} \left\| \hat{k} - \hat{k}(d_{z_i}, f_{\zeta}^-) \right\| \right) \\ &\propto \operatorname{sinc}\left(a\omega'_{\mathbf{x}} \left\| \hat{k} - \hat{k}(d_{z_i}, f_{\zeta}^-) \right\| \right), \end{aligned} \tag{32}$$

where s_h is a constant amplitude value of high frequency texture. For easier formulation, we define symbols $\hat{k_1} \triangleq \hat{k}(d_{z^-}, f_{\zeta}^-), \hat{k_2} \triangleq \hat{k}(d_{z^+}, f_{\zeta}^-)$ for derivations below. Thus $t'(z^-, \omega'_{\mathbf{x}}, \omega'_{\mathbf{u}})$ and $t'(z^+, \omega'_{\mathbf{x}}, \omega'_{\mathbf{u}})$ can be redefined as $t'(\hat{k}, \hat{k_1}, \omega'_{\mathbf{x}})$ and $t'(\hat{k}, \hat{k_2}, \omega'_{\mathbf{x}})$ respectively.

Because of the existence of absolute operator in Equation (11), the integration result relies on relative range of k compared with $\hat{k_1}$ and $\hat{k_2}$ (Intuitive illustration can be seen from Figure 2). That means this is a piece-wise integration. As an example of computation, here we let $\hat{k} \geq \hat{k_2} \geq \hat{k_1}$. Other cases can be derived similarly. Moreover, because of the symmetry of the frequency domain (Figure 2), we can just perform computation for $\omega'_{\mathbf{x}} \geq 0$ w.l.o.g. In this subspace, we have

$$t'\left(\frac{\omega'_{\mathbf{u}}}{\omega'_{\mathbf{x}}}, \hat{k_1}, \omega'_{\mathbf{x}}\right) \geq t'\left(\frac{\omega'_{\mathbf{u}}}{\omega'_{\mathbf{x}}}, \hat{k_2}, \omega'_{\mathbf{x}}\right). \tag{33}$$

The first step is to compute w_d. To equally compare different focus depths, we use same range $\Omega_x = [0, B_s]$. Because values of dynamic weight w_d are small, we estimate their terms through a polynomial approximation of sinc function. Optimal sinc function

approximation parameters $\{a_3, a_2, a_1, a_0\}$ have been studied by Qiu et al. [2010]:

$$\mathrm{sinc}(x) \approx a_3 x^3 + a_2 x^2 + a_1 x + a_0. \tag{34}$$

Thus we have

$$\begin{aligned} w'_d(\hat{k}) &\approx \int_0^{B_\mathbf{s}} \sum_{i=1}^{2} -1^{i-1} a\omega'_\mathbf{x} \left(3a_3(a\omega'_\mathbf{x}(\hat{k}-\hat{k_i}))^2 + 2a_2(a\omega'_\mathbf{x}(\hat{k}-\hat{k_i})) + a_1\right) \mathrm{d}\omega'_\mathbf{x} \\ &\propto \int_0^{B_\mathbf{s}} \omega'_\mathbf{x} \left(3a_3 a{\omega'_\mathbf{x}}^2(2\hat{k}-\hat{k_1}-\hat{k_2}) + 2a_2\omega'_\mathbf{x}\right) \mathrm{d}\omega'_\mathbf{x} \\ &\propto 9a_3 a(2\hat{k}-\hat{k_1}-\hat{k_2})B_\mathbf{s} + 8a_2. \end{aligned}$$

Using the estimation of w_d above, we obtain

$$\begin{aligned} W(\hat{k_1},\hat{k_2}) &\propto \iint \left(9a_3 a\left(2\frac{\omega'_\mathbf{u}}{\omega'_\mathbf{x}} - \hat{k_1} - \hat{k_2}\right)B_\mathbf{s} + 8a_2\right)\left(t\left(\frac{\omega'_\mathbf{u}}{\omega'_\mathbf{x}}, \hat{k_1}, \omega'_\mathbf{x}\right) - t\left(\frac{\omega'_\mathbf{u}}{\omega'_\mathbf{x}}, \hat{k_2}, \omega'_\mathbf{x}\right)\right) \mathrm{d}\omega'_\mathbf{x}\mathrm{d}\omega'_\mathbf{u} \\ &= \iint \frac{18a_3 a\omega'_\mathbf{u}}{\omega'_\mathbf{x}} B_\mathbf{s} \left(\sum_{i=1}^{2} -1^{i-1}\,\mathrm{sinc}(a\omega'_\mathbf{u} - a\hat{k_i}\omega'_\mathbf{x})\right) \mathrm{d}\omega'_\mathbf{x}\mathrm{d}\omega'_\mathbf{u} \\ &+ \iint \left(8a_2 - 9(\hat{k_1}+\hat{k_2})a_3 a B_\mathbf{s}\right)\left(\sum_{i=1}^{2} -1^{i-1}\,\mathrm{sinc}(a\omega'_\mathbf{u} - a\hat{k_i}\omega'_\mathbf{x})\right) \mathrm{d}\omega'_\mathbf{x}\mathrm{d}\omega'_\mathbf{u} \\ &= 18a_3 B_\mathbf{s} \int \left(\sum_{i=1}^{2} -1^{i-1}\left(\sin(a\omega'_\mathbf{u})(-\mathrm{Ci}(a\hat{k_i}\omega'_\mathbf{x})) + \mathrm{Si}(a\omega'_\mathbf{u} - a\hat{k_i}\omega'_\mathbf{x}) + \cos(a\omega'_\mathbf{u})\,\mathrm{Si}(a\hat{k_i}\omega'_\mathbf{x})\right)\right) \mathrm{d}\omega'_\mathbf{u} \\ &- \left(8a_2 - 9(\hat{k_1}+\hat{k_2})a_3 a B_\mathbf{s}\right) \int \left(\sum_{i=1}^{2} -1^{i-1}\frac{\mathrm{Si}(a\omega'_\mathbf{u} - a\hat{k_i}\omega'_\mathbf{x})}{a\hat{k_i}}\right) \mathrm{d}\omega'_\mathbf{u} \end{aligned} \tag{35}$$

Here Si/Ci is sine/cosine integration function can be approximated through Padé approximant. The integration over $\omega'_\mathbf{u}$ can be derived with the help of equation below

$$\int \mathrm{Si}(a\omega'_\mathbf{u} - a\hat{k_i}\omega'_\mathbf{x})\mathrm{d}\omega'_\mathbf{u} = \frac{1}{a}\left((a\hat{k_i}\omega'_\mathbf{x} - a\omega'_\mathbf{u})\,\mathrm{Si}(a\hat{k_i}\omega'_\mathbf{x} - a\omega'_\mathbf{u}) + \cos(a\hat{k_i}\omega'_\mathbf{x} - a\omega'_\mathbf{u})\right) \tag{36}$$

G SCENE INFORMATION

Table 3 shows the statistics of our test scenes.

details / scene	# vertices	# faces
fairy	96221	172669
Mars	118760	231762
Sponza	145185	262267
chess	643938	1278876
marbles	4452	8480
farm	1882270	357883
Stonehenge	9817	19362
craftsman	4182	6969
toaster	5628	11141
van Gogh	6701	11272
Viking	2555	3829

Table 3. *Geometry details of our experimental scenes.*