# *Linear view synthesis using a dimensionality gap light field prior*

# Linear View Synthesis Using a Dimensionality Gap Light Field Prior

Anat Levin
Weizmann Institute of Science

Fredo Durand
MIT CSAIL

## Abstract

*Acquiring and representing the 4D space of rays in the world (the light field) is important for many computer vision and graphics applications. Yet, light field acquisition is costly due to their high dimensionality. Existing approaches either capture the 4D space explicitly, or involve an error-sensitive depth estimation process.*

*This paper argues that the fundamental difference between different acquisition and rendering techniques is a difference between prior assumptions on the light field. We use the previously reported dimensionality gap in the 4D light field spectrum to propose a new light field prior. The new prior is a Gaussian assigning a non-zero variance mostly to a 3D subset of entries. Since there is only a low-dimensional subset of entries with non-zero variance, we can reduce the complexity of the acquisition process and render the 4D light field from 3D measurement sets. Moreover, the Gaussian nature of the prior leads to linear and depth invariant reconstruction algorithms.*

*We use the new prior to render the 4D light field from a 3D focal stack sequence and to interpolate sparse directional samples and aliased spatial measurements. In all cases the algorithm reduces to a simple spatially invariant deconvolution which does not involve depth estimation.*

## 1. Introduction

Light field or plenoptic imaging enables exciting computer vision and graphics applications such as refocusing or viewpoint changes. Light fields record the 4D set of light rays incident to the lens aperture. This can be achieved by varying the camera position on a 2D plane and capturing a 2D family of 2D images [14, 7, 21], or by placing a microlens array in front of the sensor [1, 19]. With such representation, rendering a novel view is a straightforward linear operation in the acquired data, since the scene color along any light ray one wishes to render has already been captured explicitly, and view synthesis is a simple matter of rebinning. Unfortunately, the four-dimensional nature of light fields makes their acquisition costly and often involves a spatial resolution tradeoff. This is all the more frustrating that many operations and data are 3D in nature: Lambertian scenes are three-dimensional and refocusing only involves a depth-indexed 1D family of 2D images. An alternative strategy is standard image-based rendering or novel-view generation [6, 16, 4], in which only a sparse set of images is captured. To render a novel viewpoint, depth is estimated using a variety of computer vision techniques. However, depth estimation is a complex non-linear process which can induce visual artifacts. In this work, we propose a new view synthesis approach that sits halfway between these two approaches. It is linear in the acquired images and does not involve depth estimation. It also does not require full 4D sampling.

We argue that the fundamental difference behind different capture and rendering strategies can be seen as a difference in prior assumptions on light fields. Plenoptic imaging relies on a weak but general prior where the light field is considered isotropic and fully involves four degrees of freedom. Since the data is assumed to be 4D, a 4D set of measurements is required. With depth-based view synthesis, the prior is stronger but more restrictive. It assumes that, locally, depth is constant and the object is Lambertian, which means that the light field is constant along the angular dimension and is locally 2D.

In this paper we propose a new light field prior which is a tradeoff between these two approaches. It is based on the recently-observed dimensionality gap in light fields [18, 13] which states that for Lambertian scenes with modest depth discontinuities, the 4D Fourier transform of the 4D ray space includes only a 3D subset of entries whose energy is significantly higher than zero. This observation has so far been used to analyze and improve depth of field extension [13] but we propose to use it for light field reconstruction and view synthesis. Our new prior is a Gaussian assigning a non-zero variance mostly to a 3D subset of entries. Since only three degrees of freedom are present, measurement sets which are only 3D are sufficient for interpolating the full 4D light field. Furthermore, since the prior is Gaussian, the reconstruction and rendering algorithms are simple and linear, and *do not involve depth estimation*. While the prior is simple to use, the quality of results sits halfway between the two existing approaches. The reconstruction is better than with a generic 4D prior, but the 2D MOG can produce better results when depth is successfully estimated.

We examine a number of low-dimensional acquisition schemes and show how, in conjunction with the new light field prior, they can be used to interpolate the 4D light field or to render novel viewpoints. The simplest acquisition scheme considered in Sec. 3 is a focal stack sequence, a 1D set of images focused at a varying range of depths, providing a 3D set of measurements. We show that the advantage of the focal stack is that it directly covers the non-zero entries of the 4D spectrum. We present a simple primal domain algorithm which uses the focal stack to render images from novel viewpoints inside the aperture area. In our algorithm, each focal stack image is shifted according to the disparity of its focusing distance. The shifted images are

978-1-4244-6985-7/10/$26.00 ©2010 IEEE

averaged and a spatially uniform, depth-invariant deconvolution is applied. This algorithm relates to [2] who use two defocused images of a two layers scene to generate new defocused images without segmenting the depth layers.

We also show how the dimensionality-gap prior can help interpolate the light field from measurements that are sparse or aliased, in both directional (Sec. 4) and spatial (Sec. 5) dimensions. We present simple reconstruction algorithms which also reduce to depth invariant deconvolution.

## 2. Light field priors

### 2.1. Background on light fields

The 4D light field $L(x, y, u, v)$ parameterizes each ray by its intersection with two parallel planes **uv** and **xy**, known as the *directional* and *spatial* dimensions. Usually the viewpoint (or camera aperture) is positioned and shifted along the **uv** plane, while **xy** is a scene plane.

In light field space, the set of light rays emerging from a single point lie on a plane whose slope is a function of depth. If the scene is Lambertian, all rays emerging from a point have the same color and we can express:

$$L(x, y, u, v) = L_{u_0,v_0}(x - s(u - u_0), y - s(v - v_0)), \quad (1)$$

where $L_{u_0,v_0}(x, y)$ denotes a 2D view from the point $(u_0, v_0)$, $L_{u_0,v_0}(x, y) = L(x, y, u_0, v_0)$. The slope $s$ is $s = (d - d_0)/d$ where $d$ is the object depth and $d_0$ the distance between the **uv** and **xy** planes.

A standard lens focused at slope $s$ averages rays emerging from points at the corresponding depth, all lying on a slope $s$ plane in light field space. The recorded image is:

$$B^s(x, y) = \iint_{(u,v)\in D(A,0)} L(x + us, y + vs, u, v)dudv, \quad (2)$$

where $D(r, p)$ denotes a disc of radius $r$ centered at point $p$, and $A$ is the aperture radius. If the scene depth $s_0$ is locally constant, we substitute Eq. (1) into Eq. (2) and get that $B^s(x, y)$ equals

$$\iint_{D(A,0)} L_{u_0,v_0}(x + s_0u_0 + u(s - s_0), y + s_0v_0 + v(s - s_0))dudv$$
$$= L_{u_0,v_0}(x, y) \otimes \tfrac{1}{\pi^2|s-s_0|^2A^2} D(A|s - s_0|, (s_0u_0, s_0v_0)). \quad (3)$$

That is, the recorded full aperture image $B^s(x, y)$ is a convolution of a pinhole view $L_{u_0,v_0}(x, y)$ with a PSF $\phi_s = D(A|s - s_0|, (s_0u_0, s_0v_0))$. The PSF is a disc whose radius is proportional to the difference between the focus depth $s$ and the object depth $s_0$. The disc center is shifted according to the disparity shift of that viewpoint, $(s_0u_0, s_0v_0)$.

Consider the 4D Fourier transform of the 4D light field, denoted by $\hat{L}(\omega_x, \omega_y, \omega_u, \omega_v)$[1]. For a Lambertian planar scene of slope $s$, the light field $L$ is constant along direction $s$. As a result, the Fourier transform $\hat{L}$ includes non zero entries only on the 2D plane of entries of the form [18]

$$\omega_u = s\omega_x, \omega_v = s\omega_y. \quad (4)$$

Equivalently, for an infinitely wide aperture, the 2D spectrum of a 2D image focused at depth $s$ is a slice from the 4D spectrum [18]:

$$\hat{B}^s(\omega_x, \omega_y) = \hat{L}(\omega_x, \omega_y, s\omega_x, s\omega_y). \quad (5)$$

[1] We denote by $\hat{}$ the Fourier transform of a signal.

Following [12], in an abstract way, we can express the sensor measurements $b$ as a liner projection of a ray space vector $\ell$: $b = T\ell + n$, where the matrix $T$ expresses the mapping between light rays to sensor measurements and $n$ is the imaging noise. $T$ is often rank deficient. In this framework, recovering the light field $\ell$ from $b$ is a Bayesian estimation problem which should account for prior knowledge on light fields. The choice of prior is critical because it affects the amount of required measurements, the simplicity of the estimation process, and the quality of the results.

### 2.2. Existing light field priors

We argue that the fundamental difference between capturing and rendering strategies can be seen as a difference between prior assumptions. This classifies existing research into two main categories.

The first type of prior treats the spatial and directional dimensions of the light field isotropically, and it mostly assumes that the signal is smooth. As it assumes 4 degrees of freedom, capturing a 4D measurement set is required [14, 7, 21, 1, 19]. The smoothness assumption can be expressed as a Gaussian prior on the light field which is diagonal in the frequency domain:

$$\log p(L) = -0.5 \sum_{\omega_x,\omega_y,\omega_u,\omega_v} \frac{|\hat{L}(\omega_x, \omega_y, \omega_u, \omega_v)|^2}{\sigma^2_{\omega_x,\omega_y,\omega_u,\omega_v}} + const. \quad (6)$$

The variance $\sigma^2_{\omega_x,\omega_y,\omega_u,\omega_v}$ is high for low frequencies and low for high frequencies (Fig. 1(a)). If the range of depths in the scene is bounded, plenoptic sampling theory [5, 10, 18] further allows for a smarter sampling pattern.

However, capturing the full 4D light field directly seems redundant. The second type of prior assumes that depth is locally constant, and conditioning on depth there are only two degrees of freedom in defining the surface texture. Levin *et al.* [12] express such assumptions in priors terminology and suggest a mixture of 2D Gaussians model. Conditioning on depth, $p(L|s)$ is Gaussian, and diagonal in the frequency domain. From Eq. (4), when depth is given, there is non-zero frequency content only at entries of the form $\omega_u = s\omega_x, \omega_v = s\omega_y$. Hence there is only a 2D set of entries with non-zero variance, one non-zero variance entry in each $\omega_{x,y}$-slice (Figure 1(b)). In the general model, the mixture components index over all possible depth maps: $p(L) = \int p(s)p(L|s)ds$. This prior constrains the light field tighter than the general 4D Gaussian prior above. However, inference is more complicated since the correct mixture component, or the scene depth, needs to be estimated. Given depth, one can render the scene from multiple viewpoints [6, 16, 4]. One can also improve the quality of the 2D image, for example by using the aliased plenoptic camera measurements [15, 3] for super resolution. Similarly, one can use a depth dependent PSF to partially remove defocus blur [11, 20, 13].

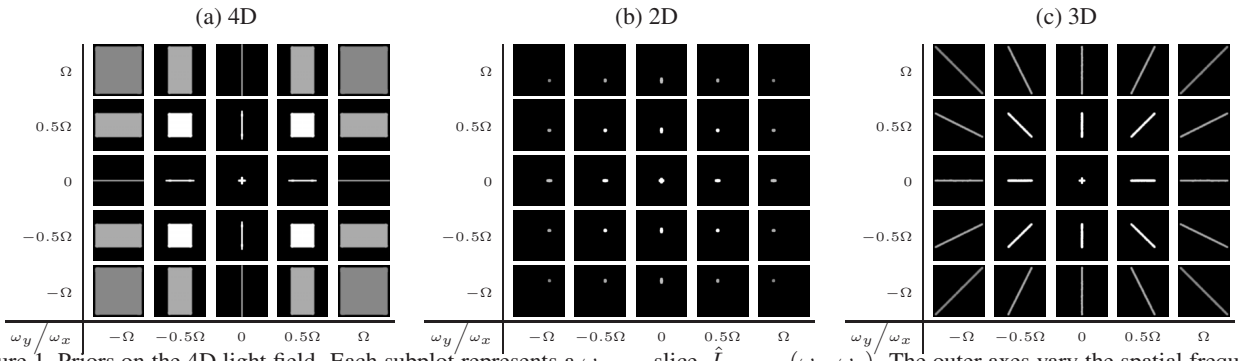



Figure 1. Priors on the 4D light field. Each subplot represents a $\omega_{x_0,y_0}$-slice, $\hat{L}_{\omega_{x_0,y_0}}(\omega_u, \omega_v)$. The outer axes vary the spatial frequency $\omega_{x_0,y_0}$, i.e., the slicing position. The inner axes of each subplot, i.e., of each slice, vary $\omega_{u,v}$. The intensity at each point visualizes the variance $\sigma^2_{\omega_x,\omega_y\omega_u\omega_v}$ at the corresponding spectrum entry. (a) Classical signal-processing prior assuming a smooth 4D signal, assigning non-zero variance to all 4D entries within a range $|\omega_u| \leq S_{max}|\omega_x|, |\omega_v| \leq S_{max}|\omega_y|$ (in this fig, $S_{max} = 1$). (b) One mixture component from a MOG prior. Conditioning on depth we have a Gaussian prior with only a 2D set of non-zero entries. (c) A Gaussian prior derived from the dimensionality gap with a 3D set of non-zero entries.

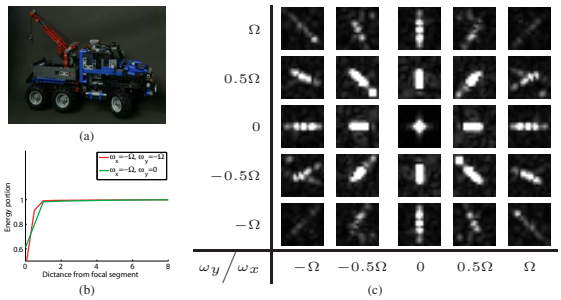



Figure 2. (a) One view from a light field. (c) The observed power spectrum. In each $\omega_{x_0,y_0}$-slice we observe energy mostly along 1D focal segments. (b) The accumulated histogram of energy as a function of distance from the focal segment, for the top left and top center subplots. Over $98\%$ of the energy is included within 1 pixel from the center.

### 2.3. The dimensionality gap as a low dimensional Gaussian prior

We suggest a new light field prior which is a tradeoff between the two priors discussed above. It is Gaussian but involves mostly 3D freedom degrees, thereby constraining the light field tighter than the 4D model. Additionally, having a Gaussian prior allows for simple linear inference, without explicit depth estimation. Our prior is based on the dimensionality gap property of the light field derived in [13, 18].

Recall that an object at slope $s$ generates frequency content only along a 2D subspace of the 4D light field (Eq 4). Since depth is only a 1D variable, a Lambertian scene with piecewise constant depth has frequency content along a 3D subset of entries of the form

$$(\omega_x, \omega_y, \omega_u, \omega_v) | \exists s : \omega_u = s\omega_x, \omega_v = s\omega_y. \tag{7}$$

Figure 2 visualizes the spectrum of a real light field. Despite the many occlusions, in each $\omega_{x_0,y_0}$-slice we can notice energy mostly along a 1D focal segment. Fig 2(b) plots cumulative histogram of energy as a function of distance from the focal segment. Over $98\%$ of the energy is concentrated up to one entry away from the focal segment. A Gaussian prior based on the dimensionality gap allows non-zero variance mostly in a 3D subset of frequencies (Fig 1(c)).

The focal segments thickness is determined by the spacing $\Delta_{\omega_{u,v}}$ between samples on the $\omega_u, \omega_v$ frequency axes, which is inversely proportional to the primal aperture width. With denser spacing the segments are thinner and the prior is tighter. In Fig 1(c) the resolution is low (only $17 \times 17$ viewpoint samples included) and the segments are thick.

This new prior does not constrain the light field as tightly as a 2D MOG. However, the Gaussianity of the prior leads to significantly simpler reconstruction algorithms which we investigate below. In contrast, in [13] the dimensionality gap was used with a 2D MOG prior, where depth is estimated, and applied for depth of field extension. This paper explores a different application of the dimensionality gap: depth-invariant light field reconstruction.

Given a Gaussian prior on the light field, estimating the light field $\ell$ from the camera data $b$ reduces to solving a linear system, or equivalently, minimizing a quadratic cost:

$$\ell = \arg\min \frac{1}{\eta^2}\|T\ell - b\|^2 + \ell^T C^{-1} \ell. \tag{8}$$

where $T$ is the measurement matrix, $\eta^2$ the noise variance and $C$ the prior covariance. Since the dimensionality gap prior permits non-zero variance to a 3D set of entries, a $T$ matrix which measures a 3D data should provide a well posed reconstruction. However, since the unknown vector in Eq. (8) is 4D, solving the system explicitly is impractical and we seek approximate strategies. As part of the approximation, the algorithms described below ignore energy off the focal segments. This restriction, however, is due to the approximate reconstruction, and is not an intrinsic limitation of the 3D Gaussian prior. The prior itself does permit low non-zero variance on off-focal-segment entries.

## 3. Novel views from a focal stack

Our first light field acquisition strategy is a 3D focal stack. We show that the advantage of a focal stack is that it directly covers the non-zero entries of the 4D spectrum. We show that the 4D light field can be rendered in a linear and depth invariant manner from a focal stack sequence. For that we describe an algorithm for rendering a 2D view $L_{u_0,v_0}(x, y)$. The rendering is limited to viewpoints within

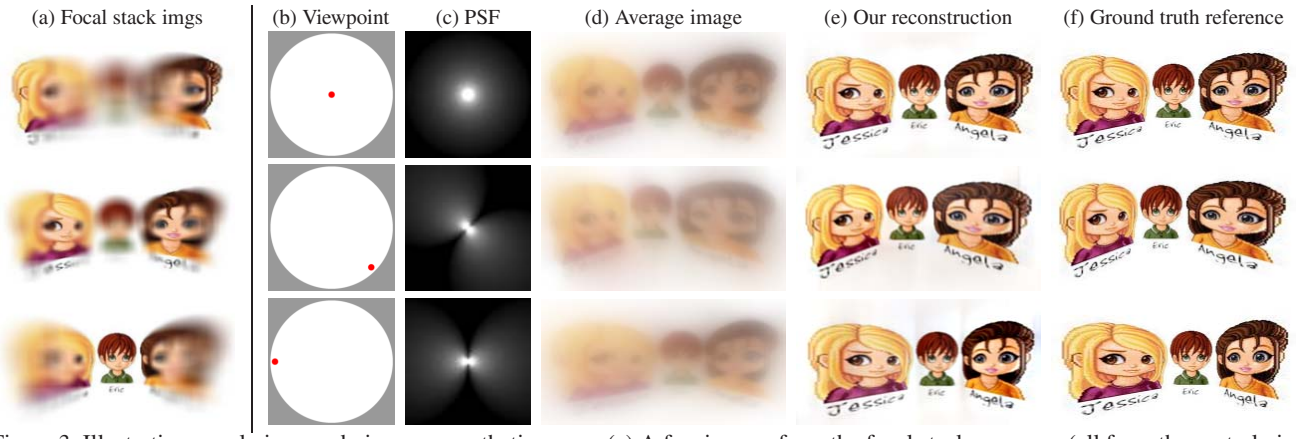



Figure 3. Illustrating novel view rendering on a synthetic scene. (a) A few images from the focal stack sequence (all from the central view point), (b-f) A few novel viewpoints rendered from the focal stack sequence without any depth estimation.

the aperture area $(u_0, v_0) \in D(A, 0)$. To simplify the derivation we start with the primal domain. In Sec. 3.2 we provide a frequency domain interpretation.

### 3.1. Primal domain derivation

Our rendering algorithm works as follows. Given a desired view point $(u_0, v_0)$, we shift each focal stack image by the disparity of its focusing depth $(su_0, sv_0)$, and compute an average image $\bar{B}(x,y) = \sum_s B^s_{su_0,sv_0}$, where $B^s_{su_0,sv_0}(x,y) = B^s(x - su_0, y - sv_0)$ is an image shifted by $(su_0, sv_0)$ (the amount of shift depends on the focus distance of the image and not on the object depth in the scene, therefore the shift is a parameter of the imaging apparatus and is independent of scene content). We show that the average image $\bar{B}(x,y)$ is an (almost) shift invariant convolution of the desired view $L_{u_0,v_0}(x,y)$. Therefore, it can be recovered with a spatially invariant deconvolution *without estimating the scene depth*.

For the central viewpoint $(u_0, v_0) = (0, 0)$ no shift happens and we sum the focal stack as it is. This average image is equivalent to the input from a focus sweep camera [9, 17] which varies the focus distance during exposure, and was shown to be a depth invariant convolution of an ideal pinhole image. Below we show that this is true for any viewpoint. Intuitively, for an infinitely wide aperture only the image focused at the right depth is sharp, all other focal stack images are flat. Therefore shifting the correct depth image by the correct disparity is sufficient.

Figure 3 visualizes average images and the depth-invariant deconvolution results, demonstrating close agreement with a ground truth reference.

**Claim 1** *For a Lambertian scene with locally-constant depth, the average image is a shift-invariant convolution of the desired view* $\bar{B}(x,y) = \phi_{u_0,v_0} \otimes L_{u_0,v_0}(x,y)$.

*The PSF* $\phi_{u_0,v_0}$ *is approximately depth invariant:*

$$\phi_{u_0,v_0}(x,y) \approx \left(\pi^2 A^2 s_{u_0,v_0}(x,y)\right)^{-1} + \left(\pi^2 A^2 s_{u_0,v_0}(-x,-y)\right)^{-1}, \tag{9}$$

*where* $s_{u_0,v_0}(x,y)$ *is the smallest* $s$ *for which* $(x,y)$ *is included in the disc* $D(sA, (su_0, sv_0))$. *Explicitly:*

$$s_{u_0,v_0}(x,y) = y/(A sin(asin(1/A(cos(\theta)v_0 - sin(\theta)u_0)) + \theta) - v_0) \tag{10}$$

*and* $\theta = phase(x + iy)$.

*Proof:* For an object at depth $s_0$ we can use Eq. (3) and express the average image as a convolution of the desired sharp one: $\bar{B}(x,y) = \phi^{s_0}_{u_0,v_0} \otimes L_{u_0,v_0}(x,y)$ with

$$\phi^{s_0}_{u_0,v_0} = \int_{S_{min}}^{S_{max}} \frac{1}{\pi^2 |s - s_0|^2 A^2} D(A|s - s_0|, ((s - s_0)u_0, (s - s_0)v_0)) ds, \tag{11}$$

where $S_{min}, S_{max}$ denote the minimal and maximal slopes in the focal stack. We change the integration variable by defining $s' = s - s_0$, $S'_{min} = S_{min} - s_0$, $S'_{max} = S_{max} - s_0$ and note that $\phi^{s_0}_{u_0,v_0}$ is independent of $s_0$ up to the exact integration boundaries:

$$\phi^{s_0}_{u_0,v_0} = \int_{S'_{min}}^{S'_{max}} \frac{1}{\pi^2 |s'|^2 A^2} D(A|s'|, (s'u_0, s'v_0)) ds'. \tag{12}$$

We show that if the slope boundaries are sufficiently far from the object depth, that is: $S_{min} << s_0 << S_{max}$, the integration boundaries are negligible and $\phi^{s_0}_{u_0,v_0}$ is nearly slope invariant. For a point $(x,y)$ the minimal slope $s$ for which $(x,y)$ is included in the disc $D(A|s'|, (s'u_0, s'v_0))$ is $s_{u_0,v_0}(x,y)$ defined in eq 10. For every $s' > s_{u_0,v_0}(x,y)$, we get an energy contribution proportional to the disc area: $1/(\pi^2 |s'|^2 A^2)$. We get a similar contribution from negative $s'$ values for which $s' < -s_{u_0,v_0}(-x,-y)$. Therefore Eq. (12) can be written as:

$$\begin{aligned} \phi^{s_0}_{u_0,v_0}(x,y) &= \int_{s(x,y)}^{S'_{max}} \frac{1}{\pi^2 |s'|^2 A^2} ds' + \int_{s(-x,-y)}^{-S'_{min}} \frac{1}{\pi^2 |s'|^2 A^2} ds' \\ &= \frac{1}{\pi^2 A^2} \left( \frac{1}{s(x,y)} + \frac{1}{s(-x,-y)} - \frac{1}{S'_{min}} - \frac{1}{S'_{max}} \right) \\ &\rightarrow \frac{1}{\pi^2 A^2} \left( \frac{1}{s(x,y)} + \frac{1}{s(-x,-y)} \right). \end{aligned} \tag{13}$$

where the last approximation is accurate when the integration boundaries $S_{min}, S_{max}$ are large relative to $s_0$. ∎

The proof of Claim 1 shows that the PSFs at different slopes are equivalent up to an additive term of $1/S'_{min} + 1/S'_{max}$. This term is small when the focal stack range $[S_{min}, S_{max}]$ is large with respect to the object depth. For the PSF to be slope-invariant we want $1/S'_{min} + 1/S'_{max}$ to be at the same order of magnitude as the imaging noise. Figure 3(c) visualizes PSFs. Note that the PSF is invariant

Figure 4. Novel viewpoints inside the aperture, rendered from a focal stack. Animation is available on the project webpage.

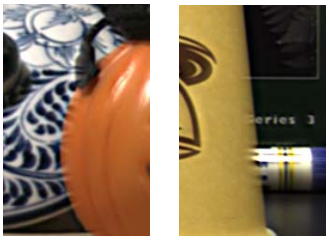


Figure 5. Rendering artifacts at occlusion boundaries, for which the spatially-invariant convolution model fails.

to depth $s_0$ but depends on the (known) viewpoint $(u_0, v_0)$. For the central view $(u_0, v_0) = (0, 0)$ the PSF is the radially symmetric kernel $\phi_{0,0}(x, y) = 1/|(x, y)|$. For other viewpoints the kernel drifts along the viewpoint direction.

**Results:** Fig. 4 shows crops from novel viewpoints generated from a focal stack sequence, consisting of 40 $f/2.0$ images. The sequence covers a slope range of $S_{min} = -0.23, S_{max} = 0.23$ while the actual objects lie in slope range $[-0.09, 0.09]$. Several sequences with viewpoint animations are available on the project webpage[2].

The dimensionality gap model is violated by occlusion boundaries and when the scene is non-Lambertian. Figure 5 zooms in some imperfect reconstructions at occlusion boundaries. In contrast, in our examples we did not detect artifacts caused by non-Lambertian objects. Apparently, most scenes are sufficiently Lambertian within the narrow angle of a camera aperture.

### 3.2. Frequency domain derivation

The Fourier version of our algorithm is straightforward because we have seen that the spectrum of a view with an infinite aperture – an image of a focal stack – is a slice of the 4D light field spectrum. This means that the set of images of the focal stack directly provides the set of slices that comprise the 3D focal manifold. That is, for infinite aperture, according to Eq. (5) $\hat{B}^s(\omega_x, \omega_y) = \hat{L}(\omega_x, \omega_y, s\omega_x, s\omega_y)$. Therefore, given a 3D focal stack data, we can construct the 4D light field spectrum. We place the focal stack spectra at entries $\hat{L}(\omega_x, \omega_y, s\omega_x, s\omega_y)$, and set the rest of the entries to zero. A finite aperture image approximates this and provides a slice from a blurred version of the 4D spectrum

$$\tilde{\hat{L}}(\omega_x, \omega_y, \omega_u, \omega_v) = \hat{L}(\omega_x, \omega_y, \omega_u, \omega_v) \otimes \hat{\psi}(\omega_u, \omega_v), \quad (14)$$

where $\hat{\psi}(\omega_u, \omega_v)$ is the 2D Fourier transform of the aperture (a disc in the primal domain $\psi(u, v) = D(A, 0)$). Below we rederive our rendering algorithm in the frequency domain.

[2] www.wisdom.weizmann.ac.il/~levina/papers/dimgap/

**Claim 2** *Let $\hat{L}_{u_0,v_0}(\omega_x, \omega_y)$ denote the 2D Fourier transform of a desired view $L_{u_0,v_0}(x, y)$. $\hat{L}_{u_0,v_0}(\omega_x, \omega_y)$ equals the average of the Fourier transforms of all shifted focal stack images multiplied by a function $\chi_{\omega_{x,y}}(u_0, v_0)$ which depends on $u_0, v_0, \omega_x, \omega_y$ but does not depend on $s$:*

$$\hat{L}_{u_0,v_0}(\omega_x, \omega_y) = \chi_{\omega_{x,y}}(u_0, v_0) \int_s \hat{B}^s_{su_0,sv_0}(\omega_x, \omega_y) ds. \quad (15)$$

Since convolution is multiplication in the frequency domain, Eq. (15) simply implies that we average the spectra of all shifted focal stack images and deconvolve with a kernel independent of $s$. This is equivalent to the primal domain algorithm described above.

*Proof:* From the definition of the Fourier transform, a 2D view from $(u_0, v_0)$ can be obtained by multiplying the 4D spectrum with the wave $e^{2\pi i(\omega_u u_0 + \omega_v v_0)}$, projecting along the $\omega_u, \omega_v$ dimensions (which provides 2D data), and then computing a 2D inverse Fourier transform. That is:

$$\hat{L}_{u_0,v_0}(\omega_x, \omega_y) = \iint e^{2\pi i(\omega_u u_0 + \omega_v v_0)} \hat{L}(\omega_x, \omega_y, \omega_u, \omega_v) d\omega_u d\omega_v. \quad (16)$$

We define new integration variables $s = (\omega_x \omega_u + \omega_y \omega_v)/|\omega_{xy}|^2$, $t = (\omega_y \omega_u - \omega_x \omega_v)/|\omega_{xy}|^2$. The dimensionality gap implies that $\hat{L}$ is zero for $t \neq 0$. Therefore Eq. (16) is equivalent to:

$$\hat{L}_{u_0,v_0}(\omega_x, \omega_y) = |\omega_{x,y}| \int e^{2\pi i(u_0\omega_x + v_0\omega_y)s} \hat{L}(\omega_x, \omega_y, s\omega_x, s\omega_y) ds. \quad (17)$$

Where the multiplicative factor $|\omega_{xy}|$ is the Jacobian of the new integration variables. The 1D segment of $\hat{L}$ over which we integrate in Eq. (17) is exactly the part covered by the focal stack sequence. For an infinite-aperture focal stack we can substitute Eq. (5) in Eq. (17) and get

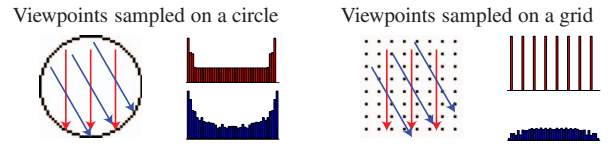
$$\begin{aligned} \hat{L}_{u_0,v_0}(\omega_x, \omega_y) &= |\omega_{xy}| \int e^{2\pi i(u_0\omega_x + v_0\omega_y)s} \hat{B}^s(\omega_x, \omega_y) ds \\ &= |\omega_{xy}| \int \hat{B}^s_{su_0,sv_0} ds, \end{aligned} \quad (18)$$

where the last equality follows from the fact that a phase change in the frequency domain is a shift in the primal domain and therefore $\hat{B}^s(\omega_x, \omega_y)$ times a phase change $e^{2\pi i(u_0\omega_x + v_0\omega_y)s}$ is the Fourier transform of the shifted version $\hat{B}^s_{su_0,sv_0}$. For the infinite aperture case, we choose $\chi_{\omega_{x,y}}(u_0, v_0) = |\omega_{xy}|$ and get the desired Eq. (15). That is, the 2D spectra of the desired view $\hat{L}_{u_0,v_0}(\omega_x, \omega_y)$ is the average spectrum of all shifted focal stack spectra, deconvolved with the slope-invariant function $\chi$.

In the finite aperture case, the focal stack images are slices from a convolved spectrum $\tilde{\hat{L}}$ (defined in Eq. (14)), $B^s(\omega_x, \omega_y) = \tilde{\hat{L}}(\omega_x, \omega_y, s\omega_x, s\omega_y)$. Consider the Fourier transform of $\hat{\psi}$ along a slice:

$$\chi'_{\omega_{xy}}(u_0, v_0) = \int e^{2\pi i(u_0\bar{\omega}_x + v_0\bar{\omega}_y)s} \hat{\psi}(s\bar{\omega}_x, s\bar{\omega}_y) ds, \quad (19)$$

where $(\bar{\omega}_x, \bar{\omega}_y) = (\omega_x, \omega_y)/|\omega_{x,y}|$. We can use the convolution theorem and express convolution with $\hat{\psi}$ as multipli-




(a) Sampling pattern projected in different directions

(b) Reconstructed images

Figure 6. Reconstructing the light field from a sparse sample of viewpoints. For good reconstruction all entries of the sampling pattern projected at any direction should be high. A sampling pattern in a circle provides higher quality reconstruction compared to a grid, because a grid projected vertically (red projection on the right) has many zero entries.

Figure 7. Quadlinear interpolation of a novel viewpoint given the sparse viewpoint sample of Fig 6-right. Objects at the reference plane are recovered well, but away from the reference plane aliasing is observed.

cation with $\chi'$. That is:

$$\int e^{2\pi i(u_0\omega_x+v_0\omega_y)s}\tilde{\hat{L}}(\omega_x,\omega_y,s\omega_x,s\omega_y)ds = \chi'_{\omega_{xy}}(u_0,v_0)\int e^{2\pi i(u_0\omega_x+v_0\omega_y)s}\hat{L}(\omega_x,\omega_y,s\omega_x,s\omega_y)ds \quad (20)$$

Defining $\chi_{\omega_{x,y}} = |\omega_{xy}|/\chi'_{\omega_{x,y}}$ we get the desired relation in Eq. (15). Eqs. (17) and (18) follow in a similar way.

In this derivation we assumed infinite integration boundaries, but as for the primal domain, there is an additional approximation here following from the fact that the focal stack sequence only covers a finite slope range. □

## 4. Novel views from a sparse set of viewpoints

The advantage of the focal stack is that it directly covers the non-zero parts of the spectrum. However, the dimensionality gap prior can help reconstruct the light field from other low dimensional sample sets. In this section we discuss sparse directional samples and in the following one aliased spatial samples. We show that a small adaptation of the algorithm from the previous section applies in these cases as well. The acquisition setup considered here captures only a sparse subset of views on the **uv** plane, instead of a full 2D set of viewpoints captured by classical systems [14, 21].

Reviewing the derivation from the previous section we note that it is valid for any aperture shape, only the definitions of $\hat{\psi}$ and $\chi'$ change. If we are given a subset of directional samples, we can treat them as holes in the aperture, and their union defines a new aperture. That is, $\psi(u,v) = 1$ iff viewpoint $(u,v)$ is captured. From the sparse set of views, we synthetically render a focal stack sequence by integrating along slope $s$ (Eq. (2)). Given the focal stack sequence we apply the exact algorithm described in the previous sections: shift each image by the desired disparity, compute the interpolated image and apply depth invariant deconvolution. However, since the aperture shape is different, we deconvolve with a different PSF.

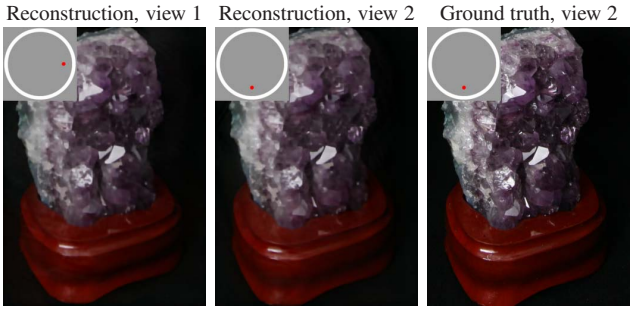



Figure 8. Rendering novel viewpoints from a circular sample of viewpoints. See the project webpage for animation.

In Sec. 3 we have shown that in the frequency domain we need to multiply the $\omega_x, \omega_y$ entry with $\chi_{\omega_{x,y}} = |\omega_{xy}|/\chi'_{\omega_{x,y}}(u_0,v_0)$. Obviously this deconvolution is well posed when $|\chi'_{\omega_{x,y}}(u_0,v_0)|$ is large. Below we derive an exact formula for $\chi'$ and attempt to understand which viewpoint sampling patterns lead to a well posed reconstruction.

**Claim 3** *Let $\rho$ denote a 1D projection of the aperture $\psi$:*

$$\rho_{(cos(\theta),sin(\theta))}(r) = \int \psi(cos(\theta)t+sin(\theta)r, sin(\theta)t-cos(\theta)r)dt. \quad (21)$$

*Then:*

$$\chi'_{\omega_{x,y}}(u_0,v_0) = \rho_{(-\bar{\omega}_y,\bar{\omega}_x)}(u_0\bar{\omega}_x + v_0\bar{\omega}_y). \quad (22)$$

*where* $(\bar{\omega}_x,\bar{\omega}_y) = (\omega_x,\omega_y)/|\omega_{x,y}|$.

*Proof:* $\chi'$ was defined in Eq. (19) as

$$\chi'_{\omega_{xy}}(u_0,v_0) = \int e^{2\pi i(u_0\bar{\omega}_x+v_0\bar{\omega}_y)s}\hat{\psi}(s\bar{\omega}_x,s\bar{\omega}_y)ds. \quad (23)$$

That is, we take a 1D slice from $\hat{\psi}$ along the direction $(\bar{\omega}_x,\bar{\omega}_y)$ and compute its inner product with a wave. This is equivalent to computing a specific entry (which depends on $(u_0,v_0)$) from its Fourier transform. According to the Fourier slice theorem [18], this is equivalent to projecting the primal aperture $\psi$ along the orthogonal direction. The aperture projection $\rho$ is defined in Eq. (21) and $\chi'_{\omega_{x,y}}(u_0,v_0)$ is simply $\rho_{(-\bar{\omega}_y,\bar{\omega}_x)}(u_0\bar{\omega}_x + v_0\bar{\omega}_y)$. □

Therefore, to achieve a stable deconvolution at all viewpoints $(u_0,v_0)$, we require that all entries of $\rho$ be high for every projection direction. This understanding allows us to compare viewpoint sampling patterns. If we distribute the sampled viewpoints on a grid (Fig 6, right), the projection has many zero entries in the harmonic directions (e.g. vertical projection illustrated in Fig 6). To obtain stable inversion, we want a sampling pattern with dense projections at every direction. Such projections can be obtained with a pseudo-random noise pattern. Another option is to sample a circle of viewpoints (Fig 6, left). We chose a circular


(a) Aliased light field spectrum (b) Bicubic (c-d) Given depth (e) Dimensionality gap

Figure 9. (a) The spectrum of an aliased light field, demonstrating replicas of focal segments. (b-e) Upsampling of a view from the aliased light field of [15]. (b) Bicubic up sampling, corresponding to a 4D prior. (c-d) The upsampling of [15], corresponding to a 2D prior conditioned on slope. High resolution is obtained at the correct depth, but other depths are blurred (tree blurred in (c), car blurred in (d)). (e) Our upsampling with a 3D prior. We improve resolution at all depths, but not as well as (c-d) which account for one given depth.

sample in our implementation because it has another useful property: it can be shown that the depth invariant PSF in the primal domain is equivalent to the one obtained from a normal focal stack sequence, which was derived in claim 1. Figure 6(b) demonstrates two novel views generated from such sampling patterns. Not surprisingly, the reconstruction from circular samples is better than from a grid.

In Figure 7 we also applied a standard quadlinear ray interpolation [14] to interpolate viewpoints from the sparse viewpoints grid of Fig 6-right. Objects on the reference plane are sharp, but away from it aliasing is observed.

**Results:** In figure 8 we used light fields from the Stanford dataset [3]. The data include grids of $17 \times 17$ views. However, we used only an outer circle of 64 views and can render any view in the interior of this circle. Viewpoint animation is available on the project webpage.

## 5. Spatial resolution enhancement

We have seen that using the dimensionality gap prior one can interpolate sparse directional samples. We now show that a similar algorithm applies to the spatial dimension. Let $\hat{L}(\omega_x, \omega_y, \omega_u, \omega_v)$ denote a high-resolution light field with spatial resolution $k\Omega$ (i.e. $|\omega_x| \leq k\Omega/2$, $|\omega_y| \leq k\Omega/2$). We measure a spatially aliased version $\hat{L}^0(\omega_x, \omega_y, \omega_u, \omega_v)$, whose spatial dimension is under-sampled by a factor of $k$, i.e., in $\hat{L}^0$, $\omega_{x,y}$ are in the range $|\omega_x| \leq \Omega/2$, $|\omega_y| \leq \Omega/2$. We want to infer $\hat{L}$ from the measured $\hat{L}^0$. While previous work demonstrated light field super resolution involving depth knowledge [15, 3], our goal is to increase resolution without depth estimation.

Aliasing due to under-sampling causes $\hat{L}^0$ to be composed of a sum of replicas from $\hat{L}$:

$$\hat{L}^0(\omega_x, \omega_y, \omega_u, \omega_v) = \sum_{i,j=0}^{k-1} \hat{L}(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v), \quad (24)$$

where the shifted frequencies $\omega_x + i\Omega$, $\omega_y + j\Omega$ are taken modulo $k\Omega/2$. Informally, aliasing means that the coefficients of the missing high frequencies are scattered somewhere in the light field. In the general case (under a 4D prior assumption) there is no way to tell them apart from the primary spectrum. However, under a 3D prior we know that most coefficients should be zero. Therefore, a non-zero coefficient observed away from the focal segments is a likely replica.

Eq. (24) defines the $T$ measurement matrix from Eq. (8), and we obtain the Bayesian reconstruction of $\hat{L}$ under a Gaussian prior as the minimization of

$$\frac{1}{\eta^2} \left| \hat{L}^0(\omega_x, \omega_y, \omega_u, \omega_v) - \sum_{i,j=0}^{k-1} \hat{L}(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v) \right|^2 + \sum_{i,j=0}^{k-1} \frac{|\hat{L}(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v)|^2}{\sigma(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v)^2} \quad (25)$$

That is, a Bayesian reconstruction redistributes the value of $\hat{L}^0(\omega_x, \omega_y, \omega_u, \omega_v)$ between the replica entries in proportion to their variance. In a nutshell, if the prior variance of one replica entry is sufficiently higher than others $(i^*, j^*) = \arg\max \sigma(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v)$, the reconstruction assigns the measured $\hat{L}^0(\omega_x, \omega_y, \omega_u, \omega_v)$ value to $\hat{L}(\omega_x + i^*\Omega, \omega_y + j^*\Omega, \omega_u, \omega_v)$ and almost zero to all other replica entries.

If a 4D Gaussian prior is used, the variance is higher for small spatial frequencies and the highest variance is obtained at $i^* = 0, j^* = 0$. Therefore, the reconstruction copies $\hat{L}^0$ in the low frequencies of $\hat{L}$ and zero at all new high frequency entries. No extra information is gained from the aliasing. Figure 10(a) visualizes the expected reconstruction error, which is indeed low in the middle and high at the periphery. On the other hand, if the depth is known and we use a 2D Gaussian prior, there is non-zero variance only for entries of the form $\omega_u = s\omega_x, \omega_v = s\omega_y$. In this case one can obtain a significantly better reconstruction [15]. If $|s| \neq 0$ the replicas do not cover each other, that is, the set of entries $\{(\omega_x + i\Omega, \omega_y + j\Omega, \omega_u, \omega_v)\}_{i,j=1}^{k}$ do not contain more than one entry of the form $\omega_u = s\omega_x, \omega_v = s\omega_y$, and the required frequency coefficients can be recovered (Fig. 10(b)). In fact, [15] extracts a focused view from the aliased samples using that exact property, but it applies the primal domain version of it: projecting, instead of slicing.

Our 3D Gaussian prior can resolve replicas better than the 4D prior, but not as accurately as a 2D prior does when depth is known. An $\omega_{x_0,y_0}$-slice of the aliased light field $\hat{L}^0$ is the sum of $k^2$ $\omega_{x_0,y_0}$-slices of $\hat{L}$. For each $\omega_{x,y}$-slice, the prior assigns non-zero variance to a 1D set of entries

[3] http://lightfield.stanford.edu/lfs.html

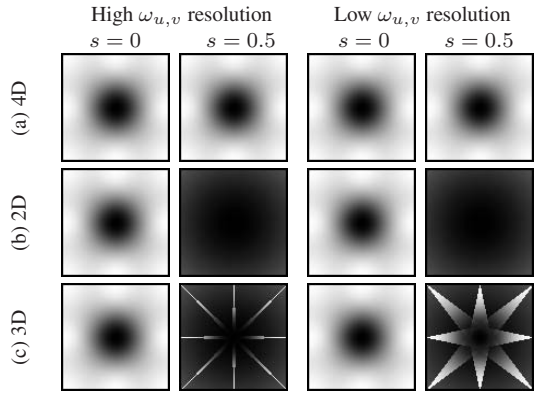



Figure 10. Expected reconstruction error for different priors, for an upsampling factor $k = 2$. We plot reconstruction error for 2D spectrum slices of the form $\hat{L}(\omega_x, \omega_y, s\omega_x, s\omega_y)$, providing the reconstruction quality for a 2D image of an object at slope $s$ (the axes of each subplot vary $\omega_x, \omega_y$). The 4D prior obtains high error at reconstructing high spatial frequencies. For slope $s = 0.5$, 2D and 3D priors can reduce the error, but our 3D prior fails to recover the harmonic directions. At slopes $s = 0$, all priors cannot resolve high frequencies. The sensitivity around the harmonic orientations reduces when the resolution on the $\omega_u, \omega_v$ axes is higher (left).

on the focal segment, whose orientation is $(\overline{\omega_x}, \overline{\omega_y})$. In the subplots of Fig. 9(a) we can clearly observe replicas of focal segments at different orientations. The replicas are well separated when for different $(i, j)$ values, the orientations of the lines $(\overline{\omega_x + i\Omega}, \overline{\omega_y + j\Omega})$ are sufficiently different. The focal segment orientations are different for most $(\overline{\omega_x}, \overline{\omega_y})$ orientations, but not at the harmonic directions. For example, if $\omega_x = 0$, the segments $(\overline{\omega_x + 0\Omega}, \overline{\omega_y + j\Omega})$ are vertical for all $j$ values, which means that the replicas cover each other and cannot be resolved. Indeed, the reconstruction error in figure 10(c) is low except at the harmonic directions. Figure 10 also illustrates that the sensitivity around the harmonic directions reduces when the focal segments are thinner. As discussed in Sec. 2.3, this happens when the $\omega_u, \omega_v$ axes resolution is higher.

While the replica analysis is carried in the frequency domain, we prefer to work in the primal domain to avoid boundary artifacts. As an approximation, we use the algorithm from the previous sections: generate a focal stack sequence (which are also the super-resolved images of [15] for given depths), average and apply depth-invariant deconvolution. Fig. 9 illustrates super-resolution on the data of [15]. Given the right depth the 2D Gaussian prior produces the best results, but objects at other depths are highly blurred. Our 3D Gaussian prior improves resolution at all depths. While the quality is worse than the 2D Gaussian, it is significantly better than naive upsampling with a 4D prior.

## 6. Discussion

In this paper we have proposed a new light field prior derived from the dimensionality gap: a Gaussian assigning non-zero variance mostly to a 3D subset of frequencies. Since only three degrees of freedom exist, capturing 3D data is sufficient and there is no need to sample the entire 4D space explicitly. The fact that the prior is Gaussian allows for simple depth invariant reconstruction algorithms.

**Acknowledgments:** The authors acknowledge B.S.F. support. A. L. acknowledges I.S.F. support. F. D. acknowledges NSF CAREER award, Shell and Quanta support.

## References

[1] E. Adelson and J. Wang. Single lens stereo with a plenoptic camera. *IEEE PAMI*, 1992.

[2] K. Aizawa, K. Kodama, and A. Kubota. Producing object-based special effects by fusing multiple differently focused images. *CirSysVideo*, 10(2):323, March 2000.

[3] T. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *ICCP*, 2009.

[4] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen. Unstructured lumigraph rendering. In *SIGGRAPH*, 2001.

[5] J. Chai, X. Tong, S. Chan, and H. Shum. Plenoptic sampling. *SIGGRAPH*, 2000.

[6] S. Chen and L. Williams. View interpolation for image synthesis. *SIGGRAPH*, 1993.

[7] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *SIGGRAPH*, 1996.

[8] S. Hasinoff and K. Kutulakos. Light-efficient photography. In *ECCV*, 2008.

[9] G. Hausler. A method to increase the depth of focus by two step image processing. *Optics Communications*, page 3842, 1972.

[10] A. Isaksen, L. McMillan, and S. Gortler. Dynamically reparameterized light fields. *SIGGRAPH*, 2000.

[11] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. *SIGGRAPH*, 2007.

[12] A. Levin, W. Freeman, and F. Durand. Understanding camera trade-offs through a Bayesian analysis of light field projections. In *ECCV*, 2008.

[13] A. Levin, S. Hasinoff, P. Green, F. Durand, and W. Freeman. 4D frequency analysis of computational cameras for depth of field extension. *SIGGRAPH*, 2009.

[14] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH*, 1996.

[15] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *ICCP*, 2009.

[16] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering approach. In *SIGGRAPH*, 1995.

[17] H. Nagahara, S. Kuthirummal, C. Zhou, and S.K. Nayar. Flexible Depth of Field Photography. In *ECCV*, 2008.

[18] R. Ng. Fourier slice photography. *SIGGRAPH*, 2005.

[19] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Stanford U. Tech Rep CSTR 2005-02*, 2005.

[20] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask-enhanced cameras for heterodyned light fields and coded aperture refocusing. *SIGGRAPH*, 2007.

[21] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Levoy, and M. Horowitz. High performance imaging using large camera arrays. *SIGGRAPH*, 2005.