A Light Transport Framework for Lenslet Light Field Cameras

Chia-Kai Liang
Lytro Inc
and
Ravi Ramamoorthi
University of California, San Diego

Light field cameras capture full spatio-angular information of the light field, and enable many novel photographic and scientific applications. It is often stated that there is a fundamental tradeoff between spatial and angular resolution, but there has been limited understanding of this tradeoff theoretically or numerically. Moreover, it is very difficult to evaluate the *design* of a light field camera, because a new design is usually reported with its prototype and rendering algorithm, all of which affect resolution.

In this paper, we develop a light transport framework for understanding the fundamental limits of light field camera resolution. We first derive the prefiltering model of lenslet-based light field cameras. The main novelty of our model is in considering the full space-angle sensitivity profile of the photosensor—in particular, real pixels have non-uniform *angular sensitivity*, responding more to light along the optical axis, rather than at grazing angles. We show that the full sensor profile plays an important role in defining the performance of a light field camera. The proposed method can model all existing lenslet-based light field cameras and allows us to compare them in a unified way in simulation, independent of the practical differences between particular prototypes. We further extend our framework to analyze the performance of two rendering methods: the simple projection-based method and the inverse light transport process. We validate our framework with both flatland simulation and real data from the Lytro light field camera.

ckliang@lytro.com, ravir@cs.ucsd.edu

1. INTRODUCTION

In recent years, plenoptic or light field cameras have increased in popularity, with multiple research systems [Veeraraghavan et al. 2007; Bishop and Favaro 2012], and commercial and consumer models becoming available from Raytrix and Lytro. These cameras capture full 4D spatio-angular information of a light field [Adelson and Bergen 1991; Levoy and Hanrahan 1996; Gortler et al. 1996], typically by placing additional optical elements between the main lens and the sensor. Light field cameras enable new applications beyond the reach of conventional 2D cameras, such as refocusing images after capture [Ng et al. 2005], or acquiring 3D depth from a single shot [Adelson and Wang 1992]. However, a significant disadvantage is the loss in image resolution, to a small fraction of the resolution provided by the camera sensor. For example the resolution of the 2D refocused image, in the basic lenslet camera [Ng et al. 2005], is reduced to the number of lenslets (reduced by a factor of the number of pixels under each lenslet), and is usually up to 100 times smaller than the number of sensor pixels.

To overcome this limitation, various light field camera designs have been proposed (see Sec. 2 for more details), but typically involve other tradeoffs. Moreover, comparing different optical designs of light field cameras remains a difficult and open problem. Most designs are presented with their own prototypes and software rendering algorithms. The cost and quality of the optics and sensors in those prototypes can differ significantly. The complexity of the rendering algorithms can also vary from simple re-sampling to computationally intensive prior-assisted deconvolution. It is hard to tell if the design of one light field camera is intrinsically better than another from the presented results, since the rendering algorithm itself affects resolution.

In this paper, we develop a theoretical framework based on light transport analysis for understanding the fundamental limits of lenslet-based light field cameras in a unified way in simulation. The main distinction to existing models, each usually derived for a specific design, is that we consider all parameters in the optical system. These include the main lens aperture size, the lenslet aperture size and spacing, and the photosensor spacing and its *full spatio-angular profile*. Real pixels have non-uniform *angular sensitivity*, responding more to light along the optical axis than at grazing angles, but this has not been taken into account in previous analyses. We first use our framework to derive the prefilter kernel for light field cameras. The key finding is that in all designs, the prefilter kernel is not only *non-perfectly bandlimited*, but also *depth-dependent* and even *spatially-variant* due to the finite pitch size and non-uniform angular sensitivity of the photosensor. Therefore, the expected resolution limit should be above the lenslet resolution, and the depth- and spatially-variant nature of the captured light field must be taken into account in algorithm designs and evaluations.

Since the light field must be further processed for display, we extend our framework to analyze two main categories of rendering algorithms: the simple projection-based algorithm [Kitamura et al. 2004; Chan et al. 2007; Perez Nava and Luke 2009; Georgiev et al. 2011] and the inverse light transport (deconvolution) process [Bishop and Favaro 2012; Shroff and Berkner 2013; Broxton et al. 2013]. For the projection-based algorithm, we derive the overall filtering kernel for the rendered image and show that it can generate high-resolution results for most designs. We show the full depth-dependent frequency response of rendered images of various designs by flatland simulation and data from the Lytro light field camera. Finally, we also show how to extend this algorithm to generate high-resolution all-in-focus images.

For the inverse light transport process, we extend our framework to construct the light transport matrix and analyze the stability of the inversion process. While most existing reports include additional priors to regularize the ill-conditioned process, we deliberately leave that out to compare different light field camera designs in a *content-independent* way. We report the depth-dependent stability for many designs by simulation.

Our goal is to understand the resolution limits from a theoretical perspective, while providing many insights into current practical systems and future designs. However, we do not claim to precisely evaluate real-world resolution profiles of light field cameras, since those also depend on the performance of the actual optics and practical software rendering algorithms. Our results are developed analytically using 2D flatland light fields for simplicity (the extension to 4D light fields is straightforward), and with numerical simulations that allow us to contrast different light field camera designs.

In summary, we make the following contributions in this paper:

- We present the first general framework to model all lenslet-based light field cameras, considering the full spatial-angular profile of the photosensor and other parameters (Sec. 4).
- We use this light transport framework to identify the non-bandlimited, depth-dependent, and spatially-variant prefiltering behavior of light field cameras (Sec. 5).
- We analyze the performance of projection-based rendering algorithms, including the theoretical filtering kernel, results from flatland simulation of various designs, and real light field data from the Lytro light field camera (Sec. 6).
- We extend the projection-based algorithm for generating all-in-focus images and show results from both flatland simulation and real light field data (Sec. 6).
- We provide the full experimental depth-dependent performance profile of the Lytro light field camera, showing its achievable resolution is above lenslet resolution across a large refocusable range, even without deconvolution (Sec. 7).
- We extend the framework to study the stability of the inverse light transport (deconvolution) process for different light field camera designs and parameters (Sec. 8).

2. RELATED WORK

Light Field Capture: Development of light field cameras, or plenoptic cameras, can be dated back to more than a century ago. Those early designs place a fly-eye lens array or a slit plate in front of the film [Lippmann 1908; Ives 1903]. This topic recently regained much attention after the theories for analyzing and processing light fields were developed [Adelson and Bergen 1991; Levoy and Hanrahan 1996; Gortler et al. 1996; Isaksen et al. 2000], and many designs and prototypes have been proposed.

In the basic lenslet-based design [Adelson and Wang 1992], a lenslet or microlens array is placed in front of the photosensor array by one focal length of the lenslet. Ng et al. built a portable prototype of this design and demonstrated various new photographic applications, such as post-capture refocusing [2005]. One noticeable limitation of such a design is that the resolution is limited to the number of lenslets, much lower than the sensor resolution. In the following discussion, we will call the resolution provided by the number of lenslets, or other elements that define the spatial sampling rate, as *lenslet resolution*.

To provide more control over resolution, Ng [2006] proposed the *generalized light field camera*, in which the lenslet-photosensor separation can be reduced, and thus the peak spatial resolution and the refocusable depth range are adjustable. Lumsdaine and Georgiev [2009] proposed the *focused light field camera* by increasing the lenslet-photosensor separation and observed similar trade-offs. Perwaß and Wietzke derived the resolution bound of focused light field cameras using a simplified model [2012]. In this work, we develop a general mathematical model for all lenslet-based designs and compare their performance in simulation.

In *heterodyne light field camera*, Veeraraghavan et al. placed a mask layer with a sum of sinusoids in front of the photosensor. The modulation due to the mask creates periodic replicas of the light field in Fourier space [2007]. The light field can be reconstructed by properly rearranging the spectral samples. Lanman et al. improved the mask design [2008], and Wetzstein et al. showed that the reconstruction can be performed in the spatial domain [2013]. However, these designs and processing are based on the bandlimit assumption, and the output is limited to the lenslet resolution.

Levin and Durand exploited the dimensionality gap of the 4D light field [2010] and proposed an efficient frequency-domain algorithm to reconstruct high-resolution light fields from the focus stack or aliased light field, without per-pixel depth information. They do not consider the depth and spatially-dependent prefiltering behavior of the light field camera as we will derive in this paper. We also analyze the performance and stability of the spatial-domain rendering algorithms which use depth information.

Light Field Processing: Bishop and Favaro first modeled the image formation process of the lenslet-based light field cameras using geometric optics [2012]. They also combined depth estimation and image priors to perform deconvolution. Wanner and Goldluecke increased both the spatial and angular resolution using depth estimation and convex optimization [2012b].

Shroff and Berkner derived the forward image formation model of the basic lenslet-based design using wave optics [2013]. They showed that each photosensor has a unique point-spread-function and further performed deconvolution to recover high-resolution images in simulation. Broxton et al. derived a similar model for light field microscopy and 3D deconvolution [2013]. In this paper, we reach a similar model using geometric optics. We use our model to compare different designs and analyze the stability of the inverse light transport process.

It is empirically found that if one directly projects the recorded light field samples to a finer grid, the perceived resolution can be higher than the lenslet resolution. This is observed in the basic lenslet-based design [Perez Nava and Luke 2009], the focused light field camera [Georgiev et al. 2011], or even camera arrays [Kitamura

Table I. Notation

Symbol	Description
$\hat{f}$	Fourier transform of f
x	Spatial coordinate
u	Angular coordinate
$\mathbf{x} = [x, u]^T$	Light field coordinate
$\mathbf{\Omega} = [\Omega_x, \Omega_u]^T$	Frequency coordinate
$\text{rect}(x/d)$	1 if $|x| < 0.5d$, 0 otherwise.
Light field transformation matrix	
$\mathbf{M}_d = \begin{bmatrix} 1 & -d \\ 0 & 1 \end{bmatrix}$	Translation by distance d
$\mathbf{R}_f = \begin{bmatrix} 1 & 0 \\ f^{-1} & 1 \end{bmatrix}$	Thin-lens refraction with focal length f

et al. 2004; Chan et al. 2007]. Yu et al. analyzed the resolution enhancement factor using the distribution of the projected samples in the 2D space, and proposed a light field aware demosaic algorithm [2012].

Venkataraman et al. [2013] constructed a camera array in the scale for a mobile module. They reduced the pixel aperture in their prototype to preserve the high-frequency details. They also use the projection-based method as the initial estimate for the following complex, iterative reconstruction process. Marwah et al. [2013] represented the local light field as a sparse combination from an overcomplete dictionary, and designed a mask-based light field camera to sample the light field in a compressive way.

Compared to existing work, our light transport framework considers all parameters in the optical system, including the full spatial-angular photosensor profile. We also characterize the performance of the projection-based algorithm and the stability of the inverse process. These theoretical analyses can be integrated into future light field designs or reconstruction algorithms.

Finally, researchers have exploited the unique structure of the light field spectrum to design effective 4D filters [Dansereau and Bruton 2007; Dansereau et al. 2013]. However, the resulting spatial resolution is limited to the lenslet resolution. We believe that our new light transport model can be combined with existing work to design more advanced filters.

Light Transport Analysis: The light field transforms as it propagates through space or interacts with elements in the scene. Light transport analysis formulates these transformations and exploits the structure of the transformed light field for various applications, including synthetic image or light field rendering [Chai et al. 2000; Durand et al. 2005; Egan et al. 2009; Egan et al. 2011; Lehtinen et al. 2011; Jarosz et al. 2012; Belcour et al. 2012], processing for light field cameras and displays [Ng 2005; Zwicker et al. 2006; Levin and Durand 2010; Wetzstein et al. 2012], image formation modeling [Levin et al. 2009; Liang et al. 2011], and inverse transport analysis [Ramamoorthi and Hanrahan 2001; Seitz et al. 2005]. Our approach leverages these foundational analyses, and focuses on analyzing the resolution of lenslet-based light field cameras.

3. BASIC FOURIER SPECTRUM ANALYSIS

In this section, we briefly review the existing light field spectrum analysis [Chai et al. 2000; Durand et al. 2005; Ng 2005; Veeraraghavan et al. 2007; Levin and Durand 2010], which motivates us to develop the complete light transport analysis in the spatial domain. This section also provides background and introduces the key notation used in the paper, shown in Table I. We analyze the
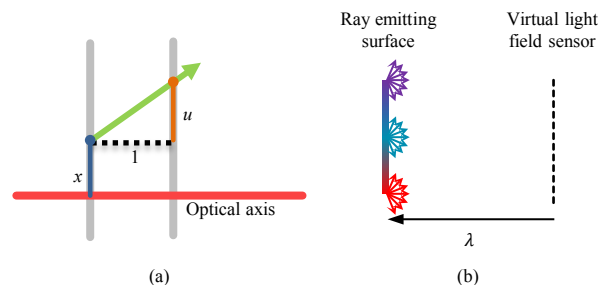

Fig. 1. (a) The local two-plane parameterization. (b) The simplified optical configuration for the spectrum analysis. A light field emitted from a Lambertian surface is sampled by a virtual light field sensor at λ units away.

flatland 2D space-angle light field (schematic is in Figure 1)—the insights carry over in a straightforward way to the 4D light field in three dimensions. Because most light field camera designs do not modify the main lens, we consider the light field *inside* the camera as in most previous work. The effects of the main lens will be considered in the full light transport analysis in the next section.

3.1 Setup

We use the local two-plane light field parameterization shown in Figure 1(a). Each light ray is represented by its intersections with two virtual parallel planes orthogonal to the main optical axis of the whole system. The second plane is one unit away from the first one. The *spatial* coordinate x measures the distance between the first intersection and the optical axis, and the *angular* coordinate u measures the offset from x. Considering the Lambertian surface at distance λ away from the light field sensor (Figure 1(b)), its surface light field is [1]

$$l_\lambda(\mathbf{x}) = t_\lambda(x), \tag{1}$$

where t_λ denotes the texture function of the surface. The surface light field would propagate by λ to reach the sensor. The observed light field and its spectrum $\hat{l}$ are:

$$l(\mathbf{x}) = l_\lambda(\mathbf{M}_\lambda \mathbf{x}) = t_\lambda(x - \lambda u), \tag{2}$$

$$\hat{l}(\mathbf{\Omega}) = \hat{l}_\lambda(\mathbf{M}_\lambda^{-T} \mathbf{\Omega}) = \hat{t}_\lambda(\Omega_x)\delta(\lambda\Omega_x + \Omega_u), \tag{3}$$

where $\mathbf{M}_\lambda$ is the light transport matrix due to translation [Gerrard and Burch 1975; Durand et al. 2005] (Table I), and δ denotes a Dirac delta function.

We can see that the energy of $\hat{l}$ only falls on a line of slope $-\lambda$ through the origin in the 2D Fourier space. When the objects in the scene lie within a depth range $\lambda \in [-\Lambda, \Lambda]$, $\hat{l}$ would be the sum of all $\hat{l}_\lambda$'s and contains energy within a double wedge bounded by two lines of slope $-\Lambda$ and Λ (Figure 2(a)). [2]

3.2 Aliasing and Prefiltering

A light field camera samples the light field along both the spatial and angular domains, which define a periodic sampling lattice over

---

[1] We allow the light field to transport backwards when $\lambda < 0$. This happens when the object is focused behind the sensor by the main lens.

[2] Occlusions can introduce discontinuities into the light field and make it non-bandlimited, but their numerical effect is generally small.

(a) Light field spectrum with depth range $[-\Lambda, \Lambda]$

(b) Sampling (a) without prefiltering

(c) Sampling (a) with bandlimit prefiltering

(d) Sampling (a) with a different bandlimit prefiltering

(e) Sampling (a) with a different spatial-angular sampling rate and prefiltering

(f) Local light field without aliasing

(g) Local light field with the aliasing of replicas onto itself

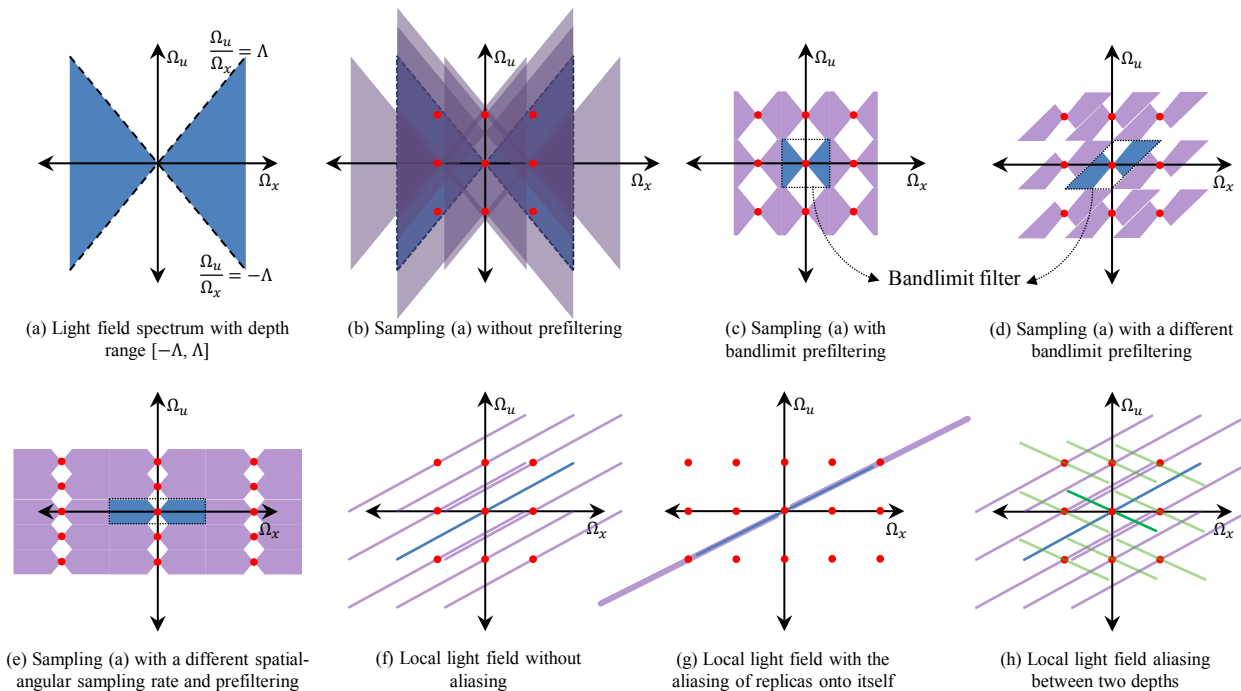(h) Local light field aliasing between two depths

Fig. 2. (a) The source light field spectrum from a scene with objects at depths ranging from $-\Lambda$ to $+\Lambda$. (b) The spectrum of the sampled light field in (a) contains severe aliasing. The red dots represent the centers of replicas. (c) The sampled spectrum with a pre-sampling bandlimit filter applied. The dotted rectangle shows the passband of the ideal bandlimit filter. (d) In the focused light field camera, the pre-sampling filter is sheared to preserve more high frequency components at certain depths. (e) Double the spatial sampling rate and half the angular sampling rate would change the required prefilter shape, compared to (c). (f) The sampled spectrum of a local light field with constant depth. (g) The sampled spectrum of a local light field at another depth. Aliasing arises since the central replica now touches others, but the distance between aliasing replicas is larger than the spatial Nyquist rate. (h) The sampled spectrum of a local light field with two depth layers. Aliasing arises at a few specific frequencies.

the 2D space. The spectrum of the sampled light field would have periodic replicas over the entire Fourier space [Chai et al. 2000]. If no proper prefiltering is applied before sampling, replicas would overlap and cause severe aliasing (Figure 2(b)).

To prevent aliasing, a prefilter must be applied before sampling to bandlimit the light field to the spatial and angular Nyquist frequencies (Figure 2(c)). If the prefilter is designed perfectly, spectral components outside the prefilter bandwidth cannot be reconstructed. Traditional analysis for the basic light field camera assumes such a perfect prefilter is applied before sampling, and thus the output resolution is limited to the lenslet resolution [Ng 2005].

A few lenslet-based designs address this problem by changing the optical configuration [Ng 2006; Lumsdaine and Georgiev 2009]. However, it is easy to show that those designs basically change the shape of the prefilter, but not the overall sampling density [Lumsdaine et al. 2012] (Figure 2(d)). The spatial bandwidth of the prefilter can increase for certain depths but decrease for others, and thus reduce the overall refocusable range. On the other hand, if we reallocate the sampling budget to increase the spatial sampling rate and decrease the angular one, we may increase the spatial bandwidth of the prefilter. However, the angular bandwidth would decrease, and the range of depths not being affected by the angular prefiltering is reduced (Figure 2(e)).

In summary, if a light field camera is designed to be entirely aliasing-free, the prefilter would eliminate most high-frequency details at the current optics and sensor technology. Simple reshaping of the prefilter kernel or reallocating the spatial and angular sampling rates cannot fundamentally improve the situation.

3.3 Localized Spectrum Analysis and Implications

If we analyze the light field locally within a small window, its depth range can be much smaller than that of the entire light field. Moreover, the depth complexity can be reduced to a patch with constant depth or a few layers with distinct depths. We illustrate a few cases in Figure 2(f)-(h). For a local region at a specific constant depth, even without prefiltering, its spectrum would not overlap with the replicas (Figure 2(f)). In this case, we may reconstruct the local light field without unmixing the aliased data. Aliasing can still arise with the constant depth, but the aliased replica is not necessarily the nearest one (e.g. Figure 2 (g)), and thus the anti-aliasing filter can have wider bandwidth. Finally, aliasing may happen when there are multiple layers in the local region (Figure 2(h)). However, depending on the local scene configuration, the aliasing can happen only at a few specific frequencies.

These examples provide a few interesting implications. First, even when the prefilter of a light field camera does not perfectly bandlimit the light field, for a local region, many frequency compo-

nents above the spatial and angular Nyquist rates may still not be corrupted by the replicas. It is possible to recover those components without sophisticated unmixing processes. Second, even when aliasing happens, one can design the anti-aliasing filter to match the local depth configurations and preserve most high-frequency details (like a notch filter rather than low-pass filter). Note that we do observe such cases in the Lytro light field camera, as shown later in Sect. 7.

The main challenge for exploiting those properties is that prior knowledge of the local depth is required, and the reconstruction processing is spatially-variant. Fortunately, there exist many algorithms to extract per-sample depth from light fields [Liang et al. 2008; Wanner and Goldluecke 2012a; Kim et al. 2013; Tao et al. 2013]. To explore these implications, we are motivated to develop a general light transport framework to model the light field camera prefiltering kernels, and show their interactions with the depth-dependent reconstruction algorithms. These are the main contributions of the paper.

4. LIGHT TRANSPORT ANALYSIS

In this section, we derive a general framework to model the prefilter kernels for existing lenslet-based light field camera designs. We consider all parameters in the optical systems, in particular, the full space-angle sensitivity profile of the photosensor, including both its *spatial support* and *angular sensitivity*. Typically, the angular sensitivity peaks for rays along the optical axis and falls off in other directions. Most previous work assumes pixels have uniform angular sensitivity, but real pixels have spatial and angular profiles depending on the particular sensor.

We use the developed framework to show that the prefilter is depth-dependent in all designs, and generally not as bandlimited as the previous studies assume. Therefore, the sampled light field would contain frequencies higher than the Nyquist rate. This is not merely theoretical; in Sec. 7, we show that real systems already exploit this feature to obtain higher resolutions. Moreover, we show that unlike for the idealized camera, the prefiltering kernel is *spatially-varying*, which precludes analytic Fourier theory, but we can still simulate prefiltering kernels numerically, and obtain insights into resolution. We present the simulation in Sec. 5.

Assumptions: We derive the light transport from a Lambertian textured surface at a fixed depth, through a lenslet light field camera. Note that the light transport analysis derives the prefilter kernel at a single photosensor pixel, as opposed to the earlier Fourier spectrum analysis which considers the global light field. Therefore, the fixed depth assumption only means depth is locally-constant within the area *seen by a single photosensor*. This assumption simplifies the derivation and is invalid only when a single photosensor receives light rays from multiple surfaces or a surface with very large depth variation in the observed region. In Sec. 6.5, we also discuss algorithmic extensions for depth variation and occlusion. Finally, the theory restricts itself to intensity light fields, and does not explicitly consider Bayer patterns or demosaicing. The color channels can simply be handled separately in the standard way. Our practical verification on real scenes in Fig. 15 includes a variety of complex surfaces and occlusions, and moderate non-Lambertian reflectance.

4.1 Derivation

Our goal is to derive a direct relationship between the recorded output light field samples at the photosensor, and the texture function t_λ that defines the in-camera input light field from the Lambertian surface at depth λ. If each output light field sample only gathers light rays from a small spatial support in t_λ, it means the texture function is weakly filtered before sampling, and thus high-frequency components are preserved—refocused images can achieve high resolution. Conversely, if there is a large spatial support, the texture will be strongly filtered, and resolution is reduced.

We consider the general lenslet-based light field camera design shown in Figure 3. The system consists of a main lens with focal length f, aperture width A, and F-number $F = f/A$. The lenslet array consists of lenslets with identical focal length f_m and aperture width d. The distance between two lenslet centers is g. The pitch size, or the active area, of a photosensor is p, and the distance between two photosensor centers is h. The source surface is λ away from the lenslet array, and the lenslet-photosensor array separation is α. Note that in our analysis, we do not require that the F-number of the main lens matches that of the lenslet. This gives us more flexibility in design.[3]

We assume the main lens is a thin-lens which transforms the light field from the world space into the camera. Therefore, the only difference of the in-camera light field (parameterized on the virtual surface) to (1) is that light rays outside the aperture are blocked (Figure 3(b)). To account for this effect, we refine (1) to[4]

$$l_\lambda(\mathbf{x}) = t_\lambda(x)\text{rect}(Fu). \qquad (4)$$

We also consider the full spatial and angular profile of the photosensor in our analysis; the sensitivity of each sensor is a spatial- and angular-variant function $s_c(\mathbf{x})$, where c is the index of the photosensor. Since spatial and angular factors are usually independent, we can decompose s_c into two one-dimensional functions:

$$s_c(\mathbf{x}) = \text{rect}\Big(\frac{x - x_c}{p}\Big)\rho(u), \qquad (5)$$

where x_c denotes the center of the photosensor and p denotes its pitch size, which can be smaller than the inter-sensor distance h (Figure 3(d)). $\rho(u)$ is the angular sensitivity function, which can be extracted from the sensor specification. Note that in practice, it is impossible to manufacture a photosensor with constant angular profile. Besides foreshortening, the photodetector in the CMOS sensor is usually buried under a deep dielectric tunnel formed by multiple metal layers, and this pixel vignetting can reduce the overall optical efficiency [Catrysse and Wandell 2002; El Gamal and Eltoukhy 2005]. We will show that the full spatial-angular profile has a strong influence on the performance of a light field camera.

Given the source light field (parameterized at the sensor), and the photosensor profile, we can describe each pixel c as the integral of

---

[3]In practice, the chief ray angle, the F-number, or even the shape of the effective aperture can be spatially variant across the sensor, but we ignore those variations in our derivation. We have not seen any light field camera design utilize those variations yet.

[4]This is strictly correct only on the optical axis, and the center of the rect function can shift at other sensor locations due to variation of the chief ray angle (similar arguments hold for photosensor angular sensitivity). However, the shift is very small in practical systems, and can usually be ignored.
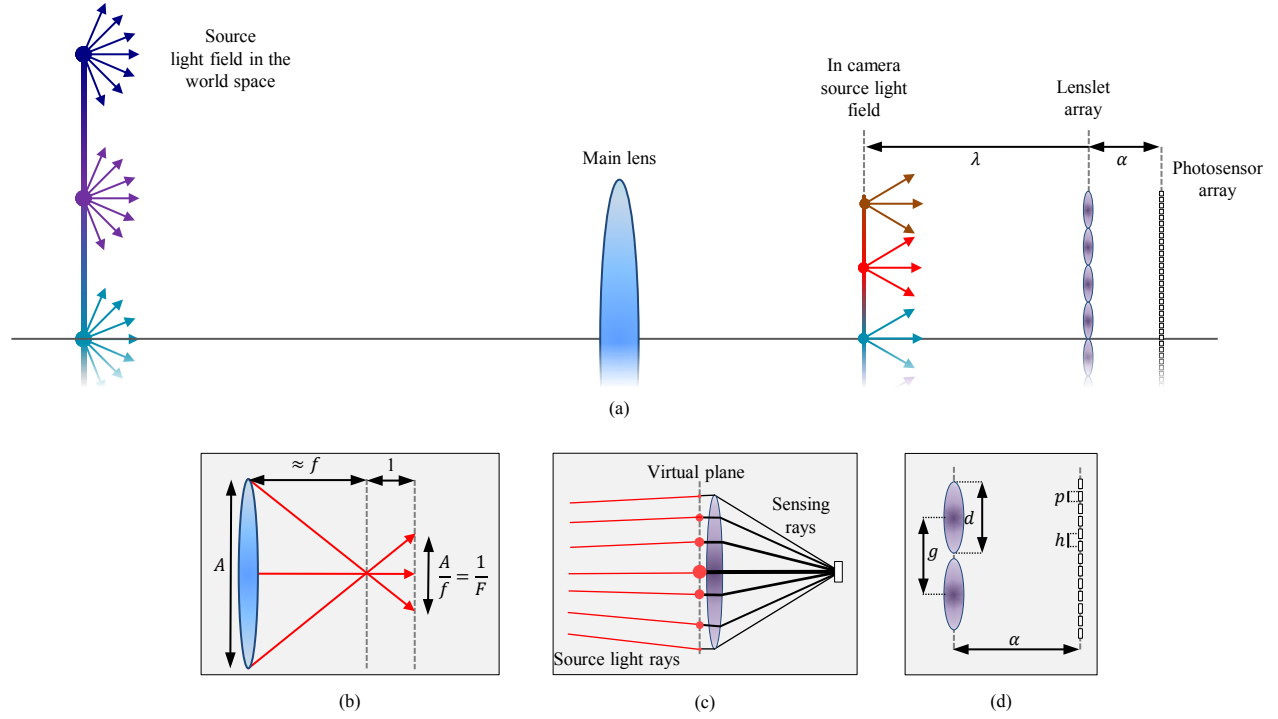
(a)

(b)

(c)

(d)

Fig. 3. Lenslet-based light field camera. (a) The configuration of the optical system and the source light field. We only show the top-half to save space. (b) The feasible ray directions from the source light field are approximately limited by F. (c) To simplify the derivation, we propagate both the source light field and the sensing rays from the photosensor to the virtual plane in front of the lenslet. (d) The closeup of (a) near the lenslet and photosensor arrays.

the product of the incoming light field and the sensor profile:

$$i[\mathrm{c}] = \int_{x=-\infty}^{\infty}\int_{u=-\infty}^{\infty} l\left(\begin{bmatrix} x \\ -u \end{bmatrix}\right)s_{\mathrm{c}}\left(\begin{bmatrix} x \\ u \end{bmatrix}\right)\mathrm{d}u\mathrm{d}x. \qquad (6)$$

The negative sign of u in l is for aligning the orientation of l and s.

However, unlike the ideal case in Sec. 3 where l is a simple linear transformation of l_λ, the light field is now propagated through a lenslet array. Because each lenslet refracts light rays independently, many discontinuities are introduced to the transformed light field, and the derivation becomes complicated. To address this issue, we borrow an idea from bidirectional path tracing [Lafortune and Willems 1993]. We define a virtual plane slightly in front of the lenslet array and define the integral on this plane (Figure 3(c)). We propagate the light field by λ to reach the lenslet array before refraction, and propagate s backward by α through the lenslet array.[5] The advantage of this approach is that while the discontinuity would be introduced to the transformed s_{c}, in most designs, we only need to consider one particular lenslet that covers this sensor.

The light field at the virtual plane is simply given by propagation,

$$l_v(\mathbf{x}) = l_\lambda(\mathbf{M}_\lambda\mathbf{x}) = t(x - \lambda u)\mathrm{rect}(Fu). \qquad (7)$$

We now seek to propagate the sensor response. Without loss of generality, we derive the sensor transformation for the lenslet with its

[5]We cannot propagate s_{c} all the way to the surface of the emitted light field because λ can be negative. However, in such cases propagating l_λ to the virtual plane is still valid because the source light field comes from the space outside the camera.

optical center aligned with that of the main lens. To obtain the sensor response on the virtual plane, we first propagate it to the microlens and then apply the lens refraction given in Table I.

$$\mathbf{M}_s = \mathbf{M}_\alpha\mathbf{R}_{f_m} = \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ f_m^{-1} & 1 \end{bmatrix} = \begin{bmatrix} (1-\frac{\alpha}{f_m}) & -\alpha \\ f_m^{-1} & 1 \end{bmatrix}, \qquad (8)$$

where $\mathbf{R}_{f_m}$ is the light transport matrix due to thin-lens refraction [Gerrard and Burch 1975]. Here the matrices are multiplied in the reverse order because we wish to transform the light field signal, not the coordinate. We obtain the propagated sensor profile $s_{\mathrm{c},v}$ by applying $\mathbf{M}_s$ to s_{c}:

$$s_{\mathrm{c},v}(\mathbf{x}) = \mathrm{rect}\left(\frac{(1-\alpha f_m^{-1})x - \alpha u - x_{\mathrm{c}}}{p}\right)\rho(f_m^{-1}x + u). \qquad (9)$$

Finally, we also consider the aperture of the lenslet as a masking function b:

$$b(\mathbf{x}) = \mathrm{rect}\left(\frac{x}{d}\right), \qquad (10)$$

where d can be smaller than the inter-lenslet distance g (Figure 3(d)). In practice, it can be implemented by depositing a black chromium mask on top of the lenslet array [Georgiev et al. 2011].

Table II. Simulation parameters

Main lens F-number (F)	1.9
Inter-sensor distance (h)	2.0 μm
Photosensor pitch size (p)	1.0 μm
Lenslet focal length (f_m)	37.0 μm
Inter-lenslet distance (g)	21.0 μm
Lenslet aperture size (d)	21.0 μm

We can now redefine (6) as the integral of the product of (7), (9), and (10) (note that $\text{rect}(-Fu) = \text{rect}(Fu)$ below):

$$
\begin{aligned}
i[\text{c}] &= \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} l_v\left(\begin{bmatrix} x \\ -u \end{bmatrix}\right) b(\mathbf{x}) s_{\text{c},v}(\mathbf{x}) \mathrm{d}u\mathrm{d}x \\
&= \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} t(x+\lambda u)\text{rect}(Fu)\text{rect}\left(\frac{x}{d}\right)\rho(f_m^{-1}x+u)\times \\
&\quad \text{rect}\left(\frac{(1-\alpha f_m^{-1})x - \alpha u - x_{\text{c}}}{p}\right) \mathrm{d}u\mathrm{d}x.
\end{aligned}
\tag{11}
$$

While the integral is over the whole spatial and angular dimensions, $\text{rect}(Fu)$ due to the main lens aperture and $\text{rect}(x/d)$ due to the lenslet aperture jointly define the effective integration range, and thus the feasible range of x_{c}. Therefore, for a single lenslet, only a finite number of photosensors would have a nonzero response.

One can further substitute $x + \lambda u$ with k in (11) and obtain:

$$
i[\text{c}] = \int_{k=-\infty}^{\infty} t(k)w_{\text{c}}(k)\mathrm{d}k = (t * \tilde{w}_{\text{c}})(0),
\tag{12}
$$

$$
w_{\text{c}}(k) = \int_{u=-\infty}^{\infty} \text{rect}(Fu)b\left(\begin{bmatrix} k-\lambda u \\ u \end{bmatrix}\right) s_{\text{c},v}\left(\begin{bmatrix} k-\lambda u \\ u \end{bmatrix}\right) \mathrm{d}u,
\tag{13}
$$

where $*$ denotes convolution and $\tilde{f}(x) = f(-x)$. We can see that a recorded light field sample is the average of $t(k)$ weighted by a *prefilter kernel function* $w_{\text{c}}(k)$, or the convolution of t and $\tilde{w}_{\text{c}}$ evaluated at the origin. *Note that the shape of $w_{\text{c}}(k)$ depends on every single parameter in the optical system*: the F-number of the main lens, the sensor pitch size and its angular sensitivity profile, the lenslet-sensor distance, the focal length of lenslet, and even the aperture size of the lenslet.

4.2 Reduction to and Verification of Simplified Model

Because w_{c} has strong dependency on x_{c}, it is no longer possible to describe the integration as a simple global convolution over the entire light field, and we have to proceed with numerical simulation in the next section. In this sub-section, we show that although the final integral (11) appears complex, it can be reduced to verify the models in previous work [Perwaß and Wietzke 2012]. Therefore, our derivation can be considered a significant generalization of those approaches. Previous work describes the light field camera only under specific settings, that may not be realistic for practical designs (for example, they omit the pixel's angular sensitivity).

Simplified Sensor Response: In previous analyses, the pitch size of the photosensor is usually ignored ($p \to 0$), and the angular sensitivity is assumed to be constant. With these two assumptions,
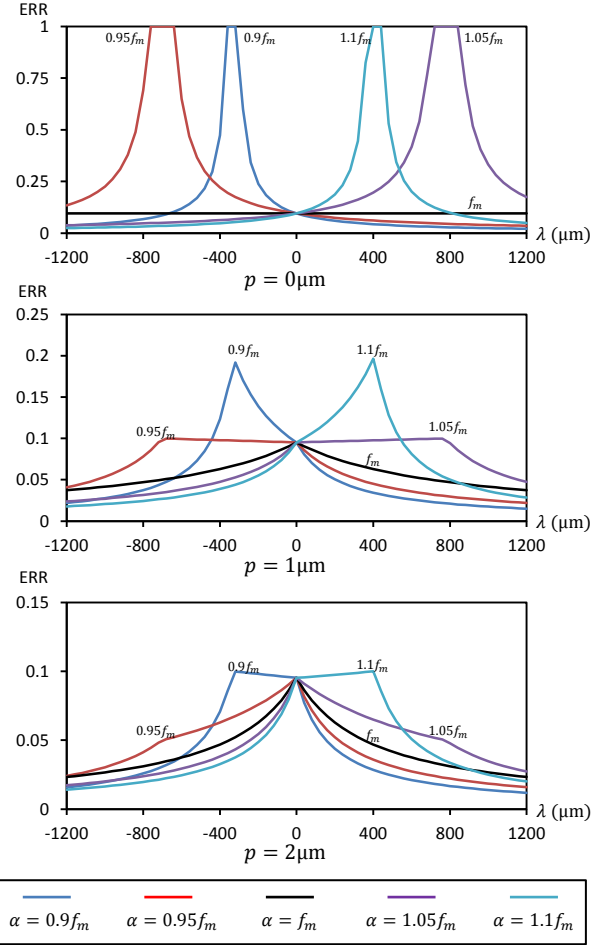

Fig. 4. Effective resolution ratios (ERR) from the simplified model with photosensor pitch size (top) 0μm (middle) 1μm and (bottom) 2μm for different α's. Other parameters are given in Table II. We can see that the basic design (black curves, $\alpha = f_m$) always has the lowest peak among all designs. Other designs can achieve a higher resolution peak. However, the resolution falloff is steeper than the basic design, and the resolution profile is asymmetric along λ. For all designs, the resolution decreases as the photo sensor pitch size increases.

the pixel sensitivity function (5) is simplified to:[6]

$$
s_{\text{c}} = \delta(x - x_{\text{c}}).
\tag{14}
$$

and (9) becomes

$$
s_{\text{c},v}(\mathbf{x}) = \delta((1-\alpha f_m^{-1})x - \alpha u - x_{\text{c}}).
\tag{15}
$$

That is, for a specific point x on the virtual plane, only the light ray from a specific direction $u = ((1-\alpha f_m^{-1})x - x_{\text{c}})/\alpha$ would be sensed by the photosensor.

Reduced Kernel: We can now use the simplified sensor response to determine a reduced kernel w_c for sensing the light field. Specifically, we obtain the simplified form of (11) by using the delta func-

---

[6]Mathematically, there should also be a factor proportional to p, accounting for the loss of light efficiency in integrating over a limited pitch size. For simplicity, since it does not affect our insights, we ignore this normalization.

tion (15) and omitting ρ,

$$i[\mathrm{c}] = \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} t(x + \lambda u)\mathrm{rect}(Fu)\mathrm{rect}\left(\frac{x}{d}\right) \times$$
$$\delta((1 - \alpha f_m^{-1})x - \alpha u - x_c)\mathrm{d}u\mathrm{d}x. \quad (16)$$

Since the integral over the δ-distribution above evaluates the integrand at $u = ((1 - \alpha f_m^{-1})x - x_c)/\alpha$, we have

$$i[\mathrm{c}] = \frac{1}{\alpha} \int_{x=-\infty}^{\infty} t\left(\frac{(\alpha + \lambda - \lambda\alpha f_m^{-1})x - \lambda x_c}{\alpha}\right) \times$$
$$\mathrm{rect}\left(F\frac{(1 - \alpha f_m^{-1})x - x_c}{\alpha}\right)\mathrm{rect}\left(\frac{x}{d}\right)\mathrm{d}x. \quad (17)$$

From this relation, we can extract the kernel function in much the same way as in (12). Substitute $k = \alpha^{-1}((\alpha + \lambda - \lambda\alpha f_m^{-1})x - \lambda x_c)$,

$$i[\mathrm{c}] = \frac{1}{\alpha + \lambda - \lambda\alpha f_m^{-1}} \int_{k=-\infty}^{\infty} t(k)\mathrm{rect}\left(F\frac{(1 - \alpha f_m^{-1})k - x_c}{\alpha + \lambda - \lambda\alpha f_m^{-1}}\right) \times$$
$$\mathrm{rect}\left(\frac{\alpha k + \lambda x_c}{d(\alpha + \lambda - \lambda\alpha f_m^{-1})}\right)\mathrm{d}k. \quad (18)$$

Finally, the reduced kernel can be written as follows (we drop the constant normalization term for simplicity),

$$w_c(k) = \mathrm{rect}\left(F\frac{(1 - \alpha f_m^{-1})k - x_c}{\alpha + \lambda - \lambda\alpha f_m^{-1}}\right)\mathrm{rect}\left(\frac{\alpha k + \lambda x_c}{d(\alpha + \lambda - \lambda\alpha f_m^{-1})}\right). \quad (19)$$

We can see that the integrand in (18) is the product of the texture function and two rect functions. The first rect function is due to the main lens aperture, and the second one is due to the lenslet aperture. In existing analysis [Georgiev et al. 2011; Perwaß and Wietzke 2012], the effect of the main lens aperture is also ignored because of the F-number matching constraint: the system is configured such that every light ray entering the main lens would pass the lenslet and eventually reach the sensor. Therefore, the integrand can be further simplified as a product of the texture function and a rect function centered at $-\lambda x_c/\alpha$ with width $|d(\alpha + \lambda - \lambda\alpha f_m^{-1})/\alpha|$.

Basic Lenslet Camera: In this simplified model, when $\alpha = f_m$, the spot size becomes constant, d, for all λ's. Therefore, the basic light field camera design cannot preserve features smaller than the size of the lenslet, and hence the effective resolution is fixed to the lenslet resolution for all depths. However, this conclusion is based on many strong assumptions: the photosensor angular sensitivity is uniform, the photosensor pitch size is negligible, and the F-number of the main lens perfectly matches that of the lenslet.

Sweet Spot: When the kernel width $|d(\alpha + \lambda - \lambda\alpha f_m^{-1})/\alpha|$ is zero, no blurring is introduced by the system, and thus all high-frequency components are preserved in the sampled data. This corresponds to a *sweet spot* where $\lambda = 1/(f_m^{-1} - \alpha^{-1})$. The sweet spot is in front of the lenslet array in the focused camera design and behind the lenslet array in the generalized design.

Finite Pixels: If we allow p to be finite, the kernel is the superposition of rects centered from $-\lambda(x_c - \frac{p}{2})/\alpha$ to $-\lambda(x_c + \frac{p}{2})/\alpha$. The overall kernel width becomes $|d(\alpha + \lambda - \lambda\alpha f_m^{-1})/\alpha| + |\lambda p/\alpha|$. The location of the sweet spot also becomes a function of p, and the minimal kernel width will no longer be zero.

Effective Resolution: Perwaß and Wietzke [2012] define the effective resolution ratio (ERR) as the ratio of sensor sampling period h and the kernel width. We plot the ERR with different p's
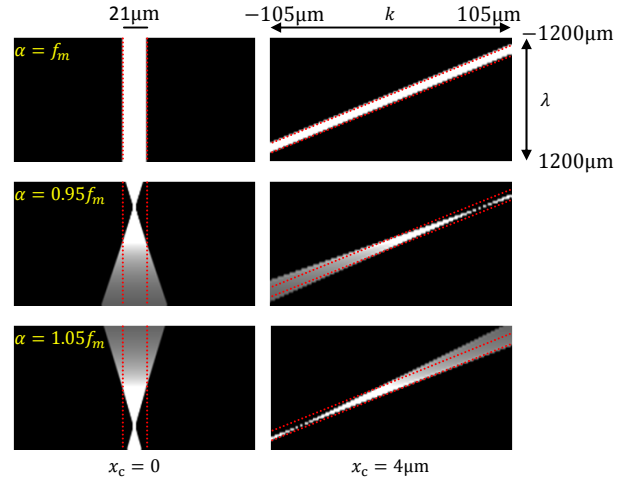
Fig. 5. The kernels for uniform ρ and $p = 0$. The red dotted lines represent the ideal box filter of width $21\,\mu$m, corresponding to the lenslet aperture.
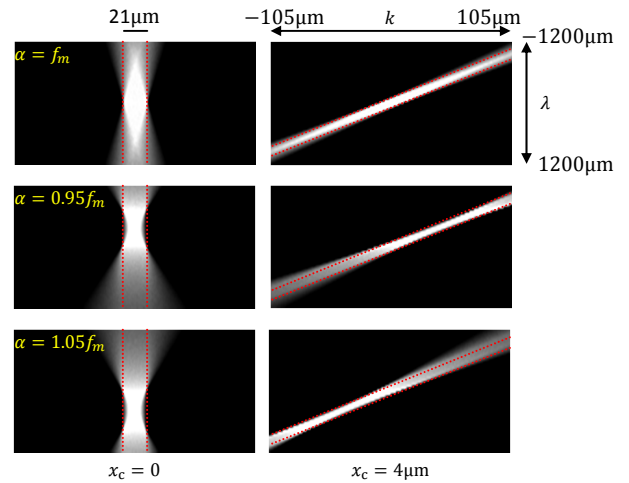

Fig. 6. The kernels for uniform ρ and $p = 1\,\mu$m (fill factor of 0.5). The red dotted lines represent the ideal box filter of width $21\,\mu$m.

in Figure 4.[7] The resulting plots match the results presented in the previous work. We can see that in this simplified model, the basic design ($\alpha = f_m$) has lowest peak resolution among all designs. When the α is away from f_m, the peak resolution would increase, but with faster falloff.

5. SIMULATION FOR PREFILTER KERNELS

As we have shown in the previous section, the kernel w_c is depth-dependent and spatially-variant, and thus we utilize numerical simulation in the following analysis.

Unless stated otherwise, we use the parameters listed in Table II in our simulation—these parameters have the same general range as current light field cameras. We use the simple power of cosine

---

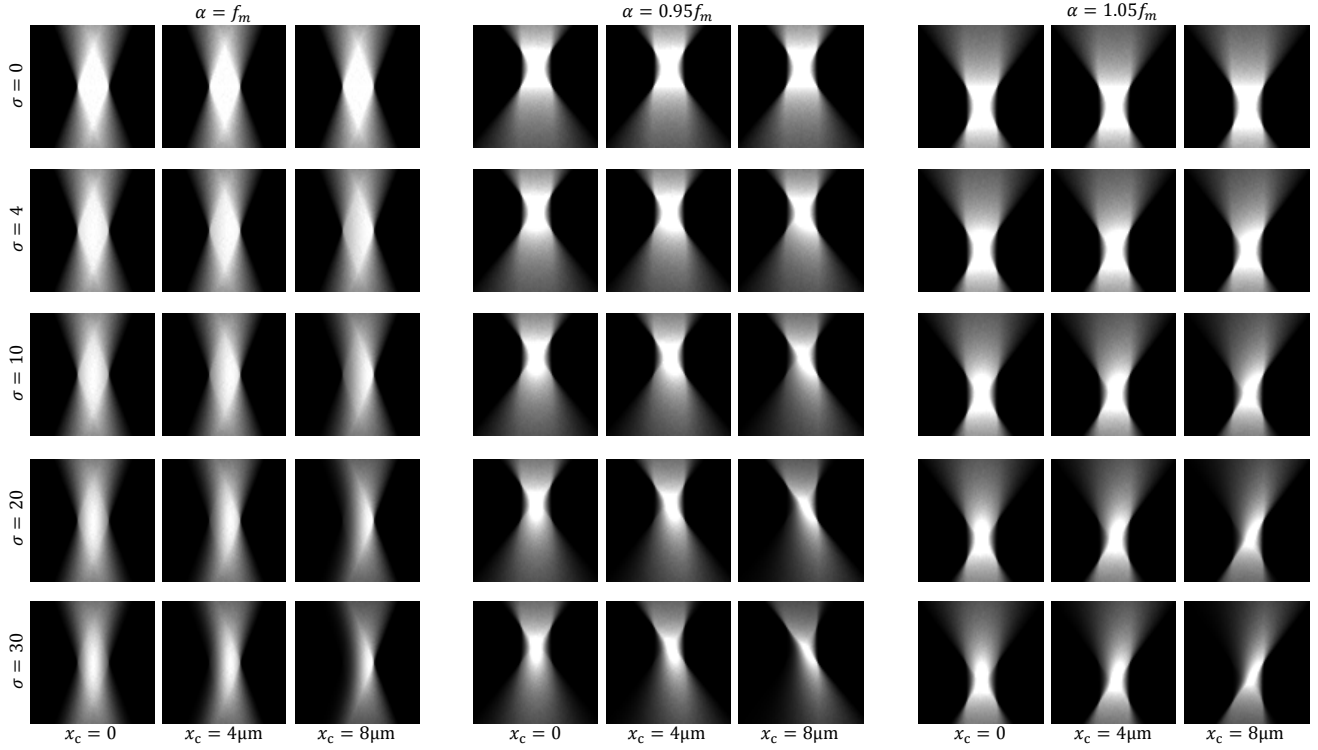[7]The value of p is not reported in [Perwaß and Wietzke 2012].

Fig. 7. The kernels for photo sensor with $p = 1\mu m$ and different angular sensitivity functions. Here, we shift each kernel by $\lambda x_c/\alpha$ and reduce the spatial range k to $\pm 41\mu m$ for clear visualization.

model to represent the angular sensitivity function $\rho(u)$, in much the same way as glossiness is represented in the popular Phong BRDF model (σ is analogous to the Phong exponent):

$$\rho(u) = \cos^\sigma(\tan^{-1}(u)). \tag{20}$$

When $\sigma = 0$, the angular sensitivity function is uniform. Higher σ means the photosensor is more "glossy" or "picky" in direction.

Uniform Sensitivity and Point Pixels: We begin with the simplification in Sec. 4.2: the sensor angular sensitivity is uniform ($\sigma = 0$ in (20)) and $p = 0$, and show the kernels at three different lenslet-sensor distances: f_m, $0.95 f_m$ and $1.05 f_m$, which correspond to the three designs in Sec. 2. The kernel functions under those settings at the center of the lenslet ($x_c = 0\mu m$) and side ($x_c = 4\mu m$) are shown in Figure 5. For each configuration, we show the kernel functions for $\lambda \in [-1200\mu m, 1200\mu m]$. This range covers all refocusable ranges presented in previous work (more than $\pm 30 f_m$).

When the sensor pitch size is zero, in the basic light field camera design [Adelson and Wang 1992; Ng et al. 2005] ($\alpha = f_m$), the kernel function is a box function of width d for all depths. If the target output resolution is the lenslet resolution, the refocusable range would be infinite. In the generalized ($\alpha < f_m$) [Ng 2006] or focused ($\alpha > f_m$) light field camera [Lumsdaine and Georgiev 2009] designs, the kernel functions can be very narrow, even smaller than h at a specific *sweet spot* depth. However, as we go away from the sweet spot, the kernel functions would grow quickly and eventually become wider than d.

For the photosensor off the optical axis of the lenslet, the kernel profile across depths is sheared in all designs, but the kernel width at each depth is identical to that of the center photosensor.

This observation matches the existing models and the derivations in Sec. 4.2. In the basic design, the signal is uniformly prefiltered by a constant box function to match the spatial Nyquist rate. In the design where α does not match f_m, the prefilter can be narrower than d, and the preserved high-frequency details can be recovered if properly processed.[8] However, the working depth range of those designs is significantly reduced.

Finite Pixels: Next, we drop the assumption that the photosensor sensing area is infinitesimal and set p to $1\mu m$ (i.e. , the photosensor fill factor is 0.5). The resulting kernel functions are shown in Figure 6. We can see that for all designs, the kernel width increases with λ when away from the sweet spot depth.[9] The basic design no longer applies a constant prefilter for all depths, and thus the refocusable range is limited.

One important finding is that in the generalized or focused designs, even for the sweet spot, the kernel width is still much larger than h and even comparable to d. Therefore, *when the sensor pitch size is finite, the resolution performance of focused light field camera designs could be much lower than what is claimed in previous work.*

---

[8]The exact frequency response also depends on the *shape* of the kernel, not only the width. Here we keep the discussion at an abstract level and will show the true frequency response by the rendering results in Sec. 6.

[9]In the basic design, we define the sweet spot at $\lambda = 0$.

Angular Sensitivity ($\sigma > 0$): Finally, we visualize the kernel functions with different σ's in Figure 7. Since our primary interest is the shape of the kernel functions, we shift each kernel by $(\lambda x_{\mathrm{c}}/\alpha)$ to align all kernels. We can see that for all designs, as σ increases, the kernels become more narrow. Therefore, more high-frequency details are preserved. However, for the sweet spot depth of each design (i.e. the one with smallest kernel), the kernel profiles do not change much with σ. This is because at the sweet spot, the rays emitted from a point on the source light field would converge to a point at the photosensor plane. In this case, the kernel width is largely affected by pitch size p, which defines the source area *seen* by the photosensor.

Another important observation is that *the kernel function for each photosensor is distinct to others*. As σ increases, the difference is more obvious. Compared to the photosensor aligned with the lenslet optical axis ($x_{\mathrm{c}} = 0$), the photosensor on the side receives fewer rays due to the aperture. Also, its main sensing direction does not align with that of most incoming rays. Therefore, its kernel function spans a smaller area, and the overall magnitude is lower.

Discussion: These observations suggest that we should *not* treat a light field camera as a device that performs uniform prefiltering and sampling on the light field signal. Instead, it behaves more like a cluster of individual light field samplers, and each has a unique prefilter kernel and sampling location. The light field is non-uniformly filtered and irregularly sampled by a light field camera, and traditional signal processing and spectrum analysis cannot be trivially applied. On the other hand, this irregular filtering and sampling suggests that proper reconstruction can recover a wider range of light field frequencies and produce higher-resolution refocused images.

Besides the sensor sensitivity profile, the kernel functions are also affected by many other parameters (F-number, lenslet aperture size, etc), and thus the design space is much larger than simply changing the lenslet-photosensor distance. Some existing prototypes may exploit these properties without realizing their importance.

6. PROJECTION-BASED RENDERING

We have shown that light field cameras often preserve frequency components above the spatial Nyquist rate. In this section, we show that those components can be utilized with the simple projection algorithm.

The projection algorithm recently gained popularity owing to its simplicity and efficiency [Kitamura et al. 2004; Perez Nava and Luke 2009; Georgiev et al. 2011; Yu et al. 2012]. In contrast, other works rely on expensive deconvolution computations, and usually require sophisticated image priors to regularize the process [Georgiev and Lumsdaine 2009; Bishop and Favaro 2012; Wanner and Goldluecke 2012b; Marwah et al. 2013]. While those methods may produce higher-quality results (we do not focus on quantitative evaluation or comparison in this paper), we show here that they are not actually required to achieve higher resolutions. Later, in Sec. 8, we briefly analyze the conditioning of a full inverse light transport algorithm for different light field camera designs.

After a brief algorithm review, we will extend the light transport analysis to derive the exact filter characteristics for the projection-based rendering results. Then, we will show the flatland simulation results for different light field camera designs. To demonstrate the flexibility of the projection-based algorithm, we also show how to extend the basic algorithm to handle depth variation and occlusion.

6.1 Algorithm Overview

The basic projection rendering algorithm is straightforward to implement as illustrated in Figure 8. To generate a refocused image at depth λ, we transform each light field sample by $\mathbf{M}_{-\lambda}$. In other words, each sample at (x, u) is moved to $(x - \lambda u, u)$ (Figure 8(b)). This corresponds to shearing the light field for refocusing at a given depth. Since we have already sheared the light field, we can simply drop the u coordinate of transformed samples, and splat their value to pixels close to $x - \lambda u$ in the target image buffer (Figure 8(c)).

It is known that in practice, this projection algorithm can create images with resolution higher than the lenslet resolution [Perez Nava and Luke 2009; Georgiev et al. 2011]. Intuitively, if the scene is Lambertian, a sample at (x, u) can represent all samples with identical x coordinates. Therefore, we can replace the angular domain integration with simple projection. The concept is also exploited to reproject samples for distribution rendering [Lehtinen et al. 2011].

Because the distribution of projected samples is much denser than the lenslet density, as shown in (Figure 8(c)), if each one is a pointwise sample of the light field, the resolution is limited by the spatial distribution of projected samples [Yu et al. 2012]. However, if the light field is properly filtered before sampling, the output resolution is bandlimited, no matter how dense the samples are (Figure 8(d)).

6.2 Derivation

We extend the proposed framework to derive the frequency response of the projected image. An output pixel m in the projected image o can be represented as the linear combination of i's:

$$o[\mathrm{m}] = \frac{\sum_{\mathrm{c}} i[\mathrm{c}] \cdot r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}{\sum_{\mathrm{c}} r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}, \tag{21}$$

where $x_{\lambda}[\mathrm{c}] = x[\mathrm{c}] - \lambda u[\mathrm{c}]$, x_{m} is the spatial coordinate of the pixel m, and r is the spatial reconstruction kernel, of which the profile is defined according to the target output resolution. In practice, r is usually a simple low-pass function (box, tent, or Gaussian).

We can combine (12) and (21) and obtain:

$$\begin{aligned} o[\mathrm{m}] &= \frac{\sum_{\mathrm{c}} \int_{k=-\infty}^{\infty} t(k) w_{\mathrm{c}}(k) \mathrm{d}k \cdot r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}{\sum_{\mathrm{c}} r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})} \\ &= \int_{k=-\infty}^{\infty} t(k) \Big(\frac{\sum_{\mathrm{c}} w_{\mathrm{c}}(k) r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}{\sum_{\mathrm{c}} r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})} \Big) \mathrm{d}k \\ &= \int_{k=-\infty}^{\infty} t(k) W_{\lambda,\mathrm{m}}(k) \mathrm{d}k = (t * \tilde{W}_{\lambda,\mathrm{m}})(0), \end{aligned} \tag{22}$$

$$W_{\lambda,\mathrm{m}}(k) = \frac{\sum_{\mathrm{c}} w_{\mathrm{c}}(k) r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}{\sum_{\mathrm{c}} r(x_{\lambda}[\mathrm{c}], x_{\mathrm{m}})}. \tag{23}$$

That is, the output pixel is a weighted average of the source t. The weight function $W_{\lambda,\mathrm{m}}$ is a weighted average of the prefiltering kernel functions w_{c}'s. As shown in Sec. 5, w_{c} varies with depth and the optical configurations, and is generally not perfectly bandlimited. Therefore, we expect the perceived resolution to be higher than the lenslet resolution. Moreover, because each w_{c} is unique due to angular sensitivity, and each output pixel combines different sets of source pixels, *each output pixel has a unique frequency response*.

While it is possible to design r to fully compensate the prefiltering effect, here we apply a low-pass and spatially-invariant kernel.
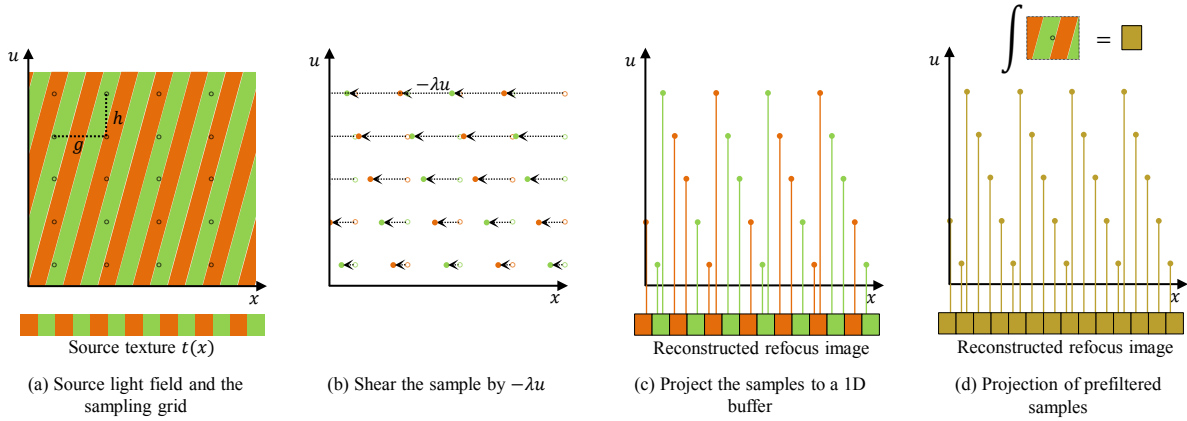
(a) Source light field and the
sampling grid

(b) Shear the sample by $-\lambda u$

(c) Project the samples to a 1D
buffer

(d) Projection of prefiltered
samples

Fig. 8. Illustration of projection-based rendering. (a) The continuous light field from a textured surface at a constant depth λ and the sampling grid. Note the source texture has higher frequency than the lenslet density $1/g$. (b) Shear the discrete light field samples by $-\lambda u$. (c) Project the samples by splatting to the target reconstruction buffer. Note again the reconstruction buffer has higher sampling rate than the lenslet density. (d) If the light field is perfectly prefiltered before sampling (top), the projection result is bandlimited (bottom).

The output generated in this way serves as the *lower bound of the resolution* of a light field camera design since the process does not amplify the attenuated high-frequency components. We leave the discussion of the inverse light transport process to Sec. 8.

Finally, the projection algorithm shares some similarity with the back-projection methods in computed tomography [Kak and Slaney 2001] or light field display synthesis [Wetzstein et al. 2011]. While those methods are developed for reconstructing semi-transparent volume or layers, the projection algorithm is used in reconstructing Lambertian and opaque scenes from the light field.

6.3 Implementation

We first normalize each sample $i[c]$ by its effective exposure $e[c]$, similar to vignetting correction in 2D images:

$$i_n[c] = \frac{i[c]}{e[c]}, \tag{24}$$

$$e[c] = \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} \text{rect}(Fu)b(\mathbf{x})s_{c,v}(\mathbf{x})\mathrm{d}u\mathrm{d}x. \tag{25}$$

In previous work, the spatial coordinate of each sample, $x[c]$, is the center of the lenslet. Here, we define the $(x[c], u[c])$ coordinate of each sample as the weighted average of coordinates of all incoming light rays; we found this refinement gave slightly better results:

$$x[c] = \frac{1}{e[c]} \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} x \cdot \text{rect}(Fu)b(\mathbf{x})s_{c,v}(\mathbf{x})\mathrm{d}u\mathrm{d}x, \tag{26}$$

$$u[c] = \frac{1}{e[c]} \int_{x=-\infty}^{\infty} \int_{u=-\infty}^{\infty} u \cdot \text{rect}(Fu)b(\mathbf{x})s_{c,v}(\mathbf{x})\mathrm{d}u\mathrm{d}x. \tag{27}$$

We compute that transformed spatial coordinate and splat $i_n[c]$ to the two nearest pixels with bilinear weighting (that is, r in (21) is a tent function). Finally, we normalize each pixel by the total weights it received, as in equation 21.
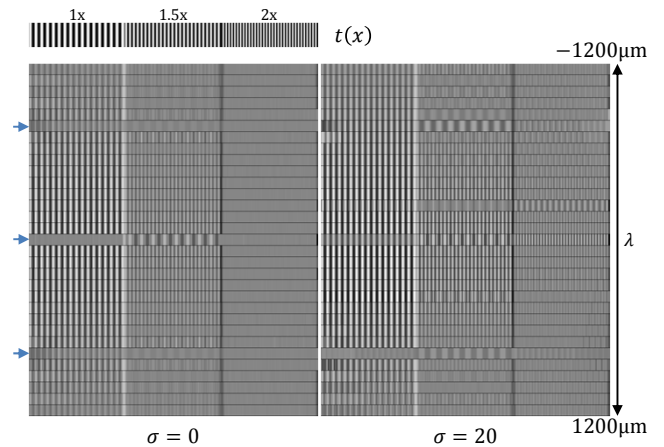

Fig. 9. Refocusing from the basic light field camera design $\alpha = f_m$ using the projection algorithm. (Left) The results from photosensor with uniform angular sensitivity. (Right) The results from photosensor with strong angular sensitivity variation. Readers are encouraged to zoom into these figures in the electronic version to see the details and avoid unrelated aliasing artifacts from resampling in their document reader or printer.

6.4 Simulation

In our simulation, we use parameters in Table II and set the photosensor count to 1050 (i.e. the sensor width is 2100μm). The input texture function $t(x)$ contains three segments of square waves (line grating) with periods 42μm, 28μm and 21μm, which correspond to 1x, 1.5x, and 2x lenslet resolution, respectively. According to the traditional analysis, the basic light field camera design cannot preserve details narrower than 21μm (period 42μm), and thus the maximal effective resolution is 100 pixels. The target image buffer size is 300 pixels for the projection algorithm.

Basic Lenslet Camera: We first show the results from the basic light field camera design (i.e. $\alpha = f_m$) in Figure 9. One interesting observation is that even when the angular sensitivity function is uni-
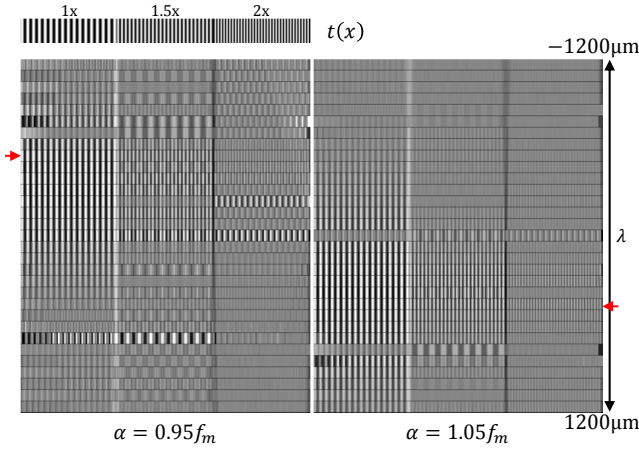
Fig. 10. Refocusing from the basic light field camera design using the projection algorithm. (Left) The generalized light field camera design. (Right) The focused light field camera design. $\sigma = 0$ in both designs. The red arrows mark the location of the sweet spot in both cases. Readers are encouraged to zoom into these figures in the electronic version.

form ($\sigma = 0$), the projection results still show details at 1.5x lenslet resolution. Although the kernel functions are wider than the lenslet pitch size, they are not perfectly bandlimited. Frequency components above the spatial Nyquist rate are attenuated, but not eliminated. We can see that the contrast of the wave functions decreases as $|\lambda|$ increases, which matches the increase of kernel width with $|\lambda|$ shown in Figure 7. We also see that for certain λ's. the rendering results are much worse than others (blue arrows on the side of Figure 9). This is because many samples coincide at those λ's. In other words, the projected coordinate ($x[c] - \lambda u[c]$) of many different c's collide with others, and thus the resolution is limited.

Angular Sensitivity: When $\sigma = 0$, we can hardly see the details at 2x lenslet resolution (almost looks uniform grey in Figure 9 left.) However, as σ is increased to 20, the details at 2x lenslet resolution start becoming visible for some λ. This matches our prediction at the end of Sec. 4.1: a more peaked photosensor response would narrow the kernel and preserve more high frequency details.

Generalized Light Field Camera: Next, we show the refocusing results for the generalized and focused camera designs in Figure 10. Here we only show the results of $\sigma = 0$. When σ increases, the contrast enhancement is similar to Figure 9. The results match our analysis of the kernel functions. At the sweet spot (red arrows in the figure), all details up to 2x lenslet resolution are preserved. However, the sharpness falls off quickly, and thus the effective refocusable range is smaller than that in the basic design.

Quantitative Resolution Plots: To have a more quantitative assessment of the achievable sharpness, we can perform simulations with different source texture frequencies (one frequency at a time), and measure the contrast of the refocused images at different depths. For each input frequency Ω and depth λ, we define the contrast of the refocused image as

$$C_{\Omega,\lambda} = (i_{max} - i_{min})/(i_{max} + i_{min}), \qquad (28)$$

where i_{max} and i_{min} are the maximal and minimal values in the refocused image. To prevent numerical error and the corruption due to aliasing at certain λ's, we detect and average all local maximal

and minimal values at the extrema of the source t across the image domain. When $C > 0$, it means the signal at this particular frequency is preserved. In practice, the resolution of an imaging system is defined as the maximal frequency with contrast above a certain threshold.

The contrast measurement over frequencies and depths constitute the full resolution profile of the light field camera as shown in Figure 11. We can see that the contrast decreases as the source frequency increases, but even in the basic design (Figure 11 left), the preserved frequencies are above the lenslet resolution. In focused or generalized designs, the frequencies up to 3x lenslet resolution can be preserved at certain depths, but also the maximal frequency with non-zero contrast can be below the lenslet resolution at others. Note that because the simulation is performed with thousands of rays per sample, the light field is noise free. Hence, the contrast measurement represents the performance under optimal imaging conditions.

Because the 2D surfaces of the resolution profile might be too cluttered for comparison, we plot the 1D slice at 1.5x lenslet resolution in Figure 12. The plots are consistent with the perceptual results in Figures 9 and 10. We can see that the achievable resolution is above the lenslet resolution for all designs, without any deconvolution or sharpness enhancement process. For all designs, the contrast increases with σ and decreases with p. Surprisingly, when $p = h$, all designs behave very similarly, in terms of both refocusable range and peak contrast. This contradicts the prediction from the existing model [Ng 2006; Perwaß and Wietzke 2012], showing that the 4D sensor profile must be considered.

6.5 Extensions for Depth Variation

The simulations so far use scenes at a constant depth. However, the analysis computes the prefilter kernels for a single photosensor pixel at a time, and so only makes the mild assumption that the depth seen by a single imager pixel is locally constant. Here we show that the projection algorithm can be extended to handle depth variations including occlusion, and thus can be used as a practical method. If we apply the simple projection algorithm to generate the defocus effect, there will be artifacts due to angular aliasing, and various rendering algorithms can be applied to address them [Lehtinen et al. 2011; Mehta et al. 2013]. We will concentrate on generating the all-in-focus (extended depth-of-field) image.

To create an all-in-focus image, we first need per-sample depth information, which can be obtained by most light field depth estimation algorithms [Liang et al. 2008; Wanner and Goldluecke 2012a; Kim et al. 2013; Tao et al. 2013]. We then replace $x_\lambda[c]$ from $x[c] - \lambda u[c]$ to $x[c] - \lambda[c]u[c]$, and hence the projected spatial coordinate of each sample is set by its own depth value.

Second, we need to handle the samples that should be occluded at the target viewpoint (that is, the center of the aperture). Since the depth map of this viewpoint is usually available as a subset of the full depth information, we can adjust the influence of a sample by its depth value:

$$o[m] = \frac{\sum_c i[c] \cdot r(x_\lambda[c], x_m) \cdot r_d(\lambda[c], \lambda_m)}{\sum_c r(x_\lambda[c], x_m) \cdot r_d(\lambda[c], \lambda_m)}, \qquad (29)$$

where r_d is the depth-aware reconstruction kernel and λ_m is the depth value for the output pixel $o[m]$. In the experiment, we choose
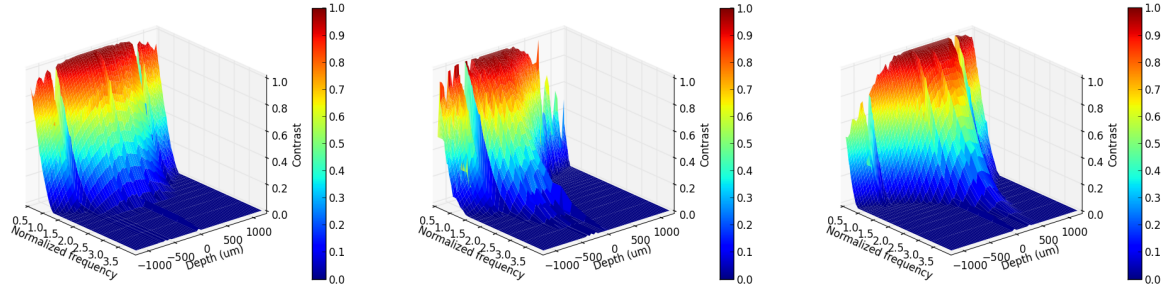
Fig. 11.   The resolution profile of the light field camera designs with $\sigma = 10$, $p = 1\,\mu\text{m}$, and $\alpha =$ (left) $1.0f_m$, (middle) $0.95f_m$, and (right) $1.05f_m$. The frequency is normalized by $2g(42\,\mu\text{m})$.
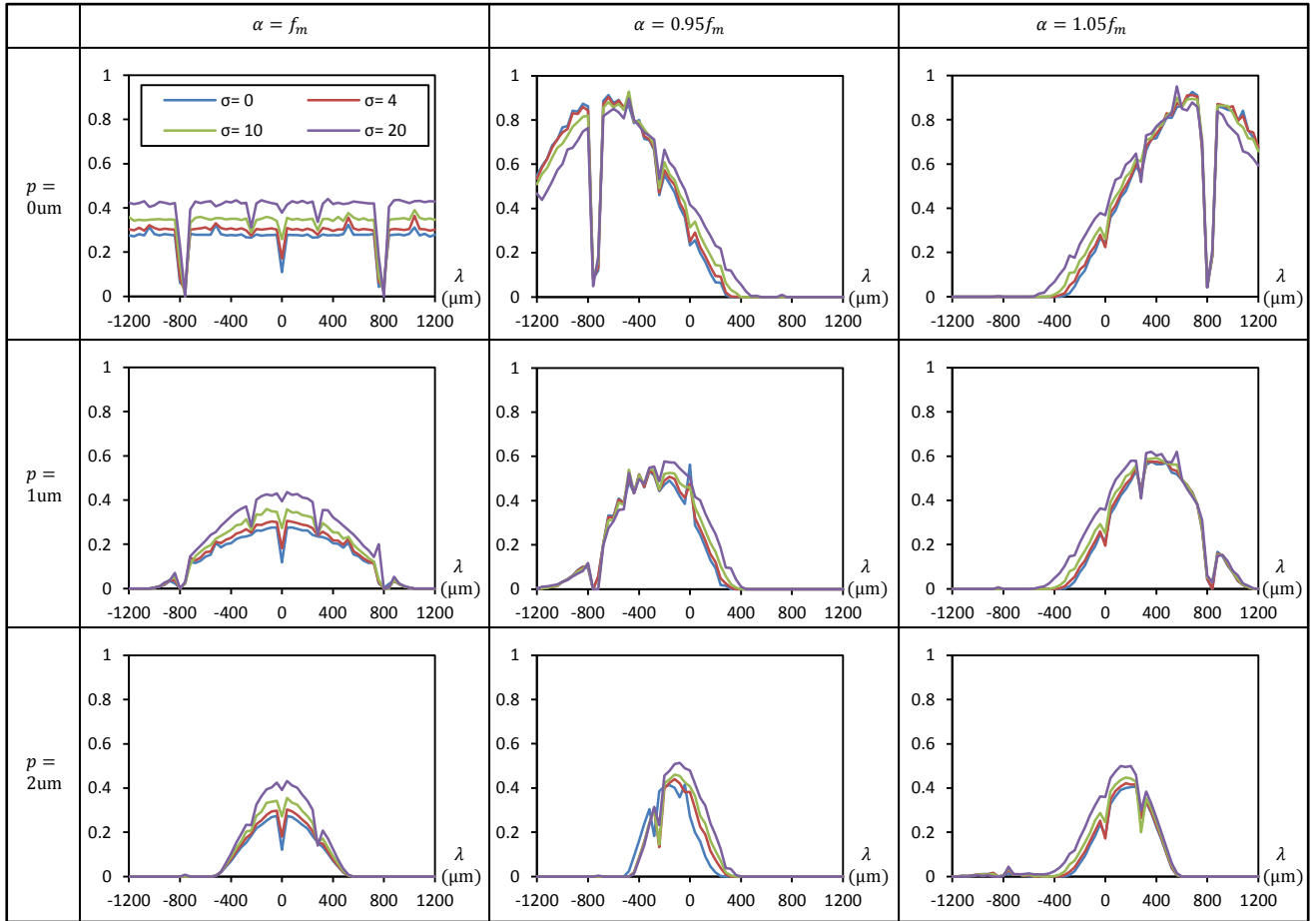

Fig. 12.   The contrast profile of light field camera designs at 1.5x lenslet resolution over λ's. The contrast is defined in equation (28). The dips correspond to the case where many projected samples collide with others (see blue arrows in Figure 9).

a simple kernel to reject occluded samples:

$$r_d(\lambda[c], \lambda_m) = \begin{cases} 1 & \text{if } |\lambda[c] - \lambda_m| \le T_\lambda \\ 0 & \text{otherwise} \end{cases}, \qquad (30)$$

where T_λ is a small constant for tolerating depth estimation error. Note that in our definition, a larger λ is farther in the real world (Figure 3(a)).

We show a few simulation results in Figure 13. We use the basic design ($\alpha = f_m$) with $\rho = 20$ and other parameters listed in Table II. When the scene is a slanted surface (Figure 13(a)), we can see the local sharpness is consistent with that of the constant depth results (Figure 9 right). For certain λ's the image quality is limited due to the collision of projected samples, and this would only affect the image quality locally. The second scene in Figure 13(b) consists

$t(x)$

−700μm   −700μm
−420μm
140μm
140μm
420μm
700μm

(a)

$t(x)$

Without occlusion handling

−700μm
−300μm λ

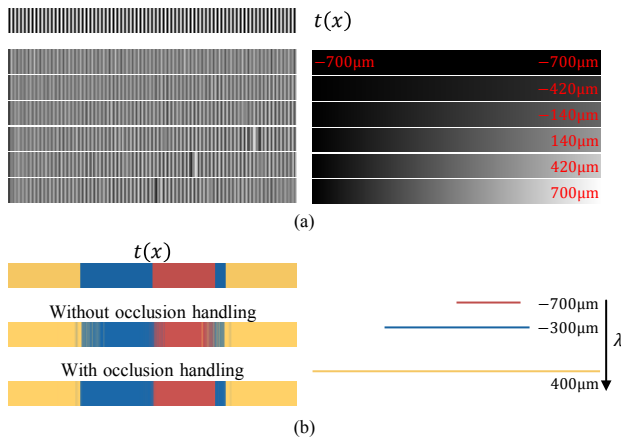With occlusion handling

400μm

(b)

Fig. 13. Projection results for scenes with depth variations. Here the basic light field camera design ($\alpha = f_m$) with $\rho = 20$ is used. (a) The scene is a single slanted surface with increasing slope. The texture is a line grating of period $28\,\mu m$ (1.5x lenslet resolution). For each row, the left is the projection results and the right is the depth map. (b) The scene consisting of three separate layers at $700\,\mu m$, $-300\,\mu m$, and $-700um$. The texture of each layer is a solid color. We can see the that rejecting the occluded samples using (29) significantly suppresses the artifact. Readers are encouraged to zoom into these figures in the electronic version.

Table III. Lytro Camera parameters

Main lens F-number (F)	2.0
Main lens focal length	50 mm
Photo sensor count	3280×3280
Inter-sensor distance (h)	1.4 μm
Photosensor pitch size (p)	1.0 μm
Lenslet focal length (f_m)	25 μm
Lenslet-sensor separation (α)	25 μm
Inter-lenslet distance* (g)	14.0 μm
Lenslet aperture size (d)	14.0 μm
Angular exponent (σ)	13

*: Hexagonal packing.

of three separate layers with solid colors. We can see that without rejecting the samples that would be occluded, the projection result contains strong artifacts around the layer boundaries. By using the modified projection, the artifacts can be greatly suppressed. We will show more projection results for real scenes captured by the Lytro camera in the next Section.

7. REAL DATA VERIFICATION

In this section, we briefly show some initial results from the first generation Lytro light field camera, to verify the 2D flatland analysis in the previous sections. It should be noted that a real device and actual data has many additional uncertainties (main lens characteristics, diffraction, manufacturing variations, depth estimation accuracy). Therefore, our goal is not to quantitatively test the predictions, but to qualitatively verify some key aspects of the theory.

The configuration of the Lytro camera is listed in Table III. It uses the basic design ($\alpha = f_m$) and the photosensor has a high directional sensitivity variation. In this configuration, the lenslet resolu-

tion is 330×330, or 330 line widths per image height (LW/PH). We mount the camera on a tripod and control the focus motor to perform a dense sweep of the ISO-12233 chart. We estimate the depth to be refocused at a few key frames, and fit the λ value versus the focus-step with a linear function.

We capture and average 32 white images as the approximation of the effective exposure in (25). We use a simple linear demosaicing algorithm and only process the green channel. As in the simulation, we use the simple projection-based algorithm for refocusing: no additional deconvolution or sharpness enhancement is applied to the images. The target output resolution is 990×990. The refocus results and sharpness measurement are shown Figure 14. The results show a reasonable agreement with the theory. We can clearly see that all images contain details above the lenslet resolution. This matches our analysis: the real device indeed does not have a perfect bandlimit prefilter, and high-frequency details are preserved. For certain depths such as $\lambda = 0$, aliasing would happen, and an additional anti-aliasing filter is required.

To measure the sharpness, we use the popular slanted-edge MTF measurement which is insensitive to framing or focus breathing [Burns 2000]. We plot the frequencies with 50% and 20% contrast (MTF50 and MTF20, respectively) in Figure 14 (right). In conventional cameras, the MTF20 value of the optical system is the reference value for choosing the corresponding photosensor. We can see that the MTF20 value is above the lenslet resolution for a wide range of depths, and the peak resolution can be as high as twice the lenslet resolution, without any deconvolution or enhancement.

All-in-focus Images of General Scenes: In Sect. 6 we show that the projection algorithm can be modified to generate all-in-focus images for non-planar scenes with depth discontinuities. Here we verify that this modification is practical for the real scenes by processing the data taken by the Lytro light field camera.

Because the modified algorithm requires per-sample depth information, we extract the depth map generated by the Lytro desktop software. The extracted depth map is at lenslet resolution at the center viewpoint, and we warp it to other viewpoints to generate the full depth map and reject all dis-occluded samples during projection. For comparison, we use the traditional algorithm of Ng [2006] to generate the sub-aperture images with equivalent aperture size ($f/20$). Again, no post-processing, such as sharpening, denoising, or deconvolution, is applied to images. For reference, for a conventional camera with the same main lens, its circle-of-confusion (defocus blur kernel) is larger than the lenslet aperture width (d) after $|\lambda| > 25\mu m$.

The representative results are shown in Figure 15, and more high-resolution ones are provided in the supplemental material. We can see that although the Lytro camera uses the basic design, the sub-aperture image still contains much aliasing due to imperfect prefiltering. The simple projection algorithm can successfully recover those details above the spatial Nyquist frequency. Even when the assumption of the Lambertian scene is moderately violated (gloss reflection, light source, etc), or when the depth map is not accurate, the projection results show few and localized artifacts.

8. INVERSE LIGHT TRANSPORT

We have shown that the projection algorithm creates images above the lenslet resolution. However, high-frequency details are attenuated, and rendering quality is spatially- and depth-variant. There-
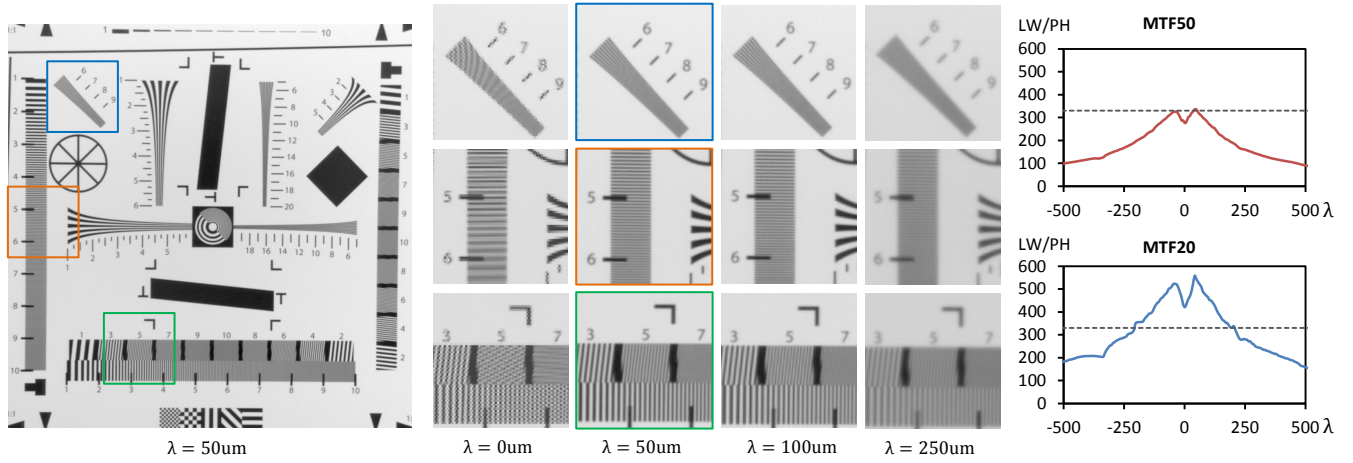
Fig. 14. The refocus results from the real device (Lytro camera) using the basic light field camera design. The parameters are given in Table III. (Left) The refocus image of the ISO-12233 chart at $\lambda = 50\mu m$ from the projection algorithm (Middle) Closeup of refocus images at various λ's. (Right) The MTF50 and MTF20 values over λ in line widths per picture height (LW/PH). The gray dotted lines represent the lenslet resolution.

fore, the projection algorithm can at best serve as a *preview tool* due to its efficiency. We should not judge the performance of light field camera designs by only comparing the projection results. The more critical question is: given the samples $\{i[c]\}$, how faithfully can $t(x)$ be reconstructed? Previous work has formulated light field rendering as solving such an *inverse problem* [Georgiev and Lumsdaine 2009; Bishop and Favaro 2012; Marwah et al. 2013]. However, those works do not consider the full camera model, and only present the results using their own prototypes. Also, since the inverse problem is considered ill-conditioned, they apply distinct image or light field priors as the regularization term, and employ different optimization algorithms. It is therefore difficult to determine the fundamental performance limit of different light field camera designs.

In contrast, we analyze the fundamental difficulty of the inverse process. Inspired by the performance analysis for super-resolution [Baker and Kanade 2002] and bandlimited light field reconstruction [Wetzstein et al. 2013], we assume the continuous texture function $t(x)$ can be represented by a piecewise constant function:

$$t(k) = t[m], \tag{31}$$

for all $k \in (k_m - P/2, k_m + P/2]$, where k_m is the center of the piece represented by $t[m]$, and P is the piece width. With this assumption, we can formulate light field capture as a linear system:

$$\mathbf{i} = \mathbf{W}\mathbf{t}, \tag{32}$$

where $\mathbf{i}$ is the vector of all captured samples $\{i[c]\}$, $\mathbf{t}$ is the vector of all piecewise constant texture elements $\{t[m]\}$, and $\mathbf{W}$ is the *forward transport matrix* that transforms the texture function to the samples. Each element in $\mathbf{W}$ can be derived by replacing $t(k)$ in (12) by $t[m]$:

$$i[c] = \int_{k=k_m-P/2}^{k_m+P/2} t[m] w_c(k) dk$$

$$= \left(\int_{k=k_m-P/2}^{k_m+P/2} w_c(k) dk \right) t[m]$$

$$= \mathbf{W}_{c,m} t[m], \tag{33}$$

where $w_c(k)$ is given in (13). Note that this $\mathbf{W}$ is constructed for the scene at a specific depth, and $\mathbf{W}_{c,m}$ depends on m due to the integration domain, and on c and all other camera parameters due to the integrand. The completeness of the linear system depends on the resolution of t. Note that even if the resolution of t is set to be less than that of i, it can still be much higher than the lenslet resolution.[10]

Signal and Noise Levels: The signal level of $i[c]$ scales with the magnitude of $\mathbf{W}_{c,m}$'s. Therefore, if the photosensor is more picky (i.e. smaller p and higher σ), the signal level would be lower. The noise of digital imaging systems consists of many components, and its variance can be approximated as an affine function of the signal level [Healey and Kondepudy 1994].

To fully evaluate the imaging systems, the tradeoff between the signal/noise levels and exposure settings should be considered. Here we study a relaxed situation: the stability of the inversion process when the signal-noise levels of different designs are compensated.

Inversion: The inverse process attempts to recover $\mathbf{t}$ from $\mathbf{i}$. When the linear system is square or over-complete, the least-squares solution can be obtained by the (pseudo-)inverse of $\mathbf{W}$. This process is called inverse light transport in graphics, or deconvolution in computer vision. Since the inversion undoes the mixing of source signals, the reconstruction quality can be much higher than that of the simple projection algorithm discussed in Sec. 6.

Stability: The feasibility and stability of inversion is related to the condition number of the matrix $\mathbf{W}$ [Horn and Johnson 2012]. We show condition numbers of different designs in Figure 16, for $\lambda \in [-1200\mu m, 1200\mu m]$, with numerical simulation used to construct $\mathbf{W}$ and the resolution of $\mathbf{t}$ set to 300 as in the projection algorithm (i.e. $P = 7\mu m$ in (33)). When the condition number is large, the noise in the observed $\mathbf{t}$ will be significantly amplified. In such ill-conditioned cases, a regularization term based on the prior of $\mathbf{t}$ is required to obtain satisfactory results. Therefore, light field camera

---

[10]It is simple to generalize the derivation here to represent $t(x)$ by other bases, such as Fourier series or higher-order splines.
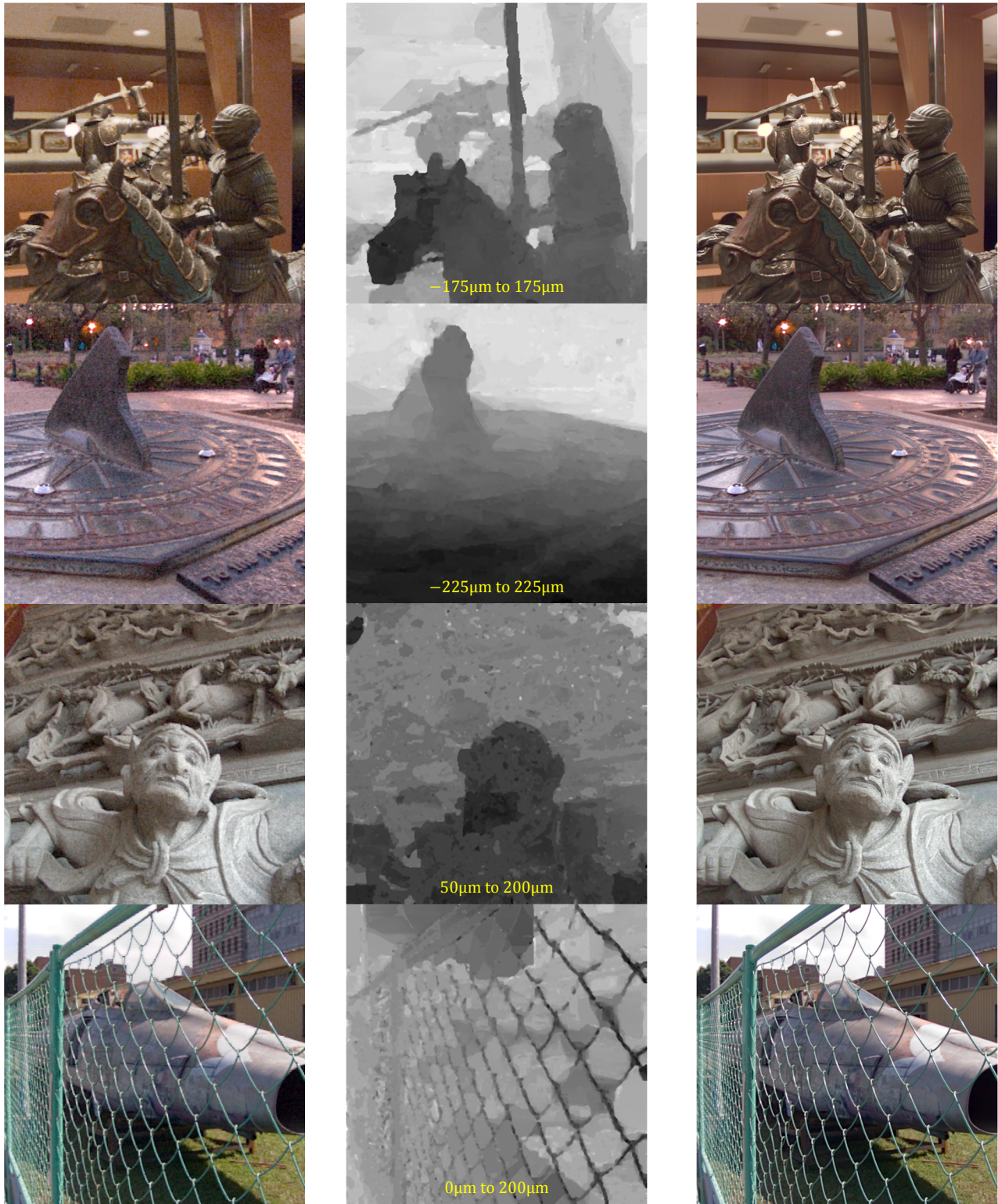
Fig. 15. The all-in-focus images from the first generation Lytro camera. (Left) The sub-aperture image (330×330), (Middle) the depth map extracted from the Lytro desktop software, and (Right) the results using the modified projection algorithm (990×990). The bottom of each row shows the estimated depth range of the scene.
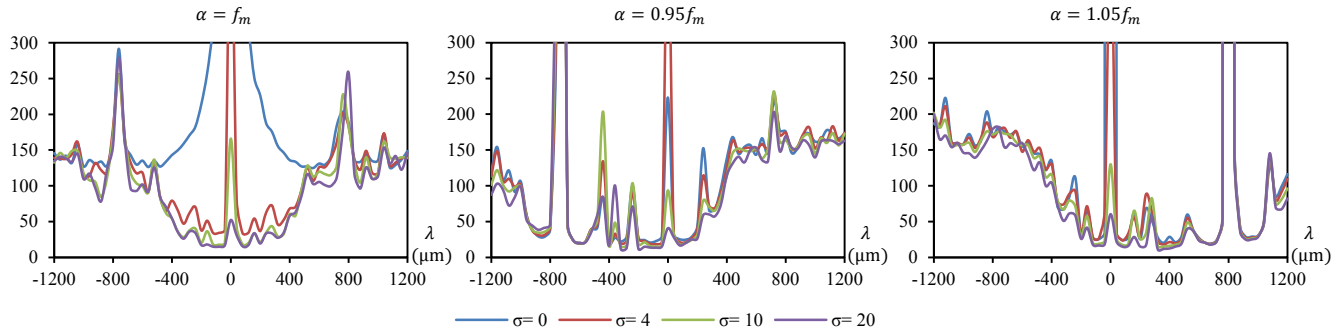
Fig. 16. Condition numbers of forward transport matrices of different light field camera designs using the parameters in Table II. The smaller the condition number, the better conditioned is the inverse process, and it requires a weaker prior or regularization. Condition numbers drop as σ increases for most λ's in all designs.
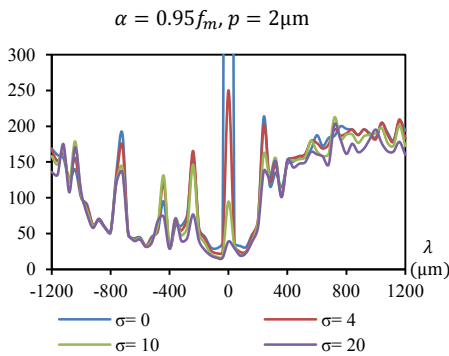

Fig. 17. The condition number of the generalized light field camera with photosensor pitch size $p = 2\mu m$ (fill factor 1.0). Compared to Figure 16 (middle), the peak of condition number is lowered, but the overall condition number increases.

designs with lower condition numbers are better. Besides the condition number, one can also evaluate the performance of a linear system by other metrics such as the structure of the covariance matrix $(\mathbf{W}^T\mathbf{W})^{-1}$ or the noise amplification factor $\mathrm{trace}((\mathbf{W}^T\mathbf{W})^{-1})$ [Wetzstein et al. 2013]. We observed that the relative performance of different designs remains similar in those metrics.

The first observation from Figure 16 is that the condition number is strongly correlated with the refocusing quality of the projection algorithm (Figures 9 and 10). At the λ's with sharp projection results, the condition number is low. In contrast, when the condition number is high, the projection results show serious artifacts.

Condition Number at $\lambda = 0$: In the basic design, the condition number is infinity at $\lambda = 0$ when $\sigma = 0$. This is because the shapes of kernel functions for all c's are identical, and there is no offset between those kernels when $\lambda = 0$. In other words, all photosensors under a lenslet collect the same information, and thus $\mathbf{W}$ is singular, unless the resolution of $\mathbf{t}$ is at lenslet resolution or lower.

However, the condition number at $\lambda = 0$ drops quickly as σ increases. This is because increasing σ not only reduces the kernel width, but also adds variation among the photosensors. Each photosensor integrates the source light field in a slightly different way, and hence the chance to have linearly-dependent rows in $\mathbf{W}$ is reduced. A similar concept has been exploited in super-resolution to replace the regular sensor layout by aperiodic Penrose tiling [Ben-Ezra et al. 2011].

Generalized/Focused Light Field Cameras: The generalized or focused light field camera designs behave as we expect. The condition numbers decrease gradually as σ increases. Compared to the basic design, the comfortable region with low condition numbers shifts toward negative λ in the generalized design, and towards positive λ in the focused one. However, neither of these two designs would effectively widen the region. At $\lambda = 0$, both designs would have difficulty in the inverse process.

Perhaps surprisingly, the condition number in both designs can be very large at certain λ, and it does not change much as σ increases! At those depths, the kernel width can be smaller than the offset of each kernel, and thus certain regions of $t(k)$ are not involved in the integration at all. In other words, for certain m's, $\mathbf{W}_{c,m}$ is zero for all c, and it is not possible to recover $t[m]$ from $\mathbf{i}$.

If we carefully trace light rays from those regions toward the photosensor array, we would find that the rays would hit the non-sensing area in-between photosensors. Therefore, one straightforward way to fix the problem is to increase the photosensor pitch size p, and the resulting condition number is shown in Figure 17. We can see that the peak condition number decreases significantly. However, because the filter kernel becomes wider, the condition number increases for most λ's. Also, the quality of rendering results from the simple projection-based algorithm would decrease significantly (see contrast profiles of the middle and bottom rows in Figure 12). Therefore, this solution basically trades off the overall system performance for improving the worst case.

9. DISCUSSION AND FUTURE WORK

In this paper, we have used analytic models and numerical simulation to provide many important insights into the resolution performance of current light field camera designs, and rendering/reconstruction algorithms. Our paper is theoretical—no new light field camera design is presented. However, the relationships of resolution to various parameters in a number of light field camera designs does provide important guidance for practitioners.

Our model shows that the prefiltering kernel in most light field camera designs is depth-dependent, and spatially variant even for constant depth. We also show that these variations in the light field would make the quality or stability of the rendering algorithms spatially or depth-dependent.

Our simulations clearly show that the photosensor profile, including both the pitch size and the angular sensitivity profile, can significantly affect the system performance. However, while reduc-

ing the pitch size or the sensitivity profile can preserve more high-frequency contents, the effective exposure is reduced and the samples would be noisier. It would be an interesting avenue to extend our framework to consider the noise-sharpness tradeoff and design the optimal photosensor profile for light field cameras. Also, the real photosensor profile is generally neither separable nor analytic, and how to accurately model and calibrate the 4D photosensor profile would be a challenging problem.

When the prefilter kernel does not fully bandlimit the signal, even the simple projection algorithm can generate output beyond the lenslet resolution. In the basic design, the prefilter can be adjusted by the photosensor pitch size or angular sensitivity profile. In the focused or generalized designs where $\alpha \neq f_m$, the peak resolution can be even higher at the expense of refocusable range reduction. Moreover, the peak resolution of these designs is lower than expected when the photosensor pitch size is finite (Figure 6), and sometimes the narrower kernel makes the inverse process ill-conditioned (Figure 16). Since a wide refocusable range with consistent resolution is as important as the peak resolution, the choice of the light field camera design is application-dependent.

Our framework can be extended to handle mask-based light field cameras [Veeraraghavan et al. 2007; Lanman et al. 2008; Marwah et al. 2013]. One can disable the lenslet by setting $f_m = \infty$, and replace the masking function b in (10) with general mask functions, while still considering all other optical parameters. Similar to the lenslet-based designs, it is possible to recover high resolution results as long as the prefilter kernel is not perfectly bandlimited. From this perspective, the mask-based light field camera is similar to the coded aperture imaging system [Levin et al. 2007]. The main difference is that each photosensor has its own aperture shape, and this diversity may provide unique advantages over the coded aperture cameras.

It is straightforward to extend the analysis to the 3D space and 4D light field. In practice, the masking function is circular instead of rectangular, but the derivation of the kernel function remains largely unchanged. However, the 2D lenslet and photosensor array can be constructed in various ways, which may affect the characteristics of the forward transport matrix. We focus on the filtering and reconstruction in this framework, and leave the analysis of 4D sampling patterns as future work. Similarly, we leave an analysis of resolution performance for non-Lambertian scenes as future work.

Finally, our light field analysis is based on geometric optics. Most steps in our derivation can be replaced with the equivalent tools (such as the Wigner distribution) in wave optics [Zhang and Levoy 2009; Levin et al. 2009]. While a more accurate model for the kernels can be obtained, we believe the main insights of our analysis would remain, and the relative performance of different light field camera designs would not be affected.

10. CONCLUSION

In this paper, we have revisited many different aspects of lenslet-based light field camera designs. We have shown that the general space-angle tradeoff usually does not matter when the scene depth is locally constant. Therefore, the prefilter design for the light field camera can be more flexible than the simple limit set by the spatial Nyquist rate.

Second, we have derived a more accurate model for light field image formation by considering the full photosensor profile and all other parameters in the system. With this model, it is possible to compare various lenslet cameras, as we have demonstrated via simulation. We have used our model to explain the success of the simple projection algorithm, and identified a few unique properties of the light field camera in the inverse light transport analysis.

We believe that our analysis provides a new way to think about how light field cameras work, and how to reconstruct images from the captured light field. We expect these new insights to inspire further research in light field imaging and computational photography.

Acknowledgements

REFERENCES

ADELSON, E. H. AND BERGEN, J. R. 1991. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, 3–20.

ADELSON, E. H. AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. *IEEE TPAMI 14,* 2, 99–106.

BAKER, S. AND KANADE, T. 2002. Limits on super-resolution and how to break them. *IEEE TPAMI 24,* 9, 1167–1183.

BELCOUR, L., SOLER, C., SUBR, K., HOLZSCHUCH, N., AND FRÉDO, D. 2012. 5D covariance tracing for efficient defocus and motion blur. MIT-CSAIL-TR-2008-049, Massachusetts Institute of Technology.

BEN-EZRA, M., LIN, Z., WILBURN, B., AND ZHANG, W. 2011. Penrose pixels for super-resolution. *IEEE TPAMI 33,* 7, 1370–1383.

BISHOP, T. E. AND FAVARO, P. 2012. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE TPAMI 34,* 5, 972–986.

BROXTON, M., GROSENICK, L., YANG, S., COHEN, N., ANDALMAN, A., DEISSEROTH, K., AND LEVOY, M. 2013. Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express 21,* 21, 25418–25439.

BURNS, P. D. 2000. Slanted-edge MTF for digital camera and scanner analysis. In *Proc. IS&T 2000 PICS Conference*. 135–138.

CATRYSSE, P. B. AND WANDELL, B. A. 2002. Optical efficiency of image sensor pixels. *JOSA A 19,* 8, 1610–1620.

CHAI, J.-X., CHAN, S.-C., AND TONG, H.-Y. S. X. 2000. Plenoptic sampling. In *SIGGRAPH '00*. 307–318.

CHAN, W.-S., LAM, E. Y., NG, M. K., AND MAK, G. Y. 2007. Super-resolution reconstruction in a computational compound-eye imaging system. *Multidimensional Systems and Signal Processing 18,* 2-3, 83–101.

DANSEREAU, D. AND BRUTON, L. T. 2007. A 4-D dual-fan filter bank for depth filtering in light fields. *IEEE Trans. Signal Processing 55,* 2, 542–549.

DANSEREAU, D. G., BONGIORNO, D. L., PIZARRO, O., AND WILLIAMS, S. B. 2013. Light field image denoising using a linear 4D frequency-hyperfan all-in-focus filter. In *IS&T/SPIE Electronic Imaging*.

DURAND, F., HOLZSCHUCH, N., SOLER, C., CHAN, E., AND SILLION, F. X. 2005. A frequency analysis of light transport. *ACM TOG 24,* 3, 1115–1126.

EGAN, K., HECHT, F., DURAND, F., AND RAMAMOORTHI, R. 2011. Frequency analysis and sheared filtering for shadow light fields of complex occluders. *ACM TOG 30,* 2, 9.

EGAN, K., TSENG, Y.-T., HOLZSCHUCH, N., DURAND, F., AND RA-MAMOORTHI, R. 2009. Frequency analysis and sheared reconstruction for rendering motion blur. *ACM TOG 28,* 3, 93.

EL GAMAL, A. AND ELTOUKHY, H. 2005. CMOS image sensors. *Circuits and Devices Magazine, IEEE 21,* 3, 6–20.

GEORGIEV, T., CHUNEV, G., AND LUMSDAINE, A. 2011. Superresolution with the focused plenoptic camera. In *Proc. SPIE.* Vol. 7873. 1105–1117.

GEORGIEV, T. AND LUMSDAINE, A. 2009. Superresolution with plenoptic 2.0 cameras. In *Signal Recovery and Synthesis.*

GERRARD, A. AND BURCH, J. M. 1975. *Introduction to matrix methods in optics.* Courier Dover Publications.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *SIGGRAPH '96.* 43–54.

HEALEY, G. E. AND KONDEPUDY, R. 1994. Radiometric CCD camera calibration and noise estimation. *IEEE TPAMI 16,* 3, 267–276.

HORN, R. A. AND JOHNSON, C. R. 2012. *Matrix analysis.* Cambridge university press.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *SIGGRAPH '00.* 297–306.

IVES, F. E. 1903. Parallax stereogram and process of making same. US Patent 725,567.

JAROSZ, W., SCHÖNEFELD, V., KOBBELT, L., AND JENSEN, H. W. 2012. Theory, analysis and applications of 2D global illumination. *ACM TOG 31,* 5, 125.

KAK, A. C. AND SLANEY, M. 2001. *Principles of computerized tomographic imaging.* Society for Industrial and Applied Mathematics.

KIM, C., ZIMMER, H., PRITCH, Y., SORKINE-HORNUNG, A., AND GROSS, M. 2013. Scene reconstruction from high spatio-angular resolution light fields. *ACM TOG 32,* 4, 73.

KITAMURA, Y., SHOGENJI, R., YAMADA, K., MIYATAKE, S., MIYAMOTO, M., MORIMOTO, T., MASAKI, Y., KONDOU, N., MIYAZAKI, D., TANIDA, J., ET AL. 2004. Reconstruction of a high-resolution image on a compound-eye image-capturing system. *Applied Optics 43,* 8, 1719–1727.

LAFORTUNE, E. P. AND WILLEMS, Y. D. 1993. Bi-directional path tracing. In *Proceedings of CompuGraphics.* Vol. 93. 145–153.

LANMAN, D., RASKAR, R., AGRAWAL, A., AND TAUBIN, G. 2008. Shield fields: modeling and capturing 3D occluders. *ACM TOG 27,* 5, 131.

LEHTINEN, J., AILA, T., CHEN, J., LAINE, S., AND DURAND, F. 2011. Temporal light field reconstruction for rendering distribution effects. *ACM TOG 30,* 4, 55.

LEVIN, A. AND DURAND, F. 2010. Linear view synthesis using a dimensionality gap light field prior. *CVPR,* 1831–1838.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM TOG 26,* 3, 70.

LEVIN, A., HASINOFF, S. W., GREEN, P., DURAND, F., AND FREEMAN, W. T. 2009. 4D frequency analysis of computational cameras for depth of field extension. *ACM TOG 28,* 3, 97.

LEVOY, M. AND HANRAHAN, P. 1996. Light field rendering. In *SIGGRAPH '96.* 31–42.

LIANG, C.-K., LIN, T.-H., WONG, B.-Y., LIU, C., AND CHEN, H. H. 2008. Programmable aperture photography: multiplexed light field acquisition. In *ACM TOG.* Vol. 27. 55.

LIANG, C.-K., SHIH, Y.-C., AND CHEN, H. H. 2011. Light field analysis for modeling image formation. *IEEE TIP 20,* 2, 446–460.

LIPPMANN, M. G. 1908. Epreuves reversible donnant la sensation du relief. *J. Phys 7,* 821–825.

LUMSDAINE, A. AND GEORGIEV, T. 2009. The focused plenoptic camera. In *ICCP.* 1–8.

LUMSDAINE, A., GEORGIEV, T. G., AND CHUNEV, G. 2012. Spatial analysis of discrete plenoptic sampling. In *IS&T/SPIE Electronic Imaging.*

MARWAH, K., WETZSTEIN, G., BANDO, Y., AND RASKAR, R. 2013. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM TOG 32,* 4, 1–11.

MEHTA, S. U., WANG, B., RAMAMOORTHI, R., AND DURAND, F. 2013. Axis-aligned filtering for interactive physically-based diffuse indirect lighting. *ACM TOG 32,* 4, 96.

NG, R. 2005. Fourier slice photography. In *SIGGRAPH '05.* 735–744.

NG, R. 2006. Digital light field photography. Ph.D. thesis, Stanford university.

NG, R., LEVOY, M., BRÉDIF, M., DUVAL, G., HOROWITZ, M., AND HANRAHAN, P. 2005. Light field photography with a hand-held plenoptic camera. CSTR 2005-02, Stanford University.

PEREZ NAVA, F. AND LUKE, J. P. 2009. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In *3DTV Conference.* 1–4.

PERWASS, C. AND WIETZKE, L. 2012. Single lens 3D-camera with extended depth-of-field. In *SPIE Electronic Imaging.* 22–26.

RAMAMOORTHI, R. AND HANRAHAN, P. 2001. A signal-processing framework for inverse rendering. In *SIGGRAPH '01.* 117–128.

SEITZ, S. M., MATSUSHITA, Y., AND KUTULAKOS, K. N. 2005. A theory of inverse light transport. In *ICCV.* Vol. 2. 1440–1447.

SHROFF, S. A. AND BERKNER, K. 2013. Image formation analysis and high resolution image reconstruction for plenoptic imaging systems. *Applied Optics 52,* D22–D31.

TAO, M. W., HADAP, S., MALIK, J., AND RAMAMOORTHI, R. 2013. Depth from combining defocus and correspondence using light-field cameras. In *ICCV.*

VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM TOG 26,* 3, 69.

VENKATARAMAN, K., LELESCU, D., DUPARRÉ, J., MCMAHON, A., MOLINA, G., CHATTERJEE, P., MULLIS, R., AND NAYAR, S. 2013. PiCam: an ultra-thin high performance monolithic camera array. *ACM TOG 32,* 6, 166.

WANNER, S. AND GOLDLUECKE, B. 2012a. Globally consistent depth labeling of 4D lightfields. In *CVPR.*

WANNER, S. AND GOLDLUECKE, B. 2012b. Spatial and angular variational super-resolution of 4D light fields. In *ECCV.* 608–621.

WETZSTEIN, G., IHRKE, I., AND HEIDRICH, W. 2013. On plenoptic multiplexing and reconstruction. *IJCV 101,* 2, 384–400.

WETZSTEIN, G., LANMAN, D., HEIDRICH, W., AND RASKAR, R. 2011. Layered 3D: tomographic image synthesis for attenuation-based light field and high dynamic range displays. In *ACM TOG.* Vol. 30. 95.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. *ACM TOG 31,* 4, 80.

YU, Z., YU, J., LUMSDAINE, A., AND GEORGIEV, T. 2012. An analysis of color demosaicing in plenoptic cameras. *CVPR,* 901–908.

ZHANG, Z. AND LEVOY, M. 2009. Wigner distributions and how they relate to the light field. 1–10.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3D displays. In *EGSR'06.*