Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing

Ashok Veeraraghavan Ramesh Raskar Amit Agrawal Ankit Mohan[†] Jack Tumblin[‡]
Mitsubishi Electric Research Labs (MERL)[*], Cambridge, MA Northwestern University, Evanston, IL

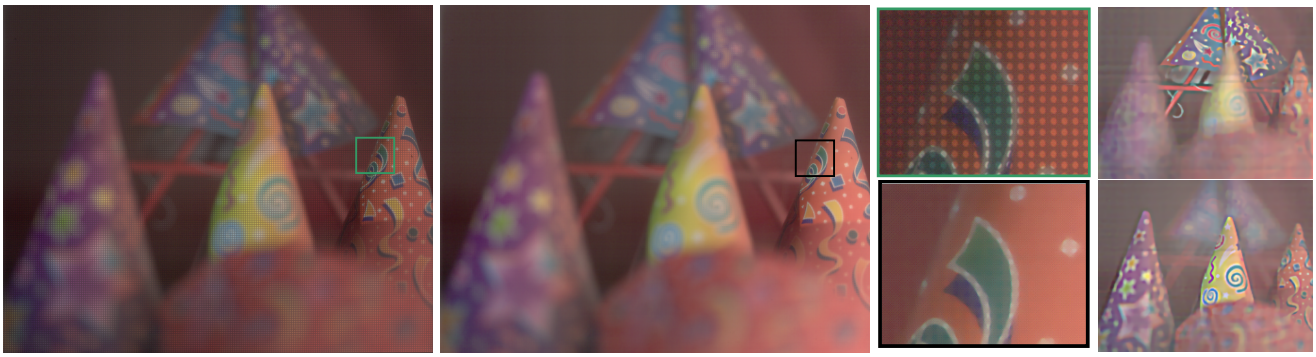

Figure 1: Our heterodyne light field camera provides 4D light field and full-resolution focused image simultaneously. (First Column) Raw sensor image. (Second Column) Scene parts which are in-focus can be recovered at full resolution. (Third Column) Inset shows fine-scale light field encoding (top) and the corresponding part of the recovered full resolution image (bottom). (Last Column) Far focused and near focused images obtained from the light field.

Abstract

We describe a theoretical framework for reversibly modulating 4D light fields using an attenuating mask in the optical path of a lens based camera. Based on this framework, we present a novel design to reconstruct the 4D light field from a 2D camera image without any additional refractive elements as required by previous light field cameras. The patterned mask attenuates light rays inside the camera instead of bending them, and the attenuation recoverably encodes the rays on the 2D sensor. Our mask-equipped camera focuses just as a traditional camera to capture conventional 2D photos at full sensor resolution, but the raw pixel values also hold a modulated 4D light field. The light field can be recovered by rearranging the tiles of the 2D Fourier transform of sensor values into 4D planes, and computing the inverse Fourier transform. In addition, one can also recover the full resolution image information for the in-focus parts of the scene.

We also show how a broadband mask placed at the lens enables us to compute refocused images at full sensor resolution for layered Lambertian scenes. This partial encoding of 4D ray-space data enables editing of image contents by depth, yet does not require computational recovery of the complete 4D light field.

1. Introduction

The trend in computational photography is to capture more optical information at the time of capture to allow greater post-capture image processing abilities. The pioneering work of Ng *et al.* [2005] has shown a hand-held plenoptic camera where the user can adjust focus and aperture settings after the picture has been taken. The key idea is to capture the entire 4D light field entering via the lens and incident on the camera sensor. In a conventional camera, the sensed 2D image is a 2D projection of the 4D light field [Ng 2005] and it is not possible to recover the entire 4D light field. Using a clever arrangement of optical elements, it is possible to re-bin the 4D rays and capture them using a 2D sensor [Georgiev et al. 2006; Ng et al. 2005]. These lens arrays perform the optical implementation of the two plane parameterization of the light field [Levoy and Hanrahan 1996; Gortler et al. 1996].

However, optical re-binning of rays forces a fixed and permanent tradeoff between spatial and angular resolution via the array of lenses. In this paper, we describe novel hybrid imaging/light field camera designs that are much more easily adjustable; users change a single attenuating mask rather than arrays of lenses. We call this *Dappled Photography*, as the mask shadows the incoming light and dapples the sensor. We exploit the fact that light rays can be linearly combined: rather than sense each 4D ray on its own pixel sensor, our design allows sensing linearly independent weighted sums of rays, rays combined in a coded fashion that can be separated by later decoding. Our mapping from 4D ray space to a 2D sensor array exploits *heterodyning* methods [Fessenden 1908] that are best described in the frequency domain. By exploiting the modulation and convolution theorems [Oppenheim et al. 1999] in the frequency domain, we derive simple attenuating mask elements that can be placed in the camera's optical path to achieve Fourier domain remapping. No additional lenses are necessary, and we can compute decoded rays as needed in software.

1.1. Contributions

We present a set of techniques to encode and manipulate useful portions of a 4D light field.

- We derive a 4D Fourier domain description of the effect of placing an attenuating mask at any position within a conven-

[*]emails: vashok@umd.edu (Currently at Univ. of Maryland), [raskar, agrawal]@merl.com Web: http://www.merl.com/people/raskar/Mask
[†]email: ankit@cs.northwestern.edu
[‡]email: jet@cs.northwestern.edu

ACM Reference Format
Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J. 2007. Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing. *ACM Trans. Graph. 26*, 3, Article 69 (July 2007), 12 pages. DOI = 10.1145/1239451.1239520 http://doi.acm.org/10.1145/1239451.1239520.

tional 2D camera.

- We identify a new class of 4D cameras that re-map the Fourier transform of 4D ray space onto 2D sensors. Previous 4D cameras used 2D lens arrays to project 4D ray-space itself rather than it's Fourier transform.
- We achieve this frequency domain re-mapping using a single transmissive mask, and our method does not require additional optical elements such as lens arrays.
- We analyze defocus blur as a special case of this frequency domain re-mapping and demonstrate that a broadband mask placed at the aperture can preserve high spatial frequencies in defocused images.

Our analysis leads to two camera designs:

Heterodyne Light Field Camera: The first design is based on the modulation theorem in the 4D frequency domain. We capture the light field using a 4D version of the method known as 'heterodyning' in radio. We create spectral tiles of the light field in the 4D frequency domain by placing high-frequency sinusoidal pattern *between* the sensor and the lens of the camera. To recover the 4D light field, we take the Fourier transform of the 2D sensed signal, re-assemble the 2D tiles into a 4D stack of planes, and take the inverse Fourier transform. Unlike previous 4D cameras that rely on lens arrays, this hybrid imaging/light field design does not force resolution tradeoffs for in-focus parts of the scene. The mask does not bend rays as they travel from scene to sensor, but only attenuates them in a fine, shadow-like pattern. If we compensate for this shadowing, we retain a full-resolution 2D image of the parts of the scene that were in focus, as well as the lower-resolution 4D light field we recover by Fourier-domain decoding. A prototype for this design is shown in Figure 2.

Encoded Blur Camera: The second design is based on the convolution theorem in the frequency domain. By placing a broadband mask in the lens aperture of a conventional 2D camera (Figure 2), we encode the defocus blur to preserve high spatial frequencies which can be recovered by image deblurring. We show how to computationally refocus the image at different depths for layered Lambertian scenes at full-resolution. We show that this computed refocusing is a special case of 4D re-mapping in the frequency domain that does not require measurement of the entire 4D light field, allowing us to avoid its huge resolution penalties.

For both designs, we show optimality criteria of the mask pattern and describe a procedure for computing highly efficient mask.

1.2. Benefits and Limitations

Mask-based hybrid imaging/light field cameras offer several advantages over previous methods. An attenuating mask is far simpler and less costly than lenses or lens arrays, and avoid errors such as spherical, chromatic aberration, coma, and mis-alignment. Simpler mounts and flexible masks may allow camera designs that offer user-selectable masks; photographers could then select any desired tradeoff in angle vs. spatial resolution. The design of Ng *et al.* [2005] matches main-lens aperture (f-stop) to the micro-lens array near the detector to avoid gaps or overlaps in their coverage of the image sensor; mask-only designs avoid these concerns.

Our mask based designs also impose limitations. Masks absorb roughly 50% of usable light that enters the lens. To counter this loss, we show that masks allow use of much larger apertures, up to a factor of 7 for the second design. Masks effectively reduce the lens aperture, inducing a proportional increase in blur from diffraction. This blur reduces our ability to compute refocused images by adding masks to diffraction-limited systems such as microscopes.

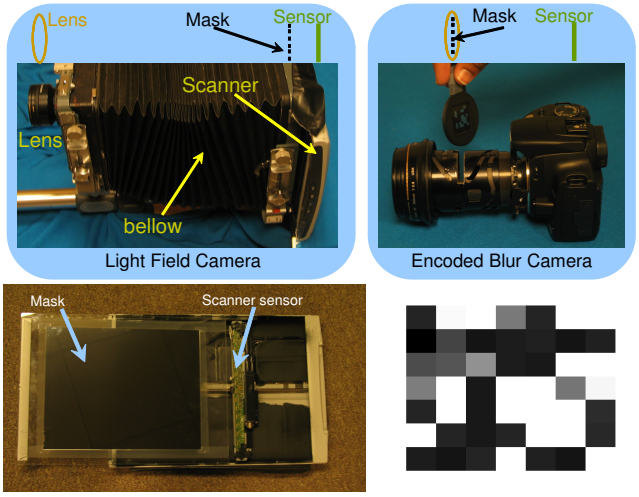

Figure 2: Prototype camera designs. (Top Left) Heterodyne light field camera holds a narrowband 2D cosine mask (shown in bottom left) near the view camera's line-scan sensor. (Top Right) Encoded blur camera holds a coarse broadband mask (shown in bottom right) in the lens aperture.

Our computed refocusing method based on image deblurring suffers from deconvolution noise, but we show that typical noise performance is 40 db better than conventional deconvolution of images taken with an open aperture of the same size.

1.3. Related Work

Light Field Acquisition: Integral Photography [Lippmann 1908] was first proposed almost a century ago to undo the directional integration of all rays arriving at one point on a film plane or sensor, and instead measure each incoming direction separately to estimate the entire 4D function. For a good survey of these first integral cameras and its variants, see [Okano et al. 1999; Martnez-Corral et al. 2004; Javidi and Okano 2002]. The concept of the 4D light field as a representation of all rays of light in free-space was proposed by Levoy and Hanrahan [1996] and Gortler *et al.* [1996]. While both created images from virtual viewpoints, Levoy and Hanrahan [1996] also proposed computing images through a virtual aperture, but a practical method for computing such images was not demonstrated until the thorough study of 4D interpolation and filtering by Isaksen *et al.* [2000]. Similar methods have also been called synthetic aperture photography in more recent research literature [Levoy et al. 2004; Vaish et al. 2004].

To capture 4D radiance onto a 2D sensor, following two approaches are popular. The first approach uses an array of lenses to capture the scene from an orderly grid of viewpoints, and the image formed behind each lens provides an orderly grid of angular samples to provide a result similar to integral photography [Ives 1928; Lippmann 1908]. Instead of fixed lens arrays, Wilburn *et al.* [2005] perfected an optically equivalent configuration of individual digital cameras. Georgiev *et al.* [2006] and Okano *et al.* [1997] place an array of positive lenses (aided by prisms in [Georgiev et al. 2006]) in front of a conventional camera. The second approach places a single large lens in front of an array of micro-lenses treating each sub-lens for spatial samples. These *plenoptic cameras* by Adelson *et al.* [1992] and Ng *et al.* [2005] form an image on the array of lenslets, each of which creates an image sampling the angular distribution of radiance at that point. This approach swaps the placement of spatial and angular samples on the image plane. Both these approaches trade spatial resolution for the ability to resolve angular differences. They require very precise alignment of microlenses with respect to

sensor.

Our mask-based heterodyne light field camera is conceptually different from previous camera designs in two ways. First, it uses *non-refractive* optics, as opposed to refractive optics such as microlens array [Ng et al. 2005]. Secondly, while previous designs sample individual rays on the sensor, mask-based design samples *linear combination* of rays in Fourier space. Our approach also trades spatial resolution for angular resolution, but the 4D radiance is captured using information-preserving coding directly in the Fourier domain. Moreover, we retain the ability to obtain full resolution information for parts of the scene that were in-focus at capture time.

Coded Imaging: In astronomy, coded aperture imaging [Skinner 1988] is used to overcome the limitations of a pinhole camera. Modified Uniformly Redundant Arrays (MURA) [Gottesman and Fenimore 1989] are used to code the light distribution of distant stars. A coded exposure camera [Raskar et al. 2006] can preserve high spatial frequencies in a motion-blurred image and make the deblurring process well-posed. Other types of imaging modulators include mirrors [Fergus et al. 2006], holograms [Sun and Barbastathis 2005], stack of light attenuating layers [Zomet and Nayar 2006] and digital micro-mirror arrays [Nayar et al. 2006]. Previous work involving lenses and coded masks is rather limited. Hiura & Matsuyama [1998] placed a mask with four pin holes in front of the main lens and estimate depth from defocus by capturing multiple images. However, we capture a single image and hence lack the ability of compute depth at every pixel from the information in defocus blur. Nayar & Mitsunaga [2000] place an optical mask with spatially varying transmittance close to the sensor for high dynamic range imaging.

Wavefront Coding [Dowski and Cathey 1995; Dowski and Johnson 1999; van der Gracht et al. 1996] is another technique to achieve extended *Depth of Field (DOF)* that use aspheric lenses to produce images with a depth-independent blur. While their results in producing extended depth of field images are compelling, their design cannot provide a light field. Our design provides greater flexibility in image formation since we just use a patterned mask apart from being able to recover the light field. Passive ranging through coded apertures has also been studied in the context of both wavefront coding [Johnson et al. 2000] and traditional lens based system [Farid and Simoncelli 1998].

Several deblurring and deconvolution techniques have also been used to recover higher spatial frequency content. Such techniques include extended DOF images by refocusing a light field at multiple depths and applying the digital photomontage technique (Agarwala *et al.* [2004]) and fusion of multiple blurred images ([Haeberli 1994]).

2. Basics

For visualization purposes, we consider a 2D light field space (LS), with one spatial dimension x and one angular dimension θ and a 1D detector as shown in Figure 3. We denote variables by lower case letters and their corresponding Fourier domain representations by upper case letters. Let $l(x,\theta)$ denote the 2D light field parameterized by the twin plane parameterization as shown in Figure 3. The θ-plane is chosen to be the plane of the main lens (or the aperture stop for cameras composed of multiple lens) of the camera. For the case of planar Lambertian object, we assume that the x-plane coincides with the object plane.

2.1. Effects of Optical Elements on the Light Field

We now discuss the effect of various optical elements such as lens, aperture and sensor to the 2D light field in frequency domain, which

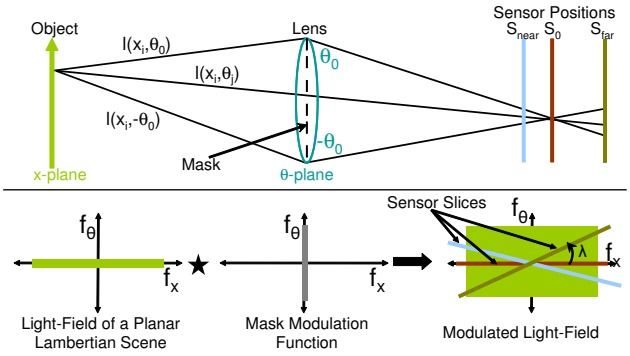

Figure 3: Encoded blur camera. (Top) In ray-space, focused scene rays from a scene point converge through lens and mask to a point on sensor. Out of focus rays imprint mask pattern on the sensor image. (Bottom) In Fourier domain. Lambertian scenes lack θ variation & form a horizontal spectrum. Mask placed at the aperture lacks x variation & forms a vertical spectrum. The spectrum of the modulated light field is a convolution of two spectrums. A focused sensor measures a horizontal spectral slice that tilts when out-of-focus.

we refer as *Fourier domain light field space* (FLS). The (x,θ) space is referred to as the primal domain.

Sensor: The image formed on a 1D sensor is a 1D projection of the 2D light field entering the camera, which also corresponds to a slice of the light field in Fourier domain. For different focus settings, the obtained images correspond to slices at different angles/trajectories [Ng 2005].

Lens: A thin lens shifts the x-plane of the light field to the conjugate plane according to the thin-lens equation. The lens also inverts the x-plane of the light field.

Aperture: The aperture of a camera acts as a limiter, allowing only the light rays that pass through the aperture to enter the camera. The light field l after passing through the aperture is given by

$$l_a(x,\theta) = l(x,\theta)a(x,\theta), \tag{1}$$

where $a(x,\theta)$ is the aperture modulation function given by $a(x,\theta) = \texttt{rect}(\frac{\theta}{2\theta_0})$, and $2\theta_0$ is the size of the aperture. From (1), the Fourier transform of the light field after the aperture is given by

$$L_A(f_x,f_\theta) = L(f_x,f_\theta) \otimes A(f_x,f_\theta), \tag{2}$$

where $\otimes$ denotes convolution. L and A are the Fourier transforms of the light field (before the aperture) and the aperture modulation function respectively. Since $a(x,\theta)$ is a `rect` function,

$$A(f_x,f_\theta) = 2a_0\,\texttt{sinc}(2a_0 f_\theta). \tag{3}$$

2.2. FLS and Information Content in the Light Field

A light field is a 4D representation of the light rays in the free-space. A 2D sensor can only sample a 2D slice of this light field. Depending on the scene, the information content in the light field is concentrated in different parts of the light field.

2.2.1. Planar Lambertian Object

Let us assume that the scene being imaged consists of a planar Lambertian object at the focus plane. Since there are no angular variations in the irradiance of rays from a Lambertian object, the information content of its light field is restricted to be along the f_x axis

(Figure 3). Thus, $L(f_x, f_\theta) = 0, \forall f_\theta \neq 0$. Since $L(f_x, f_\theta)$ is independent of f_θ and $A(f_x, f_\theta)$ is independent of f_x, from (2) and (3) we obtain,

$$\begin{aligned} L_A(f_x, f_\theta) &= L(f_x, f_\theta) \otimes A(f_x, f_\theta), & (4)\\ &= L(f_x, 0) A(0, f_\theta), & (5)\\ &= 2a_0 L(f_x, 0) \mathtt{sinc}(2a_0 f_\theta). & (6) \end{aligned}$$

The sensed image is a slice of this modulated light field. When the sensor is in focus, all rays from a scene point converge to a sensor pixel. Thus, the in-focus image corresponds to a slice of $L_A(f_x, f_\theta)$ along f_x ($f_\theta = 0$). Let $y(s)$ and $Y(f_s)$ denotes the sensor observation and its Fourier transform respectively. For an in-focus sensor

$$Y(f_s) = L_A(f_s, 0) = 2a_0 L(f_s, 0). \quad (7)$$

Thus, no information is lost when the Lambertian plane is in focus.

When the sensor is out of focus, the sensor image is a slanted slice of the modulated light field as shown in Figure 3, where the slant angle λ depends on the degree of mis-focus. Thus,

$$\begin{aligned} Y(f_s) &= L_A(f_s \cos\lambda, f_s \sin\lambda),\\ &= 2a_0 L(f_s \cos\lambda, 0) \mathtt{sinc}(2a_0 f_s \sin\lambda) \end{aligned} \quad (8)$$

Thus, for out of focus setting, the light field gets attenuated by the frequency transform of the aperture modulation function, which is a sinc function for an open aperture. This explains the attenuation of the high spatial frequencies in the captured signal when the scene is out of focus. Thus, we need to modify the aperture so that the resulting aperture modulation function has a *broadband* frequency response, ensuring that high spatial frequencies are preserved in out of focus images.

Incidentally, for a pinhole camera, the aperture function is a Dirac delta function and the aperture modulation function is broadband in f_θ. This explains why the images captured via a pinhole camera are always in-focus. However, a pinhole camera suffers from severe loss of light, reducing the signal to noise ratio (SNR) of the image. In Section 4, we show that one can use a carefully selected mask to perform the function of a broadband modulator of the light field in f_θ and realize greater DOF for Lambertian scenes, while increasing the amount of light captured as compared to a pinhole.

2.2.2. Bandlimited Light Fields

For general scenes, we assume that the light field is bandlimited to f_{x0} and $f_{\theta 0}$ as shown in Figure 5: $L(f_x, f_\theta) = 0 \quad \forall |f_x| \geq f_{x0}, |f_\theta| \geq f_{\theta 0}$. A traditional camera can only take a 2D slice of the 4D light field. To recover the entire information content of the light field, we need to modulate the incoming light field so as to redistribute the energy from the 4D FLS to the 2D sensor.

3. Heterodyne Light Field Camera

In this section, we show that the required modulation can be achieved in frequency domain by the use of an appropriately chosen 2D mask placed at an appropriate position between the lens and the sensor. Although a mask is only a 2D modulator, in tandem with the lens, it can achieve the desired 4D modulation. We believe that this is the first design of a single-snapshot light field camera that does not use any additional lenses or other refractive elements.

3.1. Modulation Theorem and its Implications

According to the modulation theorem [Oppenheim et al. 1999], when a baseband signal $s(x)$ is multiplied by a cosine of frequency f_0, it results in copies of the signal at that frequency.

$$\mathfrak{F}[\cos(2\pi f_0 x) s(x)](f_x) = \frac{1}{2}(F(f_x - f_0) + F(f_x + f_0)), \quad (9)$$

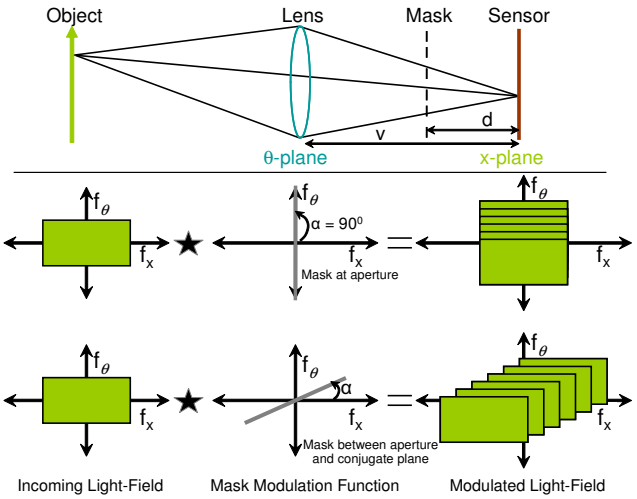

Figure 4: Heterodyne light field camera. (Top) In ray-space, the cosine mask at d casts soft shadows on the sensor. (Bottom) In Fourier domain, scene spectrum (green on left), convolved with mask spectrum (center) made of impulses creates offset spectral tiles (right). Mask spectral impulses are horizontal at $d = 0$, vertical at $d = v$, or tilted.

where $\mathfrak{F}[s(x)](f_x) = F(f_x)$ denotes the Fourier transform of $s(x)$. This principle has been widely used in telecommunications and radio systems. The baseband signal is modulated using a *carrier* signal of much higher frequency so that it can be transmitted over long distances without significant loss of energy. The receiver demodulate the received signal to recover the baseband signal. In essence, what we wish to achieve is very similar. We would like to modulate the information in the angular variations of the light field (f_θ frequencies) to higher frequencies in f_x so that the high resolution 1D sensor may be able to sense this information.

Figure 5 shows a bandlimited light field in frequency domain. For simplicity, let us assume the x plane to be the conjugate plane, so that the sensor image corresponds to a slice along f_x (horizontal slice). Now consider a modulator whose frequency response is composed of impulses arranged on a slanted line as shown in Figure 5. If the light field is modulated by such a modulator, each of these impulses will create a spectral replica of the light field at its center frequency. Therefore, the result of this convolution will be several spectral replicas of the light field along the slanted line. The elegance of this specific modulation is that the horizontal slice (dashed box) of the modulated light field spectrum now captures all the information in the original light field. Note that the angle α is designed based upon the required frequency resolution in θ and x, and the bandwidth of the incoming light field.

Heterodyne receivers in telecommunications demodulate the incoming signal to recover the baseband signal. In our case, demodulation must also redistribute the energy in the sensed 1D signal to the 2D light field space. The process of demodulation consists of rearranging the frequency response of the sensor to recover the bandlimited light field as shown in Figure 5.

3.2. Mask based Heterodyning

Now we show that the required modulation can be achieved by placing a suitably chosen attenuating mask in the optical path of a conventional camera.

Masks as Light Field Modulators: A mask is essentially a special 1D code c(y) (2D for 4D light field) placed in the optical path. In flatland, although the mask is 1D, its modulation function is 2D

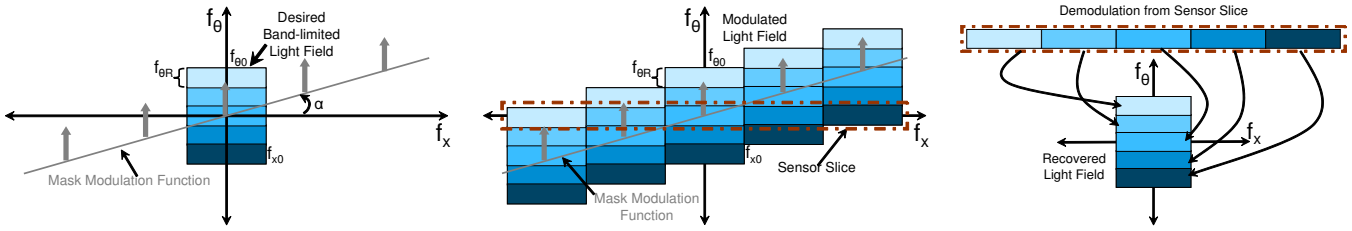

Figure 5: Spectral slicing in heterodyne light field camera. (Left) In Fourier domain, the sensor measures the spectrum only along the horizontal axis ($f_\theta = 0$). Without a mask, sensor can't capture the entire 2D light field spectrum (in blue). Mask spectrum (gray) forms an impulse train tilted by the angle α. (Middle) By the modulation theorem, the sensor light field and mask spectra convolve to form spectral replicas, placing light field spectral slices along sensor's broad $f_\theta = 0$ plane. (Right) To re-assemble the light field spectrum, translate segments of sensor spectra back to their original f_x, f_θ locations.

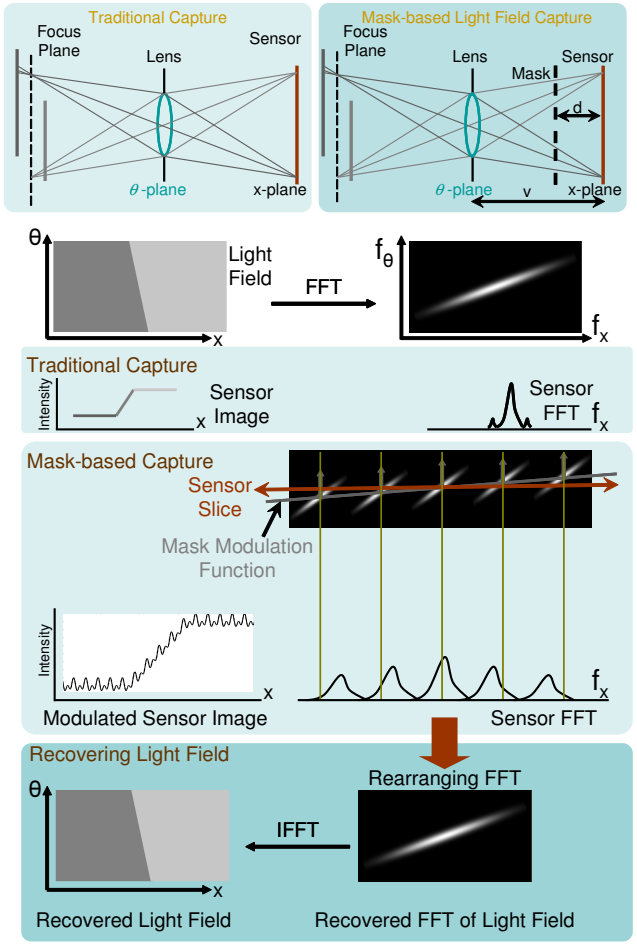

Figure 6: Ray space and Fourier domain illustration of light field capture. The flatland scene consists of a dark background planar object occluded by a light foreground planar object. In absence of a mask, the sensor only captures a slice of the Fourier transform of the light field. In presence of the mask, the light field gets modulated. This enables the sensor to capture information in the angular dimensions of the light field. The light field can be obtained by rearranging the 1D sensor Fourier transform into 2D and computing the inverse Fourier transform.

(Figure 4). The mask affects the light field differently depending on where it is placed. If the mask is placed at the aperture stop (θ plane), then the effect of mask is to multiply the aperture modulation function by the mask modulation function. The mask modulation function $m(x,\theta)$ is then given by $m(x,\theta) = c(y=\theta)$, i.e., the modulation function is independent of x. Intuitively, when placed at the θ-plane, the mask affects all rays at an angle θ in similar way, independent of the scene point from which they are originating.

If the mask is placed at the conjugate plane, it attenuates all rays (independent of θ) for same x equally. This is because at the conjugate plane, all rays originating from a point on the plane of focus converge to a single point. Thus, the mask modulation function changes to $m(x,\theta) = c(y=x)$.

Thus, we see that the modulation function corresponding to placing the *same* code at the aperture and the conjugate plane are related by a rotation of 90° in the 2D light field space. Moreover, as the 1D code is moved from the aperture plane to the plane of the sensor, the resulting mask modulation function gets rotated in 2D as shown in Figure 4.

If the mask $c(y)$ is placed at a distance d from the conjugate plane, the mask modulation function is given by

$$M(f_x, f_\theta) = C(f_x \csc(\alpha))\delta(f_\theta - f_x \tan\alpha), \tag{10}$$

where C denotes the Fourier transform of the 1D mask and v is the distance between the aperture and the conjugate plane. The angle α is given by

$$\alpha = \frac{d}{v}\frac{\pi}{2}. \tag{11}$$

In other words, the mask modulation function has all its energy concentrated on a *line* in the 2D FLS space. The angle α of this line with respect to the f_x axis depends upon the position of the mask. When the mask is placed at the conjugate plane ($d = 0$), the angle α is equal to 0. As the mask moves away from the conjugate plane towards the aperture, this angle increases linearly to 90° at the aperture plane as shown in Figure 4,

Optimal Mask Position: In order to capture the 2D light field, we need the modulation function $M(f_x, f_\theta)$ to be a series of impulses at an angle α given by

$$\alpha = \arctan\frac{2f_{x0}}{f_{\theta R}}, \tag{12}$$

where f_{x0} is the bandwidth of the light field along the f_x axis and $f_{\theta R}$ represents the desired frequency resolution along the f_θ axis. For example, in Figure 5, the frequency resolution has been depicted as being equal to $f_{\theta R} = (2/5)f_{\theta 0}$, where $f_{\theta 0}$ is the bandwidth of the light field along the f_θ axis. Thus, for capturing a light field of a given bandwidth, the physical position of the mask can be

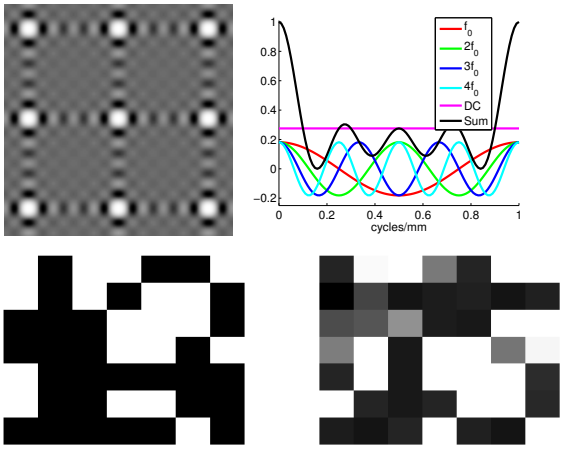

Figure 7: (Top Left) Zoom in of a part of the cosine mask with four harmonics. (Top Right) Plot of 1D scan line of mask (black), as sum of four harmonics and a constant term. (Bottom Left) 7×7 binary valued mask. (Bottom Right) 7×7 continuous valued mask with a higher minimum value of the magnitude of zero padded Fourier transform.

calculated from (12) and (11). In practice, since the spatial resolution is much larger than the angular resolution, α is very small, and therefore the mask needs to be placed close to the sensor.

Optimal Mask Pattern: To achieve $M(f_x, f_\theta)$ as a set of 1D impulses on a slanted 2D line, the Fourier transform $C(f)$ of the 1D mask should be a set of impulses. Let $2p+1$ be the number of impulses in $M(f_x, f_\theta)$. The Fourier transform of the 1D mask is then given by

$$C(f) = \sum_{k=-p}^{k=p} \delta(f - k f_0), \tag{13}$$

where f_0 denotes the fundamental frequency and is given by $f_0 = \sqrt{4{f_{x0}}^2 + {f_{\theta R}}^2}$. From Figure 5, $(2p+1) f_{\theta R} = 2 f_{\theta 0}$. The bandwidth in f_θ is discretized by $f_{\theta R}$. Hence, the number of angular samples obtained in the light field will be equal to $\frac{2 f_\theta}{f_{\theta R}} = 2p+1$. Since the Fourier transform of the optimal mask is a set of symmetric Dirac delta functions (along with DC), this implies that the physical mask is a sum of set of *cosines* of a given fundamental frequency f_0 and its harmonics. The number of required harmonics is in fact p, which depends upon the band-width of the light field in the f_θ axis and the desired frequency resolution $f_{\theta R}$.

Solving for 2D Light Field: To recover the 2D light field from the 1D sensor image, we compute the Fourier transform of the sensor image, *reshape* the 1D Fourier transform into 2D as shown in Figure 5 and compute the inverse Fourier transform. Thus,

$$l(x, \theta) = \texttt{IFT}(\texttt{reshape}(\texttt{FT}(y(s)))), \tag{14}$$

where FT and IFT represent the Fourier and inverse Fourier transforms respectively, and $y(s)$ is the observed sensor image. Figure 6 shows a simple example of light field capture where the scene consists of a dark background plane occluded by a light foreground plane.

3.3. Note on 4D Light Field Capture

Even though the analysis and the construction of mask-based heterodyning for light field capture was elucidated for 2D light fields, the procedure remains identical for capturing 4D light fields with 2D sensors. The extension to the 4D case is straightforward. In

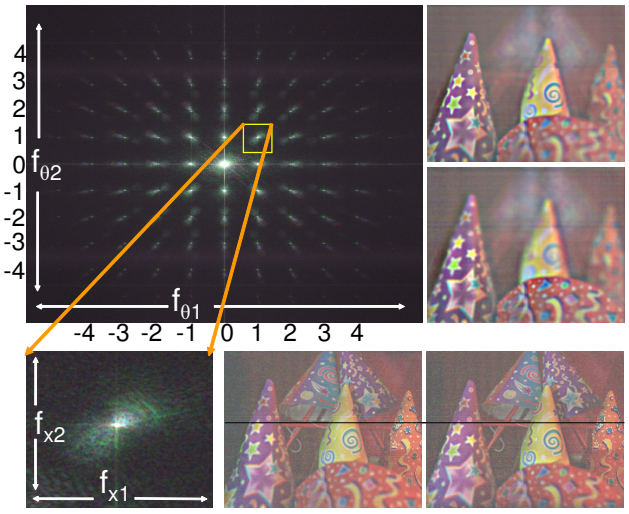

Figure 8: (Top Left) Magnitude of the 2D Fourier transform of the captured photo shown in Figure 1. θ_1, θ_2 denote angular dimensions and x_1, x_2 denote spatial dimensions of the 4D light field. The Fourier transform has 81 spectral tiles corresponding to 9×9 angular resolution. (Bottom Left) A tile of the Fourier transform of the 4D light field corresponding to $f_{\theta 1} = 1, f_{\theta 2} = 1$. (Top Right) Refocused images. (Bottom Right) Two out of 81 views. Note that for each view, the entire scene is in focus. The horizontal line depicts the small parallax between the views, being tangent to the white circle on the purple cone in the right image but not in the left image.

case of a 4D light field, the information content in the 4D light field is heterodyned to the 2D sensor space by the use of a 2D mask placed between the aperture and the sensor. The Fourier transform of the 2D mask would contain a set of impulses on a 2D plane.

$$C(f_1, f_2) = \sum_{k_1=-p_1}^{k_1=p_1} \sum_{k_2=-p_2}^{k_2=p_2} \delta(f_1 - k_1 f_0^x, f_2 - k_2 f_0^y). \tag{15}$$

Since negative values in the mask cannot be realized as required, we need to boost the DC component of $C(f_1, f_2)$ so as to make the mask positive throughout. Figure 7 shows a part of the 2D cosine mask we used for experiments, along with the plot of one of its scanline. This 2D mask consists of four harmonics in both dimensions ($p_1 = 4, p_2 = 4$) with fundamental frequencies f_0^x and f_0^y being equal to 1 cycle/mm. This allows an angular resolution of 9×9 in the 4D light field. Figure 8 shows the magnitude of the Fourier transform of the captured photo of the cones (as shown in Figure 1). The Fourier transform clearly shows 9×9 spectral tiles created due to the modulation by the mask. These spectral tiles encode the information about the angular variation in the incident light field. To recover the 4D light field, demodulation involves *reshaping* of the sensor Fourier transform in 4D. Let $t_1 = 2p_1 + 1$ and $t_2 = 2p_2 + 1$ be the number of angular samples in the light field and let the captured 2D sensor image be $N \times N$ pixels. We first compute the 2D FFT of the sensor image. Then we rearrange $t_1 \times t_2$ tiles of the 2D Fourier transform into 4D planes to obtain a $(N/t_1) \times (N/t_2) \times t_1 \times t_2$ 4D Fourier transform. Inverse FFT of this 4D Fourier transform gives the 4D light field. In Figure 1, using a $1629 * 2052$ pixel image captured with a cosine mask having four harmonics, we obtain a light field with 9×9 angular resolution and 181×228 spatial resolution.

3.4. Aliasing

Traditionally, undersampling results in masquerading of higher frequencies as lower frequencies in the *same* channel and leads to vi-

Figure 9: Our heterodyne light field camera can be used to refocus on complex scene elements such as the semi-transparent glass sheet in front of the picture of the girl. (Left) Raw sensor image. (Middle) Full resolution image of the focused parts of the scene can be obtained as described in Section 3.6. (Right) Low resolution refocused image obtained from the light field. Note that the text on the glass sheet is clear and sharp in the refocused image.

sually obtrusive artifacts like ghosting. In heterodyne light field camera, when the band-limit assumption is not valid in the spatial dimension, the energy in the higher *spatial* frequencies of the light field masquerade as energy in the lower *angular* dimensions. No purely spatial frequency leaks to other purely spatial frequency. Thus, we do not see familiar jaggies, moire-like low-frequency additions and/or blocky-ness in our results. The effect of aliasing is discussed in detail in [Veeraraghavan et al. 2007], where using the statistics of natural images, it is shown that the energy in the aliasing components is small. To further combat the effects of aliasing, we post-filter the recovered light field using a Kaiser-Bessel filter with a filter width of 1.5 [Ng 2005].

3.5. Light Field based Digital Refocusing

Refocused images can be obtained from the recovered Fourier transform of the light field by taking appropriate slices [Ng 2005]. Figure 1 and Figure 8 shows refocused cone images. The depth variation for this experiment is quite large. Notice that the orange cone in the far right was in focus at the time of capture and we are able to refocus on all other cones within the field of view. Figure 10 shows the performance of digital refocusing with varying amounts of blur on the standard ISO-12233 resolution chart. Using the light field, we were able to significantly enhance the DOF. It is also straightforward to synthesize novel views from the recovered light field. Two such views generated from the recovered light field are also shown in the bottom right part of Figure 8. The horizontal line on the images depicts small vertical parallax between the two views. Digital refocusing based on recovered light fields allow us to refocus even in the case of complicated scenes such as the one shown in Figure 9. In this example, a poster of the girl in the back is occluded by a glass sheet in front. Notice that the text 'Mask based Light Field' written on the glass sheet is completely blurred in the captured photo. By computing the light field, we can digitally refocus on the glass sheet bringing the text in focus.

3.6. Recovering High Resolution Image for Scene Parts in Focus

Our heterodyne light field camera has an added advantage that we can recover high resolution information for the *in-focus* Lambertian parts of the scene. Consider a scene point that is in sharp focus. All rays from this scene point reach the *same* sensor pixel but are attenuated differently due to the mask. Therefore, the sensor pixel value is the product of the scene irradiance and the average value of the mask within the cone of rays reaching that pixel. This attenuation $\gamma(x,y)$ varies from pixel to pixel and can either be computed analytically or recovered by capturing a single calibration image

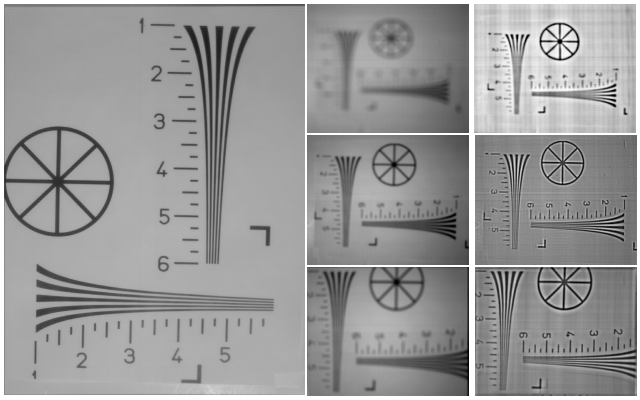

Figure 10: Analysis of the refocusing ability of the heterodyne light field camera. (Left) If the resolution chart is in focus, one can obtain a full resolution 2D image as described in Section 3.6, along with the 4D light field. (Middle) We capture out of focus chart images for three different focus settings. (Right) For each setting, we compute the 4D light field and obtain the low resolution refocused image. Note that large amount of defocus blur can be handled.

of a uniform intensity Lambertian scene. We can recover the high resolution image $I(x,y)$ of the scene points in focus as

$$I(x,y) = s(x,y)/\gamma(x,y), \tag{16}$$

where $s(x,y)$ is the captured sensor image. Parts of the scene that were not in focus at the capture time will have a spatially varying blur in $I(x,y)$. We use the image of a uniform intensity Lambertian light box as γ.

In Figure 1, zoomed in image region shows the attenuation of the sensor image due to the cosine mask. The recovered high resolution picture is also shown in Figure 1 and the inset shows the fine details recovered in the parts of the image that were in focus. Figure 10 shows the recovered high resolution picture of a resolution chart that was in focus during capture. This ability to obtain high resolution images of parts of the scene along with the 4D light field makes our approach different from previous light field cameras.

4. Encoded Blur Camera

In the previous section, we discussed a design for a light field camera with the aid of an attenuating mask. In this section, we look at a very specific sub-class of light fields; those that results from layered Lambertian scenes. For such scenes we show that using a broadband mask at the aperture is a very powerful way of achieving full-resolution digital refocusing. In conventional cameras, photographers can control the DOF by controlling the size of the aperture. As the aperture size decreases, the DOF of the camera increases proportionally, but the SNR decreases due to the loss of light. In Section 2.2.1, we showed that an open aperture suppresses high spatial frequencies in the out of focus image. To preserve high spatial frequencies, we place a physical mask at the aperture whose frequency response is broadband as shown in Figure 2.

For a mask placed at the aperture, $M(f_x, f_\theta)$ has all its energy concentrated along f_θ direction from (10). Thus, $M(f_x, f_\theta) = 0, \forall f_x \neq 0$. The frequency transform of the mask modulated light field is

$$L_M(f_x, f_\theta) = L(f_x, f_\theta) \otimes M(f_x, f_\theta). \tag{17}$$

Since for a Lambertian scene, $L(f_x, f_\theta) = 0, \forall f_\theta \neq 0$, the above equation simplifies to $L_M(f_x, f_\theta) = L(f_x, 0)M(0, f_\theta)$. Thus, the mask modulation function gets *multiplied* by the frequency transform of the light field. In primal domain, this is equivalent to a convolution of the mask and the sharp image of the scene. The scale

Figure 11: PSF estimation. (Left) Captured photo. (Middle) Initial crude segmentation of the scene based on the error maps indicate two dominant layers (person and painting) apart from homogeneous regions with scales 7 and 4 respectively. (Right) Final labeling for refocused image on the person by composing I_7, reblurred I_4 and the captured blurred image.

of the mask is dependent on the degree of defocus blur. The sharp image can be recovered by deconvolution of the blurred image with the scaled mask. The same conclusion may also be reached from ray based analysis of the captured photo.

4.1. Optimal Mask for Encoding Defocus Blur

Since the frequency transform of the light field gets multiplied by the mask modulation function, the optimal mask is the one which is broadband in frequency domain. Broadband masks popularly known as MURA codes have been used in lens-less coded aperture imaging in astronomy. However, a lens based coded aperture is significantly different from traditional lens-less coded aperture. In traditional coded aperture imaging, every scene element is *circularly* convolved with the mask. Instead, for a lens based coding, the observed image is a *linear* convolution of the sharp image with the defocus point spread function (PSF). Since linear convolution is equivalent to circular convolution with zero padded kernel, the optimal mask for lens based coded aperture is different from MURA. This was also observed by Raskar *et al.* [2006] in searching for an optimal 1D broadband code in the context of motion deblurring. Moreover, coded aperture imaging in astronomy can improve SNR only for point like sources such as stars and give no additional benefit over pin-holes for area light sources [Accorsi et al. 2001]. Hence it is not suitable for photography of natural scenes.

For the problem of motion deblurring, Raskar *et al.* [2006] performed a brute force linear search for obtaining the best 1D *binary* code. The code selection was based on maximizing the minimum of the DFT magnitudes of the zero padded code. Here we show that continuous valued codes can give superior performance compared to binary codes, with the advantage of significantly reducing the search time. We find the continuous valued code by gradient descent optimization (using Matlab function `fmincon`) based on the same selection criteria as above. A sub-optimal binary code (such as MURA) can be provided as the initial guess. Figure 7 shows a 7×7 binary mask obtained after 10 machine hours of search along with a continuous valued mask obtained within few minutes of optimization. Using the noise analysis presented in Section 5, the deconvolution noise for the continuous valued code is smaller by 7.3dB compared to the binary code.

4.2. Deconvolution based Digital Refocusing

We achieve full resolution digital refocusing from a single encoded out of focus image using image deconvolution techniques. Defocus blur in the captured photo is related to the depth of the scene. Although depth from defocus [Chaudhuri and Rajagopalan 1999] is an active area of research in computer vision, computing a depth map from a *single* defocused image is challenging, unless a priori knowledge about the scene is assumed or learning based approaches are used. Instead, we assume that the scene is made up of n distinct layers, where n is a small number and the defocus point spread function (PSF) within each layer is spatially invariant. This assumption works well for a variety of scenes. We also assume that the maximum blur diameter in the image can be T pixels.

We achieve refocusing in two steps. First, we analyze the scene and estimate the number of layers and the scale of the PSF for each layer automatically. We then generate n deblurred images, $I_1 \ldots I_n$, by deconvolving the captured blurred image by the estimated blur kernels. To refocus at a layer i, we *reblur* the remaining $n-1$ images according to the difference of their blur from the blur of layer i and then composite I_i and the reblurred images to get the refocused image.

4.2.1. PSF Estimation

Let $m(x,y)$ denote the $w \times w$ 2D mask placed in the aperture. For simplicity, assume that the entire image has a single layer with a defocus blur width of k pixels. The captured photo B is related to the sharp image I via convolution as

$$B(x,y) = I(x,y) * m(kx/w, ky/w) + \eta, \tag{18}$$

where η denote the measurement noise. Given I, the likelihood error can be written as

$$e_l(x,y) = (B(x,y) - I(x,y) * m(kx/w, ky/w))^2. \tag{19}$$

However, this error itself is not sufficient to uniquely determine I and k because $e_l(x,y)$ can be made equal to zero by assuming $B = I$. To resolve this ambiguity, we use the statistics of natural images. It has been shown that real-world images obey heavy tail distributions in their gradients [Field 1994]. In a blurred image, since high spatial gradients are suppressed, the tail of the gradient distribution will be suppressed. We use the fourth-order moment (kurtosis) of gradients as a statistic for characterizing the gradient distribution. The kurtosis will be small for blurred image gradients as compared to sharp image gradients. At every pixel, we define the gradient error, $e_g(x,y)$, using the kurtosis of gradients within a small neighborhood R around that pixel.

$$e_g(x,y) = -(kurtosis(\{I_x(x,y)\}_R) + kurtosis(\{I_y(x,y)\}_R), \tag{20}$$

where I_x, I_y denote the spatial gradients of I. However, deblurring at an incorrect scale larger than the correct scale k introduces high frequency deconvolution artifacts in I. This may increase the gradient kurtosis, thereby decreasing e_g. Thus, the two error measures compete with each other. To locate the correct scale, we minimize the combined error $e(x,y) = e_l(x,y) + \beta e_g(x,y)$, where β is a constant.

In the presence of multiple (n) layers, we deblur the given image using blur kernels of different sizes, ranging from 1 to T pixels. For each of these T deblurred images, we compute the error map $e(x,y)$. For a layer with correct scale k, the k^{th} error map should have the smallest values for the region corresponding to that layer among all the error maps. This is equivalent to a discrete labeling problem for each pixel with T labels. The labeling cost at a pixel (x,y) for a given label k is given by $e^k(x,y)$. We solve this labeling problem using the alpha-expansion graph-cut procedure (and

Figure 12: Full resolution digital refocusing using encoded blur camera. (Left) Captured photo where both the person in front and the painting are blurred. (Middle) Refocused image, the person has been brought into focus. (Right) Since the defocus PSF is made broadband by inserting a broadband mask in the aperture, deconvolution can recover fine features such as the glint in the eye and the hair strands.

software) by Boykov *et al.* [2001]. Since homogeneous regions in the image do not contain any blur information, we set the data cost for homogeneous regions to be zero, so that they get filled-in for each layer during graph cut optimization. We remove spurious layers having less than 10% of the total number of pixels in the image and perform simple morphological operations (hole filling) on the labels. Figure 11 shows a captured blurred image (person out of focus) and the resulting labeling indicating that the scene has two dominant layers (blue and green), apart from the homogeneous regions (yellow). This procedure only gives a crude segmentation of the scene in terms of layers and the corresponding scales. The exact boundaries between the layers are not obtained, but the interiors of the layers are labeled properly. This kind of labeling can also be obtained from a simple user interaction, where the user scribbles on the region corresponding to each layer in the *correct* deblurred image for that layer, given a set of T deblurred images.

4.2.2. Synthesizing Refocused Image

Since the scene has n layers, we only need to consider the n deblurred images $(I_1 \dots I_n)$ at the corresponding scales. We use the labeling in the interior from the previous step to build color histograms for each layer (each channel is treated separately) from the corresponding deblurred image. We also build histogram for homogeneous regions external to all the layers using the given blurred image. To refocus on a layer i, we *reblur* each of the $n-1$ images, $I_1, \dots I_{i-1}, I_{i+1} \dots I_n$ according to their scale difference from layer i. Finally, the refocused image is composed from I_i and the $n-1$ reblurred images. Again, this can be treated as a labeling problem and we use the procedure described in [Agarwala et al. 2004] to create the composite[1]. The data cost at each pixel is chosen as *maximum likelihood* using the color histograms and the seam objective is based on matching colors and gradients as described in [Agarwala et al. 2004]. Figure 11 also shows the final labeling.

4.3. Results

Figure 12 shows the full resolution refocused result, where the person in front has been brought into focus. Notice the glint in the eye of person and the fine hair strands have been recovered during deblurring.

Refocusing in Presence of Partial Occluders: Image completion and other hallucination techniques are used to fill in missing or unwanted regions of the image. However, such techniques may not work on out of focus blurred images. Since the hallucinated pixel values are not modeled according to the defocus blur, deblurring on such images will produce artifacts. Figure 13 shows such a scenario where the fence is in sharp focus and the person behind the fence is out of focus. Deblurring the image without taking the occluders into account will produce artifacts as shown. Since blurring distributes the information to neighboring pixels, we can recover the sharp image if the blur size is larger than the occluder size. Given a mask for the occluded pixels, we perform a weighted deconvolution of the image by solving

Figure 13: Deblurring in presence of partial occluders. (Top Left) Blurred photo of a person occluded by the in-focus fence. (Top Right) Deblurring without taking the occluders into account results in artifacts. (Bottom Left) Binary mask for the occluders. (Bottom Right) By solving the weighted deconvolution equation, one can remove these artifacts.

$$WA\mathbf{x} = W\mathbf{b}, \tag{21}$$

where $\mathbf{b}$ is the vectorized blurred image, A is the block-Toeplitz matrix representing 2D blurring and W is a weighting matrix that sets the weights corresponding to the occluded pixels in the blurred image to zero. In Figure 13, after obtaining the sharp image, we composite it with the sharp image of the fence, to bring both the person and fence in focus. Note that the mask for occluder can be over-estimated, as long as the blur is large enough.

Spatially Varying PSF: Figure 14 shows a tilted book with spatially varying defocus blur. To obtain an all in focus image, we fuse the deblurred images $I_1 \dots I_T$. We click on four points on the blurred image to estimate the homography of the book and estimate the PSF scale at each pixel using the scale at end points and the homography parameters. The deblurred images are then combined

[1]The captured blurred image is also used in the composite for homogeneous regions.

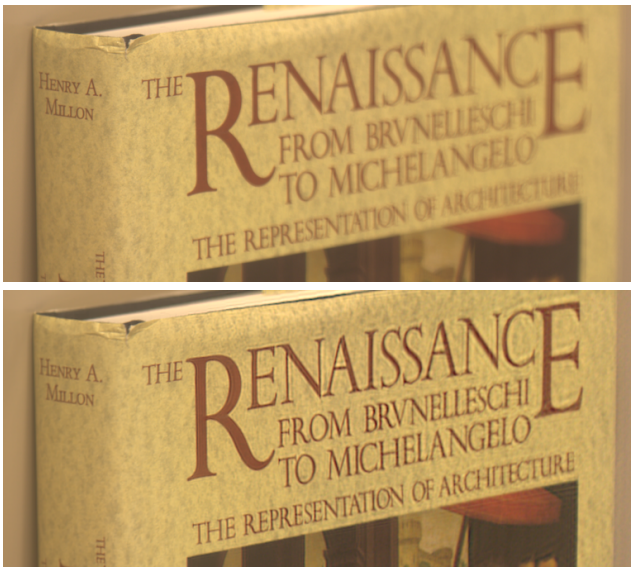

Figure 14: Spatially varying PSF can be handled for planar scenes using homography. Shown is an all focus composite obtained by fusing deblurred images at varying scales appropriately.

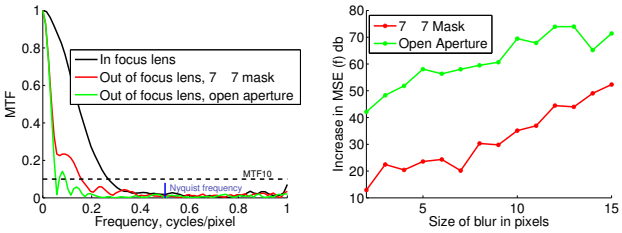

Figure 15: (Left) Comparison of the MTF of the lens in focus, and out of focus with and without the mask in the aperture. By putting the mask in the aperture, the MTF improves for out of focus lens. (Right) Noise analysis of the linear system for deblurring shows that the resulting linear system when using the mask is stable compared to that of the open aperture.

according to the spatially varying scale to obtain the all in focus image. Note that the word 'ARCHITECTURE' cannot be read in the blurred image but is sharp in the result.

5. Implementation and Analysis

Heterodyne Light Field Camera: We build a large format camera using a flatbed scanner (Canon CanoScan LiDE 70) similar to [Wang and Heidrich 2004; Yang 2000] and place a 8×10 inch^2 mask behind the scanner glass surface. The mask was printed at a resolution of 80 dots/mm using Kodak LVT continuous tone film recorder (BowHaus Inc.). The scanner itself was then placed on the back of a large format view camera fitted with a 210 mm f/5.6 Nikkor-W lens as shown in Figure 2. In practice, the motion of the scanner sensor is not smooth leading to pattern noise (horizontal/vertical lines) in the captured photo. This may lead to some artifacts in the recovered light fields. However, many of these issues will disappear with a finer mask placed inside a conventional digital camera. Calibration involves accounting for the in-plane rotation and shift of the mask with respect to the sensor which manifest as search for the rotation angle of the captured 2D image and the phase shift of the Fourier transform. Since the computation of the light field and refocusing is done in Fourier domain, computational burden is low.

Figure 16: Comparison of traditional focus deblurring with a mask in the aperture. (Top) Captured blurred photo and the deblurred image for open aperture. (Bottom) Captured blurred photo and the deblurred image using the 7×7 mask in the aperture.

Encoded Blur Camera: Figure 2 shows the prototype for encoded blur camera for extending the DOF. We use a Canon Rebel XT SLR camera with a Canon EF 100 mm f/2.8 USM Macro lens. We cut open the lens in the middle at the aperture and place a mask as shown in Figure 2. To avoid diffraction blur, we restrict ourselves to a low resolution 7×7 mask.

Figure 15 shows the effect of putting a mask in the aperture plane on the *modulation transfer function* (MTF) of the optical system. We capture out of focus images of the standard resolution chart ISO-12233 and use imatest software [Imatest] to compute the MTF. Figure 15 shows how the MTF degrades when the lens is out of focus, compared to the in-focus MTF. For an open aperture (box blur), the MTF degrades sharply and has zeros corresponding to certain spatial frequencies depending on the amount of defocus. Using a broadband mask in the aperture improves the MTF, especially for high spatial frequencies, facilitating the recovery of those frequencies.

Since deblurring can be represented as a linear system, standard covariance analysis can be used to analyze *deconvolution noise*. For a linear system $A\mathbf{x} = \mathbf{b}$, assuming zero mean Gaussian IID noise in $\mathbf{b}$ with variance σ^2, the covariance matrix Σ of the estimated $\hat{\mathbf{x}}$ is given by $\Sigma = \sigma^2 (A^T A)^{-1}$. Thus, the mean square error (MSE) increases by a factor of $f = \mathtt{trace}(\Sigma)/N^2 = \mathtt{trace}(A^T A)^{-1}/N^2$, where N^2 is the number of pixels in $\mathbf{x}$. Since A is of size $N^2 \times N^2$, it is not possible to obtain f analytically. We empirically estimate f by deblurring a noisy blurred synthetic image and comparing the result with the ground truth. The MSE was averaged over 1000 trials. Figure 15(b) shows the plot of f in db for different values of blur using the 7×7 mask and open aperture of the same size. For a 7 pixel blur, the open aperture leads to noise amplification by 58.02 dB, whereas by using the mask, it is reduced to 20.1db. Regularization algorithms such as Richardson-Lucy [Richardson 1972; Lucy 1974] are used to reduce noise in conventional deblurring but also results in loss of details as show by the comparisons in Figures 16. Note that the high spatial frequencies and details are recovered when using the mask as compared to open aperture.

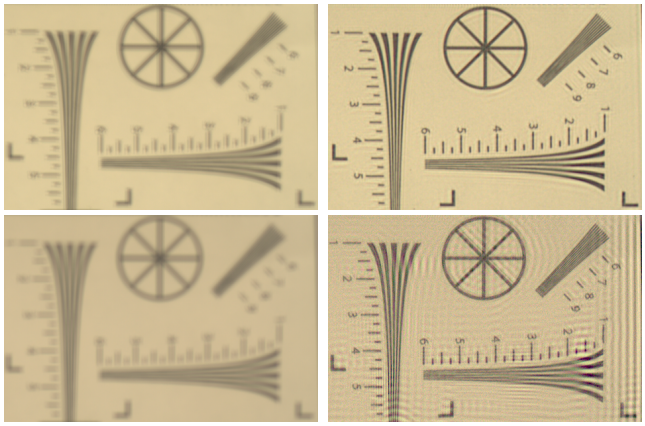

Figure 17: (Top Row) Blurred photo of the ISO-12233 chart captured using the 7×7 mask in the aperture and the corresponding deblurred image. (Bottom Row) Using a 25×25 mask for the same aperture size results in artifacts due to diffraction.

As the *mask resolution* increases, each cell of the mask becomes smaller for the same aperture size. Thus, diffraction starts playing a role and deblurring using a high resolution masks shows artifacts. Figure 17 shows the deblurred images of the ISO-12233 chart captured using masks of different resolution for the same out-of-focus lens setting. Note that the deblurred image has color artifacts due to diffraction when using 25×25 mask.

Failure Cases: The encoded blur camera assumes a layered Lambertian scene. Thus, scenes with large variation in depths and those with view dependencies and specularities cannot be handled. In practice, the 7×7 mask gives good deblurring result up to blur size of ≈ 20 pixels. To handle large defocus blur, one should use a finer resolution mask but that may lead to diffraction blur. Layers with overlapping high frequency textures shows deblurring artifacts due to the lack of accurate modeling of the seam layers. Matting based approaches may be combined with deblurring to handle such cases. The heterodyne light field camera assumes a bandlimited light field. When this assumption is not true, it leads to aliasing artifacts in the recovered light field. To recover larger angular resolution in the light field, the 2D cosine mask needs to be moved away from the sensor, which might result in diffraction.

6. Discussion

Future Directions: To capture the light field, we need to use masks that match the resolution of the sensor. It is already possible to print RGB Bayer mosaics at pixel resolution. This is ideal for the future trend of digital cameras where pixels are becoming smaller to achieve higher resolution. Such high resolution masks will support heterodyning as well as Bayer mosaic operations in a single mask. Our current masks are effectively 2D in a 4D space, but in the future one may use masks that are angle and location dependent like a hologram to achieve a complete 4D effect. We hope our work in broadband and cosine masks will also stimulate more ideas in mask functions including colored and polarized masks to estimate scene properties.

Our broadband coding can be pushed in higher dimension, for example, by coding both in time [Raskar et al. 2006] and space. The benefit of masks compared to lenses is the lack of wavelength dependent focusing and chromatic aberrations. This fact is commonly used in astronomy. Hence, masks can be ideal for hyper-spectral imaging. Shallow depth of field is a serious barrier in medical and scientific microscopy. The facility to refocus while maintaining full resolution will be a great benefit. In combination with confocal coded aperture illumination one maybe able to capture digitally refocused images in a fewer incremental steps of the focal planes.

Conclusions: We showed that two different kinds of coded masks placed inside a conventional camera will each allow us to make a different kind of computational improvement to the camera's pictures. First, if we place a fine, narrowband mask slightly above the sensor plane, then we can computationally recover the 4D light field that enters the main camera lens. The mask preserves our camera's ability to capture the focused part of the image at the full resolution of the sensor, in the same exposure used to capture the 4D light field. Second, if we place a coarse, broadband mask at the lens aperture, we can computationally refocus an out of focus image at full resolution. As this refocusing relies on deconvolution, we can correct the focusing for images that require constant or piecewise-planar focusing. These and other masks are not magical: mask reduces sensor illumination by ≈ 1 f-stop and refocusing exacerbates noise. However, high-quality masks may be less demanding than lens arrays to mount, align, or arrange in interchangeable sets and they avoid optical errors such as radial error, spherical and chromatic aberration, and coma. We believe that masks offer a promising new avenue for computational methods to substantially improve digital photography.

Acknowledgements

We thank the anonymous reviewers and several members of MERL for their suggestions. We also thank Joseph Katz, Joe Marks, John Barnwell, Katherine Yiu and Rama Chellappa for their help and support. Rebecca Witt, Aaron Stafford and Jana Matuskova posed for the photographs. We thank Keisuke Kojima, Haruhisa Okuda & Kazuhiko Sumi, Mitsubishi Electric, Japan, and Berthold K.P. Horn, MIT for helpful discussions. Jack Tumblin was supported in part by NSF under grants IIS-0535236 and CCF-0645973.

References

ACCORSI, R., GASPARINI, F., AND LANZA, R. C. 2001. Optimal coded aperture patterns for improved SNR in nuclear medicine imaging. *Nuclear Instruments and Methods in Physics Research A 474* (Dec.), 273–284.

ADELSON, T., AND WANG, J. 1992. Single lens stereo with a plenoptic camera. *IEEE Trans. Pattern Anal. Machine Intell. 14*, 99–106.

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. Graph. 23*, 3, 294–302.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization using graph cuts. *IEEE Trans. Pattern Anal. Machine Intell. 23*, 1222–1239.

CHAUDHURI, S., AND RAJAGOPALAN, A. 1999. *Depth from Defocus: A Real Aperture Imaging Approach.* Springer.

DOWSKI, E. R., AND CATHEY, W. 1995. Extended depth of field through wavefront coding. *Appl. Optics 34*, 11 (Apr.), 1859–1866.

DOWSKI, E. R., AND JOHNSON, G. E. 1999. Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems. In *SPIE Annual Meeting*.

FARID, H., AND SIMONCELLI, E. 1998. Range estimation by optical differentiation. *J. Opt. Soc. of America A 15*, 7, 1777–1786.

FERGUS, R., TORRALBA, A., AND FREEMAN, W. 2006. Random lens imaging. Tech. rep., MIT.

FESSENDEN, R. 1908. Wireless telephony. *Trans. American Institute of Electrical Engineers 27*, 553–629.

FIELD, D. 1994. What is the goal of sensory coding? *Neural Comput. 6*, 559–601.

GEORGIEV, T., ZHENG, C., NAYAR, S., CURLESS, B., SALASIN, D., AND INTWALA, C. 2006. Spatio-angular resolution tradeoffs in integral photography. In *Eurographics Symposium on Rendering*, 263–272.

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

GOTTESMAN, S. R., AND FENIMORE, E. E. 1989. New family of binary arrays for coded aperture imaging. *Appl. Optics 28*, 20 (Oct), 4344–4352.

HAEBERLI, P., 1994. A multifocus method for controlling depth of field. GraficaObscura.

HIURA, S., AND MATSUYAMA, T. 1998. Depth measurement by the multi-focus camera. In *Proc. Conf. Computer Vision and Pattern Recognition*, 953–961.

IMATEST. Image quality evaluation software. http://www.imatest.com/.

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

IVES, H. 1928. Camera for making parallax panoramagrams. *J. Opt. Soc. Amer. 17*, 435–439.

JAVIDI, B., AND OKANO, F., Eds. 2002. *Three-Dimensional Television, Video and Display Technologies*. Springer-Verlag.

JOHNSON, G. E., DOWSKI, E. R., AND CATHEY, W. T. 2000. Passive ranging through wave-front coding: Information and application. *Applied Optics 39*, 1700–1710.

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

LEVOY, M., CHEN, B., VAISH, V., HOROWITZ, M., MCDOWALL, M., AND BOLAS, M. 2004. Synthetic aperture confocal imaging. *ACM Trans. Graph. 23*, 825–834.

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

LUCY, L. 1974. An iterative technique for the rectification of observed distributions. *J. Astronomy 79*, 745–754.

MARTNEZ-CORRAL, M., JAVIDI, B., MARTNEZ-CUENCA, R., AND SAAVEDRA, G. 2004. Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays. *Applied Optics 43*, 5806–5813.

NAYAR, S., AND MITSUNAGA, T. 2000. High dynamic range imaging: spatially varying pixel exposures. In *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, 472–479.

NAYAR, S. K., BRANZOI, V., AND BOULT, T. E. 2006. Programmable imaging: Towards a flexible camera. *Int'l J. Computer Vision 70*, 1, 7–22.

NG, R., LEVOY, M., BRDIF, M., DUVAL, G., HOROWITZ, M., AND HANRAHAN, P. 2005. Light field photography with a hand-held plenoptic camera. Tech. rep., Stanford Univ.

NG, R. 2005. Fourier slice photography. *ACM Trans. Graph. 24*, 735–744.

OKANO, F., HOSHINO, H., AND YUYAMA, A. 1997. Real-time pickup method for a three-dimensional image based on integral photography. *Applied Optics 36*, 15981603.

OKANO, F., ARAI, J., HOSHINO, H., AND YUYAMA, I. 1999. Three dimensional video system based on integral photography. *Optical Engineering 38*, 1072–1077.

OPPENHEIM, A. V., SCHAFER, R. W., AND BUCK, J. R. 1999. *Discrete-Time Signal Processing*. Prentice-Hall.

RASKAR, R., AGRAWAL, A., AND TUMBLIN, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graph. 25*, 3, 795–804.

RICHARDSON, W. 1972. Bayesian-based iterative method of image restoration. *J. Opt. Soc. of Am. 62*, 1, 55–59.

SKINNER, G. K. 1988. X-Ray Imaging with Coded Masks. *Scientific American 259* (Aug.), 84.

SUN, W., AND BARBASTATHIS, G. 2005. Rainbow volume holographic imaging. *Opt. Lett. 30*, 976–978.

VAISH, V., WILBURN, B., JOSHI, N., AND LEVOY, M. 2004. Using plane + parallax for calibrating dense camera arrays. In *Proc. Conf. Computer Vision and Pattern Recognition*, 2–9.

VAN DER GRACHT, J., DOWSKI, E., TAYLOR, M., AND DEAVER, D. 1996. Broadband behavior of an optical-digital focus-invariant system. *Optics Letters 21*, 13 (July), 919–921.

VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Non-refractive modulators for coding and capturing scene appearance. Tech. Rep. UMIACS-TR-2007-21, Univ. of Maryland.

WANG, S., AND HEIDRICH, W. 2004. The design of an inexpensive very high resolution scan camera system. *Eurographics 23*, 441–450.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High performance imaging using large camera arrays. *ACM Trans. Graph. 24*, 3, 765–776.

YANG, J. C. 2000. *A Light Field Camera For Image Based Rendering*. Master's thesis, Massachussettes Institute of Technology.

ZOMET, A., AND NAYAR, S. 2006. Lensless imaging with a controllable aperture. In *Proc. Conf. Computer Vision and Pattern Recognition*, 339–346.

Appendix: Source Code

```
% Source Code for Computing 4D Light-Field from Captured 2D Photo
% Mask contains Cosines with 4 Harmonics leading to 9X9 Angular Samples

m = 2133; n=1719 % Size of Captured Image
nAngles = 9; cAngles = (nAngles+1)/2;  % Number of Angular Samples
F1Y = 237;  F1X = 191;  %Cosine Frequency in Pixels from Calibration Image
phi1 = 300;  phi2 = 150; % PhaseShift due to Mask In-Plane Transltn wrt Sensor
F12X = floor(F1X/2); F12Y = floor(F1Y/2);

%Compute Spectral Tile Centers, Peak Strengths and Phase
for i=1:nAngles ;  for j=1:nAngles
   CentY(i,j) = (m+1)/2 + (i-cAngles)*F1Y;CentX(i,j)  = (n+1)/2  + (j-cAngles)*F1X;
   Mat(i,j) = exp(sqrt(-1)*((phi1*pi/180)*(i-cAngles) + (phi2*pi/180)*(j-cAngles)));
end;  end
Mat(cAngles,cAngles) = Mat(cAngles,cAngles) * 20;

f = fftshift(fft2(imread('InputCones.png'))); %Read Photo and Perform 2D FFT

%Rearrange Tiles of 2D FFT into 4D Planes to obtain FFT of 4D Light-Field
for i = 1: nAngles; for j = 1: nAngles
        FFT_LF(:,:,i,j) =  f(CentY(i,j)-F12Y:CentY(i,j)+F12Y,...
                           CentX(i,j)-F12X:CentX(i,j)+F12X)/Mat(i,j);
end; end

LF    =    ifftn(ifftnshift(FFT_LF)); %Compute Light-Field by 4D Inverse FFT
```