High-resolution lightfield photography using two masks

Zhimin Xu, Jun Ke, and Edmund Y. Lam*

Imaging Systems Laboratory, Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

*elam@eee.hku.hk

Abstract: A major theme of computational photography is the acquisition of lightfield, which opens up new imaging capabilities, such as focusing after image capture. However, to capture the lightfield, one normally has to sacrifice significant spatial resolution as compared to normal imaging for a fixed sensor size. In this work, we present a new design for lightfield acquisition, which allows for the capture of a higher resolution lightfield by using two attenuation masks. They are positioned at the aperture stop and the optical path respectively, so that the four-dimensional (4D) lightfield spectrum is encoded and sampled by a two-dimensional (2D) camera sensor in a single snapshot. Then, during post-processing, by exploiting the coherence embedded in a lightfield, we can retrieve the desired 4D lightfield with a higher resolution using inverse imaging. The performance of our proposed method is demonstrated with simulations based on actual lightfield datasets.

OCIS codes: (110.1758) Computational imaging; (110.3010) Image reconstruction techniques; (100.3020) Image reconstruction-restoration; (100.3190) Inverse problems.

References and links

1. E. Y. Lam, "Computational photography: Advances and challenges," in *Tribute to Joseph W. Goodman*, H. J. Caulfield and H. H. Arsenault, eds., Proc. SPIE **8122**, 81220O (2011).
2. W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," Appl. Opt. **41**, 6080–6092 (2002).
3. J. Mait, R. Athale, and J. van der Gracht, "Evolutionary paths in imaging and recent trends," Opt. Express **11**, 2093–2101 (2003).
4. W.-S. Chan, E. Y. Lam, M. K. Ng, and G. Y. Mak, "Super-resolution reconstruction in a computational compound-eye imaging system," Multidim. Syst. Sign. Process **18**, 83–101 (2007).
5. T. Mirani, D. Rajan, M. P. Christensen, S. C. Douglas, and S. L. Wood, "Computational imaging systems: Joint design and end-to-end optimality," Appl. Opt. **47**, B86–B103 (2008).
6. M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of ACM SIGGRAPH* (1996), pp. 31–42.
7. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts and Company Publishers, 2004).
8. E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT Press, 1991), pp. 3–20.
9. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of ACM SIGGRAPH* (1996), pp. 43–54.
10. G. Lippmann, "Épreuves réversibles donnant la sensation du relief," J. Phys. Théor. Appl. **7**, 821–825 (1908).
11. B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *Proceedings of ACM SIGGRAPH* (2005), pp. 765–776.
12. E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 99–106 (1992).
13. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford Tech. Report CTSR (2005), pp. 1-11.
14. A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," in *Proceedings of ACM SIGGRAPH* **26**, (2007).

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

15. A. Agrawal, A. Veeraraghavan, and R. Raskar, "Reinterpretable imager: Towards variable post-capture space, angle and time resolution in photography," Comput. Graph. Forum **29**, 763–772 (2010).
16. T. Georgeiv, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoff in integral photography," in *Proceedings of Eurographics Symposium on Rendering* (2006), pp. 263–272.
17. Z. Xu and E. Y. Lam, "Light field superresolution reconstruction in computational photography," in *Signal Recovery and Synthesis*, (Optical Society of America, 2011), p. SMB3.
18. C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," in *Proceedings of ACM SIGGRAPH* **27** (2008), pp. 1–10.
19. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proceedings of IEEE International Conference on Computational Photography* (IEEE, 2009), pp. 1–8.
20. R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. (McGraw-Hill, 1999).
21. J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," in *Proceedings of ACM SIGGRAPH* **27** (2000), pp. 307–318.
22. A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offs through a Bayesian analysis of light field projections," in *Proceedings of the 10th European Conference on Computer Vision* (2008), pp. 88–101.
23. Z. Xu and E. Y. Lam, "A spatial projection analysis of light field capture," in *Frontiers in Optics*, (Optical Society of America, 2010), p. FWH2.
24. W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proceedings of IEEE/SP 14th Workshop on Statistical Signal Processing*, (IEEE, 2007), pp. 294–298.
25. W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with Toeplitz and circulant matrices," in *Visual Communications and Image Processing*, Proc. SPIE **7744**, 77440K (2010).
26. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D **60**, 259–268 (1992).
27. E. Y. Lam, X. Zhang, H. Vo, T.-C. Poon, and G. Indebetouw, "Three-dimensional microscopy and sectional image reconstruction using optical scanning holography," Appl. Opt. **48**, H113–H119 (2009).
28. X. Zhang and E. Y. Lam, "Edge-preserving sectional image reconstruction in optical scanning holography," J. Opt. Soc. Am. A **27**, 1630–1637 (2010).
29. Z. Xu and E. Y. Lam, "Image reconstruction using spectroscopic and hyperspectral information for compressive terahertz imaging," J. Opt. Soc. Am. A **27**, 1638–1646 (2010).
30. "The (new) Stanford light field archive," `http://lightfield.stanford.edu/lfs.html`.

1. Introduction

Advances in computational imaging suggest that we can capture more information than a single two-dimensional (2D) projection of a three-dimensional (3D) scene. Although the acquired picture in this manner may not be visually pleasing, via computational methods in post-processing, it can yield data that could not be obtained with the traditional methods [1–5]. In this paper, we focus on the camera design for computational photography, which allows us to capture the "lightfield". This is a term commonly used in the computer graphics literature [6], but is not a "field" in the wave optics sense [7]; instead, it is a collection of light rays in geometric optics, which takes into account not only the geometrical position of the rays but also their directions.

Generally, the radiance along all the rays in a region of 3D space is mathematically characterized by a five-dimensional (5D) plenoptic function [8], *i.e.*, three coordinates for the position and two angles for the direction. In free space, as the radiance does not change along a line unless it is occluded, such a 5D representation may be reduced to four-dimensional (4D), which is called the "lightfield" [6] or "lumigraph" [9]. With a lightfield, we can reconstruct, or render, various observations of the scene. For example, we can manipulate viewpoints and perform refocusing via ray-tracing techniques.

There are two main approaches to capturing lightfields. The first one is to sample each individual light ray directly. An early example is integral photography [10], which gathers multiple images from different perspectives by placing an array of microlenses directly before the sensor. This is optically similar to a camera array system [11]. More recently, Adelson and Wang [12], and Ng et al. [13], develop what they called plenoptic cameras. In the latter, an additional main lens is placed in front of the microlens array. Since the microlenses are located at the focal plane of this additional lens, the converging rays are separated and finally recorded by

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

the sensor behind the microlens array. A second approach is to acquire the data in the Fourier domain. Veeraraghavan et al. developed the dappled photography [14], where an attenuation mask is added to a regular camera. Its working principle will be discussed in more detail in Sections 2.1 and 2.2. After that, Agrawal et al. extend this design to the problem of capturing useful subsets of time-varying 4D lightfield in a single snapshot [15]. This "reinterpretable" imaging system adopts a design of a time-varying mask in the pupil plane and a static mask placed near the sensor, providing a variable resolution tradeoff among the spatial, angular and temporal dimensions.

Nevertheless, a common issue for different lightfield camera systems is that the spatial resolution is traded for angular information (for both angular and temporal information in [15]) because the limited sensor elements have to be allocated to all these dimensions [16, 17]. For instance, to acquire a lightfield of 144 views on a sensor of size 3072×1536, a twelvefold reduction in each spatial dimension means that the maximum resolution achievable is only 256×128. There have been attempts to overcome this tradeoff, but they come at the expense of other aspects. For example, the camera array system [11] can gain the 4D radiance information with a high resolution (*i.e.*, full sensor size of each camera) for each perspective, but the system is also known for its large size. This eventually limits its practical use. Alternatively, in a method known as programmable aperture photography [18], we need many image captures to attain the required angular resolution. This results in a long acquisition time, which is not desirable in many practical applications. In [19], Lumsdaine and Georgiev depict a new design of a plenoptic camera, called the focused plenoptic camera, where the microlens array is positioned before or behind the focal plane of the main lens. This modification samples the lightfield in a way that allows for a higher spatial resolution. However, at the same time, the angular resolution is decreased. Besides, the low angular resolution also introduces some unwanted aliasing artifacts.

In this paper, we present a camera system that collects the 4D lightfield within a single exposure. With two attenuating masks separately placed at the aperture plane and the optical path of the camera, we can encode the lightfield spectrum in the Fourier domain, and then selectively sub-sample it. We show that this economical and easily adjustable design can overcome various limitations found in other lightfield acquisition systems.

2. A lightfield camera with two masks

2.1. Lightfield mapping via mask-based multiplexing

We explain the mapping of a lightfield with mask-based multiplexing. In geometrical optics, we describe light propagation in terms of rays, which together form a lightfield [6]. We describe the light rays by their intersections with two parallel planes as shown in Fig. 1, *i.e.*, a first coordinate pair $\mathbf{u} = \{u, v\}$ (at the **u**-plane) and a second coordinate pair $\mathbf{s} = \{s, t\}$ (at the **s**-plane) [6]. The lightfield is then $\ell(u, v, s, t)$, which we abbreviate as $\ell(\mathbf{u}, \mathbf{s})$ in the rest of this paper.

Using this two-plane parametrization, we can analyze a conventional camera fitted with a mask between the **u**-plane and the **s**-plane. We depict such a camera in Fig. 2. The **u**-plane is taken to be at the aperture, while the **s**-plane at the sensor. They are separated by a distance d, while the mask is placed at a distance z in front of the sensor, where $z \leq d$. Let $m(\mathbf{u}, \mathbf{s})$ be the attenuation on a lightfield produced by the mask. The lightfield measured behind the mask is then $\ell_o(\mathbf{u}, \mathbf{s})$, given by

$$\ell_o(\mathbf{u}, \mathbf{s}) = \ell(\mathbf{u}, \mathbf{s}) m(\mathbf{u}, \mathbf{s}). \tag{1}$$

If we can capture $\ell_o(\mathbf{u}, \mathbf{s})$, we can retrieve $\ell(\mathbf{u}, \mathbf{s})$ since $m(\mathbf{u}, \mathbf{s})$ is known.

In fact, $m(\mathbf{u}, \mathbf{s})$ is completely determined by the 2D pattern $c(x, y)$ printed on the mask when the distance d is known. We denote the mask plane as the **x**-plane, with $\mathbf{x} = \{x, y\}$. With refer-

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

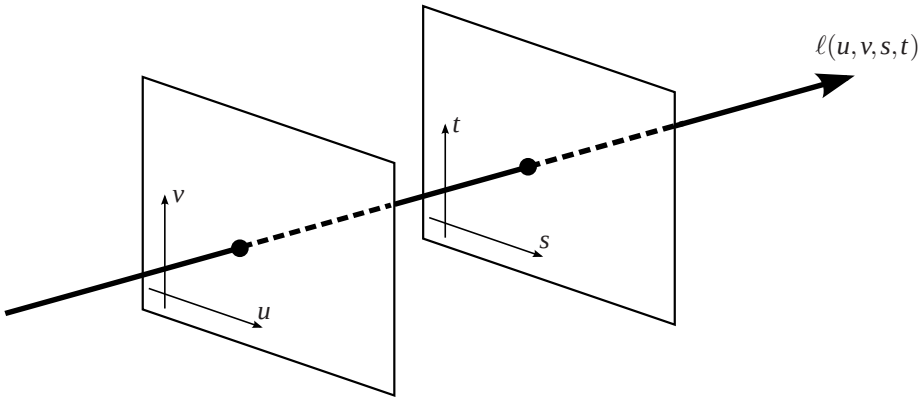

Fig. 1. The two-plane parametrization of a 4D lightfield.

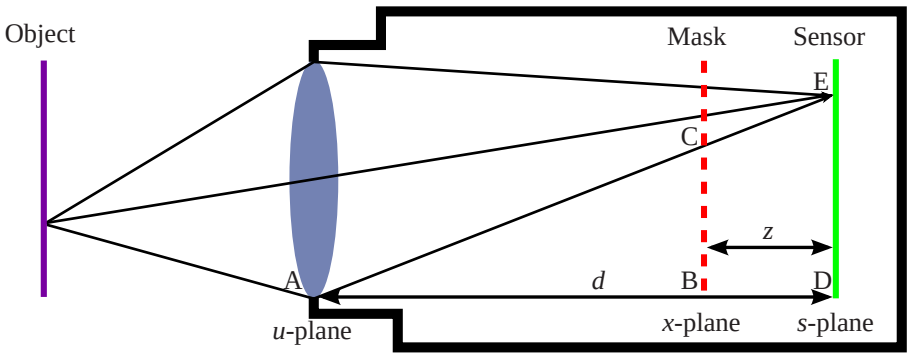

Fig. 2. Schematic diagram of a regular camera, with an attenuation mask placed inside it.

ence to Fig. 2, because $\triangle ABC$ and $\triangle ADE$ are similar triangles, we have

$$\frac{BC}{DE} = \frac{AB}{AD} \qquad \Leftrightarrow \qquad \frac{x-u}{s-u} = \frac{d-z}{d}. \tag{2}$$

Based on Eq. (2), we have $x = (1-z/d)s + (z/d)u$. But since $\mathbf{u} = \{u,v\}$ and $\mathbf{s} = \{s,t\}$,

$$\mathbf{x} = \left(1 - \frac{z}{d}\right)\mathbf{s} + \frac{z}{d}\mathbf{u}. \tag{3}$$

Thus, $m(\mathbf{u},\mathbf{s})$ can be expressed as

$$m(\mathbf{u},\mathbf{s}) = c\left[\left(1 - \frac{z}{d}\right)\mathbf{s} + \frac{z}{d}\mathbf{u}\right]. \tag{4}$$

In reality, however, we seldom directly capture the lightfield $\ell_o(\mathbf{u},\mathbf{s})$. Instead, it is instructive to consider the "lightfield-frequency" domain, which is the 4D Fourier transform applied to the lightfield in Eq. (1). Using $\mathbf{f_u}$ and $\mathbf{f_s}$ to denote the lightfield-frequency variables, we have

$$\mathscr{L}_o(\mathbf{f_u},\mathbf{f_s}) = \mathscr{L}(\mathbf{f_u},\mathbf{f_s}) * M(\mathbf{f_u},\mathbf{f_s}), \tag{5}$$

where $\mathscr{L}_o(\mathbf{f_u},\mathbf{f_s})$, $\mathscr{L}(\mathbf{f_u},\mathbf{f_s})$ and $M(\mathbf{f_u},\mathbf{f_s})$ are the respective Fourier transforms of $\ell_o(\mathbf{u},\mathbf{s})$, $\ell(\mathbf{u},\mathbf{s})$ and $m(\mathbf{u},\mathbf{s})$, and $*$ denotes the 4D convolution operation. Furthermore, we can express $M(\mathbf{f_u},\mathbf{f_s})$

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

as

$$M(\mathbf{f_u},\mathbf{f_s}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c\left[\left(1-\frac{z}{d}\right)\mathbf{s}+\frac{z}{d}\mathbf{u}\right]\exp\left[-\mathrm{j}2\pi(\mathbf{f_u}\cdot\mathbf{u}+\mathbf{f_s}\cdot\mathbf{s})\right]\,\mathrm{d}\mathbf{u}\,\mathrm{d}\mathbf{s}$$
$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} c\left[\left(1-\frac{z}{d}\right)\mathbf{s}+\frac{z}{d}\mathbf{u}\right]\exp\left(-\mathrm{j}2\pi\mathbf{f_s}\cdot\mathbf{s}\right)\,\mathrm{d}\mathbf{s}\right\}\exp\left(-\mathrm{j}2\pi\mathbf{f_u}\cdot\mathbf{u}\right)\,\mathrm{d}\mathbf{u}. \quad (6)$$

Clearly, the positioning of the mask (*i.e.*, the value of z) affects the lightfield $\ell_o(\mathbf{u},\mathbf{s})$. This effect is explained in further details as follows.

1. Generally, the mask is between the aperture and the sensor, so $0 < z < d$. According to Eq. (6), the inner integration computes the Fourier transform over the dimension of $\mathbf{s}$ with some shift and scaling, *i.e.* [20],

$$M(\mathbf{f_u},\mathbf{f_s}) = \frac{d}{d-z}\int_{-\infty}^{\infty}\left\{C\left(\frac{d}{d-z}\mathbf{f_s}\right)\exp\left[\mathrm{j}2\pi\left(\frac{z}{d-z}\mathbf{f_s}\right)\cdot\mathbf{u}\right]\right\}\exp\left(-\mathrm{j}2\pi\mathbf{f_u}\cdot\mathbf{u}\right)\,\mathrm{d}\mathbf{u}$$
$$= \frac{d}{d-z}C\left(\frac{d}{d-z}\mathbf{f_s}\right)\delta\left(\mathbf{f_u}-\frac{z}{d-z}\mathbf{f_s}\right), \quad (7)$$

   where $C(\cdot)$ represents the 2D Fourier transform of $c(\cdot)$. This means that the modulation caused by the mask in the lightfield-frequency domain happens along an inclined 2D plane, where $\mathbf{f_u}-\frac{z}{d-z}\mathbf{f_s}=0$. Its inclination angle α, if we plot $\mathbf{f_s}$ versus $\mathbf{f_u}$, is given by

$$\alpha = \arctan\frac{z}{d-z}. \quad (8)$$

2. Alternatively, the mask can be placed exactly at the aperture, where $z = d$. All the rays with the same location in the $\mathbf{u}$-plane are attenuated equally by the mask. Substitute $z = d$ into Eq. (6), then

$$M(\mathbf{f_u},\mathbf{f_s}) = C(\mathbf{f_u})\delta(\mathbf{f_s}). \quad (9)$$

   Thus, in lightfield-frequency domain, the corresponding convolution only affects the lightfield spectrum along the $\mathbf{f_u}$ axis (where $\mathbf{f_s}=0$). This observation is critical to our design, as we will explain next.

2.2. Lightfield capture and image reconstruction

The sensor at the $\mathbf{s}$-plane cannot capture the full 4D lightfield $\ell_o(\mathbf{u},\mathbf{s})$ as given in Eq. (1). Instead, all rays with the same (s,t) but different (u,v) are collected (*i.e.*, integrated together) by the same photodetector. In the lightfield-frequency domain, this means the sensor only obtains data at $\mathbf{f_u}=0$, or along the $\mathbf{f_s}$ axis.

Ref. [14] however provides a strategy to capture the 4D lightfield using a normal sensor, which we briefly review here. This will form the basis of our computational photography architecture which makes use of two masks. Assume that $c(\mathbf{x})$ is the sum of a series of cosine waves of equal amplitude; $C(\mathbf{f_x})$ is then an impulse train with even symmetry, which causes modulation along a slanted plane. Specifically, Eq. (5) suggests that $\mathscr{L}_o(\mathbf{f_u},\mathbf{f_s})$ contains replications of $\mathscr{L}(\mathbf{f_u},\mathbf{f_s})$ along a slanted plane at angle α given by Eq. (8). This is shown in Fig. 3. For ease of explanations, we depict the lightfield spectrum as one consisting of several sections along the $\mathbf{f_u}$ axis, each of which is called an angular spectral slice. By adjusting α and the distance between each consecutive replications of the lightfield spectrum along the slanted plane, we can position all the sections along the $\mathbf{f_s}$ axis. Therefore, the 2D slice of data collected by the sensor still contains all the information about the 4D lightfield.

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

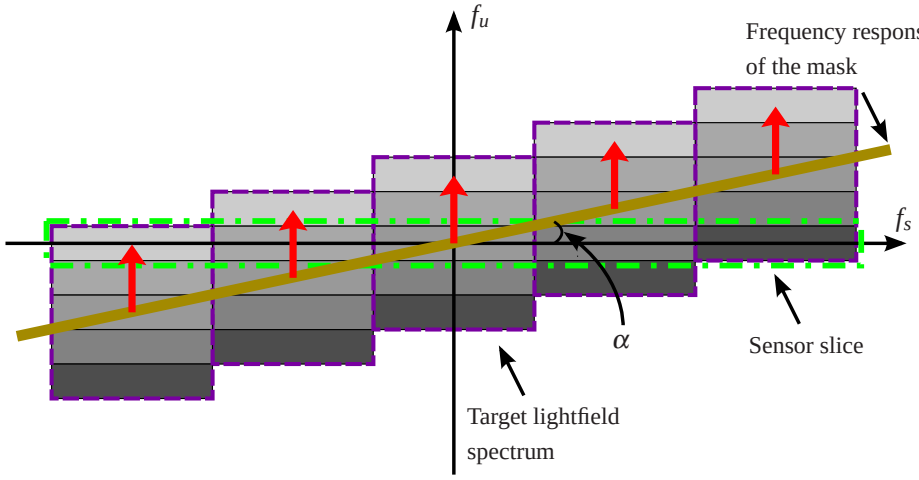

Fig. 3. The modulation in the lightfield-frequency domain.

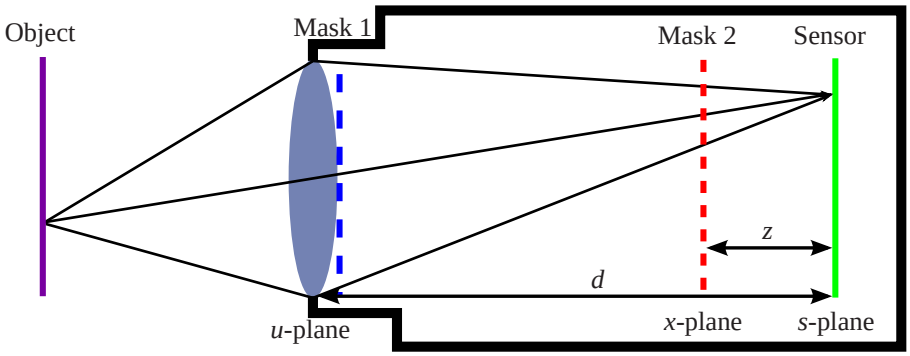

Fig. 4. Our proposed lightfield camera, with two attenuation masks respectively placed at the aperture stop and the optical path in the camera.

The tradeoff with this mode of capture is that the slice in Fig. 3 needs to be much longer than what would be needed for conventional photography; therefore, many more samples are needed to achieve the same 2D resolution for one reconstructed picture. Put another way, assume the overall number of pixels is q. Then, to resolve n different views, we only assign q/n of the pixels to sample each angular spectral slice, compared with using all q pixels for a single picture in conventional photography. This ultimately results in a loss of the spatial resolution with a scaling of $1/n$. Our design of a lightfield camera seeks to ameliorate this problem by showing that when each angular spectral slice can contain more information than merely one perspective or view, fewer replicas of the lightfield spectrum are needed. This means that effectively the sensor slice is shortened, and as a result a higher resolution lightfield can be obtained with a fixed sensor size.

2.3. Lightfield capture with a double-mask design

We propose a lightfield camera as shown in Fig. 4. We assume that the lightfield spectrum is bandlimited, *i.e.*, $\mathscr{L}(\mathbf{f_u},\mathbf{f_s}) = 0$ for $|\mathbf{f_u}| \geq B_\mathbf{u}/2$ or $|\mathbf{f_s}| \geq B_\mathbf{s}/2$. This is reasonable because the optics imposes a cutoff in the optical transfer function in the $\mathbf{f_s}$ axis. As for $\mathbf{f_u}$, Ref. [21] shows

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

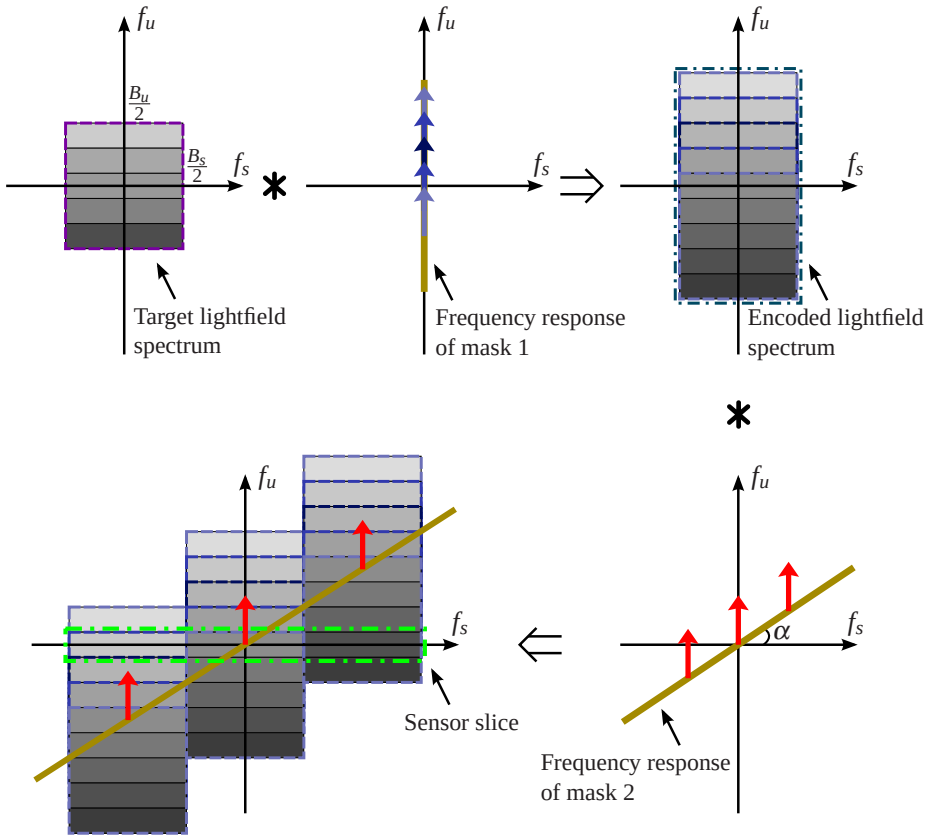

Fig. 5. The corresponding lightfield-frequency domain operations in our double-mask lightfield camera. (The asterisk pattern in the figure denotes the convolution.)

that the corresponding bandwidth is basically determined by the depth range of a scene.

We analyze the working principle of this camera by considering the operations in the lightfield-frequency domain as shown in Fig. 5. After passing through the first attenuation mask located at the aperture stop, the incoming bandwidth-limited lightfield is convolved with the mask spectrum along the $\mathbf{f_u}$ axis. If the mask frequency response is a series of impulses, the lightfield spectrum is replicated along the $\mathbf{f_u}$ axis, causing the angular spectral slices to overlay on each other. This is the lightfield spectrum encoding. Because of the second mask, the encoded lightfield spectrum is then replicated along a slanted line. By adjusting the position of the mask, we can place the desired angular spectral slices along the $\mathbf{f_s}$ axis. Thereafter, we perform the lightfield reconstruction from the 2D slice data collected by the sensor in the fashion described in Section 2.2.

The analysis in lightfield-frequency domain provides an intuitive knowledge of our design. However, for the purpose of mask design and lightfield retrieval, we need to explicitly model the acquisition process. This is expressed as

$$\begin{aligned} i(\mathbf{s}) &= \int_{-\infty}^{\infty} \ell(\mathbf{u},\mathbf{s}) m_1(\mathbf{u},\mathbf{s}) m_2(\mathbf{u},\mathbf{s})\, \mathrm{d}\mathbf{u} \\ &= \int_{-\infty}^{\infty} \ell(\mathbf{u},\mathbf{s}) c_1(\mathbf{u}) c_2\left[\left(1-\frac{z}{d}\right)\mathbf{s}+\frac{z}{d}\mathbf{u}\right]\, \mathrm{d}\mathbf{u}, \end{aligned} \tag{10}$$

where $i(\mathbf{s})$ is the 2D picture recorded by the sensor, and $m_1(\mathbf{u},\mathbf{s})$ and $m_2(\mathbf{u},\mathbf{s})$ are the respective

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

attenuation provided by the masks at the aperture stop ($c_1(\mathbf{x})$) and at the camera's optical path ($c_2(\mathbf{x})$) shown in Fig. 4. The formula for the masks are given in Eq. (4).

As indicated in Fig. 5, our design is based on a series of operations in the lightfield-frequency domain. Thus, it is rational to convert the integration of Eq. (10) into a form under the Fourier bases. After discretizing Eq. (10) and converting it into matrix form, we have

$$i = \mathbf{F}^{-1}\mathbf{M}_2\mathbf{M}_1\mathbf{F}\ell = \mathbf{F}^{-1}\mathbf{M}\mathbf{F}\ell = \mathbf{A}\ell, \tag{11}$$

where $\mathbf{F}$ and $\mathbf{F}^{-1}$ are the matrices consisting of the Fourier basis and its inverse, $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively, consist of the coefficients of the Fourier transforms of $c_1(\mathbf{x})$ and $c_2(\mathbf{x})$ and the projection matrix $\mathbf{A} = \mathbf{F}^{-1}\mathbf{M}\mathbf{F}$. Therefore, the image formation of our lightfield camera can be treated as a linear integration process in the content of geometrical optics as indicated in [22, 23]. More specifically, it is a measuring procedure in the lightfield-frequency domain through a measurement matrix $\mathbf{M} = \mathbf{M}_2\mathbf{M}_1$.

We note that the discretized lightfield ℓ is arranged into a 2D matrix of size $n \times m$, with n as the resolution in the $\mathbf{u}$ dimension and m as the resolution in the $\mathbf{s}$ dimension. Assume $\mathbf{M}_1$ and $\mathbf{M}_2$ are of size $k \times p$ and $p \times n$, respectively. Then, $\mathbf{M}$ is a $k \times n$ matrix, which means that we sample k measurements of the coefficients decomposed by n Fourier bases. The size of the final captured picture i is $k \times m$, meaning we need a sensor with km pixels. We can compare this with the design in [14], which forbids overlapping between each replicated spectrum. Consequently, the matrix $\mathbf{M}$ in their case is diagonal ($k = n$). To achieve a lightfield with the same resolution, the dappled photography system will need nm pixels. In our design, however, the measurement matrix is the product of two matrices $\mathbf{M}_2$ and $\mathbf{M}_1$. This provides us with the means to control the size of the two dimensions of $\mathbf{M}$ separately. Hence if we can achieve a measurement matrix $\mathbf{M}$ with $k < n$, fewer pixels will be used to sample the signal. In other words, we can acquire a higher spatial resolution lightfield using the same number of pixels. As discussed next, we can then realize a measurement matrix with $k < n$ in our design.

2.4. Design of the two masks

In this section, we describe the pattern design of these two attenuation masks. For clarity, only the case of 2D lightfield is carried out here, but these conclusions can be easily extended to a 4D lightfield.

The first row of Fig. 5 shows the desired frequency response of the first mask, which is actually a symmetric impulse train. The interval between each consecutive impulse is equal to the sampling interval of the lightfield spectrum along the f_u axis. Thus, the corresponding physical mask pattern is the sum of multiple cosine waves with a given amplitude, which in turn determines $\mathbf{M}_1$ completely. Specifically, assume the first mask has the following the frequency response, *i.e.*,

$$C_1(f_u) = \sum_{i=-(n-1)}^{n-1} a_i \delta(f_u - i\Delta f_u), \tag{12}$$

where n is the expected resolution along the f_u axis, a_i is the amplitude of the i-th impulse and Δf_u is the sampling interval of the lightfield spectrum along the f_u axis, which is equal to B_u/n with B_u as the bandwidth in the f_u dimension. Because the first mask is convolved with the lightfield spectrum in the lightfield-frequency domain, by converting the convolution into a

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

matrix multiplication, we have $\mathbf{M}_1$ equal to

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ a_0 & a_1 & \cdots & a_{n-1} \\ a_{-1} & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{-(n-1)} & a_{-(n-2)} & \cdots & a_0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{p\times n}. \tag{13}$$

Thus, we have constructed a matrix $\mathbf{M}_1$ with a Toeplitz-structured block inside it. Because of the second mask, only k rows of $\mathbf{M}_1$ are selected, so the other ones are marked with ellipses for simplicity. Note that we can recover the original sparse signal with a high probability from the limited observations measured by a well-designed Toeplitz-structured matrix [24, 25]. To satisfy the conditions for such a design, several methods have been recommended. As suggested in [24], we generate $\mathbf{M}_1$ with entries a_i, $i = 0,\ldots,n-1$ drawn independently from a Gaussian distribution with zero mean. Since a_i is symmetric about a_0, the values of a_i, for $i = -(n-1),\ldots,-1$, are then known. Eventually, we obtain the physical pattern of the first mask based on its frequency response in Eq. (12).

As for the second mask placed at the optical path, the second row in Fig. 5 has shown a heuristic example. That is, the frequency response of the second mask is a series of even-symmetric impulses with equal amplitudes. The number of impulses depends on how many measurements are required for reconstruction. To avoid aliasing between the adjacent spectrum replicas, the interval of this impulse train is equal to the lightfield bandwidth in the f_s dimension, *i.e.*, B_s. Specifically, the frequency response of the second mask is given by

$$C_2(f_x) = \sum_{i=-(k-1)/2}^{(k-1)/2} \delta(f_x - iB_s), \tag{14}$$

where k is the number of the measurements. Thus the corresponding mask pattern $c_2(x)$ can be obtained by computing the inverse Fourier transform of Eq. (14). That is the sum of a series of cosine waves. As regards its matrix form $\mathbf{M}_2$, it depends on the requirement of which measurements will be collected for further reconstruction. So we could realize the function of $\mathbf{M}_2$ by selecting the k rows of $\mathbf{M}_1$ according to the specific design.

2.5. Lightfield reconstruction

After constructing the two masks, we can then establish the projection matrix $\mathbf{A}$ in Eq. (11). Next we consider the reconstruction of the target 4D lightfield based on the captured 2D picture i and the projection matrix $\mathbf{A}$. We adopt two different approaches to solve such an inverse problem. The first is to find its least-norm solution, *i.e.*,

$$\ell^\star = \mathbf{A}^\dagger i = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} i, \tag{15}$$

where $\mathbf{A}^\dagger$ denotes the pseudoinverse of $\mathbf{A}$. While this is simple and fast, due to the lack of prior information about the lightfield, the solution is often not sufficiently accurate. To improve the reconstruction accuracy, we make use of the prior knowledge about a lightfield and impose regularization in the reconstruction process. One possibility is a sparse regularizer, which is a 2D total variation (TV) penalty on the **u** dimension of a lightfield to reflect the inherent correlations. We also use the 2D TV norm regularization on the **s** dimension of a lightfield to

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

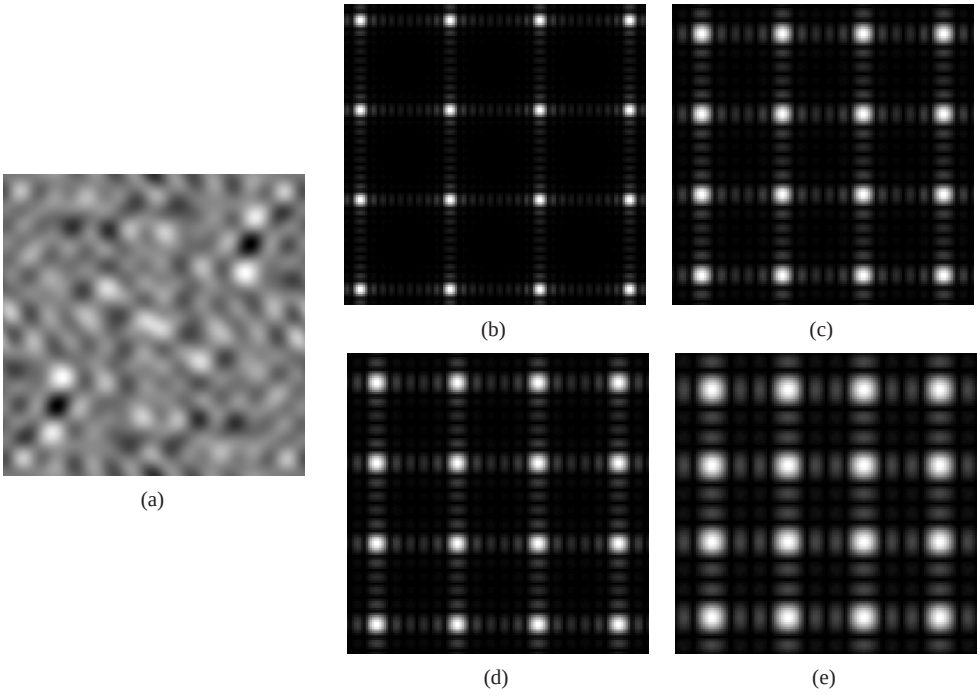

Fig. 6. (a) The pattern of the first mask; (b)-(e) are the pattern parts of the second mask, respectively in cases of using full, 64%, 36% and 16% sensor size.

preserve the edges and suppress the noise [26–28]. Thus, we reconstruct the lightfield by the optimization given by

$$\ell^{\star} = \arg\min_{\ell} \left\{ \frac{1}{2} \|\mathbf{A}\ell - i\|_2^2 + \lambda \sum_{\mathbf{u}} \|i_{\mathbf{u}}\|_{\mathrm{TV}} + \mu \sum_{\mathbf{s}} \|i_{\mathbf{s}}\|_{\mathrm{TV}} \right\}, \tag{16}$$

where λ and μ are the regularization parameters, $i_{\mathbf{u}}$ is a 2D image corresponding to the lightfield $\ell(\mathbf{u},\mathbf{s})$ at a fixed point $\mathbf{u}$, and $i_{\mathbf{s}}$ refers to the lightfield $\ell(\mathbf{u},\mathbf{s})$ at a fixed point $\mathbf{s}$.

This optimization can be solved via a nonlinear conjugate gradient method combined with backtracking line search, as adopted in [29].

3. Experimental results

To verify the ability to achieve a high-resolution lightfield, a direct way is to use a fixed number of pixels to retrieve a lightfield with a higher spatial resolution. Alternatively, one can aim at obtaining a lightfield of a fixed resolution with fewer pixels, which is the approach we take here. The following experiments are based on actual lightfield datasets from the Stanford lightfield archive [30]. For computational considerations, we choose 100 views on a 10×10 grid and resize the image to 128×256 pixels.

Figure 6 shows the corresponding mask patterns that are adopted in the experiments. According to Eq. (12) in Section 2.4, the required frequency response of the mask at the aperture stop is an even-symmetric impulse train of size 19×19 (where $n = 10 \times 10$ in our experiments). The corresponding amplitude of these impulses are drawn independently from a Gaussian distribution with zero mean. The physical pattern shown in Fig. 6(a) is the one we use here. Since

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

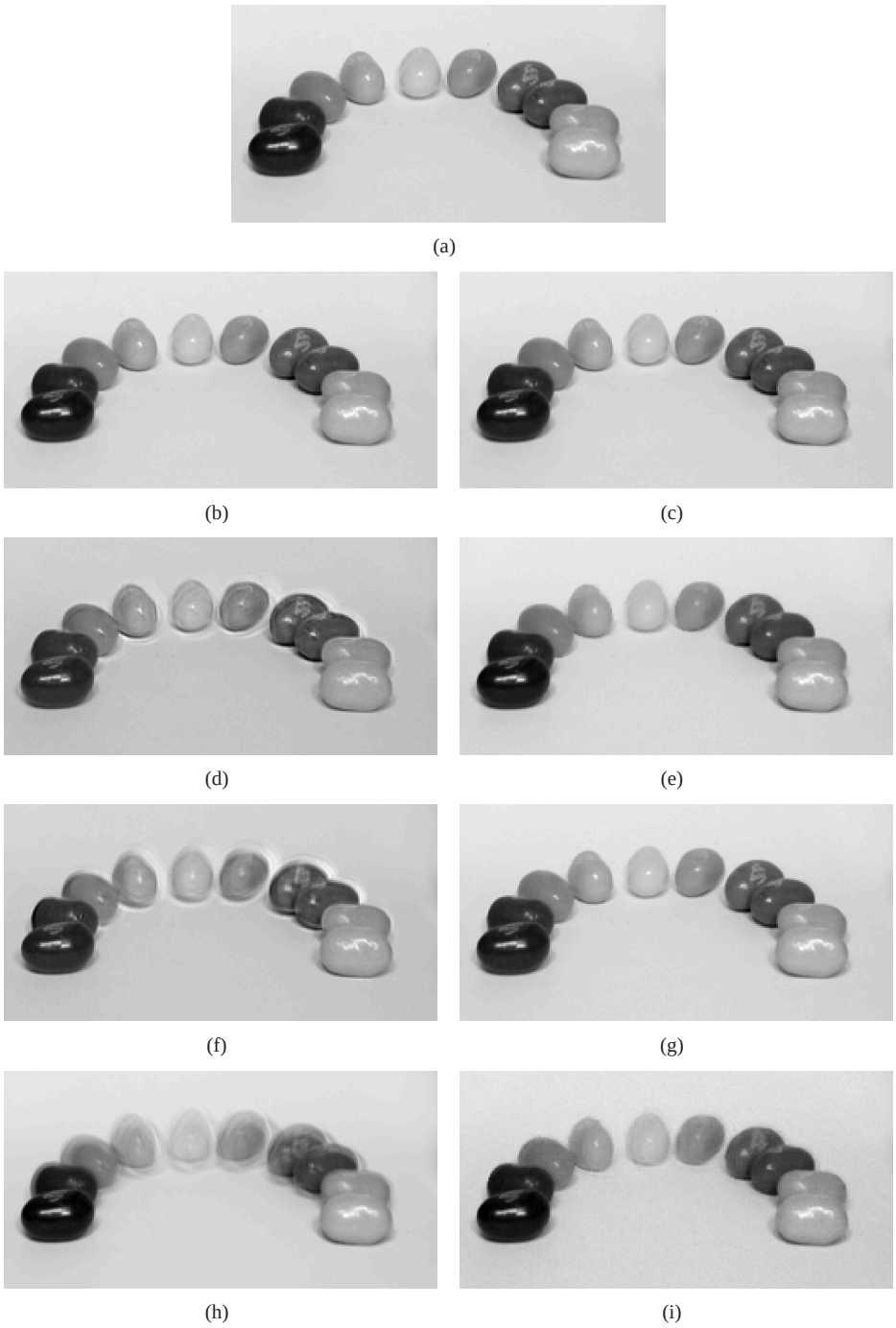

Fig. 8. The reconstructed images at one selected viewpoint by using the least-norm method (left column) and the proposed method in Section 2.5 (right column): (a) ground truth, (b) and (c) full size, (d) and (e) 64% sensor size, (f) and (g) 36% sensor size, (h) and (i) 16% sensor size.

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

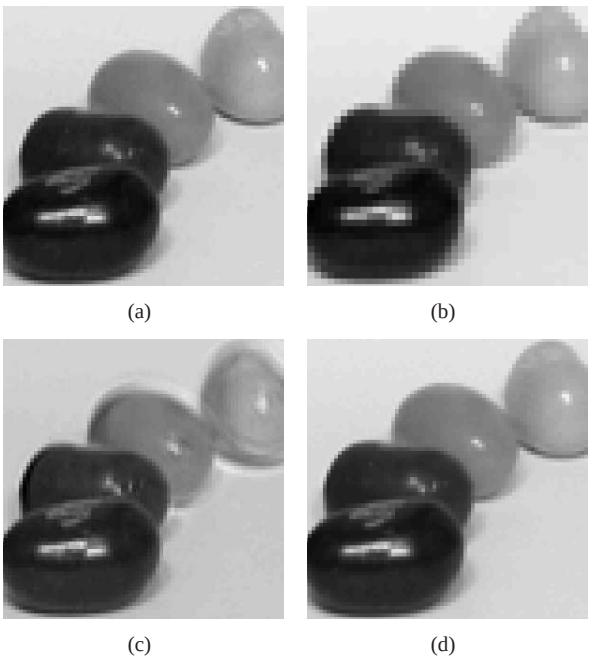

Fig. 9. Reconstructions when using 36% sensor size: (a) ground truth; (b) the best quality that can be achieved by using the traditional lightfield cameras; (c) our reconstruction with the least-norm method; (d) our reconstruction with the proposed iterative method.

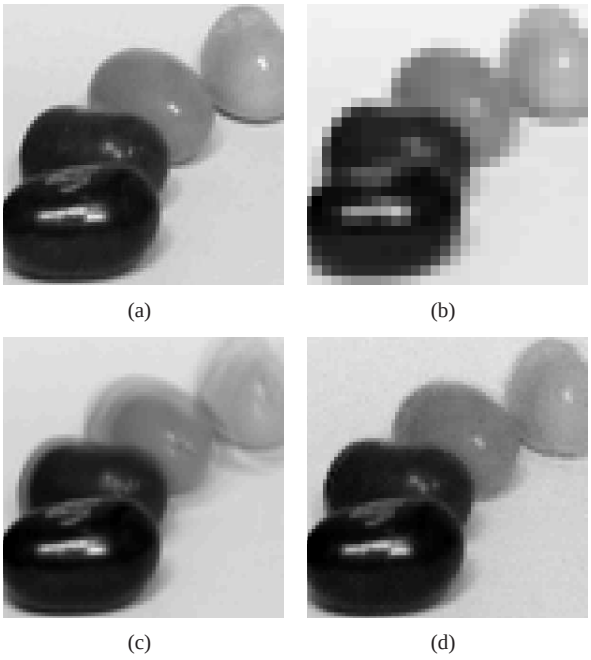

Fig. 10. Reconstructions when using 16% sensor size: (a) ground truth; (b) the best quality that can be achieved by using the traditional lightfield cameras; (c) our reconstruction with the least-norm method; (d) our reconstruction with the proposed iterative method.

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012

the mask at the aperture stop is responsible for encoding the lightfield spectrum, we keep this mask unchanged during our experiments.

For the mask placed at the optical path, its frequency response depends on the specific requirement of the measurement number. For example, for the case of using full sensor size (*i.e.*, 1280×2560), it is a 10×10 impulse train with equal amplitude based on Eq. (14). Similarly, we have 8×8 for the case of using 64% sensor size (*i.e.*, 1024×2048), 6×6 for the case of using 36% sensor size (*i.e.*, 768×1536) and 4×4 for the case of using 16% sensor size (*i.e.*, 512×1024). Figure 6(b) - 6(e) show the corresponding pattern parts in these different cases. Notice that since we cannot have negative values in the mask, we need to increase the DC component so that the values in these masks are nonnegative.

Next, we show the performance of our camera when using different sensor sizes. That is, we aim to retrieve the original lightfield of the same spatial resolution from the captured signals by using different physical sensor sizes. Figure 7 shows the captured pictures by using the proposed lightfield camera with different number of pixels. Figure 8 shows the corresponding reconstruction images at one selected viewpoint. For the sake of comparison, we use both the least-norm method in Eq. (15) and our proposed algorithm in Eq. (16) for lightfield reconstruction. In the case of using full sensor, both methods can yield perfect reconstructions as given in the ground truth. With a mild reduction in sensor size, the recoveries can still provide us good details comparable with the ground truth, such as the ones shown in the case of using 64% sensor size. With further reduction, however, the reconstruction becomes difficult, although the reconstructed images are still satisfactory with 36% and 16% pixels. Furthermore, in comparison with the reconstructions by using the least-norm method (the left column in Fig. 8), we can see that our method can preserve more details and provide better artifact control (*e.g.*, the ringing artifacts around the beans). Nevertheless, we also observe that with significant sensor size reduction, some of the details in the images are lost and the images are blurry.

Finally, we show that a higher resolution lightfield can be acquired with our proposed system than that with the conventional lightfield cameras when using the same sensor size. Figure 9 shows the case of using 36% sensor size (*i.e.*, 768×1536). If we use the conventional lightfield cameras such as the ones in [13, 14], the maximum spatial resolution that can be achieved will be 76×153. From the results shown in Fig. 9, we can see that with our proposed camera the lightfield can be recovered at a higher spatial resolution. Such a resolution enhancement effect becomes more prominent in the case of using 16% sensor size (*i.e.*, 512×1024). In this case, the best quality that can be achieved with the conventional method is 51×102. But by adopting the proposed camera, we can still reconstruct many details of the scene from the captured data. See Fig. 10 for details.

4. Conclusions

We show a system that can capture a 4D lightfield with two attenuation masks. Taking advantage of the correlations inherent in the lightfield, we develop a post-processing algorithm to reconstruct the lightfield from the captured 2D data from the sensor. The experimental results show that fewer pixels are needed to achieve the same resolution as what one can achieve with a conventional lightfield camera.

Acknowledgments

This work was supported in part by the University Research Committee of the University of Hong Kong under Project 10208648.

#164385 - $15.00 USD Received 8 Mar 2012; revised 21 Apr 2012; accepted 22 Apr 2012; published 26 Apr 2012