

Near-Invariant Blur for Depth and 2D Motion via Time-Varying Light Field Analysis

YOSUKE BANDO

TOSHIBA Corporation and MIT Media Lab

HENRY HOLTZMAN and RAMESH RASKAR

MIT Media Lab

Recently, several camera designs have been proposed for either making defocus blur invariant to scene depth or making motion blur invariant to object motion. The benefit of such invariant capture is that no depth or motion estimation is required to remove the resultant spatially uniform blur. So far, the techniques have been studied separately for defocus and motion blur, and object motion has been assumed to be 1D (e.g., horizontal). This paper explores a more general capture method that makes both defocus blur and motion blur nearly invariant to scene depth and in-plane 2D object motion. We formulate the problem as capturing a time-varying light field through a time-varying light field modulator at the lens aperture, and perform 5D (4D light field + 1D time) analysis of all the existing computational cameras for defocus/motion-only deblurring and their hybrids. This leads to a surprising conclusion that focus sweep, previously known as a depth-invariant capture method that moves the plane of focus through a range of scene depth during exposure, is near-optimal both in terms of depth and 2D motion-invariance and in terms of high frequency preservation for certain combinations of depth and motion ranges. Using our prototype camera, we demonstrate joint defocus and motion deblurring for moving scenes with depth variation.

Categories and Subject Descriptors: I.4.3 [Image Processing and Computer Vision]: Enhancement—*Sharpening and deblurring*

General Terms: Image deblurring

Additional Key Words and Phrases: Extended depth of field, defocus and motion deblurring, depth and motion-invariant photography

ACM Reference Format:

Bando, Y., Holtzman, H., and Raskar, R. YYYY. Near-invariant blur for depth and 2D motion via time-varying light field analysis. ACM Trans. Graph. VV, N, Article XXX (Month YYYY), 15 pages. DOI = 10.1145/XXXXXXXX.YYYYYY
<http://doi.acm.org/10.1145/XXXXXXXX.YYYYYY>

Authors' addresses: {bandy, holtzman, raskar}@media.mit.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 0730-0301/YYYY/12-ARTXXX \$10.00

DOI 10.1145/XXXXXXXX.YYYYYY

<http://doi.acm.org/10.1145/XXXXXXXX.YYYYYY>

1. INTRODUCTION

A conventional camera is subject to a trade-off between noise and blur. A larger aperture and longer exposure time gather more light and reduce noise, but they also result in more defocus blur and more motion blur for moving scenes with depth variation.

Removing blur from images typically involves three steps as shown in Fig. 1(a): 1) capturing an image; 2) identifying defocus/motion blur point-spread function (PSF) locally; and 3) applying (spatially-varying) deconvolution with that PSF. Both of the latter two steps are challenging as PSF identification amounts to estimating a scene depth map (for defocus blur) and/or object motion (for motion blur) from an image, and deconvolution is an ill-posed problem that tries to recover reduced/lost high frequency components of an image. Although there are a number of computational imaging methods that facilitate PSF identification while preserving image high frequencies [Agrawal and Xu 2009; Veeraraghavan et al. 2007; Levin et al. 2007; Zhou and Nayar 2009; Levin et al. 2009], defocus deblurring and motion deblurring have been dealt with separately.

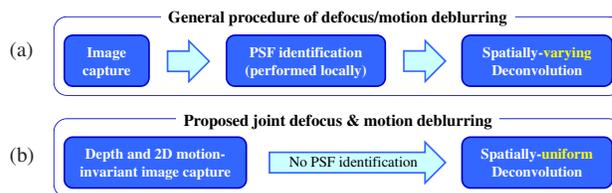


Fig. 1. Steps of defocus and motion deblurring.

To address more general cases in which objects are moving at different depths, this paper addresses the problem of joint defocus and motion deblurring. Specifically, we are inspired by a line of computational camera designs for *either* making defocus blur invariant to scene depth [Häusler 1972; Dowski and Cathey 1995; Nagahara et al. 2008; Cossairt et al. 2010] or making motion blur invariant to 1D (e.g., horizontal) object motion [Levin et al. 2008], and we seek a camera design that makes *both* defocus blur and motion blur (nearly) invariant to scene depth and in-plane 2D object motion. The benefit of such invariant image capture is that it results in spatially uniform blur, which can be removed by deconvolution with a single, *a priori* known PSF without estimating depth or motion, as shown in Fig. 1(b). This will completely eliminate the PSF identification step, which would otherwise be extremely difficult because defocus and motion blur would be combined in a spatially varying manner.

In seeking a method of depth and 2D motion-invariant capture, we formulate the problem as capturing a time-varying light field

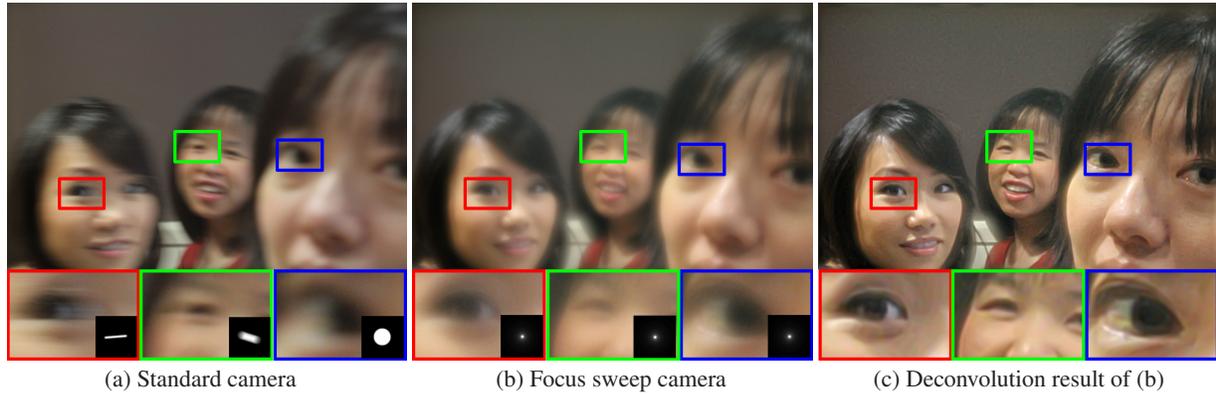


Fig. 2. (a) Standard camera image of moving people at different depths, exhibiting both defocus blur and motion blur. The insets in the magnified views below show corresponding point-spread functions (PSF). (b) Image captured while continuously varying the focused depth during exposure. The entire image exhibits nearly identical blur irrespective of scene depth and motion. (c) Deconvolution result of (b) with a single, a priori known PSF.

[Levoy and Hanrahan 1996] through a time-varying light field modulator kernel at the lens aperture, and perform 5D (4D light field + 1D time) analysis to jointly study all the existing computational cameras for defocus-only and motion-only deblurring and their hybrids. We introduce a measure of depth/motion-invariance as well as that of high frequency preservation to evaluate performance of kernels, and we found one that is fairly close to optimal in both measures for certain combinations of depth and motion ranges. Surprisingly, this corresponds to one of existing depth-invariant capture methods called *focus sweep* [Häusler 1972; Nagahara et al. 2008], which moves the plane of focus through a range of scene depth during exposure, so that every object within the depth range gets focused at a certain moment, thereby producing nearly depth-invariant defocus blur. To the best of our knowledge, the focus sweep method has been studied solely in terms of depth of field extension and manipulation, and this paper is the first to show both theoretically and experimentally that it also makes motion blur nearly invariant to 2D linear object motion up to some speed limit. We show deblurring results for challenging scenes in which objects are moving differently at different depths as shown in Fig. 2, and demonstrate joint defocus and motion deblurring using our prototype camera shown in Fig. 3(a).

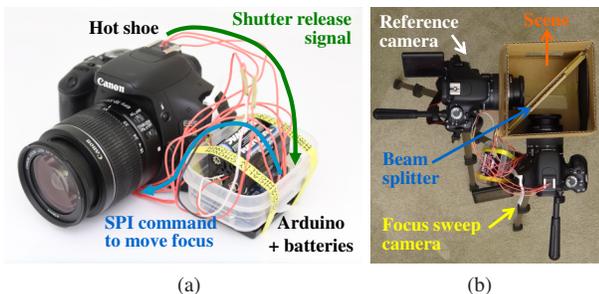


Fig. 3. Our time-varying light field analysis leads to a conclusion that focus sweep is near-optimal both in terms of depth and 2D motion-invariance and in terms of high frequency preservation. We fabricated a prototype focus sweep camera (a) to demonstrate joint defocus/motion deblurring. Experimental setup is shown in (b).

1.1 Contributions

- We perform 5D analysis of a time-varying light field in relation to combined defocus/motion blur in order to provide a framework for evaluating performance of joint defocus and motion deblurring. Specifically, we introduce performance measures of depth/motion-invariance and high frequency preservation.
- We find that the focus sweep method, which has been known as a depth-invariant capture method, also works as near 2D motion-invariant capture, and show that it achieves near-optimal performance measures both in terms of depth/motion-invariance and high frequency preservation for certain combinations of depth and motion ranges.
- We demonstrate joint defocus/motion deblurring for moving scenes with depth variation using a prototype camera.

1.2 Limitations

Limitations of our theoretical analysis include:

- Our depth/motion-invariance measure in the frequency domain provides a necessary but not sufficient condition. That is, for any method to be depth/motion-invariant, a high score for this measure needs to be achieved, but we need to confirm the invariance in the spatial domain.
- Similar to previous analyses [Levin et al. 2008; Levin et al. 2009], we assume ideal conditions for each method (such as infinite exposure assumption), and our analysis does not handle PSF *variance* coming from violation of such assumptions.

Limitations of the focus sweep method for joint defocus and motion deblurring include:

- Object speed and depth ranges need to be *a priori* bounded. Objects moving too fast and/or at the outside of the depth range will break depth/motion-invariance.
- The object motion and depth ranges cannot be controlled separately. Both are determined through a single parameter of focus sweep speed. In other words, near-optimality is achieved only for certain combinations of depth and motion ranges.
- When used solely as depth-invariant capture for static scenes, the method remains advantageous over stopping down the aperture. However, when used solely as motion-invariant capture for

scenes with little depth variation, it has no advantage over short exposure capture. When used as depth and motion-invariant capture, it is beneficial.

- Our model assumes uniform linear object motion within a small depth slab. Modest violation is tolerated in practice, but significant acceleration, rotation, and motion in the depth direction will break motion-invariance.
- The method does not handle camera shake, whose trajectory is typically too complex to be modeled as linear motion.
- Scenes need to be Lambertian and have locally constant depth. Specularity, saturation, and occlusions cause deconvolution artifacts.

2. RELATED WORK

We mainly focus on previous computational camera methods for deblurring.

Defocus deblurring (extended depth of field): Defocus blur PSFs can be made broadband in the frequency domain by introducing a patterned mask in the aperture [Levin et al. 2007; Veeraghavan et al. 2007; Zhou and Nayar 2009], or by using a lens made up of a collection of sub-lenses having different focal lengths [Levin et al. 2009]. These methods facilitate but require PSF identification. If multiple shots are available, depth of field can be extended by fusing differently focused images [Burt and Kolczynski 1993; Agarwala et al. 2004; Kubota and Aizawa 2005; Hasinoff et al. 2009a], which also requires depth estimation.

Depth-invariant capture methods eliminate PSF identification by using a cubic phase plate [Dowski and Cathey 1995], by placing an annular diffuser in the aperture [Cossairt et al. 2010], by maximizing chromatic aberration of the lens [Cossairt and Nayar 2010], or by focus sweep [Häusler 1972; Nagahara et al. 2008]. As the phase plate, diffuser, and aberrated lens are static, for moving scenes, images will exhibit motion blur in addition to the depth-invariant PSF. In contrast, having moving parts during exposure, a focus sweep camera produces a depth-invariant PSF that also remains nearly invariant to 2D object motion, which, to the best of our knowledge, this paper shows for the first time.

Motion deblurring: Motion blur PSFs can be made broadband in the frequency domain by preventing incoming light from being integrated on the sensor at several time instances during exposure [Raskar et al. 2006; Agrawal and Xu 2009; Tai et al. 2010; Ding et al. 2010]. These methods facilitate but require PSF identification. If multiple shots are available, motion deblurring benefits from differently exposed/blurred images [Rav-Acha and Peleg 2005; Yuan et al. 2007; Agrawal et al. 2009; Zhuo et al. 2010], or can be replaced by a problem of fusing noisy but unblurred short exposure images [Zhang et al. 2010]. These approaches also require motion estimation. For motion blur due to camera shake, which is out of the scope of this paper, several methods attached additional cameras [Ben-Ezra and Nayar 2004; Tai et al. 2008] or inertial measurement sensors [Joshi et al. 2010] to facilitate estimation of camera shake motion.

Levin et al. [2008] showed that 1D motion-invariance can be achieved by moving the sensor with a constant acceleration during exposure in a target object motion direction (e.g., horizontal). That is, motion blur PSF becomes nearly identical for objects moving horizontally at any speed up to some upper limit. Two methods sought to extend it to arbitrary in-plane 2D motion directions in terms of high frequency preservation, but they remained requiring PSF identification [Cho et al. 2010; Bando et al. 2011]. This paper shows that we can eliminate PSF identification for 2D object motion (up to some speed limit) by using focus sweep.

Theoretical analyses: This paper builds upon the previous analyses of computational cameras in relation to light field capture and rendering [Ng 2005; Levin et al. 2008; Levin and Durand 2010], depth of field extension [Levin et al. 2009; Hasinoff et al. 2009a; Baek 2010] and motion deblurring [Levin et al. 2008; Agrawal and Raskar 2009; Cho et al. 2010; Bando et al. 2011].

3. ANALYSIS

This section analyzes defocus and motion blur in a unified manner, so as to provide a framework for evaluating computational cameras for joint defocus and motion deblurring.

We begin by modeling combined defocus and motion blur as being introduced by capturing a time-varying light field through a time-varying light field modulator kernel at the aperture, and we show that the Fourier transform of a kernel characterizes the optical transfer function (OTF) for a set of defocus/motion blur PSFs produced by the kernel (Sec. 3.1). We then introduce performance measures of a kernel in terms of 1) depth and 2D motion-invariance and 2) high frequency preservation, and derive their upper bounds (Sec. 3.2). After that we move on to explaining a general form of kernels and how it relates to reduced kernels for either defocus-only or motion-only deblurring (Sec. 3.3). The performance measures and kernel form analysis allow us to theoretically evaluate existing computational cameras and their combinations (Sec. 3.4), which leads to our discovery of near 2D motion-invariance of the focus sweep method as will be described in the next section.

3.1 Time-Varying Light Field and Defocus/Motion Blur

Here we show how a time-varying light field translates to combined defocus and motion blur.

We denote a time-varying light field as $l(\mathbf{x}, \mathbf{u}, t)$, where $\mathbf{x} = (x, y)$ denotes locations, $\mathbf{u} = (u, v)$ viewpoint directions, and t time. As shown in Fig. 4, uv -plane is taken at the aperture, and xy -plane is at the sensor with distance d_0 from the aperture. A pinhole image $I(\mathbf{x})$ seen from the center viewpoint $\mathbf{u} = (0, 0)$ at $t = 0$ is given as:

$$I(\mathbf{x}) = l(\mathbf{x}, \mathbf{0}, 0), \quad (1)$$

which represents an ideal sharp image with no defocus or motion blur.

We consider a scene consisting of a Lambertian object moving with velocity $\mathbf{m} = (m_x, m_y)$ at depth d . We take depth parameters inside the camera as a distance from the lens aperture to the plane conjugate to the corresponding scene depth as in Fig. 4. Then we can express the time-varying light field generated by that object in

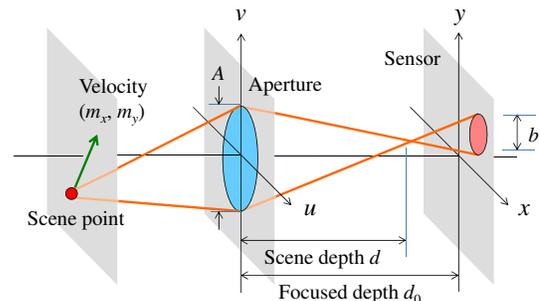


Fig. 4. Light field parameterization $xyuv$ and a moving scene point. Scene depth d is taken as a distance from the aperture towards the sensor.

terms of $I(\mathbf{x})$ as:

$$l(\mathbf{x}, \mathbf{u}, t) = I(\mathbf{x} - s\mathbf{u} - \mathbf{m}t). \quad (2)$$

where $s = (d - d_0)/d$ encodes the object depth [Levin et al. 2009] so that it works as a scaling to u and v in a similar manner as m_x and m_y are to t . The shear $-s\mathbf{u}$ comes from the similar triangles in Fig. 4 (i.e., $A : b = 1 : s$), and $-\mathbf{m}t$ is because the object motion translates the image.

When we capture a time-varying light field through a spatio-temporally modulated aperture, by extending the static light field case of [Ng 2005; Levin et al. 2009], the captured image $B(\mathbf{x})$ can be modeled as 5D convolution by a kernel $k(\mathbf{x}, \mathbf{u}, t)$:

$$B(\mathbf{x}_0) = \iiint k(\mathbf{x}_0 - \mathbf{x}, -\mathbf{u}, -t) \cdot l(\mathbf{x}, \mathbf{u}, t) d\mathbf{x}d\mathbf{u}dt, \quad (3)$$

where the integrals are taken over $(-\infty, +\infty)$ for each of the 5D parameters $xywvt$. Substituting Eq. (2) with change of variable as $\mathbf{x}' = \mathbf{x} - s\mathbf{u} - \mathbf{m}t$, we have $B(\mathbf{x}_0) = \int \phi_{s,\mathbf{m}}(\mathbf{x}_0 - \mathbf{x}') I(\mathbf{x}') d\mathbf{x}'$, which is 2D convolution of the sharp image I and a PSF $\phi_{s,\mathbf{m}}$:

$$\phi_{s,\mathbf{m}}(\mathbf{x}_0) = \iint k(\mathbf{x}_0 + s\mathbf{u} + \mathbf{m}t, \mathbf{u}, t) d\mathbf{u}dt, \quad (4)$$

where we have inverted the signs of \mathbf{u} and t , which does not affect the integral. Note that the PSF $\phi_{s,\mathbf{m}}$ includes both defocus blur due to object depth s and motion blur due to object motion \mathbf{m} .

To see the optical transfer function (OTF) introduced by the kernel k , we take the 2D Fourier transform of $\phi_{s,\mathbf{m}}(\mathbf{x})$, denoted using a hat symbol as $\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_\mathbf{x})$, where $\mathbf{f}_\mathbf{x} = (f_x, f_y)$ represents frequency in the x and y directions. By applying change of variable as $\mathbf{x} = \mathbf{x}_0 + s\mathbf{u} + \mathbf{m}t$ to Eq. (4), we can see that

$$\begin{aligned} \hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_\mathbf{x}) &= \iiint k(\mathbf{x}, \mathbf{u}, t) \cdot e^{-2\pi i \mathbf{f}_\mathbf{x} \cdot (\mathbf{x} - s\mathbf{u} - \mathbf{m}t)} d\mathbf{x}d\mathbf{u}dt \\ &= \hat{k}(\mathbf{f}_\mathbf{x}, -s\mathbf{f}_\mathbf{x}, -\mathbf{m} \cdot \mathbf{f}_\mathbf{x}). \end{aligned} \quad (5)$$

This means that the OTF is a 2D slice of the 5D Fourier transform $\hat{k}(\mathbf{f}_\mathbf{x}, \mathbf{f}_\mathbf{u}, f_t)$ of the kernel k (which is an instance of the Fourier projection-slice theorem [Bracewell 1965]).

3.2 PSF Invariance and High Frequency Preservation

The PSF $\phi_{s,\mathbf{m}}$ generally depends on object depth s and motion \mathbf{m} as apparent from Eq. (4), and it also attenuates image high frequencies. In this paper we are more interested in invariant capture, but pursuing PSF invariance alone could degrade deblurring performance, as there is a trade-off between PSF invariance and high frequency preservation [Baek 2010] (e.g., narrow-aperture short-exposure capture is a trivial solution to minimizing PSF variance by sacrificing high frequency preservation). Hence, we would like to choose kernel k that minimizes PSF variance while maximizing high frequency preservation for a range of depth and motion. Here we use constants S and M to denote depth and motion (speed) ranges as $|s| \leq S/2$ and $|\mathbf{m}| \leq M/2$. Given a raw depth range $d \in [d_{min}, d_{max}]$, we can always choose d_0 so that $|s| \leq S/2$, where $S = \frac{2(d_{max} - d_{min})}{d_{max} + d_{min}}$ [Levin et al. 2009].

In what follows, we first introduce a measure of high frequency preservation and then that of PSF invariance.

Measure of high frequency preservation: Following the previous analyses [Levin et al. 2009; Hasinoff et al. 2009a], we use the

worst-case squared MTF (modulation transfer function, the magnitude of OTF) as a measure of high frequency preservation.

$$\min_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_\mathbf{x})|^2 \quad \text{subject to } |s| \leq S/2, |\mathbf{m}| \leq M/2. \quad (6)$$

We would like to *maximize* this measure because all possible PSFs corresponding to object depth and motion within the range will have no less performance than this. On the other hand, given aperture diameter A and exposure time T , we can show that this measure is bounded as follows:

$$\min_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_\mathbf{x})|^2 \leq \frac{2A^3T}{3SM|\mathbf{f}_\mathbf{x}|^2}, \quad (7)$$

which represents the best performance we can obtain.

We prove this bound in Eq. (7) below. From Eq. (5), the equations of the OTF slice plane are given as:

$$\begin{cases} f_u = -sf_x \\ f_v = -sf_y \\ f_t = -m_x f_x - m_y f_y \end{cases}. \quad (8)$$

Thus for fixed $\mathbf{f}_\mathbf{x} = (f_x, f_y)$, this slice corresponds to a point in the 3D $f_u f_v f_t$ volume as shown as a red dot in Fig. 5. Because scene depth and object speed are bounded as $|s| \leq S/2$ and $|\mathbf{m}| \leq M/2$, the area in which this point can be located is confined to the blue rectangle in Fig. 5, defined on the slice as:

$$|\mathbf{f}_\mathbf{u}| \leq \frac{S}{2} |\mathbf{f}_\mathbf{x}|, \quad |f_t| \leq \frac{M}{2} |\mathbf{f}_\mathbf{x}|. \quad (9)$$

This is a 2D area rather than a 3D volume because the slice equations Eq. (8) have a common parameter s for the $f_u f_v$ axes. This 1D reduction is called a *dimensionality gap* in light fields [Levin and Durand 2010], and we refer to this 2D area as *focal area*.

As can be seen from Eq. (9) and Fig. 5, the focal area has an area of $SM|\mathbf{f}_\mathbf{x}|^2$. Meanwhile, by extending [Hasinoff et al. 2009b] to the time domain as in Appendix A, we can show that the integral of the squared MTF over the entire focal area (i.e., without the boundaries due to S and M as in Eq. (9)) is bounded as follows.

$$\iint |\hat{k}(\mathbf{f}_\mathbf{x}, s\bar{\mathbf{f}}_\mathbf{x}, f_t)|^2 dsdf_t \leq \frac{2A^3T}{3}, \quad (10)$$

where $\bar{\mathbf{f}}_\mathbf{x}$ is the normalized version of vector $\mathbf{f}_\mathbf{x}$. Thus, to maximize the worst-case performance, this ‘‘budget’’ must be uniformly distributed across $SM|\mathbf{f}_\mathbf{x}|^2$ area, and hence Eq. (7) results.

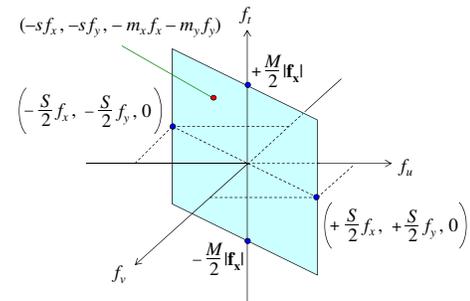


Fig. 5. 3D $f_u f_v f_t$ volume in a 5D frequency domain with a fixed (f_x, f_y) value. An OTF slice from Eq. (8) corresponds to a point (shown in red) in this volume, and the blue region shows the rectangular area which this point can lie on.

Measure of PSF invariance: For this measure we use the ratio of the worst-case squared MTF to the best-case one.

$$\frac{\min_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_{\mathbf{x}})|^2}{\max_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_{\mathbf{x}})|^2} \quad \text{subject to } |s| \leq S/2, |\mathbf{m}| \leq M/2, \quad (11)$$

which has a trivial upper bound of 1. We would like to *maximize* this measure because Eq. (11) being close to 1 indicates that the MTF has a narrow range and hence little variance. Although this measure naturally becomes 1 when the worst-case MTF is optimal, it characterizes how much an MTF varies depending on depth and motion in other cases. Some designs produce a value of 1 for this measure without achieving the optimal worst-case MTF [Dowski and Cathey 1995; Cossairt et al. 2010; Cossairt and Nayar 2010].

Working on an MTF will ignore the phase of an OTF, which can also be a source of PSF variance. However, invariant capture methods such as [Dowski and Cathey 1995; Levin et al. 2008] produce spatially shifted PSFs for different depths or motions, causing OTF phases to vary. Excluding the phase as in Eq. (11) successfully evaluates such cases as invariant PSFs. Hence this measure provides a necessary but not sufficient condition for depth/motion-invariance, and we call this measure as *MTF invariance*. For any method to be depth/motion-invariant, a value close to one needs to be achieved for MTF invariance, but we need to confirm PSF invariance in the spatial domain, as we will do in Sec. 4.2. We have considered using more elaborate measures such as [Baek 2010] by extending them to motion-invariance, but resultant complicated equations hindered tractable analysis of all of the existing camera designs.

3.3 General Form of Kernels and Its Reduced Forms

This section describes a general form of a 5D time-varying light field modulator kernel $k(\mathbf{x}, \mathbf{u}, t)$, and how it can be decomposed into two “reduced” versions of kernels that either account for defocus-only or motion-only deblurring.

Since capturing a time-varying light field amounts to 5D ray integration as in Eq. (3), a kernel can generally be written as:

$$k(\mathbf{x}, \mathbf{u}, t) = \underbrace{\delta(\mathbf{x} - \mathbf{c}(\mathbf{u}, t))}_{\text{Integ. surface}} \underbrace{W(\mathbf{u}, t)}_{\text{Integ. window}}, \quad (12)$$

where δ is a Dirac delta function representing an *integration surface* where the function \mathbf{c} maps ray direction \mathbf{u} at time t to spatial location \mathbf{x} , and W is an *integration window* that limits the range of \mathbf{u} and t . For example, a standard lens focused at depth s_0 with aperture diameter A and exposure time T can be modeled as [Levin et al. 2009]

$$\mathbf{c}(\mathbf{u}, t) = s_0 \mathbf{u}, \quad W(\mathbf{u}, t) = R(|\mathbf{u}|/A)R(t/T), \quad (13)$$

where R is a rect function such that $R(z) = 1$ for $|z| < 1/2$ and $R(z) = 0$ otherwise.

We note that most existing approaches and their combinations can be modeled by decomposing integration surface \mathbf{c} and window W into an aperture part (with subscript a) and an exposure part (e) as:

$$\mathbf{c}(\mathbf{u}, t) = \mathbf{c}_a(\mathbf{u}) + \mathbf{c}_e(t), \quad (14)$$

$$W(\mathbf{u}, t) = W_a(\mathbf{u})W_e(t). \quad (15)$$

Then, Eq. (12) can be decomposed as

$$k(\mathbf{x}, \mathbf{u}, t) = \underbrace{\delta(\mathbf{x} - \mathbf{c}_a(\mathbf{u}))W_a(\mathbf{u})}_{\equiv k_a(\mathbf{x}, \mathbf{u})} * \underbrace{\delta(\mathbf{x} - \mathbf{c}_e(t))W_e(t)}_{\equiv k_e(\mathbf{x}, t)}. \quad (16)$$

where $*$ denotes 2D spatial convolution over xy . Therefore, for most existing camera designs, a 5D time-varying light field kernel can be decomposed into a 4D light field modulator kernel $k_a(\mathbf{x}, \mathbf{u})$ for defocus deblurring and a 3D temporal modulator kernel $k_e(\mathbf{x}, t)$ for motion deblurring. The magnitude of the 5D Fourier transform of k can be trivially calculated as $|\hat{k}| = |\hat{k}_a| \cdot |\hat{k}_e|$.

3.4 Analysis of Existing Designs

The decomposition of kernels in Sec. 3.3 leads to a natural conclusion that we can evaluate high frequency preservation and MTF invariance of a combination of existing defocus deblurring and motion deblurring approaches by multiplying their individual measures. Tables II(a)(b) summarize MTFs and corresponding performance measures developed in Sec. 3.2 for existing camera designs for defocus or motion deblurring. The equations of kernels and MTFs are taken from [Levin et al. 2009; Levin et al. 2008; Bando et al. 2011; Cho et al. 2010] with some modifications including adaptation to the circular aperture. Our major contribution is in the analysis of MTF invariance. See Table I for symbol notation.

3.4.1 Existing Designs for Defocus Deblurring. We explain each row of Table II(a) below. The upper bound for the worst-case MTF is as given in [Hasinoff et al. 2009a].

Standard lens: By excluding the time-dependent component in Eq. (13), the integration surface is linear as $s_0 \mathbf{u}$, and the integration window is $R(|\mathbf{u}|/A)$, which is a disc. The MTF becomes a jinc function as the result of 2D Fourier transform of a disc [Born and Wolf 1984], where $\text{jinc}(z) = 2J_1(z)/z$, and $J_n(z)$ is the n -th order Bessel function of the first kind [Watson 1922]. A jinc function is radially symmetric but otherwise $\text{jinc}(\mathbf{f}_{\mathbf{x}})$ behaves similarly to $\text{sinc}(f_x)\text{sinc}(f_y)$, and has many zero crossings. Therefore, the worst-case MTF is zero, and the MTF invariance is also zero.

Standard lens with narrow aperture: Depth of field can be extended by stopping down the aperture. If we set the focused depth $s_0 = 0$, the defocus diameter for scene depth s can be written as As as can be seen in Fig. 4, and the maximum defocus diameter will be observed at both ends of the depth range $s = \pm S/2$ as $AS/2$. Letting the pixel size Δp , the aperture diameter needs to be reduced to satisfy $AS/2 \leq \Delta p$ so that the maximum defocus diameter is less than the sensor resolution. Meanwhile, the pixel size determines the sensor frequency bandwidth as $|f_x|, |f_y| \leq \Omega$ where $\Omega = 1/(2\Delta p)$. Combining these, we need to set $A = 1/(S\Omega)$. By doing so, the MTF becomes $\text{jinc}^2(\frac{\pi}{S\Omega} s \mathbf{f}_{\mathbf{x}})$, and the argument is less than $\pi/2$, which is well before the first zero crossing of jinc. The minimum MTF and MTF invariance become non-zero, but the MTF is significantly less than the upper bound due to light loss.

Coded aperture: Coded aperture can be modeled as a collection of square sub-apertures where each sub-aperture j is located at (u_j, v_j) within the aperture, and is open or blocked determined by a pseudo-random code γ_j [Levin et al. 2007; Veeraraghavan et al. 2007]. Since coded aperture does not alter the focused depth of the lens, the integration surface remains $s_0 \mathbf{u}$. Here we assume square aperture with side length $A' = \sqrt{\pi}A/2$ so that the aperture area is the same as the circular one. If we denote the sub-aperture size as $\varepsilon A'$, we can set $\varepsilon = 1/(A'S\Omega)$ to achieve the same level of MTF invariance as the narrow aperture case with a significantly better MTF. However, the MTF is still well below the upper bound.

Lattice-focal lens: It can be modeled as a collection of sub-apertures similar to the coded aperture case, but with varying focused depths s_j that are uniformly sampled from the depth range $[-S/2, +S/2]$. Since the number of sub-apertures is $1/\varepsilon^2$, the interval between the depth samples is $\Delta s = S\varepsilon^2$. The MTF is a

Table I. Notation of symbols.

$\mathbf{x} = (x, y)$	Spatial location	A	Diameter of circular aperture	$\varepsilon A', \varepsilon T$	Sub-aperture/exposure size
$\mathbf{u} = (u, v)$	View direction	A'	Size of square aperture $\sqrt{\pi}A/2$	(u_j, v_j)	Sub-aperture location
t	Time	T	Exposure time	t_j	Sub-exposure time instance
s	Scene depth	S	Depth range	Δs	Depth step
$\mathbf{m} = (m_x, m_y)$	Object motion	M	Motion range	a	Parabolic integration coef.
$\mathbf{f}_x = (f_x, f_y)$	Spatial frequency	Ω	Maximum spatial frequency	r	Circular motion radius
		γ_j	Pseudo-random code	w	Focus sweep speed

Table II. Kernel form, MTF, high frequency preservation measure, and MTF invariance measure for camera designs for (a) defocus-only deblurring, (b) motion-only deblurring, and (c) joint defocus and motion deblurring including combinations of existing designs and focus sweep. MTFs are calculated for optimal parameters in corresponding kernels. Notation is summarized in Table I.

(a) Defocus deblurring.

Camera design	4D kernel $k_a(\mathbf{x}, \mathbf{u})$		Squared MTF	High freq. preserv.	MTF invariance
	Integration surface	Integration window	$ \hat{\phi}_s(\mathbf{f}_x) ^2 = \hat{k}_a(\mathbf{f}_x, -s\mathbf{f}_x) ^2$	$\min_s \hat{\phi}_s(\mathbf{f}_x) ^2$	$\min_s \hat{\phi}_s ^2 / \max_s \hat{\phi}_s ^2$
Upper bound	--	--	--	$\frac{2A^3}{3S \mathbf{f}_x }$	1
Standard lens	$\delta(\mathbf{x} - s_0\mathbf{u})$	$R(\mathbf{u} /A)$	$\frac{\pi^2 A^4}{16S^4\Omega^4} \text{jinc}^2(\pi A s \mathbf{f}_x)$	0	0
Narrow aperture	$\delta(\mathbf{x} - s_0\mathbf{u})$	$R(S\Omega \mathbf{u})$	$\frac{\pi^2}{16S^4\Omega^4} \text{jinc}^2(\frac{\pi}{S\Omega} s \mathbf{f}_x)$	$\frac{\pi^2}{16S^4\Omega^4} \text{jinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$	$\text{jinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$
Coded aperture	$\sum_j \{\delta(\mathbf{x} - s_0\mathbf{u})$	$\gamma_j R(\frac{u-u_j}{\varepsilon A'}) R(\frac{v-v_j}{\varepsilon A'})\}$	$\frac{A'^2}{2S^2\Omega^2} \text{sinc}^2(\frac{\pi}{S\Omega} s f_x) \cdot \text{sinc}^2(\frac{\pi}{S\Omega} s f_y)$	$\frac{A'^2}{2S^2\Omega^2} \text{sinc}^2(\frac{\pi}{2\Omega} f_x) \cdot \text{sinc}^2(\frac{\pi}{2\Omega} f_y)$	$\text{sinc}^2(\frac{\pi}{2\Omega} f_x) \cdot \text{sinc}^2(\frac{\pi}{2\Omega} f_y)$
Lattice-focal lens	$\sum_j \{\delta(\mathbf{x} - s_j\mathbf{u})$	$R(\frac{u-u_j}{\varepsilon A'}) R(\frac{v-v_j}{\varepsilon A'})\}$	$\frac{A'^{8/3}}{(S\Omega)^{4/3}} \sum_j \text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s \Omega} f_x) \cdot \text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s \Omega} f_y)$	$ \hat{\phi}_{\Delta s/2}(\mathbf{f}_x) ^2$	$ \hat{\phi}_{\Delta s/2}(\mathbf{f}_x) ^2 / \hat{\phi}_0(\mathbf{f}_x) ^2$
Wavefront coding	$\delta(\mathbf{x} - (au^2, av^2))$	$R(u/A') R(v/A')$	$\frac{A'^2}{S^2 f_x f_y }$	$\frac{A'^2}{S^2 f_x f_y }$	1
Static focus sweep	$\frac{1}{S} \int_{-S/2}^{+S/2} \{\delta(\mathbf{x} - s_0\mathbf{u})$	$R(\mathbf{u} /A) ds_0$	$\frac{A^2}{S^2 \mathbf{f}_x ^2}$	$\frac{A^2}{S^2 \mathbf{f}_x ^2}$	1

(b) Motion deblurring.

Camera design	3D kernel $k_e(\mathbf{x}, t)$		Squared MTF	High freq. preserv.	MTF invariance
	Integration surface	Integration window	$ \hat{\phi}_m(\mathbf{f}_x) ^2 = \hat{k}_e(\mathbf{f}_x, -\mathbf{m} \cdot \mathbf{f}_x) ^2$	$\min_m \hat{\phi}_m(\mathbf{f}_x) ^2$	$\min \hat{\phi}_m ^2 / \max \hat{\phi}_m ^2$
Upper bound	--	--	--	$\frac{T}{M \mathbf{f}_x }$	1
Static/follow-shot	$\delta(\mathbf{x} - \mathbf{m}_0 t)$	$R(t/T)$	$T^2 \text{sinc}^2(\pi T(\mathbf{m} - \mathbf{m}_0) \cdot \mathbf{f}_x)$	0	0
Short exposure	$\delta(\mathbf{x})$	$R(M\Omega t)$	$\frac{1}{M^2\Omega^2} \text{sinc}^2(\frac{\pi}{M\Omega} \mathbf{m} \cdot \mathbf{f}_x)$	$\frac{1}{M^2\Omega^2} \text{sinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$	$\text{sinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$
Coded exposure	$\sum_j \{\delta(\mathbf{x})$	$\gamma_j R(\frac{t-t_j}{\varepsilon T})\}$	$\frac{T}{2M\Omega} \text{sinc}^2(\frac{\pi}{M\Omega} \mathbf{m} \cdot \mathbf{f}_x)$	$\frac{T}{2M\Omega} \text{sinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$	$\text{sinc}^2(\frac{\pi}{2\Omega} \mathbf{f}_x)$
1D motion-invariant	$\delta(\mathbf{x} - (at^2, 0))$	$R(t/T)$	$\frac{T}{M f_x } R(\frac{\mathbf{m} \cdot \mathbf{f}_x}{M f_x })$	$\frac{T}{M f_x } (f_y = 0)$ 0 ($f_y \neq 0$)	1 ($f_y = 0$) 0 ($f_y \neq 0$)
Circular motion	$\delta(\mathbf{x} - \mathbf{r}(t))$ with $\mathbf{r}(t) = (r \cos(2\pi t/T), r \sin(2\pi t/T))$	$R(t/T)$	$T^2 J_n^2(\frac{M T}{2} \mathbf{f}_x)$ with $n = T \mathbf{m} \cdot \mathbf{f}_x$	0	0
Orthogonal parabolic exposures	$\delta(\mathbf{x} - (at^2, 0))$ $\delta(\mathbf{x} - (0, at^2))$	$R(2t/T)$ (1st shot) $R(2t/T)$ (2nd shot)	$\frac{T}{2\sqrt{2}M f_x } R(\frac{\mathbf{m} \cdot \mathbf{f}_x}{\sqrt{2}M f_x })$ $\frac{T}{2\sqrt{2}M f_y } R(\frac{\mathbf{m} \cdot \mathbf{f}_x}{\sqrt{2}M f_y })$	$\frac{T}{2\sqrt{2}M f_x } R(\frac{f_y}{2 f_x })$ $\frac{T}{2\sqrt{2}M f_y } R(\frac{f_x}{2 f_y })$	1 ($ f_y \leq f_x $), 0 (o/w) 1 ($ f_x \leq f_y $), 0 (o/w)

(c) Joint defocus and motion deblurring.

Camera design	5D kernel $k(\mathbf{x}, \mathbf{u}, t)$		Squared MTF	High freq. preserv.	MTF invariance
	Integration surface	Integration window	$ \hat{\phi}_{s,m}(\mathbf{f}_x) ^2 = \hat{k}(\mathbf{f}_x, -s\mathbf{f}_x, -\mathbf{m} \cdot \mathbf{f}_x) ^2$	$\min_{s,m} \hat{\phi}_{s,m}(\mathbf{f}_x) ^2$	$\min \hat{\phi}_{s,m} ^2 / \max \hat{\phi}_{s,m} ^2$
Upper bound	--	--	--	$\frac{2A^3 T}{3SM \mathbf{f}_x ^2}$	1
Combination of existing designs	$k_a(\mathbf{x}, \mathbf{u}) * k_e(\mathbf{x}, t)$		$ \hat{\phi}_s(\mathbf{f}_x) ^2 \cdot \hat{\phi}_m(\mathbf{f}_x) ^2$	$\min_s \hat{\phi}_s ^2 \cdot \min_m \hat{\phi}_m ^2$	$(\min \hat{\phi}_s ^2 / \max \hat{\phi}_s ^2) \cdot (\min \hat{\phi}_m ^2 / \max \hat{\phi}_m ^2)$
Focus sweep	$\delta(\mathbf{x} - wt\mathbf{u})$	$R(\mathbf{u} /A) R(t/T)$	$\frac{A^3 T}{\sqrt{3}SM \mathbf{f}_x ^2} \left(1 - \frac{4 \mathbf{m} \cdot \mathbf{f}_x ^2}{3M^2 \mathbf{f}_x ^2}\right)$	$\frac{2A^3 T}{3\sqrt{3}SM \mathbf{f}_x ^2}$	$\frac{2}{3}$

summation of sinc functions with varying centers $j\Delta s$ as shown in Fig. 6(a), and the optimal sub-aperture size $\varepsilon = (A'S\Omega)^{-1/3}$ makes every depth s fall within $[-\pi/2, +\pi/2]$ from the center of

one of the sinc functions. However, as can be seen in Fig. 6(b), the MTF is oscillating, reaching maxima at $s = n\Delta s$ and minima at $s = (n+1/2)\Delta s$ for integer n (Table II(a) shows $n = 0$ case). The

oscillation amplitude becomes larger for higher frequencies, which was also observed in [Baek 2010], meaning that MTF invariance is low. The average MTF is shown to be around $\frac{A^{8/3}}{S^{4/3}\Omega^{1/3}|f_x|}$, which is to date the best of existing computational cameras for defocus deblurring, but falls short of the upper bound [Levin et al. 2009].

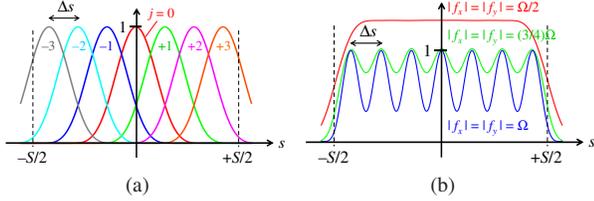


Fig. 6. Plots of the sinc functions comprising the lattice-focal lens MTF. (a) $\text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s\Omega}f_x)\text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s\Omega}f_y)$ for some j . (b) $\sum_j \text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s\Omega}f_x)\text{sinc}^2(\frac{\pi(s-j\Delta s)}{\Delta s\Omega}f_y)$ for some (f_x, f_y) .

Wavefront coding: Wavefront coding uses a cubic phase plate [Dowski and Cathey 1995] to make the integration surface quadratic as (au^2, av^2) . The MTF does not depend on depth s and MTF invariance is 1. The best performance is obtained by setting $a = S/(2A')$, but it does not achieve the MTF upper bound.

Static focus sweep: Previous analysis of focus sweep assumed static scenes and worked on a 4D kernel, where the integration surface of the standard lens is averaged over the depth range. This analysis is still valid for moving scenes if focus sweep is realized by other means than moving the focus through time, such as distributing rays with a diffuser [Cossairt et al. 2010] or sweeping through wavelength of light [Cossairt and Nayar 2010]. The MTF does not depend on depth s and MTF invariance is 1, but the MTF does not achieve the upper bound.

3.4.2 Existing Designs for Motion Deblurring. We explain each row of Table II(b) below. The upper bound for the worst-case MTF is as given in [Levin et al. 2008].

Static camera / follow-shot: A static camera has an integration surface of $\mathbf{0}$, and moving a camera to track particular object motion \mathbf{m}_0 leads to a linear surface $\mathbf{m}_0 t$. Finite exposure time creates a box function that results in a sinc function in the MTF. Due to zero crossings of the sinc, the minimum MTF is zero and MTF invariance is also zero.

Static camera with short exposure: Motion blur can be reduced by decreasing exposure time. The maximum motion blur will be observed when object speed is maximum as $|\mathbf{m}| = M/2$, and the blur size will be $TM/2$. The exposure time needs to be reduced to satisfy $TM/2 \leq \Delta p$, meaning that $T = 1/(M\Omega)$. Similar to the narrow aperture case, this makes the argument of the MTF sinc function less than $\pi/2$ to avoid zero crossings. However, the MTF is significantly less than the upper bound due to light loss.

Coded exposure: This can be considered as a 1D version of coded aperture along the time axis, where we have sub-exposure times with length εT centered at time t_j [Raskar et al. 2006]. We can set $\varepsilon = 1/(TM\Omega)$ to achieve the same level of MTF invariance as the short exposure case with a significantly better MTF. However, the MTF is still well below the upper bound.

1D motion-invariant photography: This can be considered as a 1D version of wavefront coding along the time axis. However, there is an important difference. For defocus deblurring, we have two parameters (u, v) to account for a single parameter of depth s , whereas for motion deblurring, we only have one parameter t to

account for 2D motion. This is why one has to choose 1D motion direction by setting the integration surface as $(at^2, 0)$, i.e., horizontal, for example. Note that setting (at^2, at^2) will only change the chosen direction to oblique (45°) one. The best performance is obtained by setting $a = M/(2T)$, but the worst-case MTF achieves the upper bound only when $f_y = 0$. It becomes zero in other cases and so does MTF invariance.

Circular motion: This method moves the sensor circularly about the optical axis during exposure in order to account for 2D motion directions [Bando et al. 2011]. It is shown that the average MTF asymptotically approaches to $2/\pi$ ($\approx 64\%$) of the upper bound when the circular motion radius is chosen as $r = MT/(4\pi)$, but the Bessel function in the MTF has zero crossings and hence the worst-case MTF and MTF invariance are both zero.

Orthogonal parabolic exposures: This is a two-shot approach that performs 1D motion-invariant capture in two orthogonal (horizontal and vertical) directions [Cho et al. 2010]. The two MTFs complement each other's frequency zeros if we set $a = \sqrt{2}M/T$, and the summed MTF achieves $2^{-1.5}$ ($\approx 35\%$) of the upper bound in the worst-case. Thus this method is good for high frequency preservation for 2D object motion, but MTF invariance is always zero for either shot except for $|f_x| = |f_y|$.

3.4.3 Combinations of Existing Designs. As in Sec. 3.4.1, there is no single method that achieves both the worst-case MTF and MTF invariance upper bounds for defocus deblurring. Similarly, there is no single method that does the same for motion deblurring as in Sec. 3.4.2. Therefore, combinations of existing approaches will not achieve both of the upper bounds for joint defocus and motion deblurring.

4. NEAR-OPTIMALITY OF FOCUS SWEEP

From the kernel analysis in Sec. 3.3, performance of separable kernels as in Eqs. (14-16) is determined by the product of defocus-only and motion-only deblurring methods. Therefore, in order to design a joint defocus and motion deblurring method having high scores for worst-case MTF and MTF invariance, we have two options: 1) design optimal defocus-only and motion-only methods and combine them; or 2) explore inseparable kernels and see how they perform. Considering the absence of optimal defocus/motion-only methods so far, in this paper we took option 2. Since inseparable integration window $W(\mathbf{u}, t)$ implies time-varying coded aperture, we are more interested in inseparable integration surfaces, the simplest of which may be $\mathbf{c}(\mathbf{u}, t) = w\mathbf{t}\mathbf{u}$, where w is some constant. This, in fact, represents the focus sweep method, where the focused depth s_0 is moved according to time as $s_0 = wt$, and we found that the focus sweep kernel provides the performance measures close to optimal (around 60%) for certain combinations of depth and motion ranges, which will be shown in Sec. 4.1. Since our MTF invariance measure ignores phase effects as described in Sec. 3.2, Sec. 4.2 confirms its near 2D motion-invariance in the spatial domain. The depth-invariance was confirmed previously [Nagahara et al. 2008]. Exploration of other inseparable kernels is left as future work.

4.1 Performance Measures of Focus Sweep Capture

With the integration surface $\mathbf{c}(\mathbf{u}, t) = w\mathbf{t}\mathbf{u}$ and the standard integration window, the focus sweep kernel can be written as:

$$k(\mathbf{x}, \mathbf{u}, t) = \delta(\mathbf{x} - w\mathbf{t}\mathbf{u})R(|\mathbf{u}|/A)R(t/T), \quad (17)$$

where we refer to w as *focus sweep speed*. We take the 5D Fourier transform of k while taking into account the OTF slice equations in Eq. (8) to obtain the OTF as in Eq. (5). Through the derivation in

Appendix B, we have:

$$|\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_x)|^2 = \begin{cases} \frac{A^2}{w^2|\mathbf{f}_x|^2} \left(1 - \frac{4|\mathbf{m} \cdot \mathbf{f}_x|^2}{A^2w^2|\mathbf{f}_x|^2}\right) & (\text{for } |s| \leq \frac{Tw}{2}, |\mathbf{m} \cdot \mathbf{f}_x| \leq \frac{Aw}{2}|\mathbf{f}_x|) \\ 0 & (\text{otherwise}) \end{cases} \quad (18)$$

In order to account for depth range $|s| \leq S/2$ and motion range $|\mathbf{m}| \leq M/2$, the conditions for the MTF in Eq. (18) to have non-zero value are $S \leq Tw$ and $M \leq Aw$. Although the MTF is independent of depth s , it has a fall-off term in the parentheses in Eq. (18) which depends on \mathbf{m} , as shown in the plot in Fig. 7. This reflects the deviation from strict motion-invariance, but we can maximize the worst-case performance, observed at $|\mathbf{m} \cdot \mathbf{f}_x| = (M/2)|\mathbf{f}_x|$, by making Aw sufficiently larger than M . Specifically, with $S = Tw$ and $M = \lambda Aw$, where $\lambda (< 1)$ is a margin parameter to be optimized, we compute the ratio ρ of the worst-case performance ($\min_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}|^2$) to the upper bound in Eq. (7) as:

$$\rho = \frac{\min_{s,\mathbf{m}} |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_x)|^2}{\frac{2A^3T}{3SM|\mathbf{f}_x|^2}} = \frac{3}{2}\lambda(1 - \lambda^2). \quad (19)$$

By simple arithmetic, we can show that ρ can be maximized with $\lambda = 1/\sqrt{3}$, at which we have $\rho = 1/\sqrt{3}$. Therefore, with $Aw = \sqrt{3}M$, focus sweep achieves $1/\sqrt{3} \approx 58\%$ of the worst-case MTF upper bound. MTF invariance is determined by the fall-off term and can be computed as $1 - \lambda^2 = 2/3$.

The conditions $S = Tw$ and $M = (1/\sqrt{3})Aw$ indicate that depth and motion ranges cannot be set independently, as we have only a single parameter w to tweak. Near-optimality is achieved only when S and M has a fixed ratio of $S/M = \sqrt{3}T/A$.

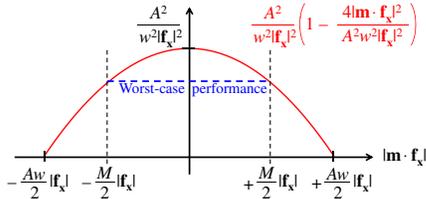


Fig. 7. Plot of the focus sweep MTF in Eq. (18). The worst-case performance at $\pm(M/2)|\mathbf{f}_x|$ can be maximized with $Aw = \sqrt{3}M$.

Discussion on optimality: Here we explain why focus sweep can be near-optimal for joint defocus/motion deblurring, even though the focus sweep MTF was previously shown to be suboptimal for defocus-only deblurring [Levin et al. 2009].

The key idea is joint optimality. For example, 1D motion-invariant photography is also suboptimal for static scene capture because a fixed frequency budget is distributed across the assumed motion range, and as a result, less portion of the budget is assigned to static scenes. Focus sweep for static scenes is optimal in the same way. The above analysis finds that, as focus sweep is a function of time, it distributes the budget over a nonzero motion range $|\mathbf{m}| \leq Aw/2$ as can be seen in Eq. (18) and Fig. 7. In other words, the budget is distributed along the f_t axis across the focal area shown in Fig. 5 (note that $f_t = -\mathbf{m} \cdot \mathbf{f}_x$ from Eq. (8)). Focus sweep for defocus-only deblurring only uses the line segment at $f_t = 0$, thereby wasting budget along the f_t (motion) direction. This makes focus sweep suboptimal for defocus-only deblurring. If defocus-only deblurring is the only concern, then the budget can

be concentrated in the $f_t = 0$ segment, thus increasing the optimal value accordingly. The lattice-focal lens is more suitable for this purpose, but it assigns significantly less budget to moving scenes.

We would like to note that, by taking the time domain into consideration, our 5D analysis revealed for the first time the fact that focus sweep's suboptimality for defocus-only deblurring comes from the budget assignment to moving scenes.

4.2 Near 2D Motion-Invariance of Focus Sweep

Here we confirm, in the spatial domain, that the focus sweep method works approximately as 2D motion-invariant capture.

Similarly to Sec. 3.1, we consider a scene in which a Lambertian object is moving at a constant velocity \mathbf{m} . Without loss of generality, we assume the object is at $s = 0$. As we sweep the focused depth during exposure as $s = wt$, the defocus blur diameter changes accordingly as $b = As = Aw t$ (Fig. 4). By modeling the instantaneous defocus blur as a disc (pillbox) function as $\psi_b(\mathbf{x}) = \frac{4}{\pi b^2} R(|\mathbf{x}|/b)$, the resulting PSF ϕ can be derived as superposition of shifting discs with varying sizes (see Fig. 8):

$$\phi(\mathbf{x}) = \int_{-T/2}^{+T/2} \psi_{Aw|t|}(\mathbf{x} - \mathbf{m}t) dt \\ = \int_{-T/2}^{t_0} \frac{4}{\pi A^2 w^2 t^2} dt + \int_{t_1}^{+T/2} \frac{4}{\pi A^2 w^2 t^2} dt, \quad (20)$$

where t_0 and $t_1 (> t_0)$ are the two roots of the equation $|\mathbf{x} - \mathbf{m}t| = Aw|t|/2$ coming from the disc boundary. We assume T to be large enough to satisfy $-T/2 < t_0$ and $t_1 < +T/2$. Straightforward integration of Eq. (20) and rearranging leads to (see Appendix C):

$$\phi(\mathbf{x}) = \frac{4}{\pi Aw|\mathbf{x}|} \left(1 - \frac{4|\mathbf{m}|^2 \sin^2 \theta}{A^2 w^2}\right)^{1/2} - \frac{16}{\pi A^2 w^2 T}, \quad (21)$$

where θ is the angle between vectors \mathbf{x} and \mathbf{m} . With large T , the second term in Eq. (21) is negligible. The first term depends on object motion \mathbf{m} , but the dependence becomes evident only when object speed $|\mathbf{m}|$ is close to $Aw/2$, which equals $(\sqrt{3}/2)M$ with the optimal focus sweep speed derived in Sec. 4.1. Hence, with $|\mathbf{m}| \leq M/2$, the radially symmetric fall-off term ($1/|\mathbf{x}|$) dominates, and the PSF is nearly 2D motion-invariant as shown in Fig. 9.

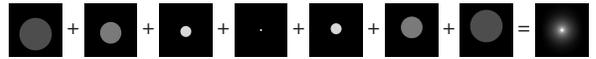


Fig. 8. Log-intensity of instantaneous PSFs during focus sweep (seven images on the left), and the resultant, time-integrated focus sweep PSF (rightmost image). The figure shows the case in which an object is moving vertically. Note that the PSF center is shifting upward through time while changing its diameter, and that the resultant PSF is still almost radially symmetric.

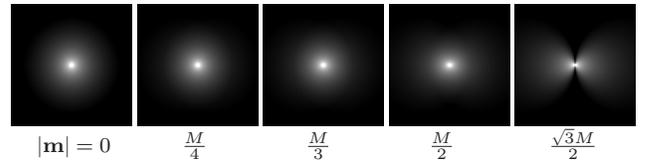


Fig. 9. Log-intensity of the focus sweep PSFs for horizontally moving objects with varying speed $|\mathbf{m}|$. The PSF profile remains nearly identical for $|\mathbf{m}| \leq M/2$, and only beyond that the deviation becomes clearly apparent. Near motion-invariance also holds for different object motion directions, as the above images will only get rotated.

5. IMPLEMENTATION

We evaluated the performance of the focus sweep method by comparing with other existing defocus/motion deblurring methods by simulation. We also fabricated a focus sweep camera to demonstrate its effectiveness for real scenes.

Simulation: To avoid clutter, we selected two combinations of existing methods for comparison, along with baseline standard camera capture as follows.

- Wide-Long:** Standard camera with wide aperture A and long exposure T .
- Wide-Short:** Standard camera with wide aperture A and short exposure $1/M\Omega$.
- Narrow-Long:** Standard camera with narrow aperture $1/S\Omega$ and long exposure T .
- LatticeFocal-OrthoParabolic:** Camera equipped with a lattice-focal lens [Levin et al. 2009] that is moved according to the orthogonal parabolic exposure method [Cho et al. 2010], which is a two-shot approach. This is the best combination in terms of high frequency preservation.
- StaticFocusSweep-1DMotionInv:** Camera equipped with an optical system according to a static focus sweep method (considered as an idealized version of [Cossairt et al. 2010; Cossairt and Nayar 2010]) that is moved according to 1D motion-invariant capture [Levin et al. 2008]. This is the best combination in terms of depth/motion-invariance.
- FocusSweep:** Focus sweep camera.

We set $A = 100\Delta p$, $T = 1$, $S = 1$, and $M = 100\Delta p$, and simulated PSFs for each of the above 6 methods for various object speeds and depths. Kernel equations for producing PSFs and how to combine existing methods can be found in Table II. Adjustable parameters were optimized for each method. We convolved a natural image $I(\mathbf{x})$ with a PSF $\phi_{s,\mathbf{m}}(\mathbf{x})$ corresponding to depth s and motion \mathbf{m} , and added Gaussian noise with standard deviation η to produce a degraded image $B_{s,\mathbf{m}}(\mathbf{x})$. Then we obtained a deblurred image $I'(\mathbf{x})$ using Wiener deconvolution as:

$$\hat{I}'(\mathbf{f}_x) = \frac{\hat{\psi}^*(\mathbf{f}_x)}{|\hat{\psi}(\mathbf{f}_x)|^2 + \eta^2/\sigma^2(\mathbf{f}_x)} \hat{B}_{s,\mathbf{m}}(\mathbf{f}_x), \quad (22)$$

where ψ is a PSF for deconvolution which will be specified below, superscript $*$ denotes complex conjugate, and $\sigma^2(\mathbf{f}_x)$ is an expected power spectrum of sharp natural images, for which we used the model in [Levin et al. 2009]. We computed the mean squared error (MSE) as $|I(\mathbf{x}) - I'(\mathbf{x})|^2/N$, where N is the number of pixels. We repeated this process for several images and took the MSE average.

To see high frequency preservation, we performed deconvolution with the PSF for the correct depth s and motion \mathbf{m} , i.e., we set $\psi = \phi_{s,\mathbf{m}}$. To see depth/motion-invariance, we always used the “center” PSF corresponding to $s = 0$ and $\mathbf{m} = \mathbf{0}$, i.e., $\psi = \phi_{0,\mathbf{0}}$, irrespective of the values (s, \mathbf{m}) used to produce $B_{s,\mathbf{m}}$. As the center PSF can produce shifted images, we register the deconvolved image with the original image before computing the MSE. To further reduce variance in reconstruction error, we also tried the Wiener filter that minimizes the average reconstruction error over the depth and motion ranges, i.e., we always used the following equation instead:

$$\hat{I}'(\mathbf{f}_x) = \frac{\iint \hat{\phi}_{s,\mathbf{m}}^*(\mathbf{f}_x) ds d\mathbf{m}}{\iint |\hat{\phi}_{s,\mathbf{m}}(\mathbf{f}_x)|^2 + \eta^2/\sigma^2(\mathbf{f}_x) ds d\mathbf{m}} \hat{B}_{s,\mathbf{m}}(\mathbf{f}_x), \quad (23)$$

where the integrals over s and \mathbf{m} are discretized over the depth and motion ranges. Note that this filter is fixed and independent of the values (s, \mathbf{m}) used to produce $B_{s,\mathbf{m}}$.

Prototype camera: We implemented a focus sweep camera as shown in Fig. 3(a) by moving a lens during exposure. We used a Canon EOS Rebel T3i DSLR and a Canon EF-S 18-55mm f/3.5-5.6 IS II lens. We programmed an Arduino board so that focus commands were sent to the lens using the serial peripheral interface (SPI), triggered by the hot shoe. We kept the lens at full aperture $f/5.6$ with 55mm focal length. Due to the limitation of the lens focusing speed, the exposure time was set to $T = 200\text{msec}$, during which the plane of focus swept from 30cm to infinity, corresponding to $d \in [55\text{mm}, 67\text{mm}]$ inside the camera, or $s \in [-0.1, +0.1]$. The zero depth plane $s = 0$ corresponds to 60cm, and the defocus diameter was 70 pixels (for a 1200×800 image) at $s = \pm 0.1$ when the lens was focused at $s = 0$. This means that we can handle motion blur of up to $70/\sqrt{3} \approx 40$ pixels, where the margin $\sqrt{3}$ is as derived in Sec. 4.1. Practically, we also need to set the depth range S less than the focus sweep range, and we set $S = 0.14$ which corresponds to the actual scene depth from 40cm to 200cm.

Fig. 10 shows PSFs of the prototype camera obtained by capturing a moving point light source with various velocities and depths. The PSFs are almost identical. When the lens was stationary, the PSF of our prototype camera showed distortion and deviated from a pillbox around the image border. Moreover, due to the lens inertia, the focus sweep speed was not perfectly constant. In spite of these facts, the focus sweep PSF remained nearly invariant, indicating that focus sweep is robust to manufacturing imperfections.

For deblurring, we used BM3D method [Dabov et al. 2008]. To obtain a reference image of a standard camera, we used a beam splitter as shown in Fig. 3(b).

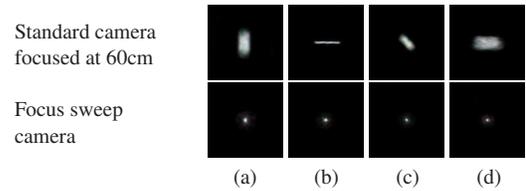


Fig. 10. PSFs of a standard camera and those of our prototype focus sweep camera. (a) 40cm depth, 30 pixels motion blur. (b) 60cm depth, 40 pixels motion blur. (c) 100cm depth, 20 pixels motion blur. (d) 200cm depth, 30 pixels motion blur.

6. RESULTS

Simulation: Fig. 11 shows the result of simulation described in Sec. 5, where we plot the performance as $\text{PSNR} = -10 \log_{10}(\text{MSE})$. As only StaticFocusSweep-1DMotionInv produced different performance depending on object motion direction, cases of horizontal and vertical motion directions are plotted for the method. The top row shows performance of high frequency preservation, where the correct PSF was used for deconvolution. The standard camera with wide aperture and long exposure produced the best PSNR for static scene at focused depth, but the performance rapidly deteriorated for other conditions. The standard camera with wide aperture and short exposure outperformed other methods when the scene is focused, but performed poorly for out of focus scenes. Stopping down the aperture always performed the

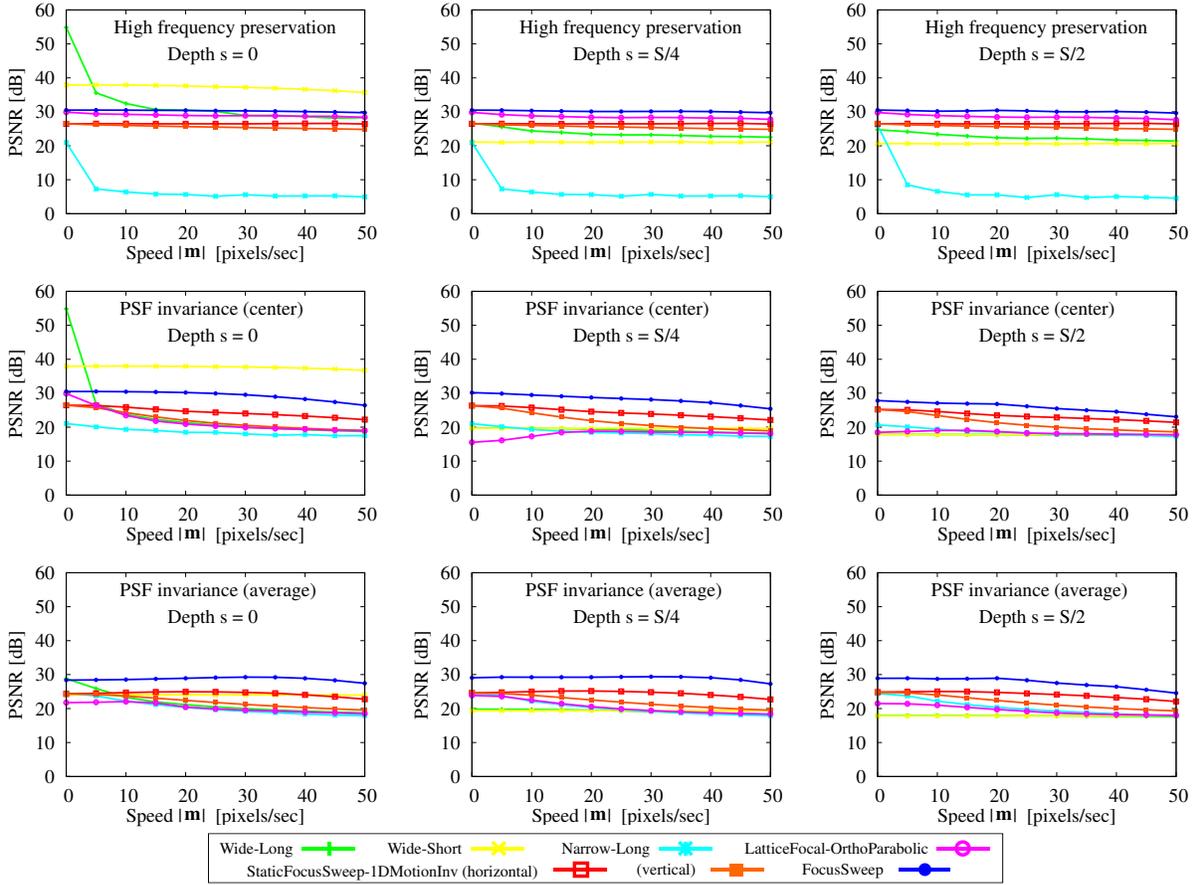


Fig. 11. PSNR of deconvolution simulation results for various scene depth $s = \{0, S/4, S/2\}$ and object speed $|\mathbf{m}| \leq M/2$, where $S = 1$ and $M = 100$. The standard deviation of the added noise is $\eta = 0.001$ for $[0, 1]$ pixel values. The top row shows performance of high frequency preservation, where the correct PSF was used for deconvolution. The middle row shows performance of PSF invariance, where the center PSF corresponding to $s = 0$ and $\mathbf{m} = \mathbf{0}$ was always used. The bottom row also shows performance of PSF invariance, where the Wiener filter that minimizes the average reconstruction error over the depth and motion ranges was always used.

worst. Both of the two hybrid approaches provided stable performance, with LatticeFocal-OrthoParabolic performing better, as expected. The focus sweep outperformed these two hybrid methods.

The middle row of Fig. 11 shows performance of PSF invariance, where the center PSF corresponding to $s = 0$ and $\mathbf{m} = \mathbf{0}$ was always used. Short exposure capture is only effective for focused scenes, which means that focus sweep is not beneficial when used solely as 2D motion-invariant capture for scenes with little depth variation. LatticeFocal-OrthoParabolic deteriorated for moving and/or out of focus scenes as it is depth and motion dependent. StaticFocusSweep-1DMotionInv performed stably for horizontal object motion, but it decayed for vertical motion. The focus sweep, although showing gradual decrease in PSNR for faster object motion, performed stably and better than the hybrid methods.

The bottom row of Fig. 11 also shows performance of PSF invariance, where the Wiener filter that minimizes the average reconstruction error over the depth and motion ranges was always used as described in Sec. 5. In this case the focus sweep always performed the best, and the gradual decrease in PSNR observed above was alleviated.

Table III(a) visually summarizes deconvolution simulation results for focus sweep capture for a resolution chart image. As can be seen in the middle column (2), when the center PSF is always used, the reconstruction error gradually increases as the scene depth s and object speed $|\mathbf{m}|$ increase. One reason for this is *tail-clipping* [Levin et al. 2008], caused by the use of finite exposure or sweep range as opposed to the theoretically assumed infinite range, which applies to all of the existing invariant capture methods. Another reason specific to focus sweep is the deviation from strict motion-invariance, which explains why the fall in PSNR is larger in $|\mathbf{m}|$ than in s . By using the Wiener filter that minimizes the average reconstruction error over the depth and motion ranges, we can alleviate the reconstruction error increase as shown in the right column (3), which has also been seen in Fig. 11. However, because the improvement is modest, and since we ultimately use a more sophisticated deconvolution method rather than Wiener filters, we keep using the center PSF for deconvolution for the rest of the paper. Note that adapting advanced deconvolution methods for the purpose of minimizing the average reconstruction error over a set of PSFs is non-trivial as opposed to the Wiener filter case.

Table III. PSNR from focus sweep simulation for various scene depth $s = \{0, S/4, S/2\}$, object speed $|\mathbf{m}| = \{0, M/4, M/2\}$, and noise level $\eta = \{0.001, 0.01\}$. Deconvolved images are also shown in (a). (1) Correct PSF was used for deconvolution. (2) Center PSF corresponding to $s = 0$ and $\mathbf{m} = \mathbf{0}$ was always used. (3) Wiener filter that minimizes the average reconstruction error was always used.

(a) Noise level $\eta = 0.001$.

	(1) High frequency preservation			(2) PSF invariance (center)			(3) PSF invariance (average)		
	$s = 0$	$s = S/4$	$s = S/2$	$s = 0$	$s = S/4$	$s = S/2$	$s = 0$	$s = S/4$	$s = S/2$
$ \mathbf{m} = 0$									
	21.7	21.7	21.8	21.7	21.6	20.2	20.4	20.8	20.7
$ \mathbf{m} = M/4$									
	21.5	21.1	21.4	21.3	19.9	18.7	20.7	20.3	20.0
$ \mathbf{m} = M/2$									
	20.8	20.7	20.6	19.2	18.2	16.4	19.7	19.3	17.5

(b) Noise level $\eta = 0.01$.

	(1) High frequency preservation			(2) PSF invariance (center)			(3) PSF invariance (average)		
	$s = 0$	$s = S/4$	$s = S/2$	$s = 0$	$s = S/4$	$s = S/2$	$s = 0$	$s = S/4$	$s = S/2$
$ \mathbf{m} = 0$	16.6	16.6	16.6	16.6	16.6	16.1	16.6	16.7	16.7
$ \mathbf{m} = M/4$	16.5	16.7	16.5	16.5	17.0	16.0	16.7	17.2	16.7
$ \mathbf{m} = M/2$	16.3	16.4	16.4	15.6	15.7	14.9	16.3	16.3	15.6

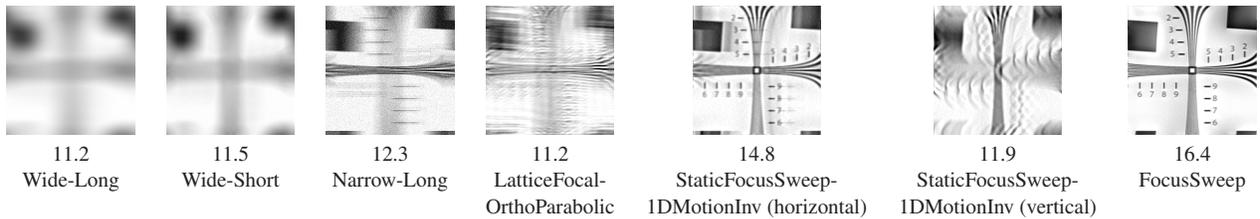


Fig. 12. Simulated comparison of invariance for various capture strategies. For each of the capture strategies, a blurred resolution chart image corresponding to $s = S/2$ and $|\mathbf{m}| = M/2$ was deconvolved with the center PSF for $s = 0$ and $\mathbf{m} = \mathbf{0}$. Corresponding PSNR values are shown under the deconvolved images. Even though focus sweep is not strictly motion-invariant, it produced a better result than a hybrid invariant capture method, i.e., StaticFocusSweep-1DMotionInv (horizontal). Note that StaticFocusSweep-1DMotionInv is not invariant to non-horizontal object motion. We included results of non-invariant capture methods only for the purpose of illustrating the difference in image quality from those of (near-)invariant capture methods.

As shown in Fig. 12, even though focus sweep deteriorates at $s = S/2$ and $|\mathbf{m}| = M/2$ when the center PSF is used, it is still better than StaticFocusSweep-1DMotionInv capture for horizontal object motion, which in theory is strictly depth and (1D) motion-invariant. Table III(b) shows PSNR values for focus sweep capture at a different noise level, showing that the observations made above remain the same.

Real examples: Fig. 2 shows an example of moving people at different depths. The standard camera (Fig. 2(a)) is focused at the person on the left, but she is blurred due to motion. The other two persons are also blurred due to motion and/or defocus. The PSFs shown in the insets in the magnified views were obtained by deconvolving each face while changing the PSF parameters (pillbox size and motion blur length/direction), and by picking the PSF that

produced a visually pleasing result. While these three PSFs for the standard camera image differ from each other significantly, focus sweep PSFs (obtained by simulation) are nearly identical as shown in Fig. 2(b). Applying non-blind, spatially uniform deconvolution produced a sharp image as shown in Fig. 2(c).

Fig. 13 shows an example of moving playing cards at various depths. The standard camera (Fig. 13(a)) is focused at the jack on the right as indicated by the green arrow, leaving other cards defocused. Some cards are also motion blurred, which is especially noticeable for the ones whose motions are indicated by the red arrows. By using the focus sweep method, we were able to capture an image which had almost identical blur everywhere across the image as in Fig. 13(b). The result of deconvolution with a single PSF is shown in Fig. 13(c). If we instead avoid defocus and mo-

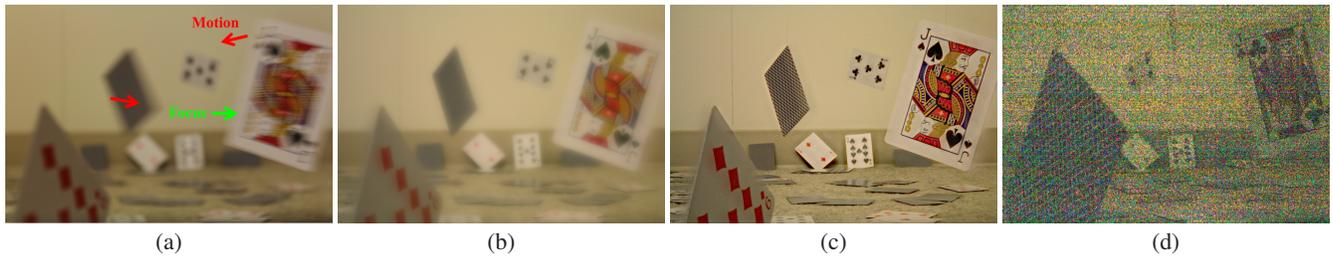


Fig. 13. Playing cards. (a) Standard camera image, focused at the jack on the right. (b) Focus sweep camera image. (c) Deconvolution result of (b). (d) Narrow aperture ($f/36$) short exposure (5msec) image (contrast-enhanced), taken at a different time instance from (a,b).

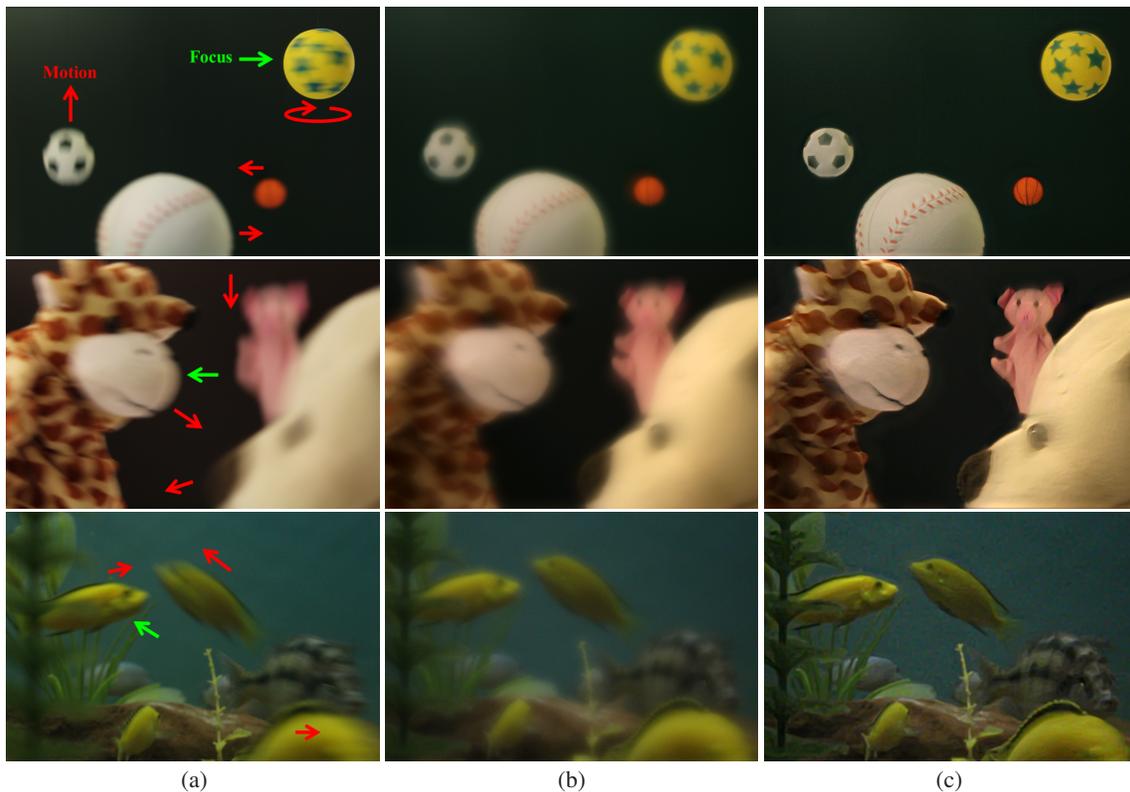


Fig. 14. Additional results. (a) Standard camera images. (b) Focus sweep camera images. (c) Deconvolution results of (b).

tion blur by stopping down the aperture and shortening exposure time, we need to do so by a factor of 40, since the maximum defocus and motion blur size our prototype focus sweep camera can handle is around 40 pixels. We reduced the exposure time by 1/40 to 5msec, but we only reduced the aperture size by 6 as the maximum f-number available for the lens was $f/36$. Even with those settings, the image got severely underexposed and extremely noisy when contrast-enhanced as shown in Fig. 13(d).

Fig. 14 shows more examples of real scenes. In the standard camera images on the left, focused objects are indicated by green arrows, meaning that other objects are defocused. Object motions are indicated by red arrows. In the top row, the yellow ball exhibits non-uniform motion blur due to rotation, but the focus sweep capture renders it to identical one because such rotation produces locally linear motion. In the middle row, the stuffed giraffe on the left is moving not only to the right but also towards the camera. However,

the focus sweep speed, going from 30cm to infinity in 200msec, is significantly faster than the object motion, and the motion can still be considered as approximately in-plane 2D, making the focus sweep capture succeed. In the bottom row, we were able to obtain a defocus and motion deblurred image of fish which were moving unpredictably in all directions, thanks to near 2D motion-invariance of the focus sweep method.

While our main focus is on low-light indoor scenes that require wide aperture and long exposure, it was easy to apply focus sweep capture to an outdoor scene as shown in Fig. 15, thanks to the portability of our prototype camera.

Failure example: Fig. 16 shows one of the typical failure modes of focus sweep, where (parts of) objects are moving faster than the assumed motion range. Textures on those parts remain smeared even after deconvolution as shown in Fig. 16(c).

7. CONCLUSION

This paper has presented 5D analysis of a time-varying light field in relation to combined defocus/motion blur that provides a framework for evaluating performance of joint defocus and motion deblurring. This has led to the conclusion that focus sweep, previously known as a depth-invariant capture method, is near-optimal both in terms of depth and 2D motion-invariance and in terms of high frequency preservation for certain combinations of depth and motion ranges. We have demonstrated removal of combined defocus and motion blur from single images using a prototype camera.

While we made a number of assumptions as stated in Sec. 1.2, we would like to note that the majority of them are common to previous work. Moreover, although we left camera shake out of scope, image stabilizing techniques can be made complementary to focus sweep: e.g., by moving the image sensor to compensate for camera shake while moving the lens to sweep focus.

Even with infinite exposure assumption, we have not reached strict 2D motion-invariance nor strictly optimal high frequency preservation for joint defocus and motion deblurring. For future work, we would like to explore other inseparable 5D kernels to address these issues.



Fig. 15. Outdoor example of pecking/walking geese. (a) Focus sweep camera image. (b) Deconvolution result of (a).

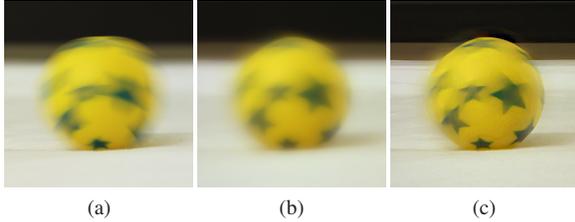


Fig. 16. Failure example. A ball is rapidly rolling. (a) Standard camera image. (b) Focus sweep camera image. (c) Deconvolution result of (b).

APPENDIX

A. FREQUENCY BOUND OVER A FOCAL AREA

We prove Eq. (10), which means that the total frequency power over a focal area is bounded. We apply a basis change to Eq. (10) so that vector \mathbf{f}_x coincides with the f_u axis, and show that

$$\iint |\hat{k}(f_x, f_y, f_u, 0, f_t)|^2 df_u df_t \leq \frac{2A^3 T}{3}. \quad (\text{A.1})$$

From [Hasinoff et al. 2009b], we have:

$$\int \left| \iiint k(x, y, u, v, t) e^{-2\pi i(f_x x + f_y y)} dx dy dv \right|^2 du \leq \frac{2A^3}{3}, \quad (\text{A.2})$$

because having an additional parameter t in k does not change the derivation in [Hasinoff et al. 2009b]. As k is non-zero only within exposure time T ,

$$\iint \left| \iiint k \cdot e^{-2\pi i(f_x x + f_y y)} dx dy dv \right|^2 dudt \leq \frac{2A^3 T}{3}. \quad (\text{A.3})$$

Since $\hat{k}(f_x, f_y, f_u, 0, f_t)$, as appearing in Eq. (A.1), can be viewed as the 2D Fourier transform of $\iiint k \cdot e^{-2\pi i(f_x x + f_y y)} dx dy dv$ with $f_v = 0$, applying Parseval's theorem to Eq. (A.3) leads to Eq. (A.1).

Note that the basis change does not affect this bound as we assume a radially symmetric (i.e., circular) aperture, and therefore we do not need an adjusting coefficient as denoted by β in the square aperture case in [Levin et al. 2009].

B. MTF OF FOCUS SWEEP CAPTURE

We derive the focus sweep MTF in Eq. (18) by taking the 5D Fourier transform of the focus sweep kernel in Eq. (17). First, we integrate the delta function over \mathbf{x} and obtain:

$$\begin{aligned} \hat{k}(\mathbf{f}_x, \mathbf{f}_u, f_t) &= \iiint \delta(\mathbf{x} - w\mathbf{t}\mathbf{u}) R(|\mathbf{u}|/A) R(t/T) \\ &\quad \cdot e^{-2\pi i(\mathbf{f}_x \cdot \mathbf{x} + \mathbf{f}_u \cdot \mathbf{u} + f_t t)} d\mathbf{x} d\mathbf{u} dt \\ &= \iint R(|\mathbf{u}|/A) R(t/T) e^{-2\pi i(\mathbf{f}_x \cdot (w\mathbf{t}\mathbf{u}) + \mathbf{f}_u \cdot \mathbf{u} + f_t t)} d\mathbf{u} dt \\ &= \iint R(|\mathbf{u}|/A) R(t/T) e^{-2\pi i((wt-s)\mathbf{f}_x \cdot \mathbf{u} + f_t t)} d\mathbf{u} dt, \end{aligned} \quad (\text{B.1})$$

where for the last line we have substituted Eq. (8) for (f_u, f_v) . Next, we integrate over \mathbf{u} . Since the 2D Fourier transform of a disc $R(|\mathbf{u}|/A)$ is a jinc: $(\pi A^2/4) \text{jinc}(\pi A |\mathbf{f}_u|)$ [Born and Wolf 1984], and Eq. (B.1) has $(wt-s)\mathbf{f}_x$ as frequency components for \mathbf{u} ,

$$\hat{k} = \int \frac{\pi A^2}{4} \text{jinc}(\pi A (wt-s) |\mathbf{f}_x|) R(t/T) e^{-2\pi i f_t t} dt. \quad (\text{B.2})$$

Finally, we integrate over t . For the moment, we omit $R(t/T)$ by assuming infinite exposure time. We rearrange Eq. (B.2) with change of variable as $t' = t - s/w$ and obtain:

$$\hat{k} = \frac{\pi A^2}{4} e^{-2\pi i f_t s/w} \int \text{jinc}(\pi A w |\mathbf{f}_x| t') e^{-2\pi i f_t t'} dt'. \quad (\text{B.3})$$

This amounts to the 1D Fourier transform of a jinc. The Fourier transform of $\text{jinc}(at)$ w.r.t. t (for $a > 0$) is given as

$$\begin{aligned} &\int_{-\infty}^{+\infty} \text{jinc}(at) e^{-2\pi i f_t t} dt \\ &= \int_{-\infty}^{+\infty} \frac{2J_1(at)}{at} (\cos(2\pi f_t t) - i \sin(2\pi f_t t)) dt \end{aligned} \quad (\text{B.4})$$

$$= \frac{4}{a} \int_0^{\infty} \frac{J_1(at) \cos(2\pi f_t t)}{t} dt \quad (\text{B.5})$$

$$= \begin{cases} \frac{4}{a} \sqrt{1 - \left(\frac{2\pi f_t}{a}\right)^2} & (2\pi |f_t| \leq a) \\ 0 & (\text{otherwise}) \end{cases}, \quad (\text{B.6})$$

where Eq. (B.4) is derived by definition. Eq. (B.5) is because the real part of Eq. (B.4) is an even function whereas the imaginary part is odd. Eq. (B.6) is due to [Watson 1922].

Applying Eq. (B.6) to Eq. (B.3) and taking the magnitude,

$$|\hat{k}|^2 = \begin{cases} \frac{A^2}{w^2|\mathbf{f}_x|^2} \left[1 - \left(\frac{2f_t}{Aw|\mathbf{f}_x|} \right)^2 \right] & \left(|f_t| \leq \frac{A}{2}w|\mathbf{f}_x| \right) \\ 0 & \text{(otherwise)} \end{cases} \quad (\text{B.7})$$

With finite exposure time, Eq. (B.3) gets convolved by the Fourier transform of $R(t/T)$, which is $T\text{sinc}(\pi T f_t)$, in the f_t axis. For tractable analysis, we continue to work on Eq. (B.7) by assuming that the exposure time T is long enough. However, we would like to note that finite exposure time introduces an additional condition for $|\hat{k}|^2$ to be non-zero. Since convolution by $\text{sinc}(\pi T f_t)$ cancels out sinusoids with higher frequencies than $T/2$, and since Eq. (B.3) has the sinusoid term $e^{-2\pi i f_t s/w}$, the additional condition is given as $|s| \leq (T/2)w$. Plugging Eq. (8) for f_t into Eq. (B.7) and rearranging leads to Eq. (18).

C. PSF OF FOCUS SWEEP CAPTURE

We derive the focus sweep PSF in Eq. (21). By computing the integrals in Eq. (20), we have:

$$\begin{aligned} \phi(\mathbf{x}) &= \frac{4}{\pi A^2 w^2} \left[-\frac{1}{t} \right]_{-T/2}^{t_0} + \frac{4}{\pi A^2 w^2} \left[-\frac{1}{t} \right]_{t_1}^{+T/2} \\ &= \frac{4}{\pi A^2 w^2} \left(\frac{1}{t_1} - \frac{1}{t_0} \right) - \frac{16}{\pi A^2 w^2 T}. \end{aligned} \quad (\text{C.1})$$

Recall that t_0 and t_1 ($> t_0$) are the two roots of $|\mathbf{x} - \mathbf{m}t| = Aw|t|/2$. They can be written explicitly by solving a quadratic equation $|\mathbf{x} - \mathbf{m}t|^2 = A^2 w^2 t^2/4$ as:

$$t_{0,1} = \frac{-\mathbf{x} \cdot \mathbf{m} \pm \sqrt{(\mathbf{x} \cdot \mathbf{m})^2 + |\mathbf{x}|^2 \left(\frac{A^2 w^2}{4} - |\mathbf{m}|^2 \right)}}{\frac{A^2 w^2}{4} - |\mathbf{m}|^2}. \quad (\text{C.2})$$

Therefore, by denoting the square-root part in Eq. (C.2) by q for brevity, we can derive as

$$\begin{aligned} \frac{1}{t_1} - \frac{1}{t_0} &= \frac{\frac{A^2 w^2}{4} - |\mathbf{m}|^2}{-\mathbf{x} \cdot \mathbf{m} + q} - \frac{\frac{A^2 w^2}{4} - |\mathbf{m}|^2}{-\mathbf{x} \cdot \mathbf{m} - q} \\ &= \left(\frac{A^2 w^2}{4} - |\mathbf{m}|^2 \right) \frac{(-\mathbf{x} \cdot \mathbf{m} - q) - (-\mathbf{x} \cdot \mathbf{m} + q)}{(\mathbf{x} \cdot \mathbf{m})^2 - q^2} \\ &= \left(\frac{A^2 w^2}{4} - |\mathbf{m}|^2 \right) \frac{-2q}{-\mathbf{x} \cdot \mathbf{m} \left(\frac{A^2 w^2}{4} - |\mathbf{m}|^2 \right)} \\ &= \frac{2}{|\mathbf{x}|^2} \sqrt{(\mathbf{x} \cdot \mathbf{m})^2 + |\mathbf{x}|^2 \left(\frac{A^2 w^2}{4} - |\mathbf{m}|^2 \right)}. \end{aligned} \quad (\text{C.3})$$

Here we note that, using θ as the angle between vectors \mathbf{x} and \mathbf{m} ,

$$\begin{aligned} (\mathbf{x} \cdot \mathbf{m})^2 - |\mathbf{x}|^2 |\mathbf{m}|^2 &= (|\mathbf{x}| |\mathbf{m}| \cos \theta)^2 - |\mathbf{x}|^2 |\mathbf{m}|^2 \\ &= -|\mathbf{x}|^2 |\mathbf{m}|^2 \sin^2 \theta. \end{aligned} \quad (\text{C.4})$$

Then we can further simplify Eq. (C.3) as:

$$\begin{aligned} \frac{1}{t_1} - \frac{1}{t_0} &= \frac{2}{|\mathbf{x}|^2} \sqrt{|\mathbf{x}|^2 \frac{A^2 w^2}{4} - |\mathbf{x}|^2 |\mathbf{m}|^2 \sin^2 \theta} \\ &= \frac{Aw}{|\mathbf{x}|} \sqrt{1 - \frac{4|\mathbf{m}|^2 \sin^2 \theta}{A^2 w^2}}. \end{aligned} \quad (\text{C.5})$$

Plugging Eq. (C.5) into Eq. (C.1) leads to Eq. (21).

ACKNOWLEDGMENTS

We thank Yusuke Iguchi, Noriko Kurachi, Matthew Hirsch, Matthew O'Toole, Douglas Lanman, and the anonymous reviewers for their constructive comments and suggestions. We also thank Cheryl Sham, Sonia Chang, Shih-Yu Sun, Jeffrey W. Kaali, Bridger Maxwell, Austin S. Lee, and Saori Bando for their help.

REFERENCES

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Trans. Graphics* 23, 3, 294–302.
- AGRAWAL, A. AND RASKAR, R. 2009. Optimal single image capture for motion deblurring. In *CVPR*. 2560–2567.
- AGRAWAL, A. AND XU, Y. 2009. Coded exposure deblurring: optimized codes for PSF estimation and invertibility. In *CVPR*. 2066–2073.
- AGRAWAL, A., XU, Y., AND RASKAR, R. 2009. Invertible motion blur in video. *ACM Trans. Graphics* 28, 3, 95:1–95:8.
- BAEK, J. 2010. Transfer efficiency and depth invariance in computational cameras. In *ICCP*. 1–8.
- BANDO, Y., CHEN, B.-Y., AND NISHITA, T. 2011. Motion deblurring from a single image using circular sensor motion. *Computer Graphics Forum* 30, 7, 1869–1878.
- BEN-EZRA, M. AND NAYAR, S. K. 2004. Motion-based motion deblurring. 26, 6, 689–698.
- BORN, M. AND WOLF, E. 1984. *Principles of Optics, sixth (corrected) edition*. Pergamon Press.
- BRACEWELL, R. N. 1965. *The Fourier transform and its applications*. McGraw-Hill.
- BURT, P. J. AND KOLCZYNSKI, R. J. 1993. Enhanced image capture through fusion. In *ICCV*. 173–182.
- CHO, T. S., LEVIN, A., DURAND, F., AND FREEMAN, W. T. 2010. Motion blur removal with orthogonal parabolic exposures. In *ICCP*. 1–8.
- COSSAIRT, O. AND NAYAR, S. 2010. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *ICCP*. 1–8.
- COSSAIRT, O., ZHOU, C., AND NAYAR, S. K. 2010. Diffusion coded photography for extended depth of field. *ACM Trans. Graphics* 29, 4, 31:1–31:10.
- DABOV, K., FOI, A., KATKOVNIK, V., AND EGIAZARIAN, K. 2008. Image restoration by sparse 3d transform-domain collaborative filtering. In *SPIE Electronic Imaging*.
- DING, Y., MCCLOSKEY, S., AND YU, J. 2010. Analysis of motion blur with a flutter shutter camera for non-linear motion. In *ECCV*. 15–30.
- DOWSKI, E. R. AND CATHEY, W. T. 1995. Extended depth of field through wave-front coding. *Applied Optics* 34, 11, 1859–1866.
- HASINOFF, S. W., KUTULAKOS, K. N., DURAND, F., AND FREEMAN, W. T. 2009a. Time-constrained photography. In *ICCV*. 333–340.
- HASINOFF, S. W., KUTULAKOS, K. N., DURAND, F., AND FREEMAN, W. T. 2009b. Time-constrained photography, supplementary material. <http://people.csail.mit.edu/hasinoff/timecon/>.
- HÄUSLER, G. 1972. A method to increase the depth of focus by two step image processing. *Optics Communications* 6, 1, 38–42.
- JOSHI, N., KANG, S. B., ZITNICK, C. L., AND SZELISKI, R. 2010. Image deblurring using inertial measurement sensors. *ACM Trans. Graphics* 29, 4, 30:1–30:8.
- KUBOTA, A. AND AIZAWA, K. 2005. Reconstructing arbitrarily focused images from two differently focused images using linear filters. *IA*, 11, 1848–1859.
- LEVIN, A. AND DURAND, F. 2010. Linear view synthesis using a dimensionality gap light field prior. In *CVPR*. 1831–1838.

- LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graphics* 26, 3, 70:1–70:9.
- LEVIN, A., FREEMAN, W. T., AND DURAND, F. 2008. Understanding camera trade-offs through a Bayesian analysis of light field projections. In *ECCV*. 88–101.
- LEVIN, A., HASINOFF, S. W., GREEN, P., DURAND, F., AND FREEMAN, W. T. 2009. 4D frequency analysis of computational cameras for depth of field extension. *ACM Trans. Graphics* 28, 3, 97:1–97:14.
- LEVIN, A., SAND, P., CHO, T. S., DURAND, F., AND FREEMAN, W. T. 2008. Motion-invariant photography. *ACM Trans. Graphics* 27, 3, 71:1–71:9.
- LEVIN, A., WEISS, Y., DURAND, F., AND FREEMAN, W. T. 2009. Understanding and evaluating blind deconvolution algorithms. In *CVPR*. 1964–1971.
- LEVOY, M. AND HANRAHAN, P. 1996. Light field rendering. In *Proc. ACM SIGGRAPH* 96. 31–42.
- NAGAHARA, H., KUTHIRUMMAL, S., ZHOU, C., AND NAYAR, S. K. 2008. Flexible depth of field photography. In *ECCV*. 60–73.
- NG, R. 2005. Fourier slice photography. *ACM Trans. Graphics* 24, 3, 735–744.
- RASKAR, R., AGRAWAL, A., AND TUMBLIN, J. 2006. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Trans. Graphics* 25, 3, 795–804.
- RAV-ACHA, A. AND PELEG, S. 2005. Two motion-blurred images are better than one. *Pattern Recog. Letters* 26, 3, 311–317.
- TAI, Y.-W., DU, H., BROWN, M. S., AND LIN, S. 2008. Image/video deblurring using a hybrid camera. In *CVPR*. 1–8.
- TAI, Y.-W., KONG, N., LIN, S., AND SHIN, S. Y. 2010. Coded exposure imaging for projective motion deblurring. In *CVPR*. 2408–2415.
- VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A., AND TUMBLIN, J. 2007. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graphics* 26, 3, 69:1–69:12.
- WATSON, G. N. 1922. *A treatise on the theory of Bessel functions*. Cambridge University Press.
- YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. *ACM Trans. Graphics* 26, 3, 1:1–1:10.
- ZHANG, L., DESHPANDE, A., AND CHEN, X. 2010. Denoising vs. deblurring: HDR imaging techniques using moving cameras. In *CVPR*. 522–529.
- ZHOU, C. AND NAYAR, S. 2009. What are good apertures for defocus deblurring? In *ICCP*.
- ZHUO, S., GUO, D., AND SIM, T. 2010. Robust flash deblurring. In *CVPR*. 2440–2447.

Received January 2012; accepted October 2012