Light-Efficient Photography

Samuel W. Hasinoff, *Member, IEEE,* and Kiriakos N. Kutulakos, *Member, IEEE*

Abstract—In this article we consider the problem of imaging a scene with a given depth of field at a given exposure level in the shortest amount of time possible. We show that by (1) collecting a sequence of photos and (2) controlling the aperture, focus and exposure time of each photo individually, we can span the given depth of field in less total time than it takes to expose a single narrower-aperture photo. Using this as a starting point, we obtain two key results. First, for lenses with continuously-variable apertures, we derive a closed-form solution for the *globally optimal* capture sequence, *i.e.*, that collects light from the specified depth of field in the most efficient way possible. Second, for lenses with discrete apertures, we derive an integer programming problem whose solution is the optimal sequence. Our results are applicable to off-the-shelf cameras and typical photography conditions, and advocate the use of dense, wide-aperture photo sequences as a light-efficient alternative to single-shot, narrow-aperture photography.

Index Terms—Computational photography, computer vision, computer graphics, shape-from-focus.

1 Introduction

Two of the most important choices when taking a photo are the photo's exposure level and its depth of field. Ideally, these choices will result in a photo whose subject is free of noise or pixel saturation [1], [2], and appears to be in focus. These choices, however, come with a severe time constraint: in order to take a photo that has both a specific exposure level and a specific depth of field, we must expose the camera's sensor for a length of time that is dictated by the lens optics. Moreover, the wider the depth of field, the longer we must wait for the sensor to reach the chosen exposure level. In practice, this makes it impossible to efficiently take sharp and well-exposed photos of a poorly-illuminated subject that spans a wide range of distances from the camera. To get a good exposure level, we must compromise something—either use a narrow depth of field (and incur defocus blur [3], [4], [5], [6]) or take a long exposure (and incur motion blur [7], [8], [9]).

In this article we seek to overcome the time constraint imposed by lens optics, by capturing a sequence of photos rather than just one. We show that if the aperture, exposure time, and focus setting of each photo is selected appropriately, we can span a given depth of field with a given exposure level *in less total time than it takes to expose a single photo* (Fig. 1), without increased imaging noise. This novel observation is based on a simple fact: even though wide apertures have a narrow depth of field (DOF), they are much more efficient than narrow apertures in gathering light from within their depth of field. Hence, even though it is not possible to span a wide DOF with a single wide-aperture photo, it is possible to span it with several of them, and do so very efficiently.

Using this observation as a starting point, we develop a general theory of *light-efficient photography* that addresses four questions: (1) under what conditions is capturing photo sequences with "synthetic" DOFs more efficient than single-shot photography? (2) How can we characterize the set of sequences that are *globally optimal* for a given DOF and exposure level, *i.e.*, whose total exposure time is the shortest possible? (3) How can we compute such sequences automatically for a specific camera, depth of field, and exposure level? (4) Finally, how do we convert the captured sequence into a single photo with the specified depth of field and exposure level?

Little is known about how to gather light efficiently from a specified DOF. To the best of our knowledge, no previous method has considered the problem of optimizing exposure time for a desired DOF and exposure level. For example, even though there has been great interest in manipulating a camera's DOF through optical [7], [10], [11], [12], [13], [14], [15], [16], [17] or computational [2], [5], [18], [19], [20], [21], [22] means, most approaches do so without regard to exposure time—they simply assume that the shutter remains open as long as necessary to reach the desired exposure level. This assumption is also used for high-dynamic range photography [2], [23], where the shutter must remain open for long periods in order to capture low-radiance regions in a scene.

In concurrent work, various computational imaging designs have been analyzed for their efficiency of capturing DOF [17], [24], however these analyses do not consider capturing multiple photos at full resolution, nor are the parameters for these designs (*e.g.*, aperture diameter) explored in detail. In contrast, here we consider capturing multiple photos, with camera settings carefully chosen to minimize total exposure time for the desired DOF and exposure level.

Most recently, capturing multiple photos has been

- *S. W. Hasinoff is with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02140. Email: hasinoff@csail.mit.edu.*
- *K. N. Kutulakos is with the Department of Computer Science, University of Toronto, Canada M5S 3G4. Email: kyros@cs.toronto.edu.*

Manuscript received Jan 1, 2009.

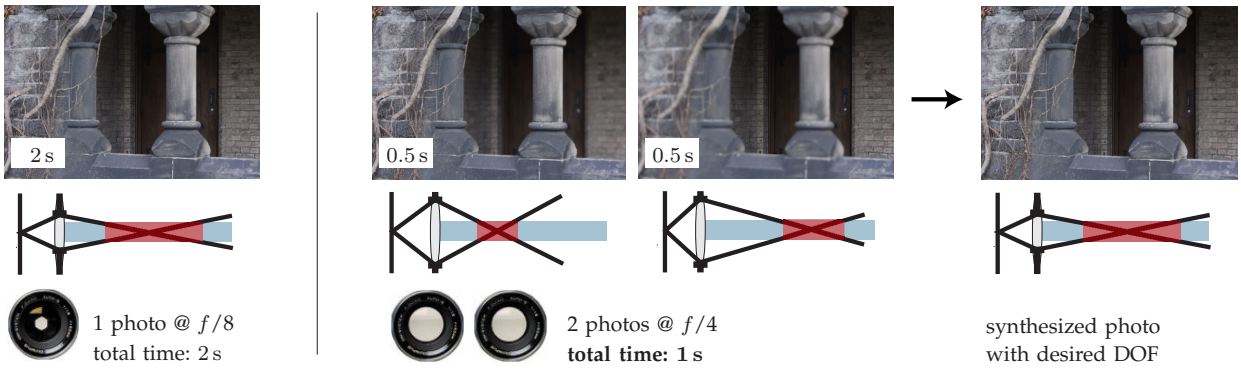


Fig. 1. *Left:* Traditional single-shot photography. The desired depth of field is shown in red. *Right:* Light-efficient photography. Two wide-aperture photos span the same DOF as a single-shot narrow-aperture photo. Each wide-aperture photo requires $1/4$ the time to reach the exposure level of the narrow-aperture photo, resulting in a $2\times$ net speedup for the total exposure time.

shown to be a generally efficient strategy for capturing DOF, even for computational cameras specifically designed for improve single-photo performance [25]. In particular, one can achieve significant efficiency gains by using the optimal number of photos to balance between noise and worst-case defocus. This article describes gains in light efficiency that are strictly orthogonal to the noise-defocus tradeoff—we only consider capture sequences that fully span the DOF, and since we hold exposure level for each photo fixed, there is no need to model noise explicitly.

Since shorter total exposure times reduce motion blur, our work can also be thought of as complementary to recent *synthetic shutter* approaches whose goal is to reduce such blur. Instead of controlling aperture and focus, these techniques divide a given exposure interval into several shorter ones, with the same total exposure (*e.g.*, N photos, each with $1/N$ the exposure time [9]; two photos, one with long and one with short exposure [8]; or one photo where the shutter opens and closes intermittently during the exposure [7]). These techniques do not increase light efficiency and do not rely on camera controls other than the shutter. As such, they can be readily combined with our work, to confer the advantages of both methods.

The final step in light-efficient photography involves merging the captured photos to create a new one (Fig. 1). As such, our work is related to the well-known technique of extended-depth-of-field imaging. This technique creates a new photo whose DOF is the union of DOFs in a sequence, and has found wide use in microscopy [22], macro photography [26], [27] and photo manipulation [26], [27]. Current work on the subject concentrates on the problems of image merging [26], [28] and 3D reconstruction [22], and indeed we use an existing implementation [26] for our own merging step. However, the problem of how to best acquire such sequences remains open. In particular, the idea of controlling aperture and focus to optimize total exposure time has not been explored.

Our work offers four contributions over the state of the art. First, we develop a theory that leads to provably-efficient light-gathering strategies, and applies both to off-the-shelf cameras and to advanced camera designs [7], [9] under typical photography conditions. Second, from a practical standpoint, our analysis shows that the optimal (or near-optimal) strategies are very simple: for example, in the continuous case, a strategy that uses the widest-possible aperture for all photos is either globally optimal or it is very close to it (in a quantifiable sense). Third, our experiments with real scenes suggest that it is possible to compute good-quality synthesized photos using readily-available algorithms. Fourth, we show that despite requiring less total exposure time than a single narrow-aperture shot, light-efficient photography provides more information about the scene (*i.e.*, depth) and allows post-capture control of aperture and focus.

2 The Exposure Time *vs.* Depth of Field Tradeoff

The *exposure level* of a photo is the total radiant energy integrated by the camera's entire sensor while the shutter is open. The exposure level can influence significantly the quality of a captured photo because when there is no saturation or thermal noise, a pixel's signal-to-noise ratio (SNR) always increases with higher exposure levels.[1] For this reason, most modern cameras can automate the task of choosing an exposure level that provides high SNR for most pixels and causes little or no saturation.

Lens-based camera systems provide only two ways to control exposure level—the diameter of their aperture and the exposure time. We assume that all light passing through the aperture will reach the sensor plane, and that the average irradiance measured over this aperture is independent of the aperture's diameter. In this case,

1. Thermal effects, such as dark-current noise, become significant only for exposure times longer than a few seconds [1].

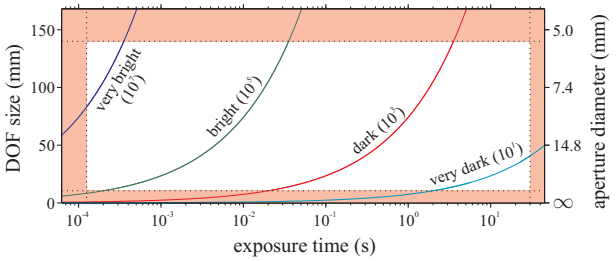


Fig. 2. Exposure time *vs*. depth of field tradeoff. Each curve represents all pairs (τ, D) for which $\tau D^2 = L^*$ in a specific scene. Shaded zones correspond to pairs outside the camera limits (valid settings were $\tau \in [1/8000\,\mathrm{s}, 30\,\mathrm{s}]$ and $D \in [f/16, f/1.2]$ with $f = 85\,\mathrm{mm}$). Different curves represent scenes with different average radiance (relative magnitude in brackets).

the exposure level L satisfies

$$L \;\propto\; \tau\, D^2 \quad , \tag{1}$$

where τ is the exposure time and D is the aperture diameter.[2]

Now suppose that we have chosen a desired exposure level L^*. How can we capture a photo at this exposure level? Eq. (1) suggests that there are only two general strategies for doing this—either choose a long exposure time and a small aperture diameter, or choose a large aperture diameter and a short exposure time. Unfortunately, both strategies have important side-effects: increasing exposure time can introduce motion blur when we photograph moving scenes [8], [9]; opening the lens aperture, on the other hand, affects the photo's *depth of field (DOF)*, *i.e.*, the range of distances where scene points do not appear out of focus. These side-effects lead to an important tradeoff between a photo's exposure time and its depth of field (Fig. 2):

> **Exposure Time *vs*. Depth of field Tradeoff:** *We can either achieve a desired exposure level L^* with short exposure times and a narrow DOF, or with long exposure times and a wide DOF.*

In practice, the exposure time *vs*. DOF tradeoff limits the range of scenes that can be photographed at a given exposure level (Fig. 2). This range depends on scene radiance, the physical limits of the camera (*i.e.*, range of possible apertures and shutter speeds), as well as subjective factors (*i.e.*, acceptable levels of motion blur and defocus blur).

Our goal is to "break" this tradeoff by seeking novel photo acquisition strategies that capture a given depth of field at the desired exposure level L^* much faster than traditional optics would predict. We briefly describe

2. More precisely, the exposure level L is proportional to the solid angle subtended by the aperture; even as $D \to \infty$ one is limited by the finite radiant power in the scene. In practice, Eq. (1) is a good approximation, since the largest apertures available for consumer photography do not exceed $0.48\,\mathrm{sr}$ ($7.7\,\%$ of the hemisphere).

thin lens law	$\frac{1}{v} + \frac{1}{d} = \frac{1}{f}$	(2)
focus setting for distance d	$v = \frac{df}{d-f}$	(3)
blur diameter for out-of-focus distance d'	$\sigma = D\,\frac{\lvert d'-d\rvert}{d'}\,\frac{v}{d}$	(4)
aperture diameter whose DOF is interval $[\alpha, \beta]$	$D = c\,\frac{\beta+\alpha}{\beta-\alpha}$	(5)
focus setting whose DOF is interval $[\alpha, \beta]$	$v = \frac{2\,\alpha\,\beta}{\alpha+\beta}$	(6)
DOF endpoints for aperture diameter D and focus v	$\alpha, \beta = \frac{Dv}{D \pm c}$	(7)

TABLE 1
Basic equations governing focus and depth of field for the thin-lens model (Fig. 3).

below the basic geometry and relations governing a photo's depth of field, as they are particularly important for our analysis.

2.1 Depth of Field Geometry

We assume that focus and defocus obey the standard thin lens model [3], [29]. This model relates three positive quantities (Eq. (2) in Table 1): the focus setting v, defined as the distance from the sensor plane to the lens; the distance d from the lens to the in-focus scene plane; and the focal length f, representing the "focusing power" of the lens.

Apart from the idealized pinhole, all apertures induce spatially-varying amounts of defocus for points in the scene (Fig. 3a). If the lens focus setting is v, all points at distance d from the lens will be in-focus. A scene point at distance $d' \neq d$, however, will be defocused: its image will be a circle on the sensor plane whose diameter σ is called the *blur diameter*. For any given distance d, the thin-lens model tells us exactly what focus setting we should use to bring the plane at distance d into focus, and what the blur diameter will be for points away from this plane (Eqs. (3) and (4), respectively).

For a given aperture and focus setting, the *depth of field* is the interval of distances in the scene $[d_1, d_2]$, whose blur diameter is below a maximum acceptable size c (Fig. 3b).

Since every distance in the scene corresponds to a unique focus setting (Eq. (3)), every DOF can also be expressed as an interval $[\alpha, \beta]$ in the space of focus settings (Fig. 3c). This alternate DOF representation gives us especially simple relations for the aperture and focus

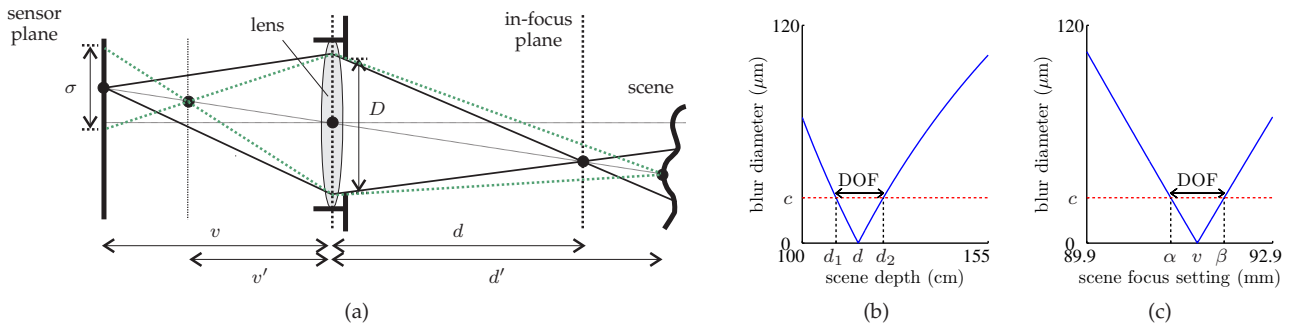


Fig. 3. (a) Blur geometry for a thin lens. (b) Blur diameter as a function of distance to a scene point. The plot is for a lens with focal length $f = 85\,\mathrm{mm}$, focused at $117\,\mathrm{cm}$ with an aperture diameter of $5.3\,\mathrm{mm}$ (*i.e.*, an f/16 aperture in photography terminology). (c) Blur diameter and DOF represented in the space of focus settings.

setting that produce a given DOF (Eqs. (5) and (6)) and, conversely, for the DOF produced by a given aperture and focus setting (Eq. (7)). We adopt this DOF representation for the rest of our analysis.

A key property of the depth of field is that it shrinks when the aperture diameter increases: from Eq. (4) it follows that for a given out-of-focus distance, larger apertures always produce larger blur diameters. This equation is the root cause of the exposure time *vs*. depth of field tradeoff.

3 The Synthetic DOF Advantage

Suppose that we want to capture a single photo with a specific exposure level L^* *and* a specific depth of field $[\alpha, \beta]$. How quickly can we capture this photo? The basic DOF geometry of Sec. 2.1 tells us we have no choice: there is only one aperture diameter that can span the given depth of field (Eq. (5)), and only one exposure time that can achieve a given exposure level with that diameter (Eq. (1)). This exposure time is[3]

$$\tau^{one} \;=\; L^* \cdot \left(\frac{\beta - \alpha}{c\,(\beta + \alpha)}\right)^2 \quad . \tag{8}$$

The key idea of our approach is that while lens optics do not allow us to reduce this time without compromising the DOF or the exposure level, we *can* reduce it by taking more photos. This is based on a simple observation that takes advantage of the different rates at which exposure time and DOF change: if we increase the aperture diameter and adjust exposure time to maintain a constant exposure level, its DOF shrinks (at a rate of about $1/D$), but the exposure time shrinks much faster (at a rate of $1/D^2$). This opens the possibility of "breaking" the exposure time *vs*. DOF tradeoff by capturing a sequence of photos that jointly span the DOF in less total time than τ^{one} (Fig. 1).

Our goal is to study this idea in its full generality, by finding capture strategies that are provably time-optimal. We therefore start from first principles, by formally defining the notion of a *capture sequence* and of its *synthetic depth of field*:

Definition 1 (Photo Tuple)**.** *A tuple* $\langle\ D,\ \tau,\ v\ \rangle$ *that specifies a photo's aperture diameter, exposure time, and focus setting, respectively.*

Definition 2 (Capture Sequence)**.** *A finite ordered sequence of photo tuples.*

Definition 3 (Synthetic Depth of Field)**.** *The union of DOFs of all photo tuples in a capture sequence.*

We will use two efficiency measures: the *total exposure time* of a sequence is the sum of the exposure times of all its photos; the *total capture time*, on the other hand, is the actual time it takes to capture the photos with a specific camera. This time is equal to the total exposure time, plus any overhead caused by camera internals (computational and mechanical). We now consider the following general problem:

> **Light-Efficient Photography**: *Given a set* $\mathcal{D}$ *of available aperture diameters, construct a capture sequence such that: (1) its synthetic DOF is equal to* $[\alpha, \beta]$*; (2) all its photos have exposure level* L^**; (3) the total exposure time (or capture time) is smaller than* τ^{one}*; and (4) this time is a global minimum over all finite capture sequences.*

Intuitively, whenever such a capture sequence exists, it can be thought of as being optimally more efficient than single-shot photography in gathering light. Below we analyze three instances of the light-efficient photography problem. In all cases, we assume that the exposure level L^*, depth of field $[\alpha, \beta]$, and aperture set $\mathcal{D}$ are known and fixed.

3.1 Noise and Quantization Properties

Because we hold exposure level constant our analysis already accounts for noise implicitly. This follows from the fact that most sources of noise (photon noise, sensor noise, and quantization noise) depend only on exposure level. The only exception is thermal or dark-current noise, which increases with exposure time [1]. Therefore,

3. The apertures and exposure times of real cameras span finite intervals and, in many cases, take discrete values. Hence, in practice, Eq. (8) holds only approximately.

all photos we consider have similar noise properties, except for thermal noise, which will be lower for light-efficient sequences because they involve shorter exposure times.

Another consequence of holding exposure level constant is that all photos we consider have the same dynamic range, since all photos are exposed to the same brightness, and have similar noise properties for quantization. Standard techniques for HDR imaging [23], [30] are complementary to our analysis, since we can apply light-efficient capture for each exposure level in an HDR sequence.

4 Theory of Light-Efficient Photography

4.1 Continuously-Variable Aperture Diameters

Many manual-focus SLR lenses as well as programmable-aperture systems [14] allow their aperture diameter to vary continuously within some interval $\mathcal{D} = [D_{min}, D_{max}]$. In this case, we prove that the optimal capture sequence has an especially simple form—it is unique, it uses the same aperture diameter for all tuples, and this diameter is either the maximum possible or a diameter close to that maximum.

More specifically, consider the following special class of capture sequences:

Definition 4 (Sequences with Sequential DOFs)**.** *A capture sequence has sequential DOFs if for every pair of adjacent photo tuples, the right endpoint of the first tuple's DOF is the left endpoint of the second.*

The following theorem states that the solution to the light-efficient photography problem is a specific sequence from this class:

Theorem 1 (Optimal Capture Sequence for Continuous Apertures)**.** *(1) If the DOF endpoints satisfy $\beta < (7+4\sqrt{3})\alpha$, the sequence that globally minimizes total exposure time is a sequence with sequential DOFs whose tuples all have the same aperture. (2) Define $D(k)$ and n as follows:*

$$D(k) = c\,\frac{\sqrt[k]{\beta}+\sqrt[k]{\alpha}}{\sqrt[k]{\beta}-\sqrt[k]{\alpha}}\quad,\quad n = \left\lfloor \frac{\log\frac{\alpha}{\beta}}{\log\left(\frac{D_{max}-c}{D_{max}+c}\right)} \right\rfloor \quad . \tag{9}$$

The aperture diameter D^* and length n^* of the optimal sequence is given by

$$(D^*, n^*) = \begin{cases} (D(n),\ n) & \text{if } D(n) > \sqrt{\frac{n}{n+1}}\,D_{max} \\ (D_{max},\ n+1) & \text{otherwise.} \end{cases} \quad , \tag{10}$$

Theorem 1 specifies the optimal sequence indirectly, via a "recipe" for calculating the optimal length and the optimal aperture diameter (Eqs. (9) and (10)). Informally, this calculation involves three steps. The first step defines the quantity $D(k)$; in our proof of Theorem 1 (see Appendix A), we show that this quantity represents the only aperture diameter that can be used to "tile" the interval $[\alpha, \beta]$ with exactly k photo tuples of the same aperture. The second step defines the quantity n; in our proof, we show that this represents the largest number of photos we can use to tile the interval $[\alpha, \beta]$ with photo tuples of the same aperture. The third step involves choosing between two "candidates" for the optimal solution—one with n tuples and one with $n+1$.

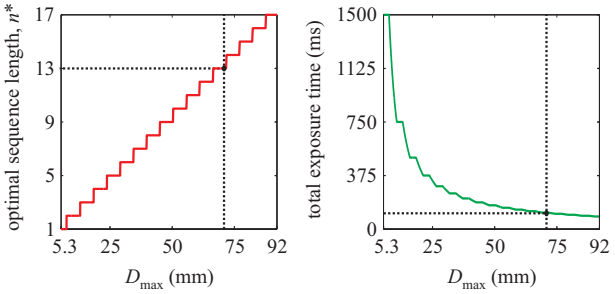


Fig. 4. Optimal light-efficient photography of a "dark" subject using a lens with a continuously-variable aperture ($f = 85\,\mathrm{mm}$). To cover the DOF ($[110\,\mathrm{cm}, 124\,\mathrm{cm}]$) in a single photo, we need a long $1.5\,\mathrm{s}$ exposure to achieve the desired exposure level. Together, the two graphs specify the optimal capture sequences when the aperture diameter is restricted to the range $[f/16, D_{\max}]$; for each value of $D_{\max}$, Theorem 1 gives a unique optimal sequence. As $D_{\max}$ increases, the number of photos (left) in the optimal sequence increases, and the total exposure time (right) of the optimal sequence falls dramatically. The dashed lines show that when the maximum aperture is $f/1.2$ ($71\,\mathrm{mm}$), the optimal synthetic DOF consists of $n^* = 13$ photos (corresponding to $D^* = 69\,\mathrm{mm}$), which provides a speedup of $13\times$ over single-shot photography.

Theorem 1 makes explicit the somewhat counter-intuitive fact that the most light-efficient way to span a given DOF $[\alpha, \beta]$ is to use images whose DOFs are very narrow. This fact applies broadly, because Theorem 1's inequality condition for α and β is satisfied for all lenses for consumer photography that we are aware of (*e.g.*, see [31]).[4] See Figs. 4 and 5 for an application of this theorem to a practical example.

Note that Theorem 1 specifies the number of tuples in the optimal sequence and their aperture diameter, but does not specify their exposure times or focus settings. The following lemma shows that specifying those quantities is not necessary because they are determined uniquely. Importantly, Lemma 1 gives us a recursive formula for computing the exposure time and focus setting of each tuple in the sequence:

Lemma 1 (Construction of Sequences with Sequential DOFs)**.** *Given a left DOF endpoint α, every ordered sequence*

4. To violate the condition, a lens must have an extremely short minimum focusing distance of under $1.077f$. Even for macro lenses that state a minimum focusing distance of 0 the condition is typically not violated, because this distance is measured relative to the front-most glass surface, while the effective lens center is deeper inside.

$D_1, \ldots, D_n$ of aperture diameters defines a unique capture sequence with sequential DOFs whose n tuples are

$$\left\langle \; D_i \,,\; \frac{L^*}{{D_i}^2} \,,\; \frac{D_i + c}{D_i}\,\alpha_i \;\right\rangle \quad , \quad i \;=\; 1, \ldots, n \quad , \tag{11}$$

with α_i given by the following recursive relation:

$$\alpha_i \;=\; \begin{cases} \alpha & \text{if } i = 1 \quad , \\ \frac{D_i + c}{D_i - c}\;\alpha_{i-1} & \text{otherwise.} \end{cases} \tag{12}$$

4.2 Discrete Aperture Diameters

Modern auto-focus lenses often restrict the aperture diameter to a discrete set of choices, $\mathcal{D} = \{D_1, \ldots, D_m\}$. These diameters form a geometric progression, spaced so that the aperture area doubles every two or three steps. Unlike the continuous case, the optimal capture sequence is not unique and may contain several distinct aperture diameters. To find an optimal sequence, we reduce the problem to integer linear programming [32]:

Theorem 2 (Optimal Capture Sequence for Discrete Apertures)**.** *There exists an optimal capture sequence with sequential DOFs whose tuples have a non-decreasing sequence of aperture diameters. Moreover, if n_i is the number of times diameter D_i appears in the sequence, the multiplicities $n_1, \ldots, n_m$ satisfy the integer program*

$$\textit{minimize} \quad \textstyle\sum_{i=1}^{m} n_i \; \frac{L^*}{{D_i}^2} \tag{13}$$

$$\textit{subject to} \quad \textstyle\sum_{i=1}^{m} n_i \log \frac{D_i - c}{D_i + c} \;\leq\; \log \frac{\alpha}{\beta} \tag{14}$$

$$n_i \geq 0 \tag{15}$$

$$n_i \;\textit{integer} \quad . \tag{16}$$

See Appendix A for a proof. As with Theorem 1, Theorem 2 does not specify the focus settings in the optimal capture sequence. We use Lemma 1 for this purpose, which explicitly constructs it from the apertures and their multiplicities.

While it is not possible to obtain a closed-form expression for the optimal sequence, solving the integer program for any desired DOF is straightforward. We use a simple branch-and-bound method based on successive relaxations to linear programming [32]. Moreover, since the optimal sequence depends only on the relative DOF size $\frac{\alpha}{\beta}$, we pre-compute it exactly for all relative sizes and store it in a lookup table (Fig. 6a).

4.3 Discrete Aperture Diameters Plus Overhead

Our treatment of discrete apertures generalizes easily to account for camera overhead. We model overhead as a per-shot constant, τ^{over}, that expresses the minimum delay between the time that the shutter closes and the time it is ready to open again for the next photo. To find the optimal sequence, we modify the objective function of Theorem 2 so that it measures for total capture time rather than total exposure time:

$$\textit{minimize} \quad \textstyle\sum_{i=1}^{m} n_i \; [\; \tau^{over} \;+\; \frac{L^*}{{D_i}^2} \;] \quad . \tag{17}$$

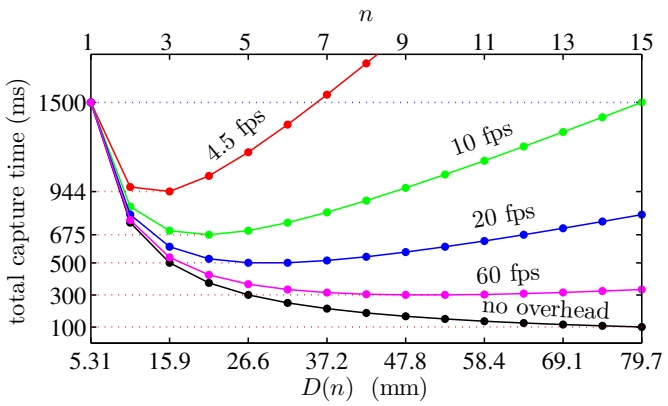


Fig. 5. The effect of camera overhead for various frame-per-second (fps) rates. Each point in the graphs represents the total capture time of a sequence that spans the DOF and whose photos all use the diameter $D(n)$ indicated. Even though overhead reduces the efficiency of long sequences, capturing synthetic DOFs is faster than single-shot photography even for low-fps rates; for current off-the-shelf cameras with high-fps rates, the speedups can be very significant.

Clearly, a non-negligible overhead penalizes long capture sequences and reduces the synthetic DOF advantage. Despite this, Fig. 6b shows that synthetic DOFs offer significant speedups even for current off-the-shelf cameras. These speedups will be amplified further as camera manufacturers continue to improve their frames-per-second rate.

5 Depth of Field Compositing and Resynthesis

While each light-efficient sequence captures a synthetic DOF, merging the input photos into a single photo with the desired DOF requires further processing. To achieve this, we use an existing depth-from-focus and compositing technique [26], and propose a simple extension that allows us to reshape the DOF, to synthesize photos with new camera settings as well.

DOF Compositing. To reproduce the desired DOF, we adopted the Photomontage method [26] with default parameters, which is based on maximizing a simple "focus measure" that evaluates local contrast according to the difference-of-Gaussians filter. In this method, each pixel in the composite has a label that indicates the input photo for which the pixel is in-focus. The pixel labels are then optimized using a Markov random field network that is biased toward piecewise smoothness [33]. Importantly, the resulting composite is computed as a blend of photos in the gradient domain, which reduces artifacts at label boundaries, including those due to misregistration.

3D Reconstruction. The DOF compositing operation produces a coarse depth map as an intermediate step.

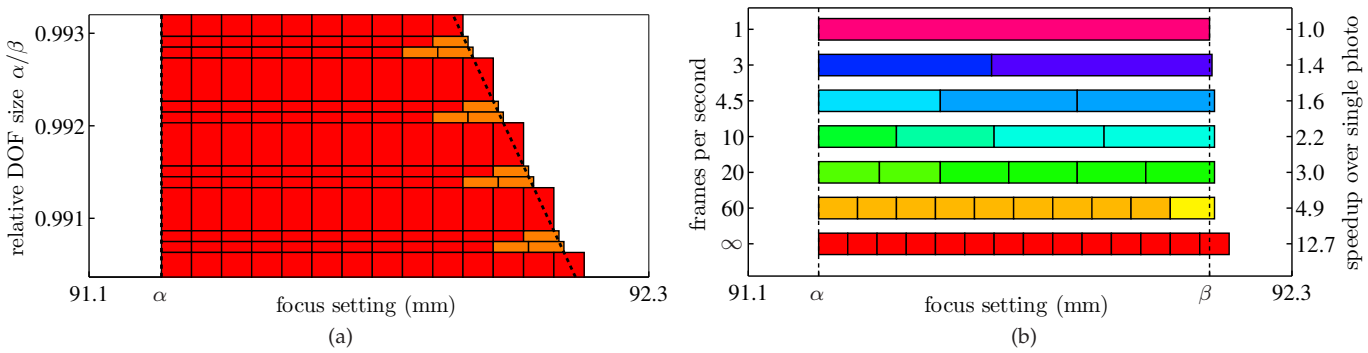


Fig. 6. Optimal light-efficient photography with discrete apertures, shown for a Canon EF85mm 1.2L lens (23 apertures, illustrated in different colors). (a) For a depth of field whose left endpoint is α, we show optimal capture sequences for a range of relative DOF sizes $\frac{\alpha}{\beta}$. These sequences can be read horizontally, with subintervals corresponding to the apertures determined by Theorem 2. The diagonal dotted line shows the minimum DOF to be spanned. (b) Visualizing the optimal capture sequence for the DOF $[\alpha, \beta]$ for differently levels of camera overhead. Note that as the overhead increases (*i.e.*, lower frames per second rates), the optimal sequence involves fewer photos with larger DOFs (*i.e.*, smaller apertures).

This is because labels correspond to input photos, and each input photo defines an in-focus depth according to the focus setting with which it was captured. As our results show, this coarse depth map is sufficient for good-quality resynthesis (Figs. 7–9). For greater depth accuracy, particularly when the capture sequence consists of only a few photos, we can apply more sophisticated depth-from-defocus analysis, *e.g.*, [6], that reconstructs depth by modeling how defocus varies over the whole sequence.

Synthesizing Photos for Novel Focus Settings and Aperture Diameters. To synthesize novel photos with different camera settings, we generalize DOF compositing and take advantage of the different levels of defocus throughout the capture sequence. Intuitively, rather than selecting pixels at in-focus depths from the input sequence, we use the recovered depth map to select pixels with appropriate levels of defocus according to the desired synthetic camera setting.

We proceed in four basic steps. First, given a specific focus and aperture setting, we use Eq. (4) and the coarse depth map to assign a blur diameter to each pixel in the final composite. Second, we use Eq. (4) again to determine, for each pixel in the composite, the input photo whose blur diameter that corresponds to the pixel's depth matches most closely. Third, for each depth layer, we synthesize a photo with the novel focus and aperture setting, under the assumption that the entire scene is at that depth. To do this, we use the blur diameter for this depth to define an interpolation between two of the input photos. Fourth, we generate the final composite by merging all these synthesized images into one photo using the same gradient-domain blending as in DOF compositing, and using the same depth labels.[5]

5. Note that given a blur diameter there are two possible depths that correspond to it, one on each side of the focus plane (Fig. 3b). We resolve this by choosing the matching input photo whose focus setting is closest to the synthetic focus setting.

To interpolate between the input photos we currently use simple linear cross-fading, which we found to be adequate when the DOF is sampled densely enough (*i.e.*, with 5 or more images). For greater accuracy when fewer input images are available, more computationally intensive frequency-based interpolation [19] could also be used. Note that blur diameter can also be extrapolated, by synthetically applying the required additional blur. There are limitations, however, to this extrapolation. While extrapolated wider apertures can model the resulting increase in defocus, we have limited ability to reduce the DOF in sharp regions of an input image. That would entail a form of super-resolution, decomposing the in-focus region into finer depth gradations [34].

6 Results and Discussion

To evaluate our technique we show results and timings for experiments performed with two different cameras—a high-end digital SLR and a compact digital camera. All photos were captured at the same exposure level for each experiment, as determined by the camera's built-in light meter. In each case, we captured (1) a narrow-aperture photo, which serves as ground truth, and (2) the optimal light-efficient capture sequence for the equivalent DOF.[6]

The digital SLR we used was the Canon EOS-1Ds Mark II (HAMSTER and FACE datasets) with a wide-angle fixed focal length lens (Canon EF85mm 1.2L). We operated the camera at its highest resolution of $16\,\mathrm{MP}$ (4992×3328) in RAW mode. To define the desired DOF, we captured a narrow-aperture photo using an aperture of $f/16$. For both datasets, the DOF we used was $[98\,\mathrm{cm}, 108\,\mathrm{cm}]$, near the minimum focusing distance of the lens, and the narrow-aperture photo required an exposure time of $800\,\mathrm{ms}$.

6. For additional results and videos, see http://www.csail.mit.edu/~hasinoff/lightefficient/.

(a) photo 3 of 14 @ $f/1.2$ exposure time: 5 ms

(b) synthetic DOF composite **total exposure time: 70 ms**

(c) 1 photo @ $f/16$ exposure time: 800 ms

(d) coarse depth map, labels from DOF composite

(e) synthesized $f/2.8$ aperture, same focus setting as (a)

(f) synthesized $f/2.8$ aperture, refocused further

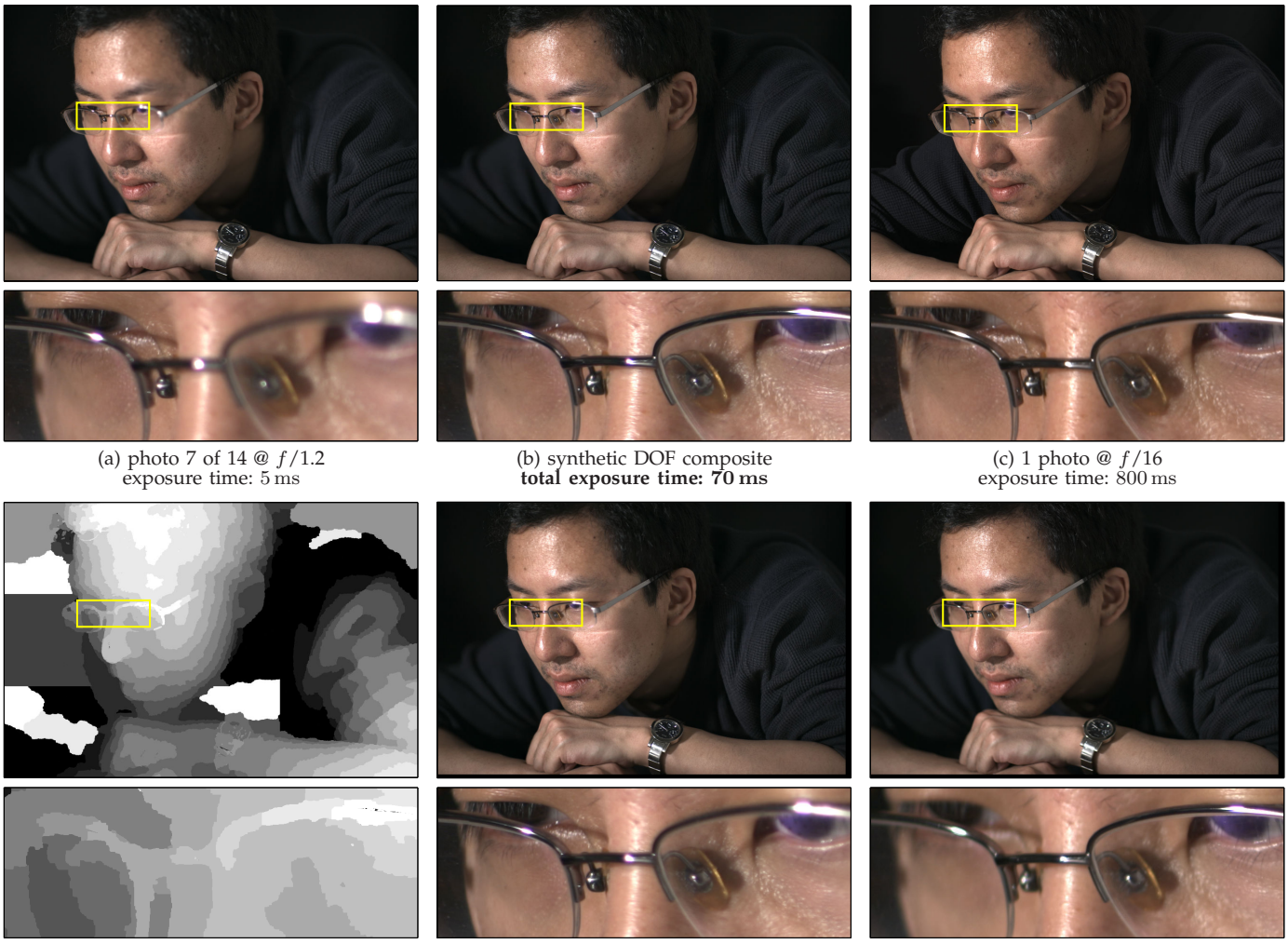
Fig. 7. HAMSTER dataset. Light efficient photography timings and synthesis, for several real scenes, captured using a compact digital camera and a digital SLR. (a) Sample wide-aperture photo from the synthetic DOF sequence. (b) DOF composites synthesized from this sequence. (c) Narrow-aperture photos spanning an equivalent DOF, but with much longer exposure time. (d) Coarse depth map, computed from the labeling we used to compute (b). (e) Synthetically changing aperture size, focused at the same setting as (a). (f) Synthetically changing focus setting as well, for the same synthetic aperture as (e).

The compact digital camera we used was the Canon S3 IS, at its widest-angle zoom setting with a focal length of 6 mm (SIMPSONS dataset). We used the camera to record 2 MP (1600×1200 pixels) JPEG images. To define the desired DOF, we captured a photo with the narrowest aperture of $f/8$. The DOF we used was $[30\,\text{cm}, 70\,\text{cm}]$, and the narrow-aperture photo required an exposure time of 500 ms.

- **HAMSTER dataset** Still life of a hamster figurine (16 cm tall), posed on a table with various other small objects (Fig. 7). The DOF covers the hamster and all the small objects, but not the background composed of cardboard packing material.
- **FACE dataset** Studio-style 2/3 facial portrait of a subject wearing glasses, resting his chin on his hands (Fig. 8). The DOF extends over the subject's face and the left side of the body closest the camera.
- **SIMPSONS dataset** Near-macro sequence of a messy desk (close objects magnified 1:5), covered in books, papers, and tea paraphernalia, on top of which several plastic figurines have been arranged (Fig. 9). The DOF extends from red tea canister to the pale green book in the background.

Implementation details. To compensate for the distortions that occur with changes in focus setting, we align the photos according to a one-time calibration method that fits a simplified radial magnification model to focus setting [35].

We determined the maximum acceptable blur diameter, c, for each camera by qualitatively assessing focus using a resolution chart. The values we used, 25 µm (3.5



(a) photo 7 of 14 @ $f/1.2$ exposure time: $5\,\mathrm{ms}$

(b) synthetic DOF composite **total exposure time: 70 ms**

(c) 1 photo @ $f/16$ exposure time: $800\,\mathrm{ms}$

(d) coarse depth map, labels from DOF composite

(e) synthesized $f/2.8$ aperture, same focus setting as (a)

(f) synthesized $f/2.8$ aperture, refocused closer

Fig. 8. FACE dataset. Light efficient photography timings and synthesis, for several real scenes, captured using a compact digital camera and a digital SLR. (a) Sample wide-aperture photo from the synthetic DOF sequence. (b) DOF composites synthesized from this sequence. (c) Narrow-aperture photos spanning an equivalent DOF, but with much longer exposure time. (d) Coarse depth map, computed from the labeling we used to compute (b). Tile-based processing leads to depth artifacts in low-texture regions, but these do not affect the quality of resynthesis. (e) Synthetically changing aperture size, focused at the same setting as (a). (f) Synthetically changing focus setting as well, for the same synthetic aperture as (e).

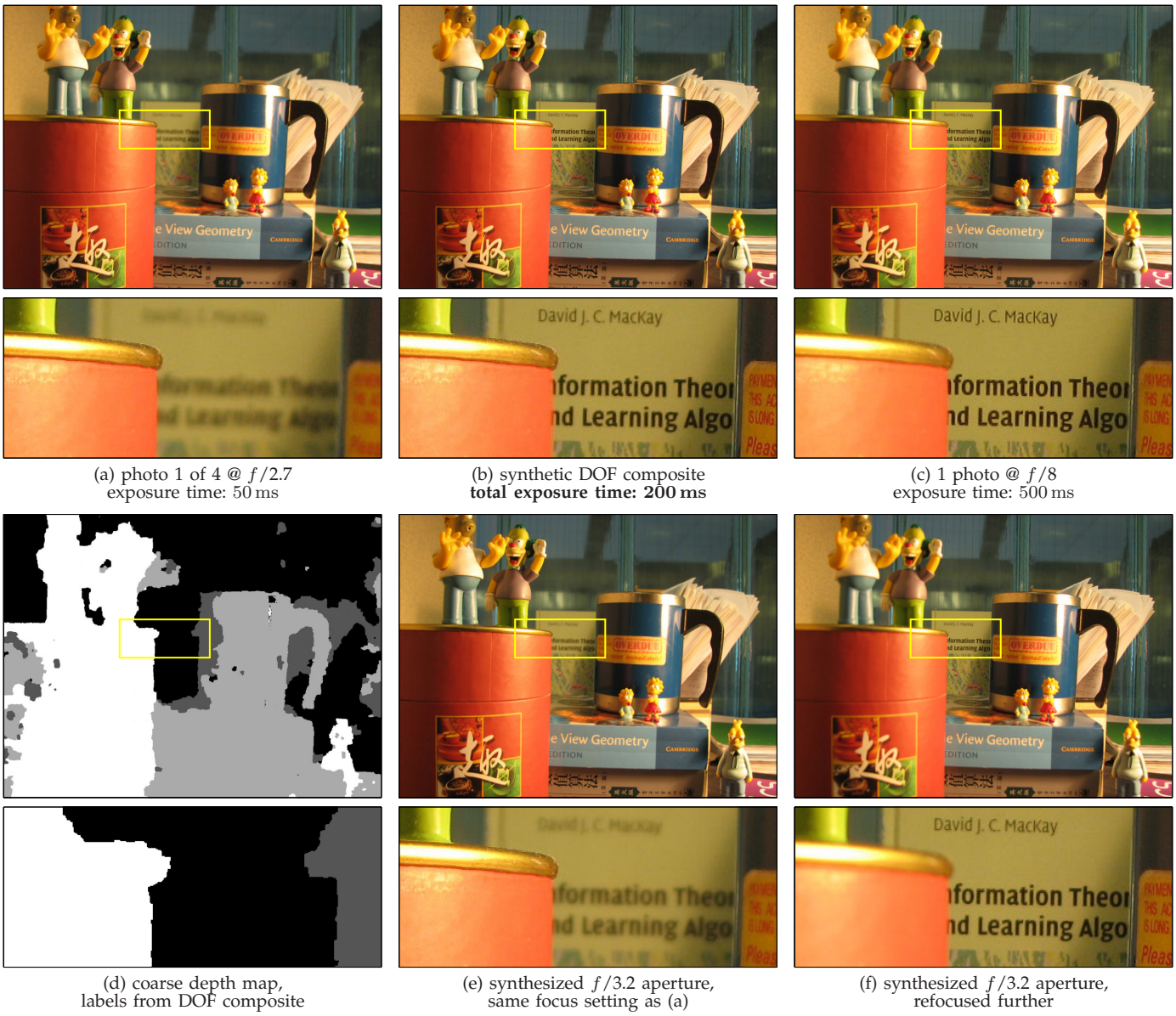
pixels) and $5\,\mu\mathrm{m}$ (1.4 pixels) for the digital SLR and compact camera respectively, agree with the standard values cited for sensors of those sizes [29].

To process the $16\,\mathrm{MP}$ synthetic DOFs captured with the digital SLR more efficiently, we divided the input photos into tiles of approximately $2\,\mathrm{MP}$ each, overlapping their neighbors by 100 pixels, so that all computation could take place in main memory. As Fig. 8d illustrates, merging per-tile results that were computed independently can introduce depth artifacts along tile boundaries. In practice, these artifacts do not pose problems for resynthesis, because they are restricted to textureless regions, whose realistic resynthesis does not depend on accurate depth.

Timing comparisons and optimal capture sequences. To determine the optimal capture sequences, we assumed zero camera overhead and applied Theorem 2 for the chosen DOF and exposure level, according to the specifications of each camera and lens. The optimal sequences involved spanning the DOF using the largest aperture in both cases. As Figs. 7–9 show, these sequences led to significant speedups in exposure time—11.9× and 2.5× for our digital SLR and compact digital camera respectively.[7]

For a hypothetical camera overhead of $17\,\mathrm{ms}$ (corresponding to a $60\,\mathrm{fps}$ camera), the optimal capture sequence satisfies Eq. (17), which changes the optimal

7. By comparison, the effective speedup provided by optical image stabilization for hand-held photography is 8–16×, when the scene is static. Gains from light efficient photography are complementary to such improvements in lens design.



(a) photo 1 of 4 @ $f/2.7$ exposure time: 50 ms

(b) synthetic DOF composite **total exposure time: 200 ms**

(c) 1 photo @ $f/8$ exposure time: 500 ms

(d) coarse depth map, labels from DOF composite

(e) synthesized $f/3.2$ aperture, same focus setting as (a)

(f) synthesized $f/3.2$ aperture, refocused further

Fig. 9. SIMPSONS dataset. Light efficient photography timings and synthesis, for several real scenes, captured using a compact digital camera and a digital SLR. (a) Sample wide-aperture photo from the synthetic DOF sequence. (b) DOF composites synthesized from this sequence. (c) Narrow-aperture photos spanning an equivalent DOF, but with much longer exposure time. (d) Coarse depth map, computed from the labeling we used to compute (b). (e) Synthetically changing aperture size, focused at the same setting as (a). (f) Synthetically changing focus setting as well, for the same synthetic aperture as (e).

strategy for the digital SLR only (HAMSTER and FACE datasets). At this level of overhead, the optimal sequence for this case takes 220 ms to capture[8]. This reduces the speedup to 3.6×, compared to 800 ms for one narrow-aperture photo.

DOF compositing. Despite the fact that it relies on a coarse depth map, our compositing scheme is able to reproduce high-frequency detail over the whole DOF, without noticeable artifacts, even in the vicinity of depth discontinuities (Figs. 7b, 8b, and 9b). The narrow-aperture photos represent ground truth, and visually they are almost indistinguishable from our composites.

The worst compositing artifact occurs in the HAMSTER dataset, at the handle of the pumpkin container, which is incorrectly assigned to a background depth (Fig. 10). This is an especially challenging region because the handle is thin and low-texture compared to the porcelain lid behind it.

Note that while the synthesized photos satisfy our goal of spanning a specific DOF, objects outside that

8. More specifically, the optimal sequence involves spanning the DOF with 7 photos instead of 14. This sequence consists of 1 photo captured at $f/2$, plus 3 photos each at $f/2.2$ and $f/2.5$.

DOF will appear more defocused than in the corresponding narrow-aperture photo. For example, the cardboard background in the HAMSTER dataset is not included in the DOF (Fig. 11). This background therefore appears slightly defocused in the narrow-aperture $f/16$ photo, and strongly defocused in the synthetic DOF composite. This effect is expected, since outside the synthetic DOF, the blur diameter will increase in proportion to the wider aperture diameter used in the capture sequence (Eq. (4)). For some applications, such as portrait photography, increased background defocus may be beneficial.

Depth maps and DOF compositing. Despite being more efficient to capture, sequences with synthetic DOFs provide 3D shape information at no extra acquisition cost (Figs. 7d, 8d, and 9d). Using the method described in Sec. 5, we also show results of using this depth map to compute novel images whose aperture and focus setting was changed synthetically (Figs. 7e–f, 8e–f, and 9e–f). As a general rule, the denser and more light-efficient a capture sequence is, the wider the range of camera settings it offers for synthetic refocusing.

Focus control and overhead. Neither of our cameras provide the ability to control focus programmatically, so we used several methods to circumvent this limitation. For our digital SLR, we used a computer-controlled stepping motor to drive the lens focusing ring mechanically [36]. For our compact digital camera, we exploited modified firmware that enables general scripting [37]. Unfortunately, both these methods incur high additional overhead, limiting us to about $1\,\mathrm{fps}$ in practice.

Note that mechanical refocusing contributes relatively little overhead for the SLR, since ultrasonic lenses, like the Canon EF85mm 1.2L we used, are fast. Our lens takes $3.5\,\mathrm{ms}$ to refocus from one photo in the sequence to the next, for a total of $45\,\mathrm{ms}$ to cover the largest possible DOF spanned by a single photo. In addition, refocusing can potentially be executed in parallel with other tasks such as transferring the previous image to memory. Such parallel execution already occurs in the Canon's "autofocus servo" mode, in which the camera refocuses continuously on a moving subject.

While light-efficient photography may not be practical using our current prototypes, it will become increasingly so, as newer cameras begin to expose their focusing API directly and new CMOS sensors increases throughput. For example, the Canon EOS-1D Mark III provides remote focus control for all Canon EF lenses, and the recently released Casio EX-F1 can capture $60\,\mathrm{fps}$ at $2\,\mathrm{MP}$. Even though light-efficient photography will benefit from the latest advances in capture speed, as Fig. 5 shows, we can still realize time savings at slower frames-per-second rates.

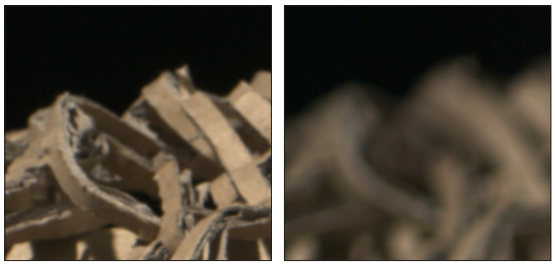
Handling motion in the capture sequence. Because of the high overhead due to our focus control mechanisms, we observed scene motion in two of our capture sequences. The SIMPSONS dataset shows a subtle change in brightness above the green book in the background, because the person taking the photos moved during acquisition, casting a moving shadow on the wall. This is not an artifact and did not affect our processing. For the FACE dataset, the subject moved slightly during acquisition of the optimal capture sequence. To account for this motion, we performed a global rigid 2D alignment between successive images using Lucas-Kanade registration [39].

Despite this inter-frame motion, our approach for creating photos with a synthetic DOF (Sec. 5) generates results that are free of artifacts. In fact, the effects of this motion are only possible to see only in the videos that we create for varying synthetic aperture and focus settings. Specifically, while each still in the videos appears free of artifacts, successive stills contain a slight but noticeable amount of motion.

We emphasize the following two points. First, had we

narrow aperture ground truth (f/16) | synthetic DOF composite | coarse depth map, from DOF composite

Fig. 10. Compositing failure for the HAMSTER dataset (Fig. 7). Elsewhere this scene is synthesized realistically. The depth-from-focus method employed by the Photomontage method breaks down at the handle of the pumpkin container, incorrectly assigning it to a background layer. This part of the scene is challenging to reconstruct because strong scene texture is visible "through" the defocused handle [38], whereas the handle itself is thin and low-texture.



narrow aperture ground truth (f/16) | synthetic DOF composite

Fig. 11. Background defocus for the HAMSTER dataset. Because the cardboard background lies outside the DOF, it is slightly defocused in the narrow-aperture photo. In the synthetic DOF composite, however, this background is defocused much more significantly. This effect is expected since the composite only produces in-focus images for objects lying *within* the DOF.

been able to exploit the internal focus control mechanism of the camera (a feature that newer cameras like the Canon EOS-1D Mark III provide), the inter-frame motion for FACE dataset would have been negligible, making the above registration step unnecessary. Second, even with fast internal focus control, residual motions would occur when photographing fast-moving subjects; our results in this sequence suggest that even in that case, our simple merging method should be sufficient to handle such motions with little or no image degradation.

7 Relation to Alternative Camera Designs

While our analysis of light-efficient capture has assumed a conventional camera, the benefits of our approach are further reaching. More specifically, whatever extended DOF can be realized using an alternative camera design, one can obtain complementary gains in efficiency by using larger apertures and capturing multiple photos to span the desired DOF.

In the following, we discuss the relationship of the standard camera to several families of computational cameras which make various tradeoffs to achieve extended DOF [17], [25].

Light field cameras. Light field cameras have commonly been portrayed as extending DOF [16], [27], however such comparisons rely on greatly increasing the number of sensor pixels relative to the standard camera (*e.g.*, by a factor of 196 [27]). When the resolution of the sensor is held constant, light field cameras must sacrifice spatial resolution for increased directional sampling. In this setting, the light field camera does not extend DOF compared to regular photography [40].

As our analysis in Appendix B shows, the DOF of light field camera is identical to that of a standard camera with the same aperture and sensor resolution. In particular, both designs share the same the depth range over which the blur diameter is less than 1 pixel, at the reduced spatial resolution of the light field camera. Interestingly, for defocused objects *outside* the DOF, the light field camera significantly reduces the degree of defocus compared to a standard camera, analogous to the comparison illustrated in Fig. 11.

As recently explored, light field photographs of a Lambertian scene can be further enhanced by applying super-resolution techniques [41], [42]. However, the same enhancements are also possible for a standard photo, for which super-resolution corresponds to deconvolving the full-resolution image.

In practice, light field cameras are less light-efficient than conventional photography, because they require stopping down the lens to avoid overlap between lenslet images [27], or they block light as a result of the imperfect packing of optical elements [43]. The above analysis also holds for the heterodyne light field camera [16], where the mask placed near the sensor blocks $70\,\%$ of the light, and the relevant sub-images are extracted in the frequency domain.

Wavefront coding and focus sweep. Wavefront coding methods rely on a special optical element that effectively spreads defocus over a larger DOF, and then recovers the underlying in-focus image using deconvolution [11]. While this approach is powerful, it exploits a tradeoff that is orthogonal to our analysis. Wavefront coding can extend perceived DOF by a factor of $N = 2$ to 10, but it suffers from reduced SNR at high frequencies [11], and it provides no 3D information. The need to deconvolve the image is another possible source of error, particularly since the point-spread function is only approximately constant over the extended DOF.

Roughly speaking, wavefront coding can be analyzed by analogy to simpler "focus sweep" methods [12], [13], which capture a single photo while continuously varying the focus setting through the DOF. Like wavefront coding, these focus sweep methods lead to a deconvolution problem that is approximately invariant to scene depth.

To compare these methods to our approach in a fair way, we fix the total exposure time, τ (thereby collecting the same number of photons), and examine the SNR of the restored in-focus photos. Wavefront coding and focus sweep both involve capturing a single photo with exposure time τ. By contrast, our approach involves capturing N in-focus photos spanning the DOF, each allocated exposure time of τ/N. The sweeping analogy suggests that these methods can do no better than our method in terms of SNR, because all of them effectively collect the same number of "in-focus" photons for a scene at a given depth.

For a more detailed analysis, *e.g.*, taking into account the implications of capturing multiple under-exposed photos, see [25], [44].

Aperture masks. Narrow apertures on a conventional camera can be thought of as masks in front of the widest aperture, however it is possible to block the aperture using more general masks. For example, ring-shaped apertures [45], [46] have a long history in astronomy and microscopy, and recent methods have proposed using coded binary masks in conjunction with regular lenses [15], [16].

Previous analysis suggests that ring-shaped apertures yield no light-efficient benefit [45], and recent work leads to similar conclusions for coded aperture masks [24], [25], [44]. While coded masks preserve high frequencies better than a standard camera, and so effectively increase DOF, this advantage is generally outweighed by the loss in light-efficiency caused by blocking $50\,\%$ of the aperture.

Unlike the wavefront coding and focus sweep case, processing a coded-aperture image depends on depth recovery, which is non-trivial from a single photo. Again, for more detailed analysis, see [25], [44].

8 Concluding Remarks

In this article we studied the use of dense, wide-aperture photo sequences as a light-efficient alternative to single-shot, narrow-aperture photography. While our emphasis has been on the underlying theory, we believe that our results will become increasingly relevant as newer, off-the-shelf cameras enable direct control of focus and aperture.

We are currently investigating several extensions to the basic approach. First, we are interested in further improving efficiency by taking advantage of the depth information from the camera's auto-focus sensors. Such information would let us save additional time, because we would only have to capture photos at focus settings that correspond to actual scene depths.

Second, we are generalizing the goal of light-efficient photography to reproduce arbitrary profiles of blur diameter *vs*. depth, rather than just reproducing the depth of field. For example, this method could be used to reproduce the defocus properties of the narrow-aperture photo entirely, including defocus for background objects as shown in Fig. 11.

Appendix A
Light-Efficiency Proofs

Theorem 1 follows as a consequence of Lemma 1 and four additional lemmas, while proving Theorem 2 is more direct. We first state Lemmas 2–5 and prove them below before addressing the theorems.

Lemma 2 (Efficiency of Sequences with Sequential DOFs)**.** *For every sequence $\mathcal{S}$, there is a sequence $\mathcal{S}'$ with sequential DOFs that spans the same synthetic DOF and has a total exposure time no larger than that of $\mathcal{S}$.*

Lemma 3 (Permutation of Sequences with Sequential DOFs)**.** *Given the left endpoint, α, every permutation of $D_1, \ldots, D_n$ defines a capture sequence with sequential DOFs that has the same synthetic depth of field and the same total exposure time.*

Lemma 4 (Optimality of Maximizing the Number of Photos)**.** *Among all sequences with up to n tuples whose synthetic DOF is $[\alpha, \beta]$, the sequence that minimizes total exposure time has exactly n of them.*

Lemma 5 (Optimality of Equal-Aperture Sequences)**.** *If $\beta < (7 + 4\sqrt{3})\alpha$, then among all capture sequences with n tuples whose synthetic DOF is $[\alpha, \beta]$, the sequence that minimizes total exposure time uses the same aperture for all tuples. Furthermore, this aperture is equal to*

$$D(n) \;=\; c\,\frac{\sqrt[n]{\beta} + \sqrt[n]{\alpha}}{\sqrt[n]{\beta} - \sqrt[n]{\alpha}} \quad . \tag{18}$$

Proof of Lemma 1. We proceed inductively, by defining photo tuples whose DOFs "tile" the interval $[\alpha, \beta]$ from left to right. For the base case, the left endpoint of the first tuple's DOF must be $\alpha_1 = \alpha$. Now consider the i-th tuple. Eq. (5) implies that the left endpoint α_i and the aperture diameter D_i determine the DOF's right endpoint uniquely:

$$\beta_i \;=\; \frac{D_i + c}{D_i - c}\,\alpha_i \quad . \tag{19}$$

The tuple's focus setting in Eq. (11) now follows by applying Eq. (6) to the interval $[\alpha_i, \beta_i]$. Finally, since the DOFs of tuple i and $i+1$ are sequential, we have $\alpha_{i+1} = \beta_i$. □

Proof of Lemma 2. Let $\langle\ D,\ \tau,\ v\ \rangle$ be a tuple in $\mathcal{S}$, and let $[\alpha_1, \beta_1]$ be its depth of field. Now suppose that $\mathcal{S}$ contains another tuple whose depth of field, $[\alpha_2, \beta_2]$, overlaps with $[\alpha_1, \beta_1]$. Without loss of generality, assume that $\alpha_1 < \alpha_2 < \beta_1 < \beta_2$. We now replace $\langle\ D,\ \tau,\ v\ \rangle$ with a new tuple $\langle\ D',\ \tau',\ v'\ \rangle$ whose DOF is $[\alpha_1, \alpha_2]$ by setting D' according to Eq. (5) and v' according to Eq. (6). Since the DOF of the new tuple is narrower than the original, we have $D' > D$ and, hence, $\tau' < \tau$. Note that this tuple replacement preserves the synthetic DOF of the original sequence. We can apply this operation repeatedly until no tuples exist with overlapping DOFs. □

Proof of Lemma 3. From Eq. (11) it follows that the total exposure time is

$$\tau \;=\; \sum_{i=1}^{n} \frac{{L^*}}{{D_i}^2}\ , \tag{20}$$

which is invariant to the permutation. To show that the synthetic DOF is also permutation invariant, we apply Eq. (19) recursively n times to obtain the right endpoint of the synthetic DOF:

$$\beta_n \;=\; \alpha \prod_{i=1}^{n} \frac{D_i + c}{D_i - c} \quad . \tag{21}$$

It follows that β_n is invariant to the permutation. □

Proof of Lemma 4. From Lemma 2 it follows that among all sequences up to length n whose DOF is $[\alpha, \beta]$, there is a sequence $\mathcal{S}^*$ with minimum total exposure time whose tuples have sequential DOFs. Furthermore, Lemmas 1 and 3 imply that this capture sequence is fully determined by a sequence of n' aperture settings, $D_1 \le D_2 \le \cdots \le D_{n'}$, for some $n' \le n$. These settings partition the interval $[\alpha, \beta]$ into n' sub-intervals, whose endpoints are given by Eq. (12):

$$\alpha = \alpha_1 < \overbrace{\alpha_2 < \cdots < \alpha_{n'}}^{\text{determined by } \mathcal{S}^*} < \beta_{n'} = \beta \quad . \tag{22}$$

It therefore suffices to show that placing $n' - 1$ points in $[\alpha, \beta]$ is most efficient when $n' = n$. To do this, we show that splitting a sub-interval always produces a more efficient capture sequence.

Consider the case $n = 2$, where the sub-interval to be split is actually equal to $[\alpha, \beta]$. Let $x \in [\alpha, \beta]$ be a splitting

point. The exposure time for the sub-intervals $[\alpha, x]$ and $[x, \beta]$ can be obtained by combining Eqs. (5) and (1):

$$\tau(x) \;=\; \frac{L}{c^2}\left(\frac{x-\alpha}{x+\alpha}\right)^2 + \frac{L}{c^2}\left(\frac{\beta-x}{\beta+x}\right)^2 \quad , \qquad (23)$$

Differentiating Eq. (23) and evaluating it for $x = \alpha$ we obtain

$$\left.\frac{d\tau}{dx}\right|_{x=\alpha} \;=\; -\frac{4L}{c^2}\,\frac{(\beta-\alpha)\,\beta}{(\beta+\alpha)^3} \;<\; 0 \quad . \qquad (24)$$

Similarly, it is possible to show that $\frac{d\tau}{dx}$ is positive for $x = \beta$. Since $\tau(x)$ is continuous in $[\alpha, \beta]$, it follows that the minimum of $\tau(x)$ occurs strictly inside the interval. Hence, splitting the interval always reduces total exposure time. The general case for n intervals follows by induction. □

Proof of Lemma 5. As in the proof of Lemma 4, we consider the case where $n = 2$. From that lemma it follows that the most efficient sequence involves splitting $[\alpha, \beta]$ into two sub-intervals $[\alpha, x]$ and $[x, \beta]$. To prove Lemma 5 we now show that the optimal split corresponds to a sequence with two identical aperture settings. Solving for $\frac{d\tau}{dx} = 0$ we obtain four solutions:

$$x = \left\{ \pm\sqrt{\alpha\beta}\,,\; \frac{(8\alpha\beta + \Delta) \;\pm\; (\beta-\alpha)\sqrt{\Delta}}{2(\beta+\alpha)} \right\} \quad , \qquad (25)$$

where $\Delta = \alpha^2 - 14\alpha\beta + \beta^2$. The inequality condition of Lemma 5 implies that $\Delta < 0$. Hence, the only real and positive solution is $x = \sqrt{\alpha\beta}$. From Eq. (5) it now follows that the intervals $[\alpha, \sqrt{\alpha\beta}]$ and $[\sqrt{\alpha\beta}, \beta]$ both correspond to an aperture equal to $c\,\frac{\sqrt{\beta}+\sqrt{\alpha}}{\sqrt{\beta}-\sqrt{\alpha}}$. To prove the Lemma for $n > 2$, we replace the sum in Eq. (23) with a sum of n terms corresponding to the sub-divisions of $[\alpha, \beta]$, and then apply the above proof to each pair of adjacent endpoints of that subdivision. This generates a set of relations, $\{\alpha_i = \sqrt{\alpha_{i-1}\alpha_{i+1}}\}_{i=2}^{n}$, which combine to define Eq. (18) uniquely. □

Proof of Theorem 1. We first consider the most efficient capture sequence, $\mathcal{S}'$, among all sequences whose synthetic DOF is identical to $[\alpha, \beta]$. Lemmas 4 and 5 imply that the most efficient sequence (1) has maximal length and (2) uses the same aperture for all tuples. More specifically, consider such a sequence of n photos with diameter $D_i = D(n)$, for all i, according to Eq. (18). This sequence satisfies Eq. (21) with $\beta_n = \beta$, and we can manipulate this equation to obtain:

$$n \;=\; \frac{\log\frac{\alpha}{\beta}}{\log\left(\frac{D(n)-c}{D(n)+c}\right)} \quad . \qquad (26)$$

Note that while n increases monotonically with aperture diameter, the maximum aperture diameter D_{max} restricts the maximal n for which such an even subdivision is possible. This maximal n, whose formula is provided by Eq. (9), can be found by evaluating Eq. (26) with an aperture diameter of D_{max}.

While $\mathcal{S}'$ is the most efficient sequence among those whose synthetic DOFs equal to $[\alpha, \beta]$, there may be sequences whose DOF strictly contains this interval that are even more efficient. We now seek the most efficient sequence, $\mathcal{S}''$ among this class. To find it, we use two observations. First, $\mathcal{S}''$ must have length at most $n + 1$. This is because longer sequences must include a tuple whose DOF lies entirely outside $[\alpha, \beta]$. Second, among all sequences of length $n+1$, the most efficient sequence is the one whose aperture diameters are all equal to the maximum possible value, D_{max}. This follows from the fact that any choice of $n+1$ apertures is sufficient to span the DOF, so the most efficient such choice involves the largest apertures possible.

From the above considerations it follows that the optimal capture sequence will be an equal-aperture sequence whose aperture will be either $D(n)$ or D_{max}. The test in Eq. (10) comes from comparing the total exposure times of the sequences $\mathcal{S}'$ and $\mathcal{S}''$ using Eq. (20). The theorem's inequality condition comes from Lemma 5. □

Proof of Theorem 2. The formulation of the integer linear program in Eqs. (13)–(16) follows from our objective of minimizing total exposure time, plus the constraint that the apertures used in the optimal capture sequence must span the desired DOF.

First, note that the multiplicities n_i are non-negative integers, since they correspond to the number of photos taken with each discrete aperture D_i. Second, we can rewrite the total exposure time given by Eq. (20) in terms of the multiplicities:

$$\tau \;=\; \sum_{i=1}^{m} n_i\,\frac{L^*}{{D_i}^2} \quad , \qquad (27)$$

This corresponds directly to Eq. (13), and is linear in the multiplicities being optimized. Finally, we can rewrite the expression for the right endpoint of the synthetic DOF provided by Eq. (21) in terms of the multiplicities:

$$\beta_m \;=\; \alpha \prod_{i=1}^{m}\left(\frac{D_i + c}{D_i - c}\right)^{n_i} \quad . \qquad (28)$$

Because all sequences we consider are sequential, the DOF $[\alpha, \beta]$ will be spanned without any gaps provided that the right endpoint satisfies $\beta_m \geq \beta$. By combining this constraint with Eq. (28) and taking logarithms, we obtain the inequality in Eq. (14), which is linear in the multiplicities as well. □

APPENDIX B
ANALYSIS OF DOF FOR THE LIGHT FIELD CAMERA

When aperture and sensor resolution are fixed, the light field camera does not have the ability to extend the DOF compared to a standard camera [40]. To see why, consider a 1D conventional camera with a P-pixel sensor, whose aperture is set to diameter D. For comparison,

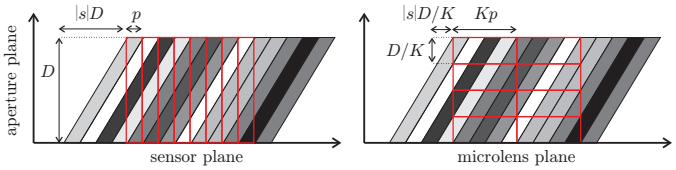


Fig. 12. Light field for a 1D Lambertian scene, overlaid with geometry describing a standard camera (left), and a light field camera with $K = 4$ directional samples and $1/K$ the spatial resolution (right). The axes of the light field correspond to the two-plane parameterization of the 2D space of rays, and the shades of grey indicate varying scene radiance (see [27] for more details). Photography can be viewed as integrating rays in this space; the red rectangles denote boundaries of the integration domain for pixels in each camera design.

consider a light field camera built by placing K lenslets in front of the same sensor, yielding reduced resolution (P/K)-pixel sub-images from K different viewpoints [27]. Each sub-image corresponds to a smaller effective aperture with diameter D/K.

To analyze DOF, we examine how sensor elements of size p integrate light from a fronto-parallel Lambertian scene. As shown in Fig. 12, such a scene corresponds to a sheared version of a light field with no variation over the aperture coordinate [27]. The slope of the shear, $1/s$, is directly related to the depth of scene. Geometrically, we can see that pixels for the standard camera integrate $p + |s|D$ of the scene radiance, while pixels for the light field camera integrate $Kp + |s|D/K$.

To define a threshold for DOF, we adopt the reduced spatial resolution of the light field camera. More specifically, we consider the scene to be within the DOF if each pixel integrates less than $(K + 1)p$ scene radiance. This criterion yields an *identical* DOF for both the standard and light field camera, *i.e.*, the depth range $|s| < Kp/D$.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada under the RGPIN program, and of the Ontario Ministry of Research and Innovation under the PREA program.

REFERENCES

[1] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, 1994.
[2] S. W. Hasinoff and K. N. Kutulakos, "A layer-based restoration framework for variable-aperture photography," in *Proc. International Conference on Computer Vision*, 2007, pp. 1–8.
[3] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, Jul. 1987.
[4] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, 1987.
[5] S. Hiura and T. Matsuyama, "Depth measurement by the multi-focus camera," in *Proc. Computer Vision and Pattern Recognition*, 1998, pp. 953–959.
[6] M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *International Journal of Computer Vision*, vol. 27, no. 3, pp. 203–225, 1998.
[7] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," in *Proc. ACM SIGGRAPH*, 2006, pp. 795–804.
[8] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," in *Proc. ACM SIGGRAPH*, 2007.
[9] J. Telleen, A. Sullivan, J. Yee, P. Gunawardane, O. Wang, I. Collins, and J. Davis, "Synthetic shutter speed imaging," in *Proc. Eurographics*, 2007, pp. 591–598.
[10] H. Farid and E. P. Simoncelli, "Range estimation by optical differentiation," *J. Optical Society of America A*, vol. 15, no. 7, pp. 1777–1786, 1998.
[11] W. T. Cathey and E. R. Dowski, "New paradigm for imaging systems," *Applied Optics*, vol. 41, no. 29, pp. 6080–6092, Oct. 2002.
[12] G. Häusler, "A method to increase the depth of focus by two step image processing," *Optics Communications*, vol. 6, no. 1, pp. 38–42, 1972.
[13] H. Nagahara, S. Kuthirummal, C. Zhou, and S. Nayar, "Flexible depth of field photography," in *Proc. European Conference on Computer Vision*, vol. 4, 2008, pp. 60–73.
[14] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *Proc. Computer Vision and Pattern Recognition*, vol. 1, Jun. 2006, pp. 339–346.
[15] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. ACM SIGGRAPH*, 2007.
[16] 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 *Proc. ACM SIGGRAPH*, 2007.
[17] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman, "4D frequency analysis of computational cameras for depth of field extension," in *Proc. ACM SIGGRAPH*, 2009.
[18] K. Aizawa, K. Kodama, and A. Kubota, "Producing object-based special effects by fusing multiple differently focused images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 323–330, Mar. 2000.
[19] S. Chaudhuri, "Defocus morphing in real aperture images," *J. Optical Society of America A*, vol. 22, no. 11, pp. 2357–2365, Nov. 2005.
[20] S. W. Hasinoff and K. N. Kutulakos, "Confocal stereo," in *Proc. European Conference on Computer Vision*, vol. 1, 2006, pp. 620–634.
[21] R. Ng, "Fourier slice photography," in *Proc. ACM SIGGRAPH*, 2005, pp. 735–744.
[22] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," in *Proc. ACM SIGGRAPH*, 2006, pp. 924–934.
[23] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, 1997, pp. 369–378.
[24] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offs through a Bayesian analysis of light field projections," in *Proc. European Conference on Computer Vision*, vol. 4, 2008, pp. 88–101.
[25] S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman, "Time-constrained photography," in *Proc. International Conference on Computer Vision*, 2009, pp. 1–8.
[26] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proc. ACM SIGGRAPH*, 2004, pp. 294–302.
[27] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Dept. Computer Science, Stanford University, Tech. Rep. CTSR 2005-02, 2005.
[28] J. Ogden, E. Adelson, J. R. Bergen, and P. Burt, "Pyramid-based computer graphics," *RCA Engineer*, vol. 30, no. 5, pp. 4–15, 1985.
[29] W. J. Smith, *Modern Optical Engineering*, 3rd ed. New York: McGraw-Hill, 2000.
[30] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Proc. Computer Vision and Pattern Recognition*, 1999, pp. 1374–1380.
[31] Canon lens specifications, http://www.usa.canon.com/eflenses/pdf/spec.pdf.
[32] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

[33] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[34] Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?" *International Journal of Computer Vision*, vol. 39, no. 2, pp. 141–162, Sep. 2000.
[35] R. Willson and S. Shafer, "What is the center of the image?" *J. Optical Society of America A*, vol. 11, no. 11, pp. 2946–2955, Nov 1994.
[36] Technical Innovations, http://www.robofocus.com/.
[37] CHDK, http://chdk.wikia.com/.
[38] P. Favaro and S. Soatto, "Seeing beyond occlusions (and other marvels of a finite lens aperture)," in *Proc. Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 579–586.
[39] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–25, 2004.
[40] M. Levoy, May 2008, personal communication.
[41] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proc. International Conference on Computational Photography*, 2009.
[42] T. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Proc. International Conference on Computational Photography*, 2009.
[43] T. Georgiev, C. Zheng, S. Nayar, D. Salesin, B. Curless, and C. Intwala, "Spatio-angular resolution tradeoffs in integral photography," in *Proc. Eurographics Symposium on Rendering*, 2006, pp. 263–272.
[44] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman, "4D frequency analysis of computational cameras for depth of field extension," MIT, Tech. Rep. MIT-CSAIL-TR-2009-019, 2009.
[45] W. T. Welford, "Use of annular apertures to increase focal depth," *J. Optical Society of America A*, vol. 50, no. 8, pp. 749–753, Aug. 1960.
[46] J. Ojeda-Castaneda, E. Tepichin, and A. Pons, "Apodization of annular apertures: Strehl ratio," *Applied Optics*, vol. 27, no. 24, pp. 5140–5145, Dec. 1988.

Samuel W. Hasinoff received the BS degree in computer science from the University of British Columbia in 2000, and the MS and PhD degrees in computer science from the University of Toronto in 2002 and 2008, respectively. He is currently an NSERC Postdoctoral Fellow at the Massachusetts Institute of Technology. In 2006, he received an honorable mention for the Longuet-Higgins Best Paper Award at the European Conference on Computer Vision. He is also the recipient of the Alain Fournier Award for the top Canadian dissertation in computer graphics in 2008. He is a member of the IEEE.

Kiriakos N. Kutulakos received the BA degree in computer science at the University of Crete, Greece in 1988, and the MS and PhD degrees in computer science from the University of Wisconsin, Madison in 1990 and 1994, respectively. Following his dissertation work, he joined the University of Rochester where he was an NSF Postdoctoral Fellow and later an assistant professor until 2001. He is currently an associate professor of computer science at the University of Toronto. He won the Best Student Paper Award at CVPR'94, the Marr Prize in 1999, a Marr Prize Honorable Mention in 2005 and a Best Paper Honorable Mention at ECCV'06. He is the recipient of a CAREER award from the US National Science Foundation, a Premier's Research Excellence Award from the government of Ontario, and an Alfred P. Sloan Research Fellowship. He served as program co-chair of CVPR 2003 and is currently an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence. He is a member of the IEEE.