

# Square Coded Aperture: A Large Aperture With Extended Depth Of Field

Ruojun He; China University of Political Science and Law School; Beijing, China

Yi Zhang; Argo AI, LLC; Mountain View, California, USA

Keigo Hirakawa; University of Dayton; Dayton, Ohio, USA

## Abstract

We propose a square aperture as a simple and practical alternative to existing coded aperture. A spatial derivative converts sensor measurements taken with a square aperture mask into measurements taken with a pinhole or slit aperture mask. Thus square aperture shares the properties of both large and small apertures, yielding excellent light efficiency while an artificial small aperture results in an infinite depth of field. We developed a prototype lens to confirm the feasibility of our blur size estimation and image deblurring approach.

## Introduction

Coded aperture refers to the technique of modifying the defocus blur kernel by replacing the standard near-circular aperture with an *aperture mask* [1, 2]. When the blur kernel and image features are dissimilar (e.g. pseudo-random blur kernel), it becomes possible to disambiguate them in post-capture processing. Naturally, this improves image deblurring and depth-from-defocus [3–5], since the recovery of in-focus image and blur kernel is made easier.

In this paper, we propose *square aperture* as an alternative to previously proposed aperture masks. Square aperture is desirable for computational photography framework because it shares the properties of both large and small apertures. The aperture by itself may be a large (with shallow depth of field and ample light). As illustrated in Figs. 1 and 2, however, a derivative operator in (4) below converts sensor measurements taken with a square aperture mask into measurements taken with a slit or pinhole aperture mask. This “sparsified” blur kernel is essentially an artificial small aperture resulting in an infinite depth of field (i.e. limited only by diffraction), making it easy to recover the sharp image from the captured blurry image. One may also understand Figure 2(b-c) as “superimposed slit cameras” and “superimposed pinhole cameras” where the two or four images are captured at different positions of the camera aperture, respectively. We also gain the ability to infer defocus blur size by measuring the displacements of the four sharp pinhole images or the two slit sharp images—precisely in the way that phase detection autofocus sensors in DSLR cameras work, but on a pixel-to-pixel basis—and recover the scene depth from the reconstructed defocus blur size indirectly via depth-from-defocus.

In this sense, square aperture can be interpreted as an aperture mask hyper-optimized for gradient image modeling. As such, we emphasize that the contribution of this work is an aperture mask designed that takes maximal advantages of a mature area of computer vision and computational photography. That is, the *algorithm* for how we recover blur size and sharp image from gra-

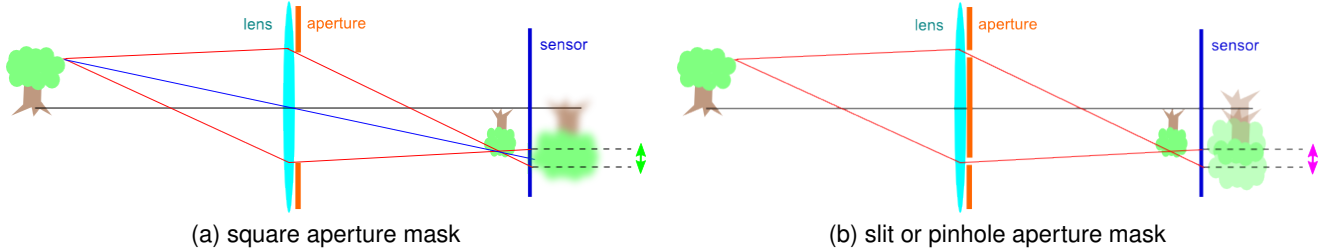
dient images may be fairly standard in and of themselves. However, it is the direct correspondence of gradient images to multiple pinhole/slit cameras (along with its desirable characteristics described above) that makes the square geometry a particularly appropriate choice for derivative-processing-enabled coded aperture. In particular, the proposed sparsifiable aperture provides a contrary perspective to coded aperture work that has thus far been dominated by blur shapes that do not have sparse representations [2, 6, 7]. (See Section .)

Square aperture is also pragmatic—with only a minor change to a standard commercial camera design, it is possible to design a “dual purpose” camera system capable of imaging with both circular and square apertures. With the potential of enabling fast autofocus enjoyed by phase detection (i.e. without extra mirror, autofocus sensor, and the bulk associated with them), recovering scene depth, and a simple image deblurring scheme without interrupting the existing regiments of camera design, square aperture has better chances of being incorporated into everyday cameras than the alternatives.

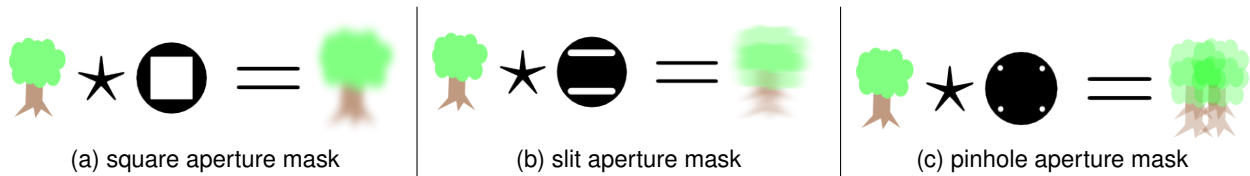
## Related Work

The light combined at the detector take the shape of the scaled versions of the aperture mask—one can interpret this as a defocus blur whose point spread function (blur kernel) is of known shape but of unknown size (plus diffraction that is assumed to be negligent). The existing coded aperture designs are aimed at discriminating the blur kernel size for depth recovery [2, 8], reducing the negative effects of “zeros” in the Fourier domain for deblurring [2, 9], and capitalizing on the compressed sensing principles [6, 10]. Many of the previously proposed aperture masks tended to resemble pseudo random sampling—whether by design [6, 10] or by a brute force optimization that converged to a non-sparse shape [2, 7]. The fact that the pseudo random blur kernel are dissimilar to image signals (a notion commonly referred to as “incoherence”) has been attributed to the reason why they can be disambiguated in post-capture processing [6]. With detected coded defocus blur kernel size inferring the scene depth and the sharp image recovery effectively extending the depth of field of the camera, research in coded aperture legitimized the idea that blur can be seen as a cue for scene understanding rather than degradation.

Coded apertures are not without their limitations, however. Three major shortcomings that have been met with limited success are light efficiency, deblurring quality, and practicality. Specifically, the light efficiency of coded aperture lens systems is poor because less light is allowed to pass through the optics. One exception to this is the recently introduced color coded aper-



**Figure 1.** The mathematical relationship in (4) converts sensor measurements taken with a square aperture mask in (a) into measurements taken with a slit or pinhole aperture mask in (b). Green arrow in (a) corresponds to an “square of confusion” for a single large square aperture. The image at the detector is therefore blurry. With the pinhole aperture mask in (b), the detector sees a few copies of the latent sharp images that are superimposed. The relative displacement of images/edges—indicated by the magenta arrow in (b)—is also determined by the scene depth. Clearly, it is far easier to infer the scene depth and to recover the latent sharp image from the system in (b).



**Figure 2.** Two dimensional representation of square, slit, and pinhole aperture masks in Figure 1.

ture by Chakrabarti *et al.* [8], which overcomes this problem by allowing the diameter of the radial aperture be smaller in green channel than the red and blue diameters. Second, the claims of extended depth of field is limited by the deblurring quality. Though the recovered sharp image is certainly an improvement over the captured blurry image, deblurring cannot resolve the fine image details to the extent that a small aperture camera does. An exception here is the tri-lens mask (aperture mask + lens) proposed in [9], which has a long depth of field thanks to the pinhole-like aperture mask, but at the severe sacrifice of light efficiency and the additional complexity of three small lenses covering the three pinholes. Third, consumer imaging industry has also been slow to embrace significant changes to acquisition hardware (especially randomized ones). Coded aperture designs are often incompatible with existing camera designs—in the sense that a mask must be inserted mechanically in the optical pathway, effectively replacing the near-circular aperture commonly used in today’s cameras. Since circular blurs (a.k.a. *bokeh*) are regarded as aesthetically desirable by photographers, coded apertures require disruptive hardware modifications that few manufacturers would be willing to incorporate. Contrast this to the proposed square aperture, which boasts favorable light efficiency, a mathematical correspondence to a slit/pinhole camera, and implementation with minimally disruptive hardware modification to standard optics.

## Square Aperture Camera System Design

### Proposed Aperture Design

We first describe a camera with a large square aperture. As illustrated by raytracing in Figure 1(a), image formed at the detector is blurred when the light originating from the same point in the scene does not converge. The severity of the blur is determined by the “square of confusion”—the extent that light passing through the aperture opening is allowed to deviate from the chief ray (blue ray in Figure 1(a)) at the detector, determined by the light rays passing the aperture boundaries (red rays in Figure 1(a)). Assuming Gaussian optics, the size of the square of confusion  $s$  and the

distance to the object  $z$  have the relation:

$$s = a|z_0(f^{-1} - z^{-1}) - 1|, \quad (1)$$

where  $z_0$  is the distance from the lens to the detector,  $f$  is the focal length, and  $a$  is the aperture opening size.

Let  $I(x, y)$  denote the latent sharp image formed *only* by the chief ray of the square aperture. Supposing for the moment that a scene is comprised of Lambertian fronto-parallel objects, the light combined at the detector  $J(x, y)$  is a convolution of  $I(x, y)$  and the blur kernel  $H_{(x,y)}(x, y)$ :

$$J(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} I(\epsilon, \tau) H_{(\epsilon, \tau)}(x - \epsilon, y - \tau) d\epsilon d\tau. \quad (2)$$

The square aperture takes the form

$$H_{(x,y)}(\epsilon, \tau) = \begin{cases} \frac{1}{s(x,y)^2} & \text{if } |\epsilon| < \frac{s(x,y)}{2} \text{ and } |\tau| < \frac{s(x,y)}{2} \\ 0 & \text{else,} \end{cases}$$

where  $s(x, y)$  is the blur size for location  $(x, y)$  (constant for a fronto-parallel object). This simplifies (2) as

$$J(x, y) = \int_{y-\frac{s}{2}}^{y+\frac{s}{2}} \int_{x-\frac{s}{2}}^{x+\frac{s}{2}} \frac{I(\epsilon, \tau)}{s^2} d\epsilon d\tau, \quad (3)$$

where we omitted  $(x, y)$  from  $s(x, y)$  for simplicity. The blurring in (3) is illustrated in Figure 2(a). We emphasize that noise is an uncertainty attributed to making sensor measurements on  $J(x, y)$ . The topic of noise modeling (Section ) and handling (Section ) are left to later sections.

## Relationship To Slit/Pinhole Aperture Masks

Consider now the derivative images of  $J(x, y)$ . By the fundamental theorem of calculus, we have

$$\begin{aligned} J_x(x, y) &= \frac{\partial}{\partial x} J(x, y) = \frac{K(x + \frac{s}{2}, y) - K(x - \frac{s}{2}, y)}{s} \\ J_y(x, y) &= \frac{\partial}{\partial y} J(x, y) = \frac{L(x, y + \frac{s}{2}) - L(x, y - \frac{s}{2})}{s} \\ J_{xy}(x, y) &:= \frac{\partial^2}{\partial x \partial y} J(x, y) \\ &= \frac{I(x + \frac{s}{2}, y + \frac{s}{2})}{s^2} - \frac{I(x + \frac{s}{2}, y - \frac{s}{2})}{s^2} \\ &\quad - \frac{I(x - \frac{s}{2}, y + \frac{s}{2})}{s^2} + \frac{I(x - \frac{s}{2}, y - \frac{s}{2})}{s^2}, \end{aligned} \quad (4)$$

where  $K(x, y)$  and  $L(x, y)$  are vertically and horizontally blurred images of  $I(x, y)$ , respectively:

$$K(x, y) = \int_{y-\frac{s}{2}}^{y+\frac{s}{2}} \frac{I(x, \tau)}{s} d\tau, \quad L(x, y) = \int_{x-\frac{s}{2}}^{x+\frac{s}{2}} \frac{I(\epsilon, y)}{s} d\epsilon. \quad (5)$$

The key is to interpret  $J_x(x, y)$  and  $J_y(x, y)$  as an aperture mask with two slits, and  $J_{xy}(x, y)$  as a mask with four pinholes as illustrated in Figures 1(b-c) and 2(b-c). One difference between (4) and the standard slit/pinhole masks is that the *sign* is negative for some terms in the derivative images.

## Detector Measurement And Noise Model

Recalling that detector has a finite pixel dimension, let

$$J[m, n] = J(m\Delta, n\Delta) + N[m, n] \quad (6)$$

be the discretized sensor measurements on the continuous image  $J(x, y)$ , where  $\Delta$  is the detector pitch and  $N[m, n]$  is the measurement noise. In the pixel domain, the derivative operators required to convert the square aperture mask into pinhole aperture mask must be approximated. We regard the discrete derivative images  $J_x[m, n]$ ,  $J_y[m, n]$ , and  $J_{xy}[m, n]$  as a result of the convolution (denoted by  $\star$ )

$$J_o[m, n] = J[m, n] \star G_o[m, n], \quad o \in \{x, y, xy\} \quad (7)$$

where  $G_o[m, n]$  is the discrete Gaussian derivative filter:

$$\begin{aligned} G_x[m, n] &= G_y[n, m] = \frac{-m}{\Delta\sigma^2\sqrt{2\pi}\sigma^2} e^{-\frac{m^2-n^2}{2\sigma^2}} \\ G_{xy}[m, n] &= G_x[m, n] \star G_y[m, n], \end{aligned} \quad (8)$$

with scale parameter  $\sigma^2$ . Examples of numerically computed gradient images are shown in Figure 4.

## Circular-Square Aperture Prototyping

A square aperture is a minimally disruptive hardware modification to the standard camera optics. Specifically, square aperture can be incorporated into a camera without “replacing” the circular aperture—meaning a “dual purpose” (i.e. circular-square) aperture can be designed with only a minor change to the standard hardware. As shown by Figure 3(a), variable size aperture mechanism has a diaphragm (or a collection of “blades” arranged in a circular pattern) that controls the diameter of an approximately circular lens opening. Figure 3(b) shows how a square

aperture can be formed by strategically rearranging a subset of the diaphragm blades (design #1). Alternatively, Figure 3(c) shows a lens system with circular aperture diaphragm blades housed inside a square aperture—diaphragm can open maximally to make way for the square aperture (design #2).<sup>1</sup> Pragmatic approach to implementation such as these gives square aperture a better chance of being incorporated into commercial products than the alternatives.

We developed a prototype for square aperture mask design #2 by modifying Nikkor 50mm f/1.8D lens. As shown in Figure 3(d), the 14mm × 14mm aperture mask was cut out of a paper and inserted next to the existing aperture diaphragm within the lens housing. One can verify that the prototype is behaving according to our specification by imaging an approximate point source (e.g. Figure 3(e)).

## Aperture Mask Comparisons

How appropriate are the choices of aperture masks for derivative-based processing? The relative advantage of square aperture is evident in Figure 4. While other aperture masks yield complex gradient images, where the details of the image are difficult to see, the gradient image of square aperture-blurred image clearly show superimposed sharp images. Drawing on the analysis in [11], gradient image of a blurred image is a sharp image blurred by “gradient blur.” By this account, the complex of the gradient image is directly commensurate to the complexity of the “gradient blur kernel,” also shown in Figure 4. Clearly, the square aperture enjoys the sparsest gradient blur, while the alternative aperture shapes and masks do not yield straightforward interpretation. Hence the square geometry is highly desirable for derivative-enabled coded aperture.

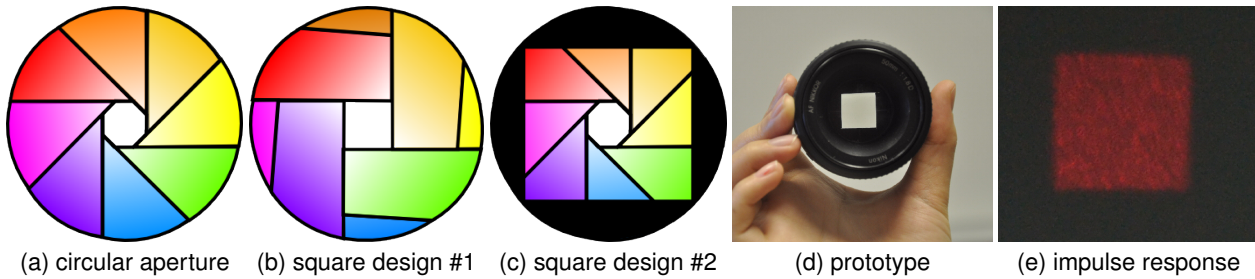
## Blur Size Estimation And Deblurring Noise Robust Blur Size Estimation

The key to recovering depth of an object is to determine the square of confusion size  $s$  from the pinhole or slit aperture images, obtained indirectly from the gradient images  $J_x[m, n]$ ,  $J_y[m, n]$ , and/or  $J_{xy}[m, n]$  in (4). Figure 5 shows a real-camera example of the square aperture image  $J[m, n]$  and the corresponding gradient images. The degree of blur is difficult to assess directly from  $J[m, n]$  since blur significantly degrades image features. As is expected from analysis in (4), however, the gradient operators restore the sharp transitions of the original image  $I[m, n]$ .

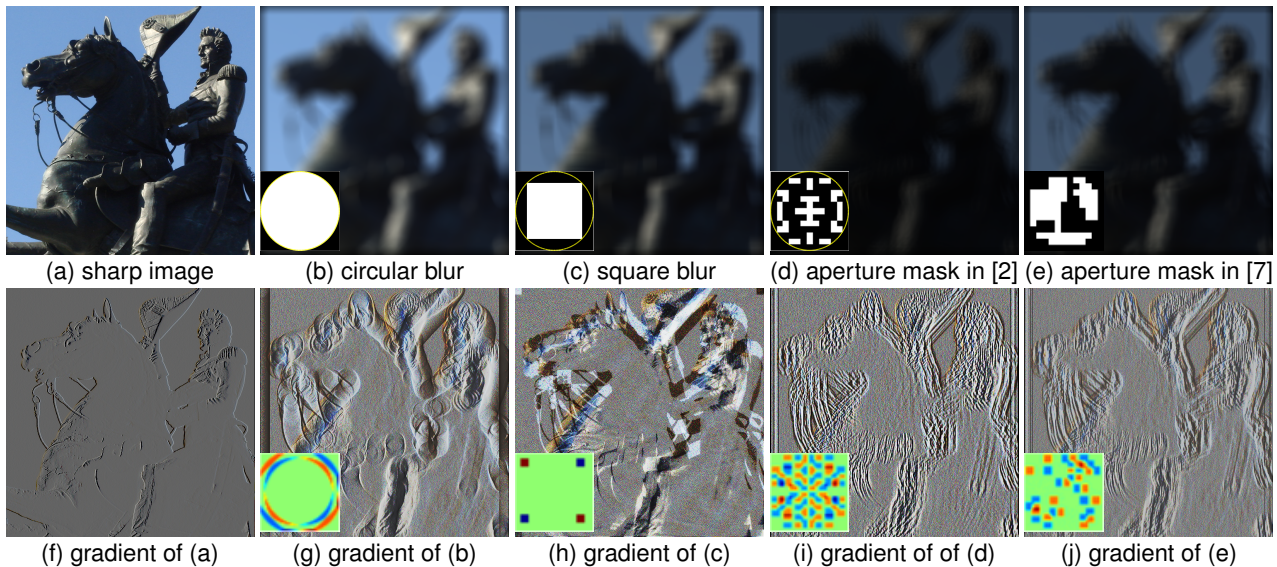
For the task of recovering  $s$ , we draw on the phase detection autofocus principles—a proven strategy for recovering blur size in DSLR cameras today. As illustrated in Figure 6, light rays passing at the top and the bottom of the lens are captured by autofocus sensors 2 and 1, respectively. When out-of-focus, captured sensor images are shifted (a.k.a. disparity in stereo matching) by an amount proportional to the blur size. Hence the task of recovering  $s$  simplifies to the problem of correlating the displaced edges.

We implemented a simple scheme to enable phase detection in square aperture images. Recalling (4),  $J_x[m, n]$  corresponds to two slit aperture images  $K[m \pm \frac{s}{2}, n]$  displaced in the horizontal direction by  $s$ . Since slit aperture camera has an long depth of field in the direction orthogonal to the slit direction, vertical edges in  $K[m, n]$  is preserved (see Figure 5(c)). To identify the horizontal

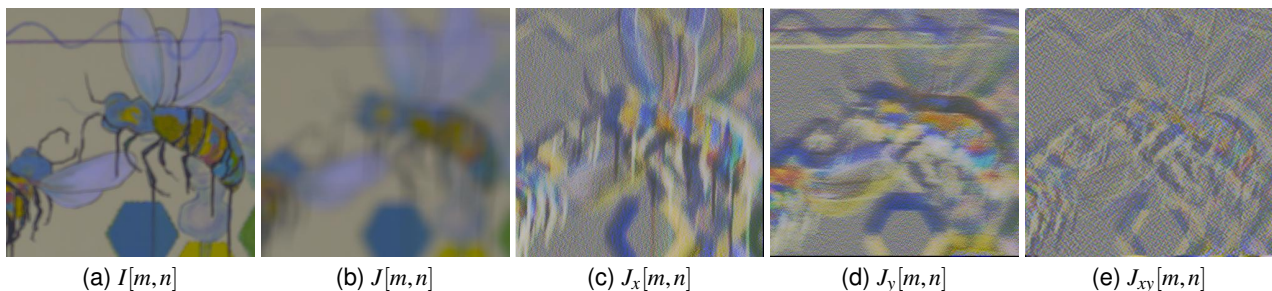
<sup>1</sup>Note that aperture mask of [8] is also compatible with design #2.



**Figure 3.** A minor modification to variable size aperture mechanism in a standard camera optics can form a square aperture. (a) Traditional circular aperture using a diaphragm. (b) Square aperture formed by strategically rearranging a subset of the diaphragm blades. (c) Square aperture revealed when circular aperture diaphragm is fully opened. (d) Prototype for design #2. (e) Sensor response to a laser pointer dot pointed on a wall verifies that the blur kernel is indeed square.



**Figure 4.** How appropriate is the aperture mask for gradient-based image deblurring? Sharp image in (a) is blurred by (b) conventional circular blur, (c) proposed square blur, (d) aperture mask in [2], and (e) aperture mask in [7]. (f-j) gradient image of (a-e) are shown with gradient aperture. It is clearly that sparser the gradient aperture, the better image features are represented in the gradient blurred image.



**Figure 5.** Example image captured by prototype camera. (a) Sharp image  $I[m,n]$  taken in focus. (b) Blurred image  $J[m,n]$  taken by prototype camera with square aperture mask. (c) Horizontal gradient image  $\frac{\partial}{\partial x} J[m,n] = J_x[m,n]$ . (d) Vertical gradient image  $\frac{\partial}{\partial y} J[m,n] = J_y[m,n]$ . (e) Horizontal-vertical gradient image  $\frac{\partial^2}{\partial x \partial y} J[m,n] = J_{xy}[m,n]$ . It is clear that gradient operator restores the sharp transitions of the original image  $I[m,n]$ .



disparity of vertical edges of  $K[m, n]$  in  $J_x[m, n]$ , we first apply a vertical derivative/edge filter (such as  $G_x[m, n]$  in (8)) on  $J_x[m, n]$  to eliminate non-vertical features:

$$\begin{aligned} J_{xx}[m, n] &:= J_x[m, n] * G_x[m, n] \\ &= \frac{K_x[m - \frac{s}{2}, n] - K_x[m + \frac{s}{2}, n]}{s} + N_{xx}[m, n], \end{aligned} \quad (9)$$

where  $K_x := \frac{d}{dx}K = K * G_x$  and  $N_{xx} := \frac{d^2}{dx^2}N = N * G_x * G_x$ . We then compute the *mean absolute sum* (MAS):<sup>2</sup>

$$\begin{aligned} \Phi[i, m, n] &= \mathbb{E} \left| J_{xx}[m, n] + J_{xx}[m + i, n] \right| \\ &= \left| \frac{K_x[m - \frac{s}{2}, n] - K_x[m + \frac{s}{2}, n] + K_x[m + i - \frac{s}{2}, n] - K_x[m + i + \frac{s}{2}, n]}{s} \right. \\ &\quad \left. + N_{xx}[m, n] + N_{xx}[m + i, n] \right|. \end{aligned} \quad (10)$$

Clearly, MAS attains minimum at  $i = \pm s[m, n]: \forall i \in \mathbb{Z}$ ,

$$\begin{aligned} \Phi[i, m, n] &\geq \Phi[\pm s, m, n] = \\ &\mathbb{E} \left| \frac{K_x[m \pm \frac{3s}{2}, n] - K_x[m \mp \frac{s}{2}, n]}{s} + N_{xx}[m, n] + N_{xx}[m \pm s, n] \right| \end{aligned} \quad (11)$$

because some  $K_x$  terms cancel. Hence we estimate the size of square of confusion  $s$  by

$$\hat{s}_x[m, n] = \arg \min_i \Phi[i, m, n]. \quad (12)$$

Similarly, estimation for the vertical gradient takes the form

$$\hat{s}_y[m, n] = \arg \min_i \Psi[i, m, n], \quad (13)$$

where  $\Psi$  is MAS for  $J_{yy}$ . The combined estimate is:

$$\begin{aligned} \hat{s}_\Phi[m, n] &= \\ &\begin{cases} \min(\hat{s}_x[m, n], \hat{s}_y[m, n]) & \text{if } |\hat{s}_x[m, n] - \hat{s}_y[m, n]| < \tau \\ 0 & \text{else,} \end{cases} \end{aligned} \quad (14)$$

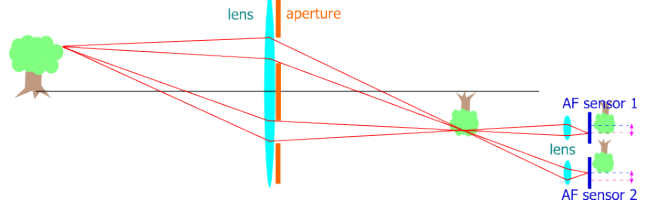
where  $\tau$  is a predefined threshold. As a post processing, we smooth the blur estimates by the graph cut method of [12] to regularizes the initial estimation of square confusion size  $\hat{s}_\Phi$ . We seek this regularized estimation  $\hat{s}$  by minimizing an energy function, similar to the one proposed in [12], [8], [2]:

$$\begin{aligned} \hat{s} &= \arg \min_s \sum_{[m, n] \in \mathbb{Z}^2} D_1(\hat{s}[m, n], \hat{s}_\Phi[m, n]) \\ &\quad + \gamma \sum_{[m, n] \in \mathbb{Z}^2, [i, j] \in \{\pm 1\}^2} D_2(\hat{s}[m, n], \hat{s}[m + i, n + j]) \end{aligned} \quad (15)$$

Where the  $\gamma$  is a weighting factor, and  $D_1$  and  $D_2$  are:

$$\begin{aligned} D_1(\hat{s}[m, n], \hat{s}_\Phi[m, n]) &= |\hat{s}[m, n] - \hat{s}_\Phi[m, n]| \\ D_2(\hat{s}[m, n], \hat{s}[m', n']) &= \\ &\begin{cases} \exp\left(\frac{-(J[m, n] - J[m', n'])^2}{\sigma^2}\right) & \text{if } J[m, n] \neq J[m', n'] \\ 0 & \text{else,} \end{cases} \end{aligned} \quad (16)$$

<sup>2</sup>Analogous to the *mean absolute difference* used in video compression.



**Figure 6.** Schematic for phase detection autofocus. Disparity between the images captured by AF sensors 1 and 2 (corresponding to light rays passing at the top and the bottom of the lens) is proportional to the blur size.

We solve this minimize problem by using the centers of clusters of initial estimation  $\hat{s}_\Phi$  as candidates for  $\hat{s}$ .

The proposed blur estimation enjoys robustness to noise for a number of reasons. Besides the fact that large square aperture allow more light than most coded aperture alternatives,  $J_{xx}$  and  $J_{yy}$  are essentially Gaussian second derivatives—we found that standard deviation of Gaussian function in (8) can be increased to nearly eliminate  $N_{xx}$  and  $N_{yy}$  without significant sacrifices to the quality of blur size estimation. Furthermore, the *expectation* in MAS defined in (10) is implemented as a spatial averaging, which further reduce the influence of noise. As a side note, the use of higher order derivative images ( $J_{xx}$  and  $J_{yy}$ ) is a unique aspect of the square aperture processing.

### Noise Robust Deblurring

We describe a method to recover the sharp image  $I$  from the small aperture camera in (4). At a high level, square blur  $H_{(x,y)}(x, y)$  is conceptually equivalent to a combination of horizontal blur and a vertical blur. Hence our strategy is to deblur in horizontal and vertical directions separately: we follow example of the double discrete wavelet transform (DDWT) of [11] which seamlessly combines any type of wavelet-based denoising algorithms with deblurring of 1D horizontal/vertical blurs. Combining (4) with 1D (horizontal) wavelet transform, we have:  $J_x^{x\ell}[m, n] =$

$$\frac{K^{x\ell}[m + \frac{s}{2}, n] - K^{x\ell}[m - \frac{s}{2}, n]}{s} + N_x^{x\ell}[m, n], \quad (17)$$

where superscript  $x\ell$  denotes  $\ell$ th subband horizontal ( $x$ ) wavelet. Then the horizontal direction wavelet coefficients  $K^{x\ell}$  is recovered by:<sup>3</sup>

$$\hat{K}^{x\ell}[m, n] = \hat{s} \cdot \text{absmin}\{J_x^{x\ell}[m - \frac{s}{2}, n], -\hat{s}J_x^{x\ell}[m + \frac{s}{2}, n]\}, \quad (18)$$

where  $\text{absmin}\{\cdot, \cdot\}$  assigns output to whichever is smaller:

$$\text{absmin}\{a, b\} = \begin{cases} a & \text{if } \|a\| < \|b\| \\ b & \text{otherwise.} \end{cases} \quad (19)$$

Inverse horizontal wavelet transform of  $\hat{K}^{x\ell}$  yields the image  $\hat{K}$ , an estimate of vertically blurred image  $K$  in (5):

$$\hat{K}[m, n] = K(m\Delta, n\Delta) + \hat{N}[m, n], \quad (20)$$

where  $\hat{N}$  is now a combination of measurement noise and residual error from estimation. Suppose we take a derivative  $\hat{K}_y = \frac{d}{dy}K$

<sup>3</sup>Optimal deblurring for sparse images with noise, according to [11].

(not to be confused with  $K_x$  in Section ):

$$\hat{K}_y[m, n] = \frac{I[m, n + \frac{s}{2}] - I[m, n - \frac{s}{2}]}{s} + \hat{N}_y[m, n],$$

where sharp image  $I[m, n]$  emerges from the fundamental theorem of calculus (recall (4) and (5)). We recover the vertical wavelet transform coefficients  $I^{y\ell}[m, n]$  (superscript  $y\ell$  denotes  $\ell$ th sub-band of vertical wavelet):

$$\hat{I}_{\text{noisy}}^{y\ell}[m, n] = \hat{s} \cdot \text{absmin}\{\hat{K}_y^{y\ell}[m, n - \frac{s}{2}], -\hat{K}_y^{y\ell}[m, n + \frac{s}{2}]\} \quad (21)$$

where  $\hat{\lambda}(\cdot)$  is any denoising scheme to suppress noise and residual error  $\hat{N}_y^{y\ell}$  (we used [13] in our work). Inverse wavelet transform of  $\hat{I}_K^{y\ell}$  recovers  $\hat{I}_K$ . We repeat the procedure (17)-(21), reversing the direction (use  $J_y^{y\ell}$  to recover  $L^{y\ell}$  in(18); use  $\hat{L}_x^{x\ell} = \frac{\partial}{\partial x} L_x$  to estimate  $\hat{I}_L^{x\ell}$ ; denoising and inverse wavelet transform yields  $\hat{I}_L$ ). We arrive at the final estimate  $\hat{I}$  by combining  $\hat{I}_K$  and  $\hat{I}_L$  in the wavelet domain:

$$\begin{aligned} \hat{I}^{x\ell}[m, n] &= \text{absmin}(\hat{I}_K^{x\ell}[m, n], \hat{I}_L^{x\ell}[m, n]) \\ \hat{I}^{y\ell}[m, n] &= \text{absmin}(\hat{I}_K^{y\ell}[m, n], \hat{I}_L^{y\ell}[m, n]) \end{aligned} \quad (22)$$

The above procedure is repeated for all wavelet levels  $\ell \in \{1, 2, \dots\}$ . In addition, the scaling coefficient of  $J$  is used as a proxy for scaling coefficient of  $I$ . Taking the inverse wavelet transform of  $\{\hat{I}^{x\ell}, \hat{I}^{y\ell}\}$  together with the scaling coefficients recovers the latent sharp image  $I[m, n]$ .

## Experiment

Figure 7(a) shows images captured by Nikon D90 with our square aperture prototype lens. All images were taken in a raw sensor mode and processed with demosaicking [14], white balancing, and color correction. The image was downsampled by factor of 4 after  $4 \times 4$  pixel neighborhood were averaged to suppress any demosaicking artifacts that may interfere with blur detection or deblurring. Gamma correction was applied after blur detection and deblurring.

The proposed blur size detection algorithm yields a dense map of blur size, shown in Figure 7(b). The estimated blur clearly separates the out-of-focus foreground from the in-focus background (or vice versa), making it possible to infer the distance between the camera and the objects in the scene by (1). Thanks in part to graph cut, object boundary shapes are well preserved in the blur size estimation.

The proposed deblurring is able to reproduce fine image details from a severely blurred input image. As evidenced in Figure 7(c), the reconstructed image is free of noticeable ringing artifacts and oversmoothing that plague most deblurring methods and coded apertures. Instabilities may still arise, however, if there are inaccuracies in the blur estimation boundaries (though not as pervasive as ringing).

Figure 8 compares the proposed square aperture to coded apertures in [2, 8]. Here, Figures 8(a-b) were acquired by Nikon D90 with the same circular-square aperture prototype lens—sharp image in Figure 8(a) was obtained with small circular aperture ( $f/22$ ), while blurry image of Figure 8(b) was with the square

aperture. Figures 8(a-b) are shown with  $4 \times 4$  downsampling after  $4 \times 4$  pixel averaging. Figure 8(e) is the result of deblurring the *actual* blurry image in Figure 8(b) directly. Blurry images of Figures 8(c-d) were simulated by convolving the aperture masks in [2, 8] with the *full resolution* version of Figures 8(a), and then downsampled in the same way as before (no noise added). Figures 8(f-g) are the result of deblurring the images in Figures 8(c-d). We considered the “aperture/blur size” for each coded aperture to be the radius of the smallest circle that fits around the coded aperture—radius in Figures 8(b-d) was 13 pixels (after downsampling). Though far from perfect, the deblurring results in Figures 8(e-g) suggests that the square aperture clearly resolves finer image details than the alternatives.

Figure 9 shows a simulation study to understand the accuracy of depths estimated from recovered blur size. Assumed optics matched our prototype modified Nikkor lens and Nikon D90: aperture size of 19.7mm ( $=14\text{mm} \cdot \sqrt{2}$ ), pixel pitch of  $5.5\mu\text{m}$ , focal length of 50mm. The camera was assumed to be “in-focus” at 2m from the camera and out of focus objects were assumed to be behind the depth of field. To simulate image capture under various lighting conditions, we simulated the blurred image as:

$$J[m, n] = (\alpha I[m, n]) \star H[m, n] + N[m, n], \quad (23)$$

where the sharp image  $I$  is in the ranges of  $[0, 1]$ ,  $N$  is additive white Gaussian noise ( $\sigma = 0.01$ ), and  $H$  is coded aperture blur with radius equal to 4.4mm on the sensor plane. Smaller value of  $\alpha$  implies higher noise. Figure 9 clearly shows that the square aperture results in smaller depth estimation error even when the influence of noise is high.

## Conclusion

We proposed a square aperture as a simple and practical alternative to existing coded aperture patterns. Square aperture shares the properties of both large and small apertures, yielding excellent light efficiency while preserving fine image details. It follows from the idea that aperture can be made sparse by a derivative operator, which simplifies the blur size estimation and image deblurring. Testing with prototype lens confirmed the feasibility of our approach. The square aperture was superior to the previously proposed coded aperture patterns in terms of depth estimation and image deblurring, even under the influence of noise.

## References

- [1] Edward E Fenimore and Thomas M Cannon, “Coded aperture imaging with uniformly redundant arrays,” *Applied optics*, vol. 17, no. 3, pp. 337–347, 1978.
- [2] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman, “Image and depth from a conventional camera with a coded aperture,” *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 70, 2007.
- [3] Alex Paul Pentland, “A new sense for depth of field,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 4, pp. 523–531, 1987.
- [4] Muralidhara Subbarao and Natarajan Gurumoorthy, “Depth recovery from blurred edges,” in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR’88., Computer Society Conference on*. IEEE, 1988, pp. 498–503.
- [5] Paul Grossmann, “Depth from focus,” *Pattern Recognition Letters*, vol. 5, no. 1, pp. 63–69, 1987.

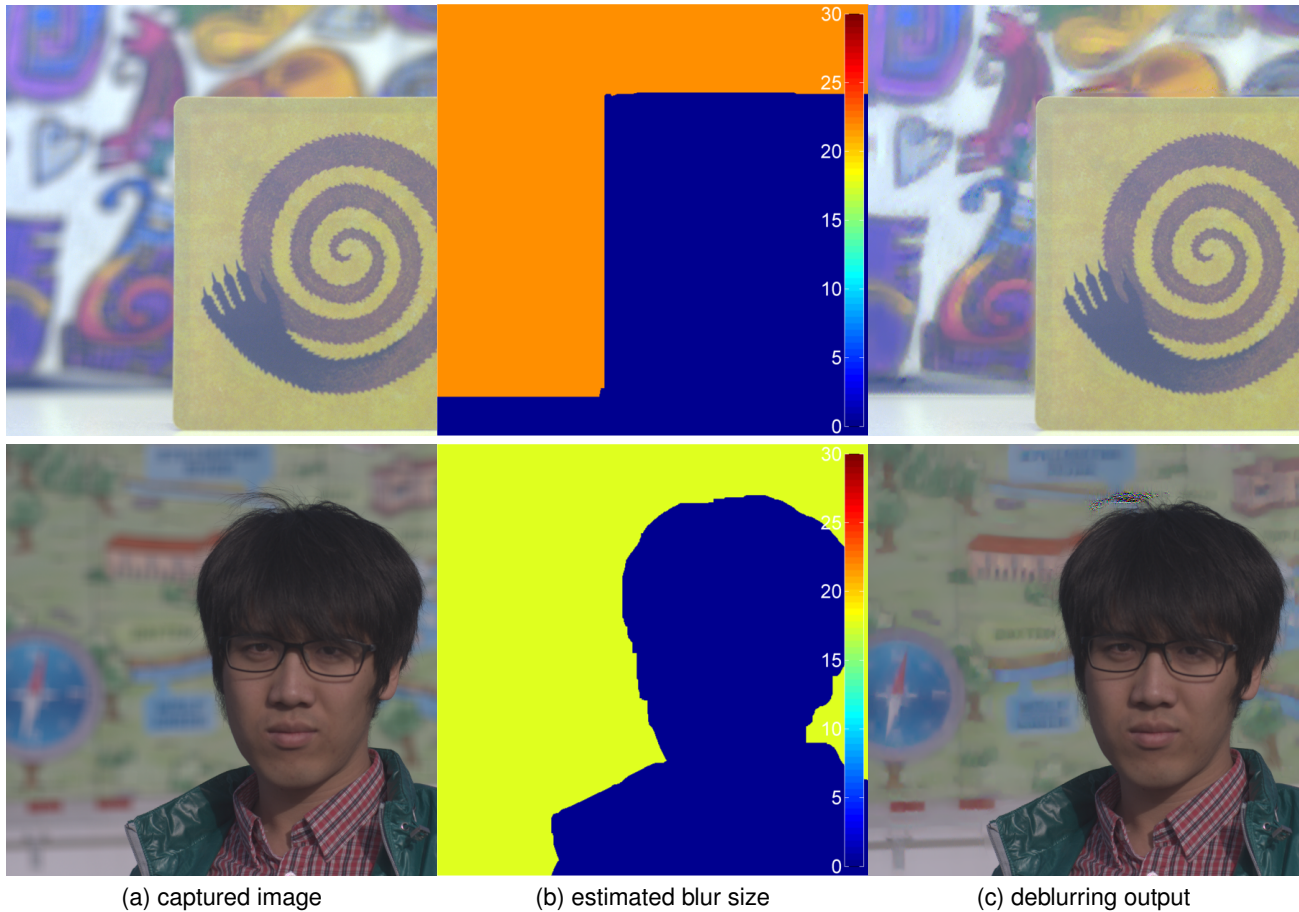


Figure 7. Real camera experiment using prototype square aperture lens system.

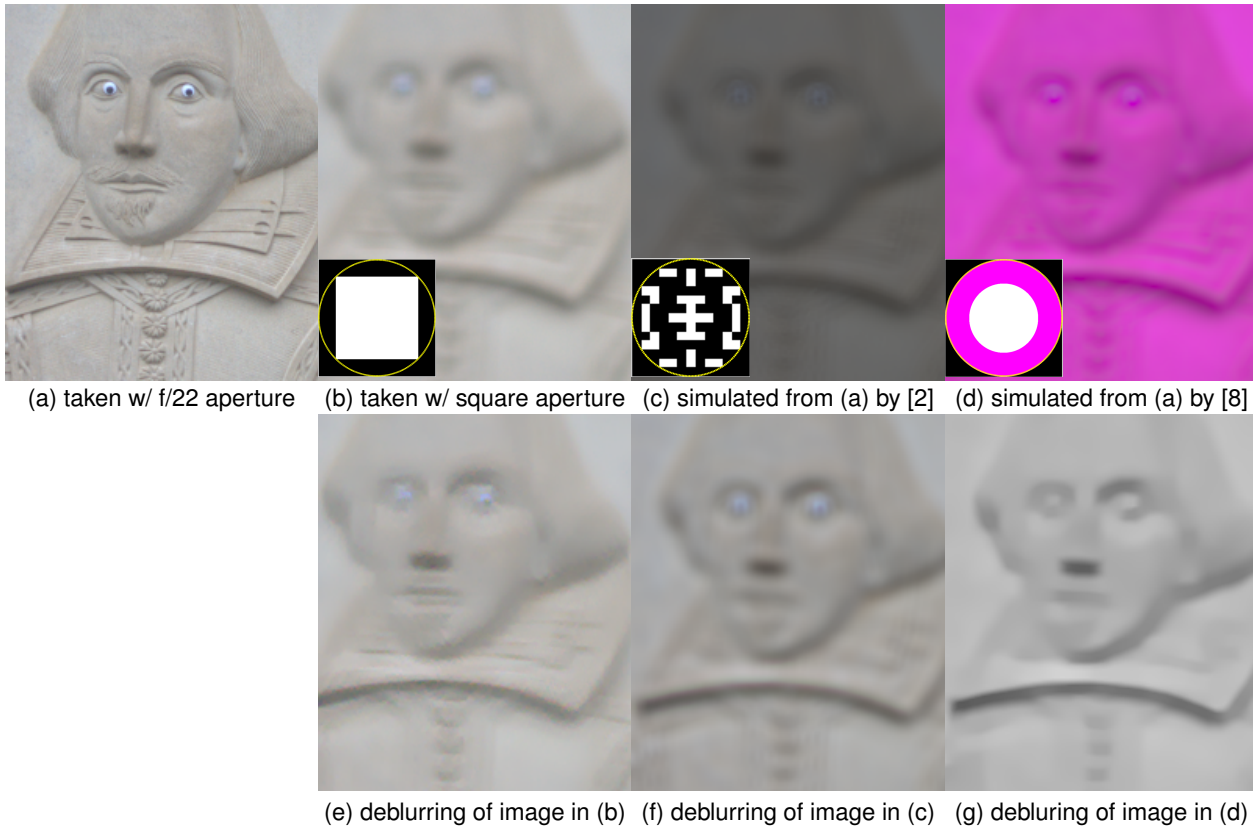
- [6] Roummel F Marcia and Rebecca M Willett, "Compressive coded aperture superresolution image reconstruction," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 833–836.
- [7] Changyin Zhou, Stephen Lin, and Shree Nayar, "Coded aperture pairs for depth from defocus," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 325–332.
- [8] Ayan Chakrabarti and Todd Zickler, "Depth and deblurring from a spectrally-varying depth-of-field," in *Computer Vision—ECCV 2012*, pp. 648–661. Springer, 2012.
- [9] Horacio E Fortunato and Manuel M Oliveira, "Coding depth through mask structure," in *Computer Graphics Forum*. Wiley Online Library, 2012, vol. 31, pp. 459–468.
- [10] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 69, 2007.
- [11] Yi Zhang and Keigo Hirakawa, "Blur processing using double discrete wavelet transform," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1091–1098.
- [12] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [13] Keigo Hirakawa and Patrick J Wolfe, "Skellam shrinkage: Wavelet-based intensity estimation for inhomogeneous poisson data," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1080–1093, 2012.
- [14] Keigo Hirakawa and Thomas W Parks, "Adaptive homogeneity-directed demosaicing algorithm," *Image Processing, IEEE Transactions on*, vol. 14, no. 3, pp. 360–369, 2005.

## Author Biography

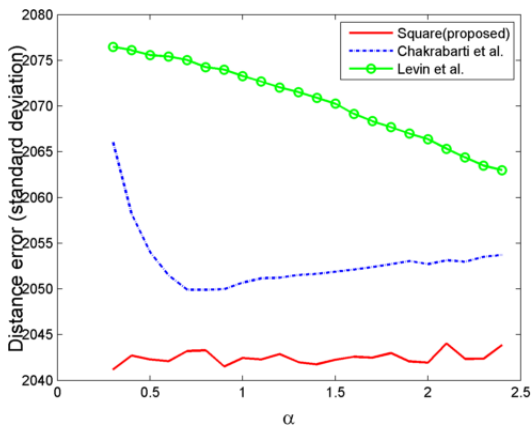
Ruojun He received B.E. from Nanjing University of Science and Technology in 2010 and M.S. in Electrical and Computer Engineering from the University of Dayton in 2014.

Yi Zhang received B.E. in Electronic Science and Technology from Nanjing University of Information Science and Technology in 2008, M.S. in Electrical and Computer Engineering from the University of Dayton in 2010, and PhD from the University of Dayton in 2015.

Keigo Hirakawa is an Associate Professor at the University of Dayton and the Director for the Computer Engineering Program. Prior to joining UD in 2010, he was with Harvard University as a Research Associate of Department of Statistics and School of Engineering and Applied Sciences. He received Ph.D in Electrical and Computer Engineering from Cornell University, B.S. in Electrical Engineering from Princeton University, and M.M. in Jazz Performance from New England Conservatory of Music.



**Figure 8.** Comparisons of the proposed square aperture to the coded apertures of [2, 8]. (a) Real camera image taken with small ( $f/22$ ) circular aperture. (b) Real camera image captured by the prototype square aperture. (c-d) Blurry images simulated from (a). (e-g) Deblurring results of (b-d). The smallest circle fitting around apertures in (b-d) has the radius of 13 pixels.



**Figure 9.** Light efficiency vs. accuracy of scene depth estimation for proposed and aperture masks of [2, 8]. The error standard deviation is in millimeters. Smaller  $\alpha$  implies higher noise.