Coded Aperture Pairs for Depth from Defocus

3 authors, including:

Changyin Zhou
Columbia University
15 PUBLICATIONS   **1,111** CITATIONS

Coded Aperture Pairs for Depth from Defocus

Changyin Zhou
Columbia University
New York City, U.S.
changyin@cs.columbia.edu

Stephen Lin
Microsoft Research Asia
Beijing, P.R. China
stevelin@microsoft.com

Shree Nayar
Columbia University
New York City, U.S.
nayar@cs.columbia.edu

Abstract

The classical approach to depth from defocus uses two images taken with circular apertures of different sizes. We show in this paper that the use of a circular aperture severely restricts the accuracy of depth from defocus. We derive a criterion for evaluating a pair of apertures with respect to the precision of depth recovery. This criterion is optimized using a genetic algorithm and gradient descent search to arrive at a pair of high resolution apertures. The two coded apertures are found to complement each other in the scene frequencies they preserve. This property enables them to not only recover depth with greater fidelity but also obtain a high quality all-focused image from the two captured images. Extensive simulations as well as experiments on a variety of scenes demonstrate the benefits of using the coded apertures over conventional circular apertures.

1. Introduction

Recent advances in computational photography have given rise to a new breed of digital imaging tools. By acquiring greater or more informative scene data, various forms of post-capture photo processing can be applied to improve image quality or alter scene appearance. This approach has made operations such as depth-based image editing, refocusing and viewpoint adjustment feasible. Many of these operations rely on the explicit or implicit recovery of 3D scene geometry.

One approach to recovering 3D scene geometry that has received renewed attention in recent years is depth from defocus (DFD). For a given camera setting, scene points that lie on a focal plane located at a certain distance from the lens will be correctly focused onto the sensor, while points at greater distances away from this focal plane will appear increasingly blurred due to defocus. By capturing two images at camera settings with different defocus characteristics, one can infer the depth of each point in the scene from their relative defocus. Relative to other image-based shape reconstruction approaches such as multi-view stereo, structure from motion, range sensing and structured lighting,

depth from defocus is more robust to occlusion and correspondence problems [12].

Since defocus information was first used for depth estimation in the early 1980's [8][13], various techniques for DFD have been proposed based on changes in camera settings. Most commonly, DFD is computed from two images acquired from a fixed viewpoint with different aperture sizes (e.g., [7] [10] [16] [3]). Since the lens and sensor are fixed, the focal plane remains the same for both images. The image with a larger aperture will exhibit greater degrees of defocus with respect to given distances from the focal plane, and this difference in defocus is exploited to estimate depth.

The relative defocus is fundamentally influenced by the shape of the camera aperture. Though most DFD methods employ conventional lenses whose apertures are circular, other aperture structures can significantly enhance the estimation of relative defocus and hence improve depth estimation. In this work, we propose a comprehensive framework of evaluating aperture pairs for DFD, and use it to solve for an optimized pair of apertures. First, we formulate DFD as finding a depth d that minimizes a cost function $E(d)$, whose form depends upon the aperture patterns of the pair. Based on this formulation, we then solve for the aperture pair that yields a function $E(d)$ with a more clearly defined minimum at the ground truth depth d^*, which leads to higher precision and stability of depth estimation. Note that there exist various other factors that influence the depth estimation function $E(d)$, including scene content, camera focus settings, and even image noise level. Our proposed evaluation criterion takes all these factors into account to find an aperture pair that provides improved DFD performance.

Solving for an optimized aperture pattern is a challenging problem – for a binary pattern of resolution $N \times N$, the number of possible solutions for an aperture is $2^{N \times N}$. This problem is made harder by the fact that the aperture evaluation criterion is formulated in the Fourier domain and the transmittance values of the aperture patterns are physically constrained to lie between 0 and 1. To make this problem more tractable, existing methods [18][15][5] have limited
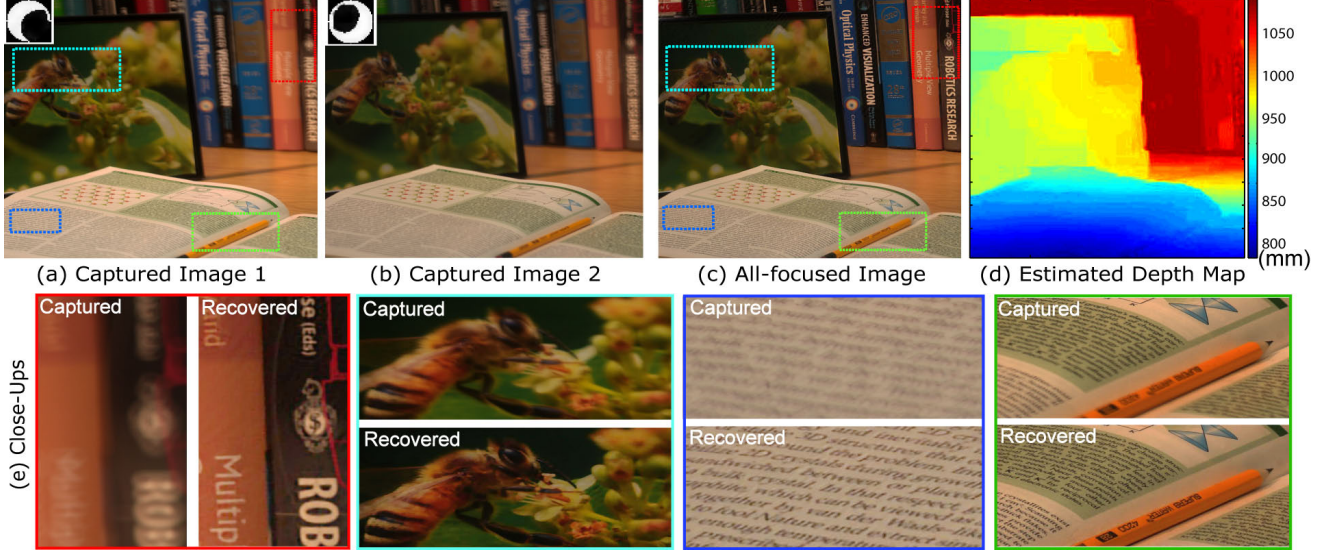
Figure 2. Depth from defocus and out-of-focus deblurring using coded aperture pairs. (a-b) Two captured images using the optimized coded aperture pair. The corresponding aperture pattern is shown at the top-left corner of each image. (c) The recovered all-focused image. (d) The estimated depth map. (e) Close-ups of four regions in the first captured image and the corresponding regions in the recovered image. Note that the bee and flower within the picture frame (light blue box) are out of focus in the actual scene and this blur is preserved in the computed all-focused image. For all the other regions (red, blue, and green boxes) the blur due to image defocus is removed.
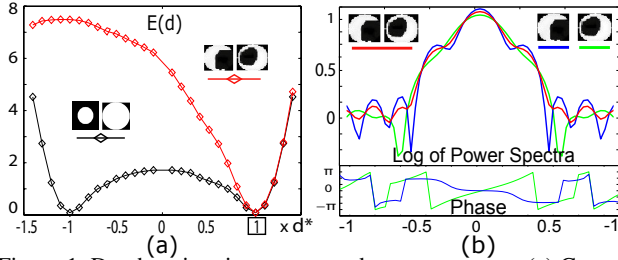

Figure 1. Depth estimation curves and pattern spectra. (a) Curves of $E(d)$ for the optimized coded aperture pair (red) and the conventional large/small circular aperture pair (black). The sign of the x-axis indicates if a scene point is farther or closer than the focus plane. (b) Top: Log of combined power spectra of the optimized coded aperture pair (red), as well as the power spectra of each single coded aperture (green and blue). Bottom: Phases of the Fourier spectra of the two coded apertures.

the pattern resolution to 13×13 or lower. However, solutions at lower resolutions are less optimal due to limited flexibility. To address the aperture resolution issue, we propose a novel recursive pattern optimization strategy that incorporates a genetic algorithm [18] with gradient descent search. This algorithm yields optimized solutions with resolutions of 33×33 or higher within a reasonable computation time. Although higher resolutions usually mean greater diffraction effects, in this particular case, we find that a high-resolution pattern of 33×33 suffers less from diffractions than other lower resolution patterns do.

Figure 1(a) displays profiles of the depth estimation function $E(d)$ for the optimized pair and for a pair of conventional circular apertures. The optimized pair exhibits a profile with a more pronounced minimum, which leads to depth estimation that has lower sensitivity to image noise and greater robustness to scenes with subtle texture. In addition, our optimized apertures are found to have complementary power spectra in the frequency domain, with zero-crossings located at different frequencies for each of the two apertures, as shown in Figure 1(b). Owing to this property, the two apertures thus jointly provide broadband coverage of the frequency domain. This enables us to also compute a high quality all-focused image from the two captured defocused images.

We demonstrate via simulations and experiments the benefits of using an optimized aperture pair over other aperture pairs, including circular ones. Our aperture pair is able to not only produce depth maps of significantly greater accuracy and robustness, but also produces high-quality all-focused images (see Figure 2 for an example.)

2. Related Work

Single Coded Apertures Coded apertures have recently received much attention. In [15] and [18], coded apertures are used to improve out-of-focus deblurring. To achieve this goal, the coded apertures are designed to be broadband in the Fourier domain. In [18], a detailed analysis of how aperture patterns affect deblurring is done. Based on this analysis, a closed-form criterion for evaluating aperture patterns is proposed. In our work, we employ a methodology similar to [18], but our goal is to derive an aperture pair that is optimized for depth from defocus.

To improve depth estimation, Levin *et al.* [5] proposed using an aperture pattern with a more distinguishable pattern of zero-crossings in the Fourier domain than that of the conventional circular apertures. Similarly, Dowski [1] designed a phase plate that has responses at only a few frequencies, which makes their system more sensitive to depth variations. These methods specifically target depth estimation from a single image, and rely heavily on specific frequencies and image priors. A consequence of this strong dependence is that they become sensitive to image noise and cannot distinguish between a defocused image of a sharp texture and a focused image of smoothly varying texture. Moreover, these methods compromise frequency content during image capture, which degrades the quality of image deblurring.

A basic limitation of using a single coded aperture is that aperture patterns with a broadband frequency response are needed for optimal defocus blurring but are less effective for depth estimation [5], while patterns with zero-crossings in the Fourier domain yield better depth estimation but exhibit a loss of information for deblurring. Figure 3 exhibits this trade-off using the aperture designed for depth estimation in [5] and the aperture for deblurring in [18]. Since high-precision depth estimation and high-quality defocus deblurring generally cannot be achieved together with a single image, we address this problem by taking two images with different coded apertures optimized to jointly obtain a high-quality depth map and an all-focused image, as shown in Figure 2.
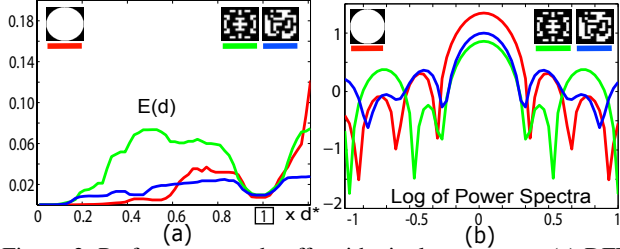

Figure 3. Performance trade-offs with single apertures. (a) DFD energy function profiles of three patterns: circular aperture (red), coded aperture of [5] (green), and coded aperture of [18] (blue). (b) Log of power spectra of these three aperture patterns. The method of [5] provides the best DFD, because of its distinguishable zero-crossings and its clearly defined minimum in the DFD energy function. On the other hand, the aperture of [18] is best for defocus deblurring because of its broadband power spectrum, but is least effective for DFD due to its less pronounced energy minimum, which makes it more sensitive to noise and weak scene textures.

Multiple Coded Apertures Multiple images with different coded apertures were used for DFD in [2] [4]. In [2], two images are taken with two different aperture patterns, one being Gaussian and the other being the derivative of a Gaussian. These patterns are such designed so that depth

estimation involves only simple arithmetic operations, making it suitable for real-time implementation. Hiura and Matsuyma[4] aims for more robust DFD by using a pair of pinhole apertures within a multi-focus camera. The use of pinhole pairs facilitates depth measurement. However, this aperture coding is far from optimal. Furthermore, small apertures significantly restrict light flow to the sensor, resulting in considerable image noise that reduces depth accuracy. Long exposures can be used to increase light flow but will result in other problems such as motion blur.

In related work, Liang *et al.* [6] proposed to take tens of images by using a set of Hadamard-code based aperture patterns for high-quality light field acquisition. From the parallax effects present within the measured light field, a depth map is computed by multi-view stereo. In contrast, our proposed DFD method can recover a broad depth range as well as a focused image of the scene by only capturing two images.

3. Aperture Pair Evaluation

3.1. Formulation of Depth from Defocus

For a simple fronto-planar object, its out-of-focus image can be expressed as

$$f = f_0 \otimes k(d) + \eta, \tag{1}$$

where f_0 is the latent in-focus image, η is the image noise which is assumed to be Gaussian white noise $N(0, \sigma^2)$, and k is the point spread function (PSF) whose shape is determined by the aperture and whose size d is related to the depth. In this paper, the sign of blur size d indicates if a scene point is farther or closer than the focal plane. For a specific setting, there is a one-one mapping from the blur size to the depth. By estimating the size of defocus blur from the image, we can infer the depth. The above equation can written in the frequency domain as $F = F_0 \cdot K(d) + \zeta$, where F_0, K, and ζ are the discrete Fourier transforms of f_0, k, and η, respectively.

A single defocused image is generally insufficient for inferring scene depth without additional information. For example, one cannot distinguish between a defocused image of sharp texture and a focused image of smoothly varying texture. To resolve this ambiguity, two (or more) images $F_i, i = 1, 2$ of a scene are conventionally used, with different defocus characteristics or PSFs for each image:

$$F_i = F_0 \cdot K_i^{d^*} + \zeta_i, \tag{2}$$

where $K_i^{d^*}$ denotes the Fourier transform of the ith PSF with the actual blur size d^*. Our objective is to find the size $\hat{d}$ and deblurred image $\hat{F}_0$ by solving a maximum a posteriori (MAP) problem:

$$
\begin{aligned}
< \hat{d}, \hat{F}_0 > \quad &\propto \quad \arg\max P(F_1, F_2 | \hat{d}, \hat{F}_0, \sigma) P(\hat{d}, \hat{F}_0) \\
&= \quad \arg\max P(F_1, F_2 | \hat{d}, \hat{F}_0, \sigma) P(\hat{F}_0). \quad (3)
\end{aligned}
$$

According to Equation 2, we have

$$P(F_1, F_2|\hat{d}, \hat{F}_0, \sigma) \propto \exp\{-\frac{1}{2\sigma^2}\sum_{i=1,2}\|\hat{F}_0 \cdot K_i^{\hat{d}} - F_i\|^2\},$$
(4)

and our prior assumes the weighted latent focused image $\Psi \cdot \hat{F}_0$ follows a Gaussian distribution with zero mean:

$$P(\hat{F}_0) \propto \exp\{-\frac{1}{2}\|\Psi \cdot \hat{F}_0\|^2\},$$
(5)

where Ψ is the matrix of weights. Note that different choices of Ψ lead to different image priors. For example, it is a simple Tikhonov regularization when Ψ takes a constant scalar value; and it becomes the popular Gaussian prior of image derivatives when Ψ is the derivative filter in the Fourier domain.

Then, blur size is estimated as the $\hat{d}$ that maximizes:

$$P(\hat{d}|F_1, F_2, \sigma) = \max_{\hat{F}_0} P(\hat{F}_0, \hat{d}|F_1, F_2, \sigma).$$
(6)

Expressed as a logarithmic energy function, the problem becomes the minimization of

$$E(\hat{d}|F_1, F_2, \sigma) = \min_{F_0} \sum_{i=1,2}\|\hat{F}_0 \cdot K_i^{\hat{d}} - F_i\|^2 + \|C \cdot \hat{F}_0\|^2,$$
(7)

where $C = \sigma \cdot \Psi$. Rather than assigning a specific value, we will optimize C by making use of the $1/f$ law [17].

3.2. Generalized Wiener Deconvolution

For a given $\hat{d}$, solving $\partial E / \partial \hat{F}_0 = 0$ yields

$$\hat{F}_0 = \frac{F_1 \cdot \bar{K}_1^{\hat{d}} + F_2 \cdot \bar{K}_2^{\hat{d}}}{|K_1^{\hat{d}}|^2 + |K_2^{\hat{d}}|^2 + |C|^2},$$
(8)

where $\bar{K}$ is the complex conjugate of K and $|X|^2 = X \cdot \bar{X}$. As in [18], C can be optimized as $\sigma/A^{\frac{1}{2}}$, where A is defined over the power distribution of natural images according to the $1/f$ law [17]: $A(\xi) = \int_{F_0} |F_0(\xi)|^2 \mu(F_0)$. Here, ξ is the frequency and $\mu(F_0)$ is the possibility measure of the sample F_0 in the image space.

Equation (8) can be regarded as a generalized Wiener deconvolution which takes two input defocused images, each with a different PSF, and outputs one deblurred image. This algorithm yields much better deblurring results than only deconvolving one input image[14][5][11]. Note that a similar deconvolution algorithm was derived using a simple Tikhonov regularization in [9]. In addition, this deconvolution method can be easily generalized for the multiple-image case as:

$$\hat{F}_0 = \frac{\Sigma_i F_i \cdot \bar{K}_i^{\hat{d}}}{\Sigma_i |K_i^{\hat{d}}|^2 + |C|^2},$$
(9)

3.3. Selection Criterion

Based on the above formulation of DFD, we seek a criterion for selecting an aperture pair that yields precise

and reliable depth estimates. For this, we first derive $E(d|K_1^{d^*}, K_2^{d^*}, \sigma, F_0)$ by substituting Equations (2) and (8) into Equation (7). Note that the estimate d is related to the unknown F_0 and the noise level σ. We can integrate out F_0 by using the $1/f$ law of natural images as done in [18]:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) = \int_{F_0} E(d|K_1^{d^*}, K_2^{d^*}, \sigma, F_0)\mu(F_0).$$

This equation can be rearranged and simplified to get

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) = \sum_\xi \frac{A \cdot |K_1^d \cdot K_2^{d^*} - K_2^d \cdot K_1^{d^*}|^2}{\sum_i |K_i^d|^2 + C^2}$$
$$+ \sigma^2 \cdot \sum_\xi [\frac{C^2}{\sum_i |K_i^d|^2 + C^2} + 1],$$
(10)

which is the energy corresponding to a hypothesized depth estimate given the aperture pair, focal plane and noise level.

The first term of Equation (10) measures inconsistency between the two defocused images when the estimated blur size d deviates from the ground truth d^*. This term will be zero if $K_1 = K_2$ or $d = d^*$. The second term relates to exaggeration of image noise.

Depth can be estimated with greater precision and reliability if $E(d|K_1^{d^*}, K_2^{d^*}, \sigma)$ increases significantly when the estimated blur size d deviates from the ground truth d^*. To ensure this, we evaluate the aperture pair (K_1, K_2) at d^* and noise level σ using

$$R(K_1, K_2|d^*, \sigma)$$
$$= \min_{d \in \mathcal{D}/d^*} E(d|K_1^{d^*}, K_2^{d^*}, \sigma) - E(d^*|K_1^{d^*}, K_2^{d^*}, \sigma)$$
$$= \min_{d \in \mathcal{D}/d^*} \sum_\xi A \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{\sum_i |K_i^d|^2 + C^2}$$
$$+ \frac{\sigma^4}{A} \cdot \frac{\sum_i |K_i^{d^*}|^2 - \sum_i |K_i^d|^2}{(\sum_i |K_i^d|^2 + C^2) \cdot (\sum_i |K_i^{d^*}|^2 + C^2)}$$
(11)
$$\approx \min_{d \in \mathcal{D}/d^*} \sum_\xi A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + K_2^d|^2 + C^2},$$
(12)

where $\mathcal{D} = \{c_1 d^*, c_2 d^*, \ldots, c_l d^*\}$ is a set of blur size samples. In our implementation, $\{c_i\}$ is set to $\{0.1, 0.15, \ldots, 1.5\}$.

According to the derivations, this criterion for evaluating aperture pairs is dependent on ground truth blur size d^* (or object distance) and noise level σ. However, this dependence is actually weak. Empirically, we have found Equation (11) is dominated by the first term, and C to be negligible in comparison to the other factors. As a result, Equation (11) can be approximated by (12) and is relatively insensitive to the noise level, such that the dependence on σ can be disregarded in the aperture pair evaluation (σ is taken to be 0.005 throughout this paper). Also, we note that differences in d^* correspond to variations in PSF size, which can be regarded as equivalent to scaling the image itself. Since the matrix A is basically scale-invariant according to the $1/f$ law [17], aperture pair evaluation is also insensitive to d^*. This insensitivity to d^* indicates that our evaluation criterion works equally well for different scene depths.

Discussion For optimal DFD performance with an aperture pair, the pair must maximize the relative defocus between the two images. The relative defocus depends on differences in amplitude and phase in the spectra of the two apertures. DFD is most accurate when the two power spectra are complementary, such that their phases are orthogonal and a zero-crossing (R1) for one aperture corresponds to a large response (R2) at the same frequency for the other aperture. Intuitively, this is because the ratio of their spectra (R2/R1) would have a more significant peak, which can be detected more robustly in the presence of noise and weak textures. The position of this detected peak indicates the scale of defocus blur, which in turn is related to depth.

With the selection criterion given by Equation (12), our method accounts for the following properties. Equation (12) is maximized when K_1 and K_2 have complementary power spectra in both magnitude and phase. Optimizing the aperture patterns according to this criterion maximizes DFD performance.

4. Optimization of Aperture Pair Patterns

Solving for optimal aperture patterns is known to be a challenging problem [5][15][18]. Our problem is made harder since we are attempting to solve for a pair of apertures rather than a single aperture. For a binary pattern pair of resolution $N \times N$, the number of possible solutions is $4^{N \cdot N}$. To solve this problem, we propose a two-step optimization strategy.

In the first step, we employ the genetic algorithm proposed in [18] to find the optimized aperture according to Equation (12) at a low resolution of 11x11. The result is shown in the first column of Figure 4. Despite the high efficiency of this genetic algorithm, we found it to have difficulties in converging at higher resolutions.

As mentioned in Section 3.3, the optimality of an aperture pair is invariant to scale. Therefore, scaling up the optimized pattern pair yields an approximation to the optimal pattern pair at a higher resolution. This approximation provides a reasonable starting point for gradient descent search. Thus, in the second step we scale up the 11×11 solution to 13×13 and then obtain a solution of resolution 13×13 by gradient descent optimization. This process is repeated until reaching a resolution of 33×33. The evolution of the optimized aperture pair through this process is shown in Figure 4. The final optimized aperture pair of size 33×33 is not only superior to the solution at 11×11 in terms of the aperture pair criterion in Equation (12), but also produces less diffraction because of greater smoothness in the pattern.

Figure 1(a) shows the depth estimation curves $E(d|K_1, K_2)$, for our optimized pair and a pair of conventional circular apertures. We can see the curves for the optimized pair are much steeper. This leads to depth estimation that is more precise and more robust to noise

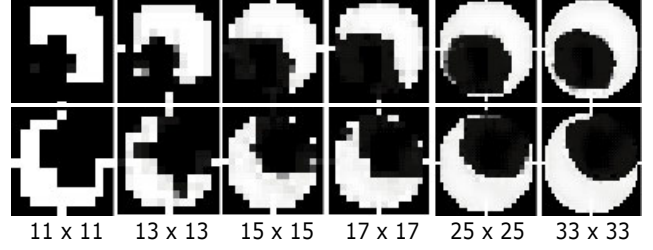| 11 x 11 | 13 x 13 | 15 x 15 | 17 x 17 | 25 x 25 | 33 x 33 |

Figure 4. Increasing the resolution of an optimized aperture pair by upsampling and gradient search.

and scene variations in practice. It is also confirmed that the curve $E(d)$ is insensitive to the blur size.

As we have shown in Figure 1(b), each of our optimized coded apertures has a distinct pattern of zero-crossings. Moreover, there is a large phase displacement between the two apertures that aids DFD. At the same time, the two apertures together preserve the full range of frequencies, which is essential for precise deblurring.

5. Recovery of Depth and All-Focused Image

With the optimized aperture pair, we use a straightforward algorithm to estimate the depth map U and recover the latent all-focused image I. For each sampled depth value $d \in \mathcal{D}$, we compute $\hat{F}_0^{(d)}$ according to Equation (8) and then reconstruct two defocused images. At each pixel, the residual $W^{(d)}$ between the reconstructed images and the observed images gives a measure of how close d is to the actual depth d^*:

$$W^{(d)} = \sum_{i=1,2} |IFFT(\hat{F}_0^{(d)} * K_i^{\hat{d}} - F_i)|, \qquad (13)$$

where IFFT is the 2D inverse Fourier transform. With our optimized aperture pairs, the value of $W^{(d)}(x,y)$ reaches an obvious minimum for pixel (x,y) if d is equal to the real depth. Then, we can obtain the depth map U as

$$U(x,y) = \arg\min_{d \in \mathcal{D}} W^{(d)}(x,y), \qquad (14)$$

and then recover the all-focused image I as

$$I(x,y) = \hat{F}_0^{(U_{x,y})}(x,y). \qquad (15)$$

The most computationally expensive operation in this algorithm is the inverse Fourier transform. Since it is $O(N \log(N))$, the overall computational complexity of recovering U and I is $O(l \cdot N log(N))$, where l is the number of sampled depth values and N is the number of image pixels. With this complexity, real-time performance is possible. In our Matlab implementation, this algorithm takes 15 seconds for a defocused image pair of size 1024×768 and 30 sampled depth values. Greater efficiency can be gained by simultaneously processing different portions of the image pair in multiple threads.

6. Performance Analysis

To quantitatively evaluate the optimized coded aperture pair, we conducted experiments on a synthetic staircase
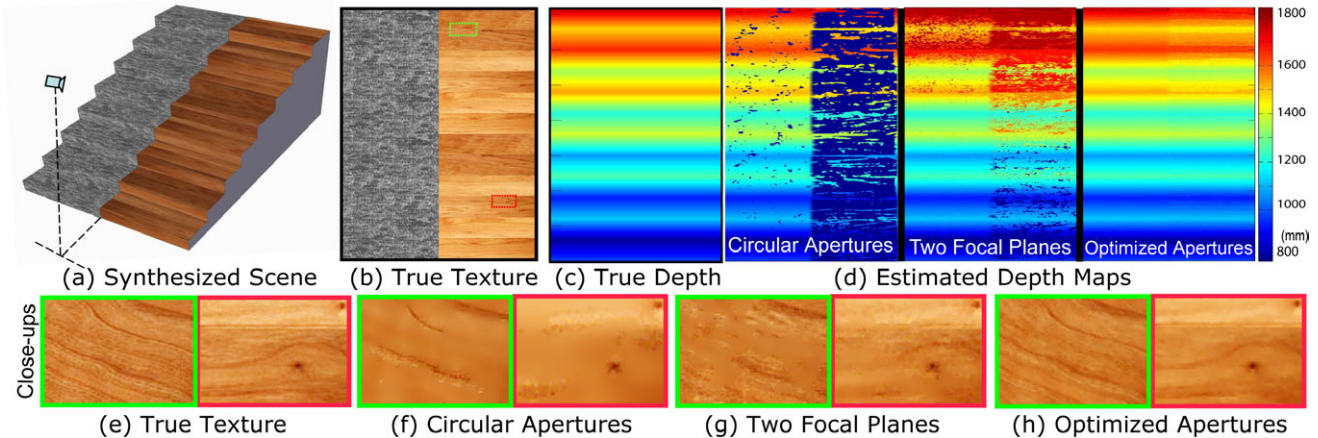
Figure 5. Comparison of depth from defocus and defocus deblurring using a synthetic scene. (a) 3D structure of a synthesized stairs. (b) Ground truth of texture map. (c) Ground truth of the mapped texture. (d) Estimated depth maps using three different methods. From left to right: small/large circular aperture pair, two focal planes, and the proposed coded aperture pair. (e) Close-ups of two regions in the ground truth texture. (f-h) The corresponding recovered all-focused patches using small/large circular aperture pair, two focal planes, and the proposed coded aperture pair.

scene with two textures, one with strong and dense patterns, and another of natural wood with weak texture. Comparisons are presented with two other typical aperture configurations: a small/large circular aperture pair, and a circular aperture with two sensor locations (shift of focus plane rather than change in aperture radius). The virtual camera (focal length = 50mm) is positioned with respect to the stairs as shown in Figure 5(a). The corresponding ground truth texture and depth map are shown in (b) and (c), respectively.

For the DFD algorithm using our optimized aperture pair, the focal plane is set near the average scene depth (1.2m) so that the maximum blur size at the nearest/farthest points is about 15 pixels. For the conventional method using a small/large circular aperture pair, the focal plane is set at the nearest scene point to avoid front/behind ambiguity with respect to the focal plane and yet capture the same depth range. This leads to a maximum blur size of about 30 pixels at the farthest point. For the DFD method with two sensor positions, the two images are synthesized with focal planes set at the nearest point (0.8m) and the farthest point (1.8m). Identical Gaussian noise ($\sigma = 0.005$) is added to all the synthesized image.

Figure 5(d) shows results of the three DFD methods. Note that no post-processing is applied in this evaluation. By comparing to (c), we can see that the depth precision of our proposed method is closest to the ground truth. At the same time, our proposed method generates an all-focused image of higher quality than the other two methods, as illustrated in (f)-(h).

A quantitative comparison among the dual-image DFD methods is given in Table 1. Using the optimized coded aperture pair leads to considerably lower root-mean-squared errors (RMSE) for both depth estimation and defocus deblurring in comparison to the conventional circular aperture pair and the two focal planes. The difference in performance is particularly large for the natural wood texture with weaker texture, which indicates greater robustness of the optimized pair.

Table 1. Quantitative evaluation of depth and deblurring error

	Strong Texture (RMSE)		Wood Texture (RMSE)	
	Depth (mm)	Grayscale	Depth (mm)	Color
Circular apertures	27.28	0.028	464.04	0.060
Two focal planes	6.32	0.027	124.21	0.045
Proposed coded apertures	4.03	0.016	18.82	0.036

For an intuitive understanding of this improvement, we refer to the analysis in [12]. In [12], it is shown that DFD can be regarded as a triangulation method, with the aperture size corresponding to the stereo baseline in determining depth sensitivity. Instead of directly increasing the depth sensitivity, our aperture patterns are optimized such that the DFD will be more robust to image noise and scene variation. Furthermore, the complementary power spectra and large phase displacement between the two optimized apertures essentially help to avoid matching ambiguity of the triangulation. Because of these, our DFD method using the optimized aperture pair can estimate depth with higher precision as shown in Table 1 without increasing the physical dimensions of the aperture.

7. Experiments with Real Apertures

We printed our optimized pair of aperture patterns on high resolution (1 micron) photomasks, and inserted them into two Canon EF 50mm $f/1.8$ lenses (See Figure (6)). These two lenses are mounted to a Canon EOS 20D cam-
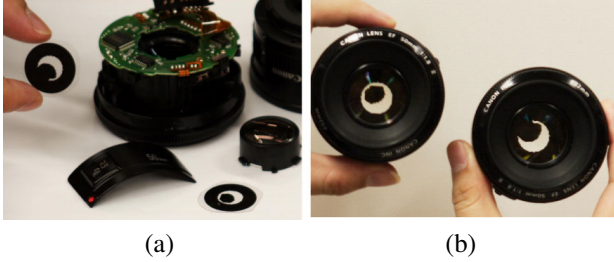
Figure 6. Implementation of aperture pair. (a) Lenses are opened. (b) Photomasks with the optimized aperture patterns are inserted.

era in sequence to take a pair of images of each scene. The camera is firmly attached to a tripod and no camera parameter is changed during the capturing. Switching the lenses often introduce a displacement of around 5 pixels between the two captured images. We correct for this with an affine transformation.

This setting was used to capture real images of several complex scenes. Figure 7 shows a scene inside a bookstore with a depth range of about 2-5 m. Two images (a,b) were taken using the optimized coded aperture pair with the focus set to 3m. From these two images, we computed a high-quality depth map as shown in (d). Note that no post-processing was applied here to the depth map. A high-quality all-focused image was also produced by using the proposed deconvolution method. By comparison with the ground truth, which was captured with a tiny aperture ($f/16$) and long exposure time, we can see that the computed all-focused image exhibits accurate deblurring over a large depth of field.

Figure 8 shows another scene with large depth variation, ranging from 3 meters to about 15 meters. We intentionally set the focus to the nearest scene point so that the conventional DFD method, which uses a circular aperture, can be applied and compared against. For the conventional method, the f-Number was set to $f/2.8$ and $f/4.5$, respectively, such that the radius ratio is close to the optimal value determined in Section 4. For a fair comparison, all of the four input images were captured with the same exposure time.

The results are similar to those from our simulation. We can see clearly from Figure 8(b) that depth estimation using the conventional circular apertures only works well in regions with strong texture or sharp edges. On the contrary, depth estimation with the optimized coded apertures is robust to scenes with subtle texture. Note that the same depth estimation algorithm as described in Section 5 is used here for both settings, and no post-processing of the depth map has been applied.

8. Discussion and Perspectives

We presented a comprehensive criterion for evaluating aperture patterns for the purpose of DFD. This criterion is used to solve for an optimized pair of apertures that complement each other both for estimating relative defocus and for preserving frequency content. This optimized aperture pair enables more robust depth estimation in the presence of image noise and weak texture. This improved depth map is then used to deconvolve the two captured images, in which frequency content has been well preserved, and yields a high-quality all-focused image.

We did not address the effects of occlusion boundaries in this paper, as it is not a central element of this work. As a result, some artifacts or blurring along occlusion boundaries might be observed in the computed depth maps and all-focused images.

There exist various ways in which coded aperture pairs may be implemented. Though it is simple to switch lenses as described in this paper, implementations for real-time capture with coded aperture pairs are highly desirable. One simple implementation is to co-locate two cameras using a half-mirror. A more compact implementation would be to use a programmable LCD or DMD aperture within a single camera to alternate between the two aperture patterns in quick succession.

In this paper, the proposed evaluation criterion was presented for optimizing the patterns of coded aperture; however, it can be applied more broadly to other PSF coding methods, such as wave-front coding which does not occlude light as coded apertures do. How to use this criterion to optimize wave-front coding for DFD would be an interesting direction for future work.

References

[1] E. Dowski. Passive ranging with an incoherent optical system. *Ph. D. Thesis, Colorado Univ., Boulder, CO.*, 1993.

[2] H. Farid and E. Simoncelli. Range estimation by optical differentiation. *Journal of the Optical Society of America A*, 15(7):1777–1786, 1998.

[3] P. Favaro and S. Soatto. A Geometric Approach to Shape from Defocus. *IEEE PAMI*, pages 406–417, 2005.

[4] S. Hiura and T. Matsuyama. Depth measurement by the multi-focus camera. In *CVPR*, pages 953–959, 1998.

[5] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. In *Proc. ACM SIGGRAPH*, 2007.

[6] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *Proc. ACM SIGGRAPH*, 2008.

[7] S. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE PAMI*, 18(12):1186–1198, 1996.

[8] A. Pentland. A New Sense for Depth of Field. *IEEE PAMI*, 9(4):423–430, 1987.

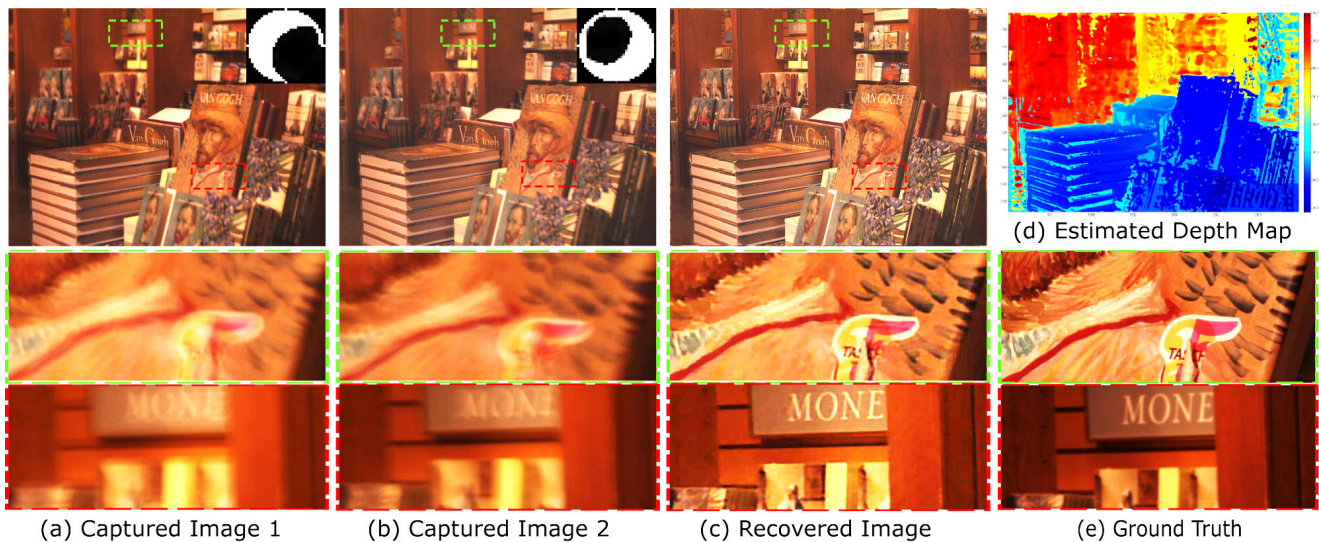(a) Captured Image 1   (b) Captured Image 2   (c) Recovered Image   (e) Ground Truth

Figure 7. Inside a book store. (a-b) Captured Images using the coded aperture pair with close-ups of several regions. The focus is set at the middle of depth of field. (c) The recovered image with close-ups of the corresponding regions. (d) The estimated depth map without post-processing. (e) Close-ups of the regions in the ground truth image which was captured by using a small aperture $f/16$ and a long exposure time.

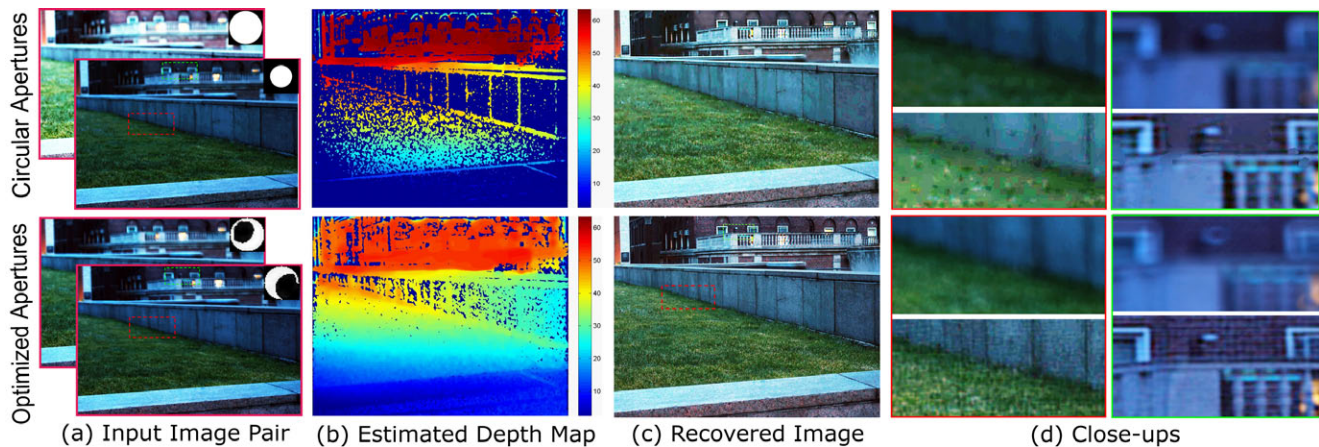(a) Input Image Pair   (b) Estimated Depth Map   (c) Recovered Image   (d) Close-ups

Figure 8. Campus view. First row: Conventional DFD method using circular apertures of different size. The two input images are captured with $f/2.8$ and $f/4.5$, respectively. Second row: DFD method using the optimized coded aperture pair. All the images are captured with focus set to the nearest point.

[9] M. Piana and M. Bertero. Regularized deconvolution of multiple images of the same object. *Journal of the Optical Society of America A*, 13(7):1516–1523, 1996.

[10] A. Rajagopalan and S. Chaudhuri. Optimal Selection of Camera Parameters for Recovery of Depth from Defocused Images. In *CVPR*, 1997.

[11] A. Rav-Acha and S. Peleg. Two motion-blurred images are better than one. *Pattern Recognition Letters*, 26(3):311–317, 2005.

[12] Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *IJCV*, pages 1784–1786, 1998.

[13] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *CVPR*, pages 498–503, 1988.

[14] M. Subbarao, T. Wei, and G. Surya. Focused image recovery from two defocused images recorded with different camera settings. *IEEE Trans. Image Processing*, 4(12):1613–1628, 1995.

[15] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing. *ACM Trans. Graphics*, 26(3):69, 2007.

[16] M. Watanabe and S. Nayar. Rational Filters for Passive Depth from Defocus. *IJCV*, 27(3):203–225, 1998.

[17] Y. Weiss and W. Freeman. What makes a good model of natural images? In *CVPR*, pages 1–8, 2007.

[18] C. Zhou and S. Nayar. What are good apertures for defocus deblurring? In *International Conference of Computational Photography*, 2009.