Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring

Changyin Zhou Stephen Lin Shree K. Nayar

Abstract— The classical approach to depth from defocus (DFD) uses lenses with circular apertures for image capturing. We show in this paper that the use of a circular aperture severely restricts the accuracy of DFD. 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. These 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 real scenes demonstrate the benefits of using the coded apertures over conventional circular apertures.

Index Terms—Depth from Defocus, Coded Aperture, Defocus Deblurring, Deconvolution

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 3-D scene geometry.

Depth from defocus (DFD) is one of the typical approaches to recovering 3-D scene geometry. 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. Therefore, by estimating the blur size of a point in the image, one can estimate its depth. Relative to other passive image-based shape reconstruction approaches such as multi-view stereo and structure from motion, DFD is more robust to image noise, occlusion and correspondence problems because of the two dimensionality of the lens aperture (in contrast to the one dimension stereo or motion baseline) (Schechner and Kiryati, 1998).

Changyin Zhou and Shree K. Nayar are with the Department of Computer Science, Columbia University. Email: {changyin, nayar}@cs.columbia.edu.
Steve Lin is with Microsoft Research Asia. Email: stevelin@microsoft.com

Since defocus information was first used for depth estimation in the 1980s (Pentland, 1987; Subbarao and Gurumoorthy, 1988), various DFD techniques have been proposed based on changes in camera settings (e.g., Favaro and Soatto, 2005; Nayar et al., 1996; Rajagopalan and Chaudhuri, 1997; Subbarao and Surya, 1994; Watanabe and Nayar, 1998). Several methods were proposed to compute DFD using a single image (Dowski, 1993; Levin et al., 2007; Pentland, 1987). As estimating blur sizes from a single image is an ill-posed problem, these single-image methods have to rely heavily on image priors or user interaction. For more reliable and robust depth estimation, most DFD methods use two or more images captured at camera settings with different defocus blurs. One way to change the defocus blur is to change the focus setting (e.g., axially move the sensor). Another approach is to change the aperture pattern (e.g., change the aperture size). In this paper, we restrict our discussion to the latter implementation.

Although DFD has been studied for decades, most work presupposes the camera apertures to be circular. A typical two-image DFD technique captures a pair of images from a fixed viewpoint, one with a large circular aperture and one with a small circular aperture (a large/small circular aperture pair). The image captured with a larger aperture will exhibit greater degrees of defocus than the one captured with a small aperture, and this relative defocus is exploited to estimate depth. Note that the relative defocus is fundamentally influenced by the shape or pattern of the camera apertures, and that aperture patterns do not have to be circular. By restricting ourselves to circular apertures, we may have severely limited the ability of DFD. Once we lift the restriction, we can now attempt to answer two fundamental questions related to DFD:

1) How do the aperture patterns affect the performance of DFD?
2) What is the optimal coded aperture pair for DFD?

In this work, we propose a comprehensive framework of evaluating aperture pairs for DFD. 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

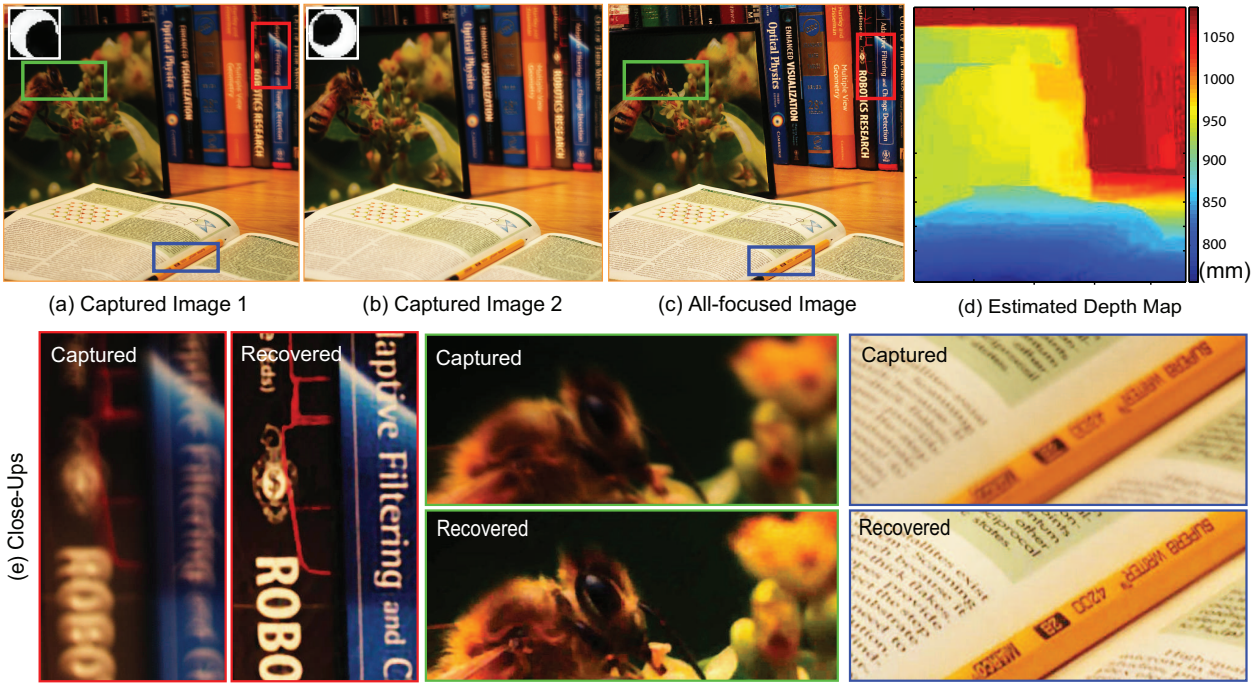


Fig. 2. Depth from defocus and out-of-focus deblurring using a coded aperture pair. (a-b) Two captured images using the optimized coded aperture pair. The corresponding aperture patterns are shown at the top-left corner of each image. (c) The recovered all-focused image. (d) The estimated depth map. (e) Close-ups of four regions in the first captured image and the corresponding regions in the recovered image. Note that the flower within the picture frame (green box) is out of focus in the actual scene and this blur is preserved in the computed all-focused image. For all the other regions, the blur due to image defocus is removed.

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.

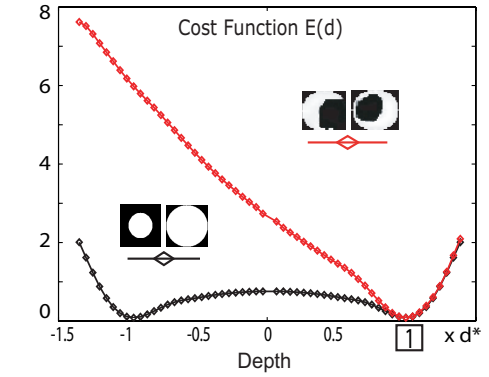
The derived evaluation criterion is first used to optimize the ratio of aperture radii of a large/small circular aperture pair. A numerical optimization based on our criterion shows that for a large/small aperture pair with Gaussian patterns, the optimal ratio is 1.7. Interestingly, Rajagopalan and Chaudhuri (1997) previously derived a similar ratio 1.73 in an analytic manner. While their analytical optimization only applies to Gaussian aperture patterns, pattern optimizations based on our proposed criterion can be performed for any form of coded aperture patterns.

As discussed in several previous works (Levin et al., 2007; Zhou and Nayar, 2009), solving for the optimal 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 in the spatial domain. To make this problem more tractable, existing methods have limited the pattern resolution to 13×13 or even lower (Levin et al., 2007; Veeraraghavan et al., 2007; Zhou and Nayar, 2009). However, solutions at lower resolutions are less optimal due to limited flexibility.

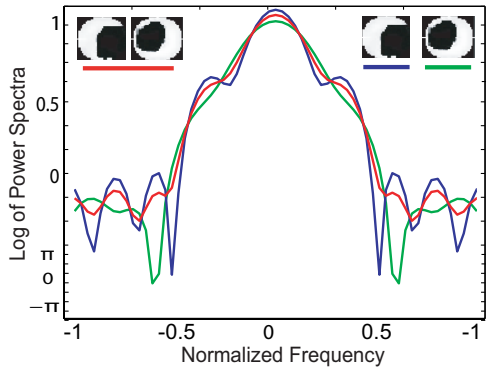
To address this resolution issue, we propose a novel recursive strategy for pattern optimization that incorporates a genetic algorithm (Zhou and Nayar, 2009) with gradient descent search. This algorithm yields optimized solutions with resolutions of 33×33 or even higher within a reasonable computation time. Although higher resolutions usually generate greater diffraction effects, in this particular case we find that a high-resolution pattern of 33×33 actually suffers less from diffractions than other lower resolution patterns, likely due to the smoother features of the optimized high-resolution pattern.

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.

Besides lower robustness, DFD methods using conventional circular apertures (or any other point-symmetric



(a) Energy Curves of Depth Estimation



(b) Log of Power Spectra

Fig. 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 whether a scene point is farther or closer than the focus plane. (b) 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).

apertures) also cannot distinguish whether an object lies in front of or behind the focus plane, since defocus kernels could be identical in these two cases. The asymmetric apertures derived in our work easily avoid this ambiguity.

We demonstrate via extensive 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

2.1 Circular Aperture for Depth from Defocus

Depth from defocus has been studied extensively, at least for circular apertures, in the past decades (a few examples are Favaro and Soatto (2005); Nayar et al. (1996); Pentland (1987); Rajagopalan and Chaudhuri (1997); Subbarao and Gurumoorthy (1988); Subbarao and Surya (1994); Watanabe and Nayar (1998)). These works assume that the point-spread- functions (PSFs) of an imaging system are either pillbox (or cylindrical) functions or Gaussian. Partly due to the good mathematical properties of these functions, people have been able to develop a variety of effective DFD algorithms.

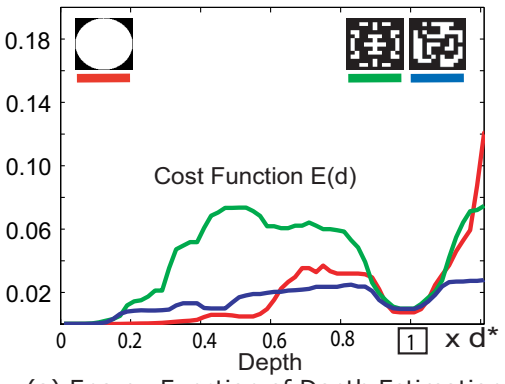
Also, a lot of analysis and optimization on these DFD algorithms and camera settings were conducted based on the assumption of pillbox or Gaussian PSFs. Subbarao and Tyan (1997) study how image noise affects the performance of a spatial-domain DFD approach proposed in (Subbarao and Surya, 1994). Schechner and Kiryati (1993) analyze the effect of focus setting on the depth-from-defocus method implemented by axially moving the sensor, and prove that the change in focus setting should be less than twice the depth of field. Rajagopalan and Chaudhuri (1997) discuss what effect the degree of relative blurring has on the accuracy of the depth estimation and propose a criterion for optimal selection of camera parameters. Most relevantly, they show that for a Gaussian aperture pair, the optimal radius ratio is 1.73, which is very close to the optimization result in this paper.

2.2 Single Coded Aperture Techniques

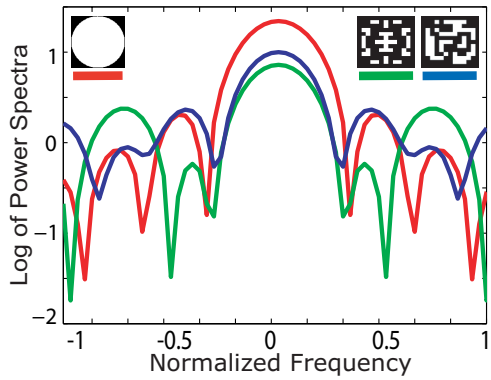
Coded aperture techniques were introduced several decades ago in the field of high energy astronomy to address the issues of signal-to-noise ratio related to lensless imaging of x-ray and γ-ray sources (Caroli et al., 1987; Gottesman and Fenimore, 1989), and in the field of optics to increase depth of field (Mino and Okano, 1971; Welford, 1960). In recent years, coded apertures have received increasing attention in the computer vision community. Coded apertures have been used to improve out-of-focus deblurring (Veeraraghavan et al., 2007; Zhou and Nayar, 2009). To achieve this goal, the coded apertures are designed to be broadband in the Fourier domain. In the work (Zhou and Nayar, 2009), a detailed analysis is presented on how aperture patterns affect deblurring. Based on this analysis, a closed-form criterion for evaluating aperture patterns is proposed. Note that a high-precision depth estimation is prerequisite for high-quality defocus deblurring, since depth determines the size of blur kernel. In this work, they either assume the depth is known or rely on user interaction to specify the depth.

To improve depth estimation, Levin et al. (2007) proposed using an aperture pattern with a more distinguishable pattern of zero-crossings in the Fourier domain than that of conventional circular apertures. Similarly, Dowski (1993) 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



(a) Energy Function of Depth Estimation



(b) Log of Power Spectra

Fig. 3. Performance trade-offs with single apertures. (a) DFD energy function profiles of three patterns: circular aperture (red), coded aperture of (Levin et al., 2007) (green), and coded aperture of (Zhou and Nayar, 2009) (blue). (b) Log of power spectra of these three aperture patterns. The method of (Levin et al., 2007) 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 (Zhou and Nayar, 2009) 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.

needed for optimal defocus blurring but are less effective for depth estimation (Levin et al., 2007), 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 (Levin et al., 2007) and the aperture for deblurring in (Zhou and Nayar, 2009). 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.

2.3 Multiple Coded Apertures Techniques

Multiple images with different coded apertures are used for DFD in previous works (Farid and Simoncelli, 1998; Hiura and Matsuyama, 1998). In the work (Farid and Simoncelli, 1998), two images are taken with two different aperture patterns, Gaussian and the derivative of a Gaussian. These patterns are designed so that depth estimation involves only simple arithmetic operations, making it suitable for real-time implementation. Hiura and Matsuyama (1998) aim 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.

Liang et al. (2008) proposed to take tens of images by using a set of Hadamard-like coded 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 capturing only two images.

2.4 Other Related Techniques

Several techniques have been proposed to computationally extend depth of field by either using custom optical elements in the aperture plane (Dowski and Johnson, 1999; Levin et al., 2009), or by moving the sensor/object during the exposure (Hausler, 1972; Nagahara et al., 2008). The purpose of these methods is to create depth-invariant PSFs so that one can recover an all-in-focus image using deconvolution without knowing scene depth.

Greengard et al. (2006) exploits 3-D diffraction effects to make spatially rotating PSFs by using a 3-D optical phase plate. The PSF rotates as the depth changes and is used for depth estimation. Hasinoff and Kutulakos (2009) propose to capture a large set of images of a scene with predetermined foci and apertures of the lens. From these images, one can reconstruct the scene with high geometric complexity and fine-scale texture.

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 assumed to be Gaussian white noise $N(0, \sigma^2)$, and k^d 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 whether a scene point is farther or closer than the focal plane. For a specific setting, there is a one-to-one mapping from the blur size to the depth. By estimating the size of defocus blur from the image, we can calculate the depth. The above equation can be 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. (Throughout this paper, we use lower case notation for variables in the spatial domain and upper case for those in the spatial domain.)

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 i^{th} 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:

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

According to Equation 2, we have

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

The image prior term $P(\hat{F}_0)$ can be written as $P(\hat{F}_0) \propto \exp\{-\frac{1}{2}\|\phi(\hat{F}_0)\|^2\}$, where the potential function ϕ varies with the choice of image priors. To make the later analytical discussion possible, we choose a linear potential function $\phi(\hat{F}_0) = \Psi \cdot \hat{F}_0$ and then have

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

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

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

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

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

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

where $C = \sigma \cdot \Psi$. Rather than assigning a specific weighting matrix C, we can optimize C by making use of the $1/f$ law of natural images (Van der Schaaf and Van Hateren, 1996; Weiss and Freeman, 2007) and have $C = \sigma/A$, where $|A^2|$ is the expected power spectrum of natural images: $|A(\xi)|^2 = \int_{F_0} |F_0(\xi)|^2 \mu(F_0)$. Here, ξ is the frequency and $\mu(F_0)$ is the possibility measure of the sample F_0 in the image space. A similar optimization of the weighting matrix C is given in (Zhou and Nayar, 2009).

3.2 Generalized Wiener Deconvolution

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

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

where $\bar{K}$ is the complex conjugate of K and $|X|^2 = X \cdot \bar{X}$. This can be regarded as a generalized Wiener deconvolution algorithm which takes two input defocused images, each with a different PSF, and outputs one deblurred image. This deconvolution method can be easily extended to the multiple-image case as:

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

To obtain a good deblurring result, the summed power spectra of all aperture patterns must be broadband. Since this property is much easier to satisfy by using more than one aperture pattern, multiple defocused images with different PSFs are better.

We note that similar deconvolution algorithms were derived without a regularization term (Klarquist et al., 1995) or with a simple Tikhonov regularization (Piana and Bertero, 1996). Our algorithm significantly improves the deblurring quality by making use of the $1/f$ law. More importantly, the generalized Wiener deconvolution algorithm and its important implications have been overlooked in the computer vision community, as some recent works still do deblurring only using one input image even if they have two or more, or solve the two-image deconvolution problem in an inefficient iterative manner. This algorithm will be further discussed in Section 6.

3.3 Selection Criterion for Depth from Defocus

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 (Zhou and Nayar, 2009):

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

where $\mu(F_0)$ is the possibility measure of the sample F_0 in the image space. This equation can be rearranged and simplified to get

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

which is the energy corresponding to a hypothesized depth estimate given the aperture pair, focal plane and noise level. A proof of Equation 10 is given in Appendix A.

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

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

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 turns out to be weak. We have found Equation 11 to be 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).

We then standardize Equation 12 and get

$$R(K_1, K_2|d^*) \approx \min_{d\in\mathcal{D}/d^*} \left[\frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2} \right]^{1/2}, \qquad (13)$$

where n is the pixel number of the PSF. Letting

$$M(K_1, K_2, d, d^*) = \left[\frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1^d K_2^{d^*} - K_2^d K_1^{d^*}|^2}{|K_1^d|^2 + |K_2^d|^2 + C^2} \right]^{1/2}, \qquad (14)$$

we have

$$R = \min_{d\in\mathcal{D}/d^*} M(K_1, K_2, d, d^*). \qquad (15)$$

A larger R value indicates the energy function for DFD is steeper and therefore the estimation will be more robust to image noise and weak texture.

3.3.1 Analysis

When the ratio $c = d/d^*$ approaches 1, we have

$$\begin{aligned} & M(K_1, K_2, d, d^*) \\ &= \left[\frac{1}{n} \sum_{\xi} A \cdot \frac{(|c-1|d^*)^2 |K_1'^{d^*} K_2^{d^*} - K_2'^{d^*} K_1^{d^*}|^2}{|K_1^{d^*}|^2 + |K_2^{d^*}|^2 + C^2} \right]^{1/2} \\ &= |c-1|d^* \cdot \left[\frac{1}{n} \sum_{\xi} A \cdot \frac{|K_1'^{d^*} K_2^{d^*} - K_2'^{d^*} K_1^{d^*}|^2}{|K_1^{d^*}|^2 + |K_2^{d^*}|^2 + C^2} \right]^{1/2}, \qquad (16) \end{aligned}$$

where $K_i'^{d^*}$ is the derivative of $K_i^{d^*}$ with respect to the blur size. See Appendix B for the detailed derivation. It indicates that the M curve is linear in c when $|c| \to 1$. For a specific d^* and frequency ξ, the slope is determined by $\frac{|K_1'^{d^*} K_2^{d^*} - K_2'^{d^*} K_1^{d^*}|^2}{|K_1^{d^*}|^2 + |K_2^{d^*}|^2 + C^2}$. Figure 4 (a) shows M curves of a circular aperture pair at three different depths. We can see that the M curves are linear when $d \to d^*$.

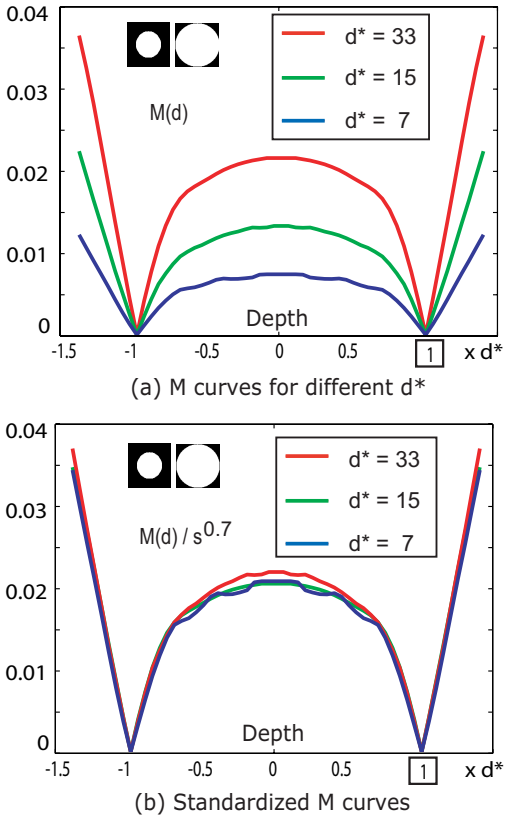


Fig. 4. M curves. (a) Three M curves of a circular aperture pair at $d^* = 33, 15,$ and 7 pixels, plotted as red, green, and blue lines, respectively. When $d \to d*$, the M curves are linear in d. (b) Three standardized M curves. Note that the normalization factor $s^{0.7}$ does not rely on specific aperture patterns (Equation 18). The three standardized M curves are quite consistent. It indicates that the proposed evaluation criterion works equally well for different scene depths. Once an aperture pair is optimized for a specific blur size d^* (i.e. a specific object depth), it will also be optimal for other depths.

Intuitively, for optimal DFD performance with an aperture pair, the pair must maximize the relative defocus between the two images. Equation 16 reveals that defocus depends on differences in amplitude and phase in the spectra of the two apertures. DFD is most accurate when the two Fourier spectra are complementary in both magnitude and phase, such that their phases are orthogonal and a zero-crossing for one aperture corresponds to a large response at the same frequency for the other aperture. For example, if $K_1 = 0$ at a specific frequency ξ, the slope

$$\frac{|K_1'^{d^*} K_2^{d^*} - K_2'^{d^*} K_1^{d^*}|^2}{|K_1^{d^*}|^2 + |K_2^{d^*}|^2 + C^2} = |K_1'^{d^*}|^2 \cdot \frac{|K_2^{d^*}|^2}{|K_2^{d^*}|^2 + C^2}. \qquad (17)$$

Then, a larger derivative of K_1 and a larger $|K_2|$ are preferred at this frequency to maximize the slope. As a result, although our main objective is to compute optimal apertures for DFD, the complementary power spectra yielded by our approach also enables the capture of a broad range of scene frequencies and hence is effective for defocus deblurring.

Differences in d^* correspond to variations in the size of

ground truth PSF, which is in turn determined by the depth. To assess how the depth variation affects the aperture pair evaluation, consider two PSF scales d_1^* and d_2^* with a ratio $s = d_2^*/d_1^*$. By assuming that the ratio $c = d/d^*$ approaches 1 as we derive Equation 16, we are able to get

$$M(K_1, K_2, c \cdot d_2^*, d_2^*) \approx M(K_1, K_2, c \cdot d_1^*, d_1^*) \cdot s^{\alpha/2}, \quad (18)$$

where α is a constant related to the power order in the $1/f$ law (Van der Schaaf and Van Hateren, 1996). See Appendix C for the detailed derivation. Note the factor $s^{\alpha/2}$ is dependent on the choice of aperture patterns. Figure 4 (b) shows three standardized M curves of the circular aperture pair by factors $s^{\alpha/2}$. In our implementation, α is found to be 1.4. We can see the three M curves are quite consistent after the standardization. This indicates our evaluation criterion works equally well for all scene depths. This property ensures that once an aperture pair is optimized for a specific blur size d^* (i.e. a specific object depth), it will also be optimal for other depths.

In this analysis, the proposed criterion (Equation 15) is simplified by assuming $d/d^* \rightarrow 1$. While this helps us better understand the criterion in an intuitive way, it is not accurate when d is significantly different from d^*. For example, as shown in Figure 1, M is no longer linear in c when $|c|$ deviates far away from 1. Because of this, we will still use Equation 15 as the criterion for aperture pair evaluation.

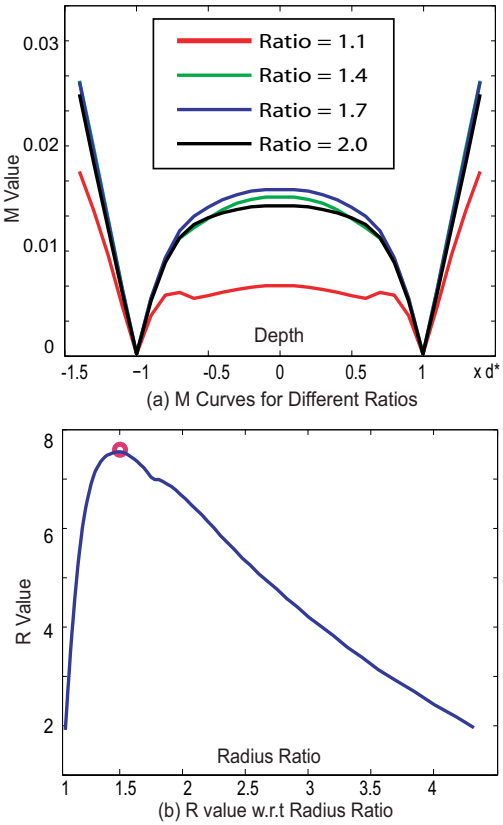


Fig. 5. Using M and R to determine optimal radius ratio for DFD in the case of the conventional circular aperture. (a) M curves of the circular aperture pairs with four different radius ratios. (b) R values of circular aperture pairs with respect to radius ratio. The R value is maximized at a radius ratio of 1.5.

4 Optimization of Aperture Pairs

4.1 Circular Aperture Pair

We first use our derived evaluation criterion to determine the optimal radius ratio of circular aperture pairs for DFD. In Figure 5 (a), we show curves of the M energy function from Equation 14 for four different ratios. These plots highlight the well-known ambiguity with circular aperture pairs of whether a scene point lies in front of or behind the focal plane. This problem exists for any point-symmetric apertures (e.g., the one optimized in (Levin et al., 2007)). Figure 5 (b) shows a plot of our evaluation measure R with respect to the radius ratio. R is maximized at the ratio 1.5, which indicates that 1.5 is the optimal radius ratio for DFD.

A related analysis specifically for Gaussian aperture patterns has been previously performed in (Rajagopalan and Chaudhuri, 1997) and an optimal ratio of 1.73 was derived based on information theory. For Gaussian PSFs, our numerical optimization yields a similar ratio of 1.70. This shows the consistency between the theoretical approach and our numerical approach. While this theoretical approach requires Gaussian PSFs, our method can be applied to optimize arbitrary patterns.

4.2 Coded Aperture Pair

We then use the evaluation criterion to solve for optimal coded aperture patterns. Pattern optimization is known to be a challenging problem (Levin et al., 2007; Veeraraghavan et al., 2007; Zhou and Nayar, 2009). First, it is difficult to solve the optimal pattern analytically, because the evaluation criterion is defined in the Fourier domain and the pattern is strictly constrained in the spatial domain - all the transmittance values of the pattern must lie between 0 and 1. Secondly, it is computationally infeasible to do brute force search. For $N \times N$ binary patterns, the number of possible solutions is huge – $2^{N \times N}$. If we use gray-level patterns, the space will be even larger. Our problem is made harder since we are attempting to solve for a pair of apertures rather than a single aperture. To solve this problem, we propose a two-step optimization strategy.

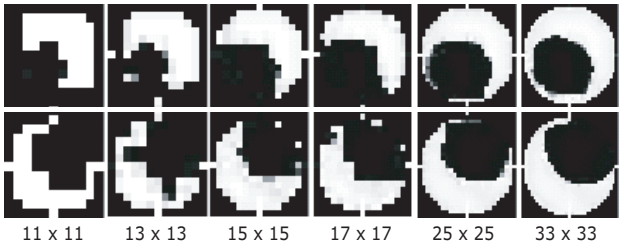


Fig. 6. Increasing the resolution of an optimized aperture pair by up-sampling and gradient search.

In the first step, we employ the genetic algorithm proposed in (Zhou and Nayar, 2009) to find the optimized binary aperture at a low resolution of 11×11 according to Equation 15. The optimized aperture pair at 11×11 is shown in the first column of Figure 6. Despite the high

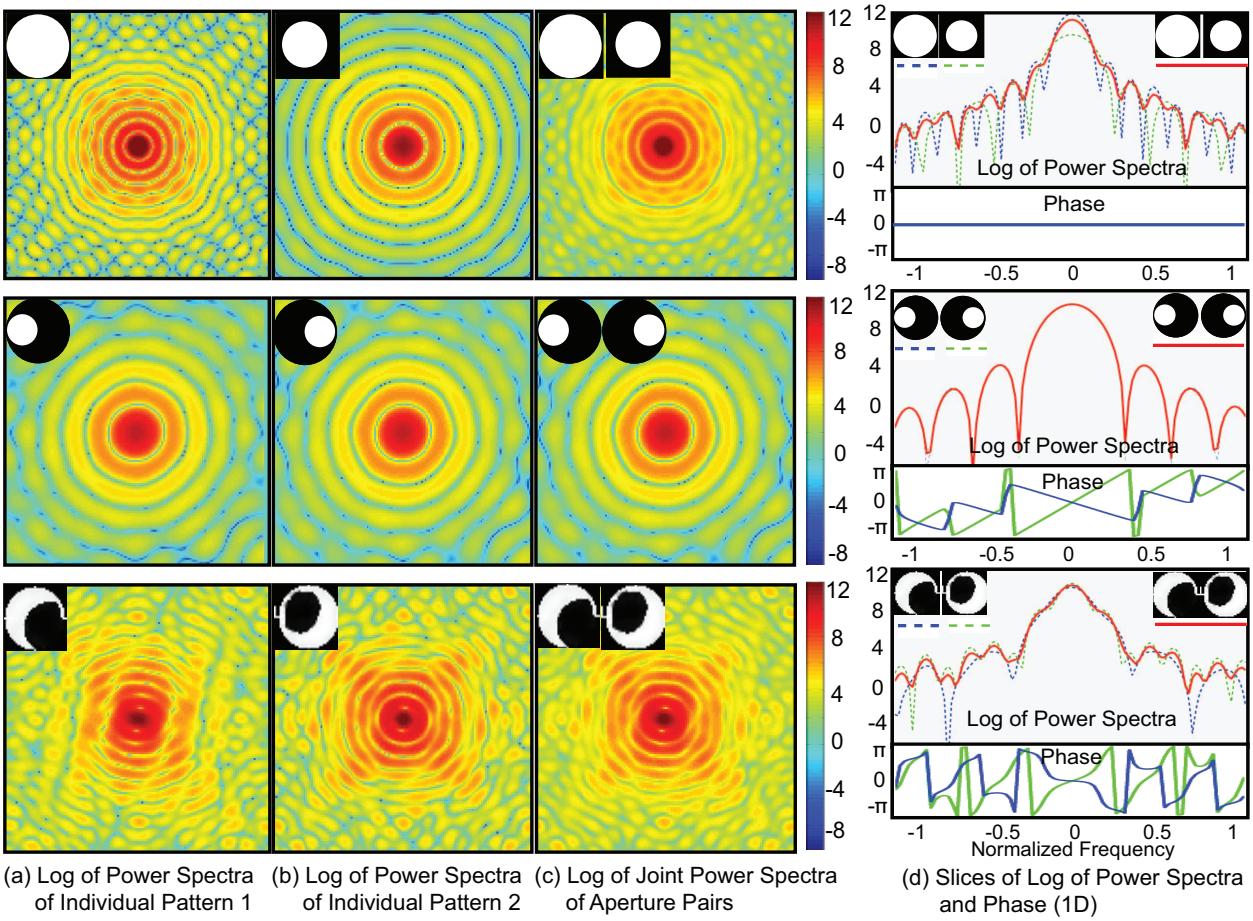


Fig. 7. Pattern spectra of three different aperture pairs, including the optimized large/small circular aperture pair (Row 1), a pair of circular apertures with shifted centers (Row 2), and our optimized coded aperture pair (Row 3). The log of power spectra of each single pattern in the aperture pairs is illustrated in (a) and (b); and the log of joint power spectra of the aperture pairs is illustrated in (c). For a clearer illustration, one 1-D slice of each 2-D power spectra is plotted in (d). In addition, one 1-D slice of phase of each single pattern is also plotted in (d). We can see that the two patterns in the optimized coded aperture pair compensate each other in both power spectra and phase.

efficiency of this genetic algorithm, we found it to have difficulties converging at higher resolutions.

As discussed in Section 3.3.1, 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. Therefore, in the second step, we scale up the 11×11 solution to 13×13 and then refine the solution using gradient descent optimization. This scale-and-refine process is repeated until reaching a resolution of 33×33. Figure 6 shows the evolution of this pattern optimization from 11×11 to 33×33, from left to right. The far right aperture pair is our final optimized coded aperture pair for DFD.

4.3 Discussion

4.3.1 On Depth from Defocus

The optimal radius ratio of a large/small aperture pair is shown to be 1.5 in Section 4.1. For an intuitive visualization of this ratio's optimality, we illustrate the large/small aperture pair with radius ratio 1.5 in the Fourier domain (Figure 7 (a, b), Row 1). One slice of the log of power spectrum of the large circular pattern ($log(|K_1|^2)$) is plotted as a dashed blue line in the first row of Figure 7 (d); the corresponding slice of the small circular pattern ($log(|K_2|^2)$) is plotted as a dashed green line. We can see that due to the optimized ratio 1.5, these two power spectra compensate each other with respect to the zero-crossing frequencies. This compensation intuitively increases the relative defocus between the two PSFs and benefits the depth estimation.

One can also increase the relative defocus by designing a pair of patterns whose spectra compensate each other in phase. One example is a pair of small circular patterns with shifted centers (a stereo-like aperture pair) as shown in Figure 7, Row 2. These two patterns share the same power spectra, but compensate each other in phase (Figure 7 (d), Row 2). This compensation in phase yields a stereo-like effect in the captured images and increases the performance of DFD.

Remarkably, our optimized coded aperture pairs exhibit significant compensations in both power spectra and phase as shown in Figure 7 (d), Row 3. Intuitively, this compensation maximizes the score defined in Equation 15, greatly enhances the relative defocus, and improves the performance of DFD.

Figure 8 (a) shows the depth estimation curves $M(d, d^*, K_1, K_2)$ for the optimized circular aperture pair (green), a pair of shifted circular apertures (blue), and our optimized coded aperture pair (red). We can see the optimized coded aperture pair exhibits a more pronounced minimum in the profile of M than the other two pairs. This leads to depth estimation that is more precise and more robust to noise and scene variations.

4.3.2 On Defocus Deblurring

Equation 8 implies broadband joint power spectra will bring great improvements in the quality of defocus deblurring. Although the aperture pairs are optimized for best DFD, the resulting complementary power spectra enable us to also compute a high quality all-focused image from the two captured defocused images. This is because, with zero-crossings located at different frequencies for each of the two apertures, the two apertures jointly provide broadband coverage of the frequency domain. Logs of the joint power spectra of the aperture pairs, $log(|K_1|^2/2 + |K_1|^2/2)$, are shown in Figure 7 (c). For both the optimized circular aperture pair and the optimized coded aperture pair, the joint pattern pairs are much more broadband than the individual patterns. 1-D Slices of the power spectra of three single aperture patterns are shown in Figure 7 (d) for a clearer illustration.

Two defocused images with different blur kernels can thus be much better than each single image. This is an important implication of Equation 8. Rav-Acha and Peleg (2005) discussed a similar idea in the context of motion-blur deblurring, but do not provide detailed reasoning or a closed-form deblurring algorithm.

For the stereo-like pair with shifted circular patterns, its power spectra does not have any compensation, and thus thus contains many zero-crossings as shown in Figure 7 (b), Row 2. Therefore, the aperture pair is not ideal for defocus deblurring. The joint power spectra of the three aperture pairs are compared in Figure 8 (c).

4.3.3 On Diffraction

The final optimized aperture pair of resolution 33×33 is not only superior to the solution at 11×11 in terms of the evaluation criterion defined in Equation 15, but also produces less diffraction because of greater smoothness in the pattern. In Figure 8 (c), the in-focus diffraction pattern of one of our optimized apertures is compared to three other aperture patterns, including a large circular aperture, a small circular aperture, and an optimized pattern at a lower resolution (the first pattern in Figure 6). We can see that the diffraction pattern of the optimized pattern at a high resolution is more compact than the small circular aperture and the optimized pattern at a low resolution.

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^*:

$$W^{(d)} = \sum_{i=1,2} |IFT(\hat{F_0}^{(d)} * K_i^{\hat{d}} - F_i)|, \qquad (19)$$

where IFT is the 2-D 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

$$U(x, y) = \arg\min_{d \in \mathcal{D}} W^{(d)}(x, y), \qquad (20)$$

and then recover the all-focused image I as

$$I(x, y) = \hat{F_0}^{(U_{x,y})}(x, y). \qquad (21)$$

The most computationally expensive operation in this algorithm is the inverse Fourier transform. Since it is $O(N \cdot \log N)$, the overall computational complexity of recovering U and I is $O(l \cdot\ Nlog(N))$, 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.

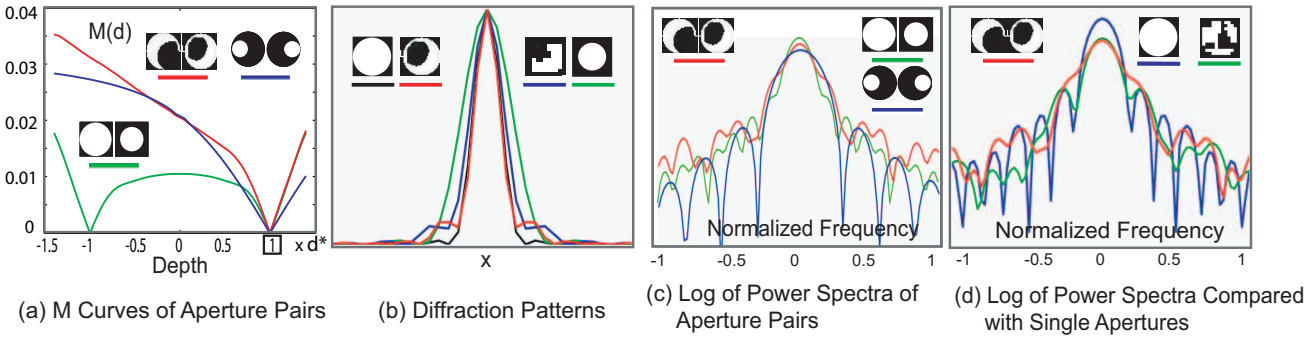
From the sparsely sampled depth values, we can increase the depth resolution at a location (x, y) by fitting the sequence of residuals $\{W_{xy}^{(d-2)}, W_{xy}^{(d-1)}, W_{xy}^{(d)}, W_{xy}^{(d+1)}, W_{xy}^{(d+2)}\}$ with a 3^{rd}-order polynomial curve: $v = a_1 d^3 + a_2 d^2 + a_3 d + a_4$. With this interpolating polynomial, a continuous-valued depth estimate can be obtained by solving $\delta v / \delta d = 0$.

6 Performance Analysis

6.1 Defocus Deblurring with Known Depth

Equation 8 shows that a high-quality deblurring requires both a broadband joint power spectra and an accurate depth estimation d. For simplification, we first assume the ground truth depth d is given and evaluate via simulation the performance of the optimized coded aperture pair in defocus deblurring. The evaluation is performed in comparison to several other two-aperture or single-aperture methods, including the optimized circular aperture pair, a single circular aperture, a single coded aperture optimized for defocus deblurring (Zhou and Nayar, 2009), and a single coded aperture optimized for DFD (Levin et al., 2007).

For each aperture pattern configuration, we simulate a defocused image or image pair of an IEEE resolution chart using Equation 1 (Figure 9 (b-g)). The same budget of exposure time is given to each configuration. Therefore, the brightness of simulated defocused images varies with



(a) M Curves of Aperture Pairs (b) Diffraction Patterns (c) Log of Power Spectra of Aperture Pairs (d) Log of Power Spectra Compared with Single Apertures

Fig. 8. (a) Comparison of M curves among the optimized coded aperture pair, optimized circular aperture pair and the stereo-like aperture pair. (b) The in-focus diffraction patterns of four apertures, including a large circular aperture, a small circular aperture, one of our optimized coded apertures at high resolution, and one of our optimized coded apertures at low resolution. (c) Comparison of the joint power spectra of the optimized coded aperture pair with those of the other two aperture pairs. (d) Comparison of the joint power spectra of the optimized coded aperture pair with the power spectra of several single aperture patterns, including a conventional circular aperture and one coded aperture optimized for defocus deblurring in (Zhou and Nayar, 2009).

aperture patterns, and the captured images appear darker for aperture pair methods, since each image only gets half of the time budget.

The deconvolution algorithm (Equation 8) is then applied to deblur the simulated defocused image. By comparing the close-ups of the recovered images, we see that the optimized coded aperture pair (b), the coded aperture of (Zhou and Nayar, 2009) (f), and the optimized circular aperture pair (d) are able to recover much more image detail and produces less noise and artifacts compared to the other three pattern configurations (c, e, g). This is because the power spectra (or joint power spectra) of the aperture patterns in (b, d, f) are broadband and have few zero-crossing frequencies, as shown in Figure 8 (c, d).

6.2 Depth from Defocus and Defocus Deblurring

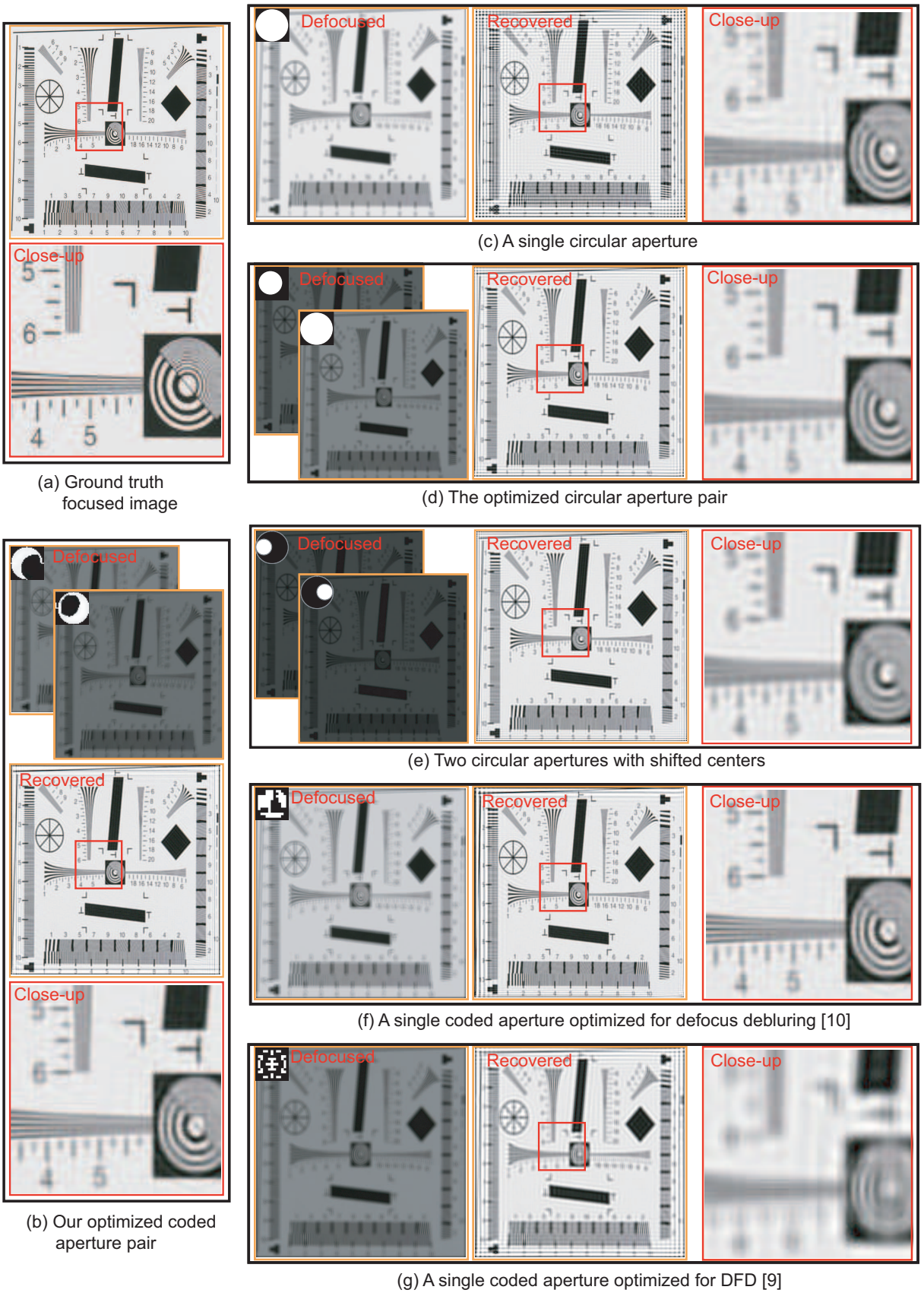
In this work, we are interested in recovering an accurate depth map and a high-quality focused image simultaneously from the captured defocused images. Although the optimized single aperture of (Zhou and Nayar, 2009) showed outstanding deblurring performance in the previous evaluation, it is not ideal for this purpose since it cannot work well for depth estimation. Also, note that an accurate depth estimation is a pre-requisite for deblurring.

To quantitatively evaluate the optimized coded aperture pair, we conducted experiments on a synthetic staircase scene with two textures, one with strong and dense patterns, and another of natural wood with weak texture. The virtual camera (focal length = 50mm, pixel size = 10 μ m) is positioned with respect to the stairs as shown in Figure 10 (a). The corresponding ground truth texture and depth map are shown in (b) and (c), respectively. 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).

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 the 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. The radius ratio of the two circular apertures is set to 1.5, the optimal value. For the DFD method with two sensor positions, (Schechner and Kiryati, 1993) reveals that moving the sensor in a DOF interval is optimal with respect to estimation robustness, and the depth estimation can be unstable if the interval is larger than the DOF by a factor of 2 or higher. However, in many scenes, including this simulated one, the depth range is often far larger than the DOF and therefore the optimal interval is practically not achievable. In this simulation, the two defocused 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 images.

Figure 10 (d) shows results of the three DFD methods. Note that no post-processing is applied in this estimation. By comparing to (c), we can see that the depth precision of our proposed method is closest to the ground truth. For a clearer comparison, depth residuals are plotted in (f) for vertical slices of the computed depth maps, with the strong texture in the top plot and the wood texture at the bottom. At the same time, our proposed method generates an all-focused image of higher quality than the other two methods, as illustrated in (e).

A quantitative comparison among these 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 methods. The difference in performance is particularly large for the natural wood with weaker texture, which indicates greater robustness of the



(a) Ground truth focused image

(b) Our optimized coded aperture pair

(c) A single circular aperture

(d) The optimized circular aperture pair

(e) Two circular apertures with shifted centers

(f) A single coded aperture optimized for defocus debluring [10]

(g) A single coded aperture optimized for DFD [9]

Fig. 9. Defocus deblurring with known depth. (a) The ground truth focused image of an IEEE resolution chart and its close-up. (b) to (g) shows six sets of defocused image or image pair, recovered image, and close-up image, simulated using six different aperture configurations. We can see that the recovered images in (b, d, f) include much more details and suffer fewer artifacts and noise than those in (c, e, g). (To be fair, it should be noted that the pattern of Levin et al. (2007) was not designed for defocus deblurring.)

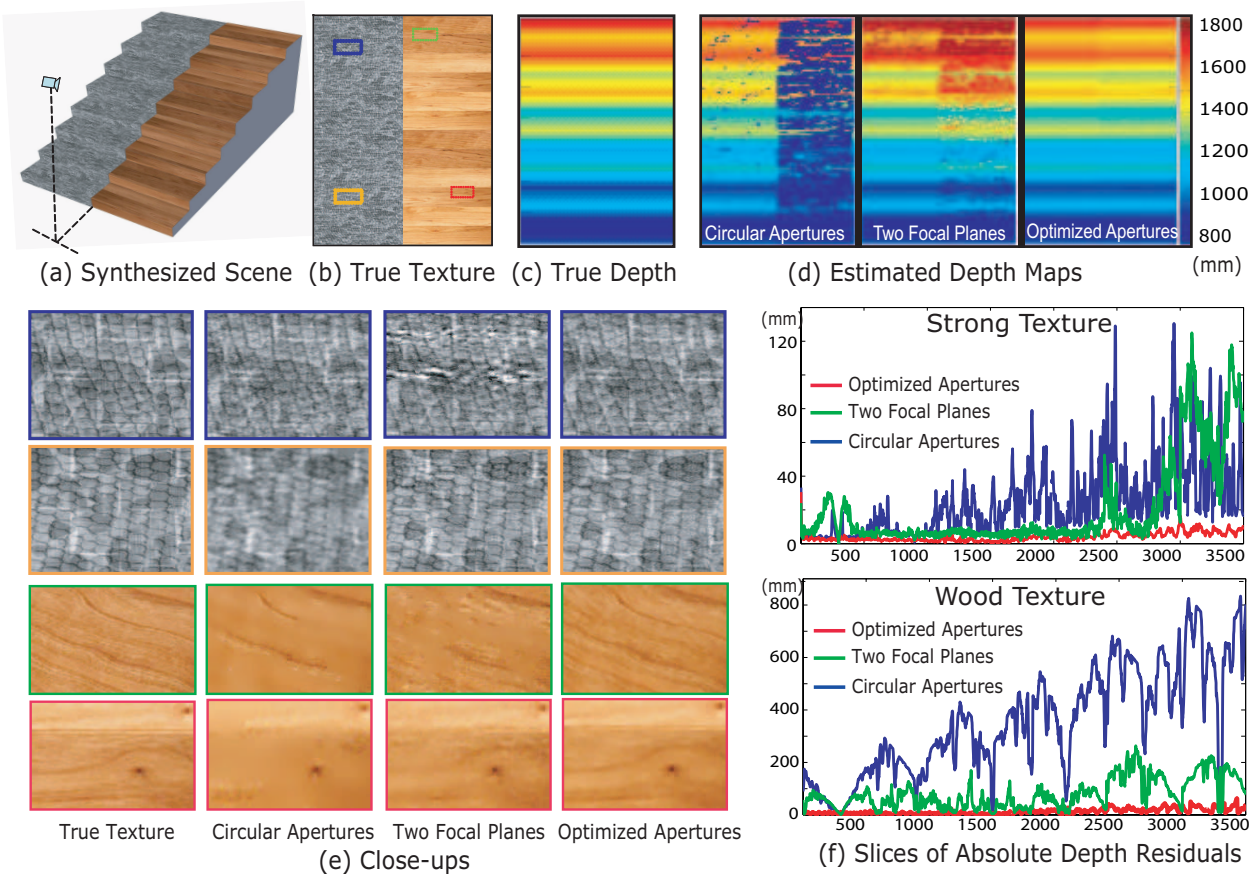


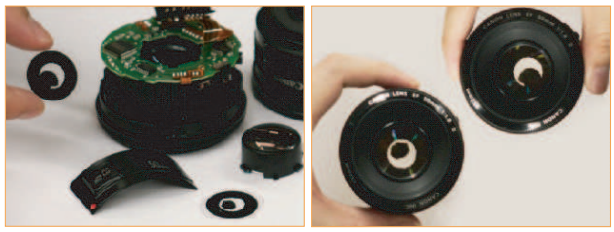
Fig. 10. Comparison of depth from defocus and defocus deblurring using a synthetic scene. (a) 3-D structure of synthesized stairs. (b) Ground truth of texture map. (c) Ground truth of the depth map. (d) Estimated depth maps using three different methods. From left to right: small/large circular aperture pair, two focal planes, and the proposed coded aperture pair. (e) Close-ups of four regions in the ground truth texture and the images recovered using the four different methods. (f) Top: The depth residuals of the four depth estimation methods on the strong texture; bottom: the depth residuals on the wood texture.

optimized pair.

For an intuitive understanding of this improvement, we refer to the analysis in (Schechner and Kiryati, 1998). In (Schechner and Kiryati, 1998), it is shown that DFD can be regarded as a triangulation-based 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 this, 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.

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



(a) A disassembled Canon EF 50mm f/2.8 lens (b) Two lenses with the optimized patterns inserted

Fig. 11. Implementation of aperture pair. (a) Lenses are opened. (b) Photomasks with the optimized aperture patterns are inserted.

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 (11)). These two lenses are mounted to a Canon EOS 20D camera 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 introduces a displacement of around 5 pixels between the

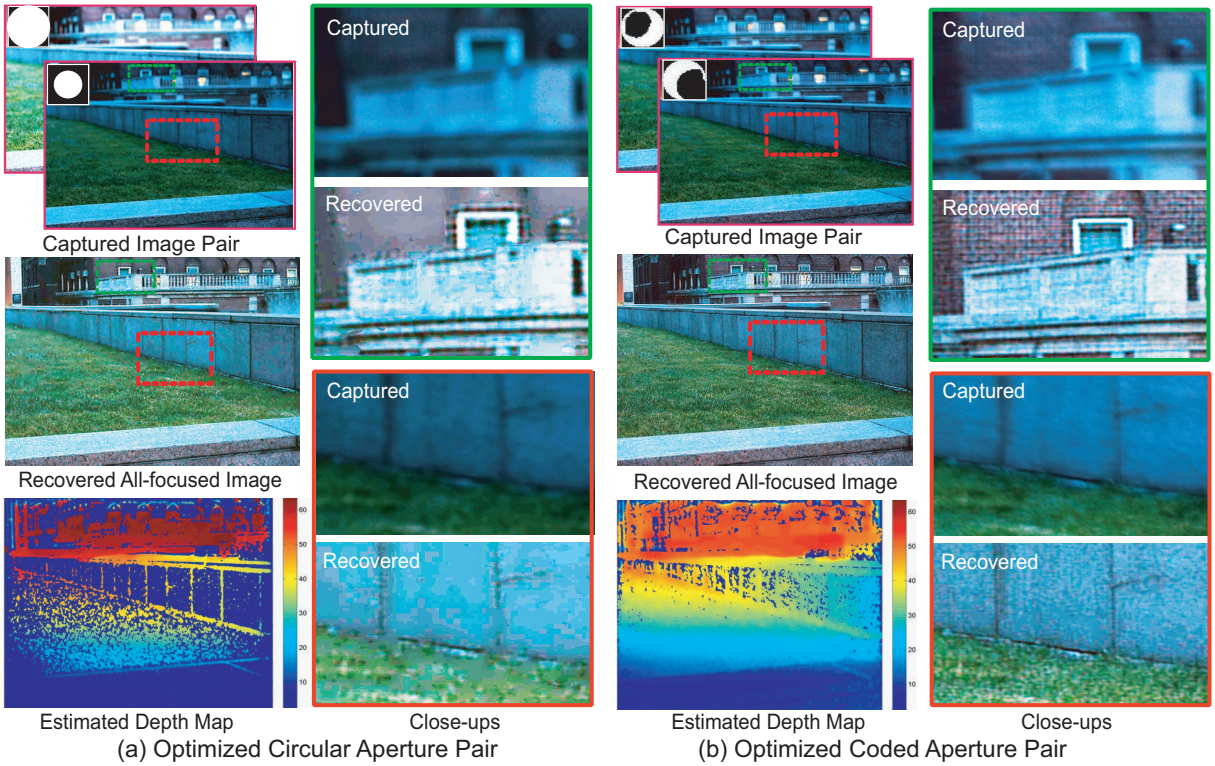


Fig. 12. Campus view example. (a) Conventional DFD method using circular apertures at the optimal ratio of 1.5. The two input images are captured with $f/2.8$ and $f/4.5$, respectively. (b) DFD method using the optimized coded aperture pair. All images are captured with focus set to the nearest point. Note that the only difference between (a) and (b) is the choice of the aperture patterns.

two captured images. We correct for this with an affine transformation.

This setting was used to capture real images of several complex scenes. Figure 2 shows a scene on a study desk with a depth range of about 0.8-1.5m. Two images (a, b) were taken using the optimized coded aperture pair with the focus set to the middle of the depth of field. From these two captured defocused images, we computed a high-quality all-focused image and a high-quality depth map simultaneously as shown in (c, d) using the algorithm described in Section 5. No user interaction was involved throughout the whole estimation precedure. By comparing the close-ups of the captured image with the recovered image (e), we can see that lots of image details were recovered. Note that the flower within the picture frame (green box) is out of focus in the actual scene and this blur is preserved in the computed all-focused image. In addition, the defocus of the yellow pencil on the book (blue box) changes gradually since it is not parallel to the focus plane, and our algorithm, although is simple and straightforward, able to achieve good recovery with little artifacts.

Figure 12 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 1.5 as determined in Section 4.1. For a fair comparison, all four of the input images were captured with the same exposure time.

The results are similar to those from our simulation. We can see clearly from Figure 12 that depth estimation using the conventional circular apertures only works well in regions with strong texture or sharp edges. On the other hand, depth estimation with the optimized coded apertures is robust to scenes with subtle texture. As a result, we are able to recover many more image details by using the coded aperture pair, as shown in the close-ups.

Figure 13 shows a scene inside a bookstore. The depth range is about 2-5 m. Two images (a,b) were taken using the optimized coded aperture pair with the focus set to 3m. The computed all-focused image and depth map are shown in (c) and (d). The ground truth images (e) were 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 and appears very similar to the ground truth image.

8 Discussion and Perspectives

We have presented a comprehensive criterion for evaluating aperture patterns for the purpose of depth from defocus (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

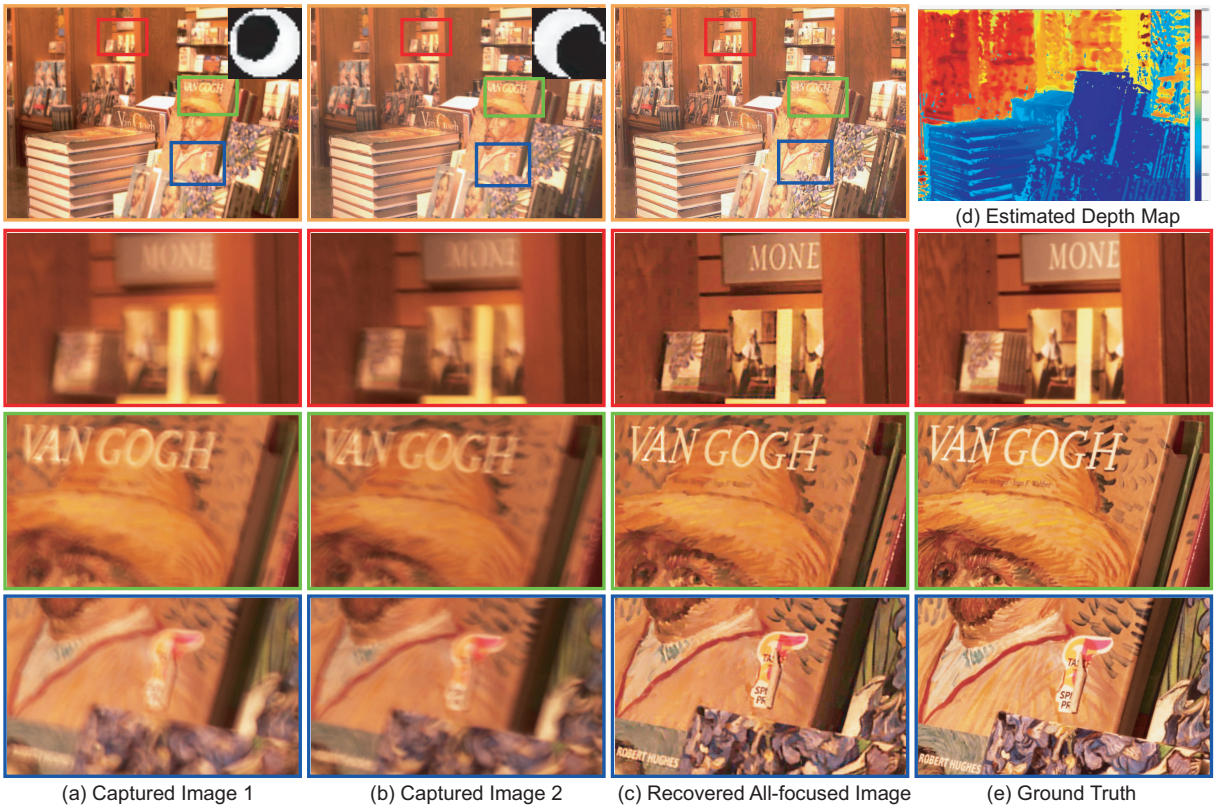


Fig. 13. Inside a book store. (a-b) Captured images using the coded aperture pair with close-ups of several regions. The focus is set at the middle of depth of field. (c) The recovered image with close-ups of the corresponding regions. (d) The estimated depth map without post-processing. (e) Close-ups of the regions in the ground truth image, captured by using a small aperture $f/16$ and a long exposure time.

depth estimation in the presence of image noise and weak texture. The 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 patterns of coded aperture; however, it can be applied more broadly to other PSF coding methods, such as wavefront coding, which does not occlude light as coded apertures do. How to use this criterion to optimize wavefront coding for DFD is an interesting direction for future work.

APPENDIX A: PROOF OF EQUATION 10

Given a coded aperture pair (K_1, K_2), a ground truth blur size d^*, and a noise level σ, the energy E corresponding to a hypothesized blur estimate d is as follows:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{22}$$

$$= \sum_{\xi} \frac{A \cdot |K_1^d \cdot K_2^{d^*} - K_2^d \cdot K_1^{d^*}|^2}{\sum_i |K_i^d|^2 + C} \tag{23}$$

$$+ \sum_{\xi} \frac{\sigma^2 \cdot (\sum_i |K_i^{d^*}|^2 + C)}{\sum_i |K_i^d|^2 + C} + n \cdot \sigma^2. \tag{24}$$

Proof:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{25}$$

$$= \mathop{\mathbb{E}}_{F_0} E(d|K_1^{d^*}, K_2^{d^*}, \sigma, F_0) \tag{26}$$

$$= \mathop{\mathbb{E}}_{F_0, F_1, F_2} E(d|K_1^{d^*}, K_2^{d^*}, F_1, F_2, F_0) \tag{27}$$

$$= \mathop{\mathbb{E}}_{F_0, F_1, F_2} \left[\sum_{i=1,2} \|\hat{F_0} \cdot K_i - F_i\|^2 + \|C \cdot \hat{F_0}\|^2 \right], \tag{28}$$

where $\mathbb{E}(x)$ is the expectation of x, and F_i is the i^{th} captured image. Substituting $\hat{F_0}$ with Equation 8, we get:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{29}$$

$$= \mathop{\mathbb{E}}_{F_0,F_1,F_2}\left[\sum_{i=1,2} \|\frac{F_1\cdot\bar{K}_1^d + F_2\cdot\bar{K}_2^d}{|K_1^d|^2+|K_2^d|^2+|C|^2}\cdot K_i - F_i\|^2\right. \tag{30}$$

$$\left. + \|C\cdot\frac{F_1\cdot\bar{K}_1^d + F_2\cdot\bar{K}_2^d}{|K_1^d|^2+|K_2^d|^2+|C|^2}\|^2\right]. \tag{31}$$

Then, by substituting F_i with Equation 2, we have:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{32}$$

$$= \mathop{\mathbb{E}}_{F_0,\zeta_1,\zeta_2}\left[\sum_{i=1,2} \|(F_0\cdot K_i^{d^*} + \zeta_i) - \right. \tag{33}$$

$$\frac{(F_0\cdot K_1^{d^*}+\zeta_1)\cdot\bar{K}_1^d + (F_0\cdot K_2^{d^*}+\zeta_2)\cdot\bar{K}_2^d}{|K_1^d|^2+|K_2^d|^2+|C|^2}\cdot K_i\|^2 + \tag{34}$$

$$\left.\|C\cdot\frac{(F_0\cdot K_1^{d^*}+\zeta_1)\cdot\bar{K}_1^d + (F_0\cdot K_2^{d^*}+\zeta_2)\cdot\bar{K}_2^d}{|K_1^d|^2+|K_2^d|^2+|C|^2}\|^2\right]. \tag{35}$$

Since ζ_1 and ζ_2 are independent Gaussian white noise $N(0,\sigma)$, we have $\mathbb{E}\,\zeta_i^2 = \sigma^2$, $\mathbb{E}\,\zeta_i = 0$, and $\mathbb{E}\,\zeta_1\zeta_2 = 0$. Let $B = K_1^2 + K_2^2 + C$. Then, Equation 35 can be rearranged to be:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{36}$$

$$= \mathop{\mathbb{E}}_{F_0,\zeta_1,\zeta_2}\sum_{i=1,2}\left[\|\frac{F_0[(K_1^{d^*}\bar{K}_1 + K_2^{d^*}\bar{K}_2^d)\cdot K_i^d - K_i^{d^*}B]}{B}\|^2\right. \tag{37}$$

$$\left. + \|\frac{(\zeta_1\bar{K}_1^d + \zeta_2\bar{K}_2^d)K_i^d}{B} - \zeta_i\|^2\right] \tag{38}$$

$$+ \|C\cdot\frac{F_0\cdot(K_1^{d^*}\bar{K}_1^d + K_2^{d^*}\bar{K}_2^d)}{B} + \frac{\zeta_1\cdot\bar{K}_1^d + \zeta_2\cdot\bar{K}_2^d}{B}\|^2 \tag{39}$$

$$= \mathop{\mathbb{E}}_{F_0}\sum_{i=1,2}\|\frac{F_0[(K_1^{d^*}\bar{K}_1^d + K_2^{d^*}\bar{K}_2^d)\cdot K_i^d - K_i^{d^*}B]}{B}\|^2 \tag{40}$$

$$+ \sigma^2\cdot(\|\frac{K_i^{d^2}+C}{B}\|^2 + \|\frac{K_1^dK_2^d}{B}\|^2 + \|C\cdot\frac{K_i^d}{B}\|^2) \tag{41}$$

$$+ \|C\cdot\frac{F_0\cdot(K_1^{d^*}\bar{K}_1^d + K_2^*\bar{K}_2^d)}{B}\|^2. \tag{42}$$

According to the $1/f$ law, we define the expectation of the power spectrum of F_0 as A, where $A(\xi) = \int_{F_0}|F_0(\xi)|^2\,\mu(F_0)$. In addition, it is known that $C = \sigma^2/A$. Then, Equation 42 can be further re-arranged and simplified as:

$$E(d|K_1^{d^*}, K_2^{d^*}, \sigma) \tag{43}$$

$$= \sum_\xi \frac{A\cdot|K_1^d\cdot K_2^{d^*} - K_2^d\cdot K_1^{d^*}|^2}{\sum_i|K_i^d|^2 + C} \tag{44}$$

$$+ \sum_\xi \frac{\sigma^2\cdot(\sum_i|K_i^{d^*}|^2 + C)}{\sum_i|K_i^d|^2 + C} + n\cdot\sigma^2. \tag{45}$$

xyz

APPENDIX B: PROOF OF EQUATION 16

When the ratio $c = d/d^*$ approaches 1, we have

$$M(K_1, K_2, d, d^*) \tag{46}$$

$$= \left[\frac{1}{n}\sum_\xi A\cdot\frac{(|c-1|d^*)^2|K_1'^{d^*}K_2^{d^*} - K_2'^{d^*}K_1^{d^*}|^2}{|K_1^{d^*}|^2+|K_2^{d^*}|^2+C^2}\right]^{1/2} \tag{47}$$

$$= |c-1|d^*\cdot\left[\frac{1}{n}\sum_\xi A\cdot\frac{|K_1'^{d^*}K_2^{d^*} - K_2'^{d^*}K_1^{d^*}|^2}{|K_1^{d^*}|^2+|K_2^{d^*}|^2+C^2}\right]^{1/2}, \tag{48}$$

where $K_i'^{d^*}$ is the derivative of $K_i^{d^*}$ with respect to the blur size.

Proof: A kernel K can be regarded as a function of both the frequency ξ and the scale d^*. Assume the derivative of K with respect to d^* exists and is denoted by K'^{d^*}, we have $K^d = K^{d^*} + \delta d\cdot K'^{d^*}$ when $\delta d = d - d^* = (c-1)d^*$ approaches zero. Then, we get

$$M(K_1, K_2, d, d^*) \tag{49}$$

$$= \left[\frac{1}{n}\sum_\xi A\cdot\frac{|K_1^dK_2^{d^*} - K_2^dK_1^{d^*}|^2}{|K_1^d|^2+|K_2^d|^2+C^2}\right]^{1/2} \tag{50}$$

$$= \left[\frac{1}{n}\sum_\xi A\cdot\frac{|(K_1'^{d^*}+\delta d\cdot K_1'^{d^*})K_2^{d^*} - (K_2^{d^*}+\delta d\cdot K_2'^{d^*})K_1^{d^*}|^2}{|K_1^{d^*}+\delta d\cdot K_1'^{d^*}|^2+|(K_2^{d^*}+\delta dK_2'^{d^*})|^2+C^2}\right]^{1/2} \tag{51}$$

$$= |\delta|d^*\cdot\left[\frac{1}{n}\sum_\xi A\cdot\frac{|K_1'^{d^*}K_2^{d^*} - K_2'^{d^*}K_1^{d^*}|^2}{|K_1^{d^*}|^2+|K_2^{d^*}|^2+C^2}\right]^{1/2}. \tag{52}$$

$$= |c-1|d^*\cdot\left[\frac{1}{n}\sum_\xi A\cdot\frac{|K_1'^{d^*}K_2^{d^*} - K_2'^{d^*}K_1^{d^*}|^2}{|K_1^{d^*}|^2+|K_2^{d^*}|^2+C^2}\right]^{1/2}. \tag{53}$$

APPENDIX C: PROOF OF EQUATION 18

Consider two scales d_1^* and d_2^* with a ratio $s = d_2^*/d_1^*$. When $c = d/d^*$ approaches 1, we have

$$M(K_1, K_2, c\cdot d_2^*, d_2^*) \approx M(K_1, K_2, c\cdot d_1^*, d_1^*)\cdot s^{\alpha/2}, \tag{54}$$

where α is a constant related to the power order in the $1/f$ law (Van der Schaaf and Van Hateren, 1996).

Proof: According to Equation 16, we have

$$M(K_1, K_2, c\cdot d_2^*, d_2^*) \tag{55}$$

$$= |c-1|d_2^*\cdot\left[\frac{1}{n}\sum_\xi A\cdot\frac{|K_1'^{d_2^*}K_2^{d_2^*} - K_2'^{d_2^*}K_1^{d_2^*}|^2}{|K_1^{d_2^*}|^2+|K_2^{d_2^*}|^2+C^2}\right]^{1/2}. \tag{56}$$

Since K_2^* is a scaled K_1^* of factor s, $K_2^*(\xi) = K_1^*(s\xi)$.

Therefore,

$$M(K_1, K_2, c \cdot d_2^*, d_2^*) \tag{57}$$

$$= s \cdot |c-1| d_1^* \cdot \left[\frac{1}{n} \sum_{\xi=1}^{n} A(\xi) \cdot \frac{|{K_1}'^{d_1^*}(s\xi) K_2^{d_1^*}(s\xi) - {K_2}'^{d_1^*}(s\xi) K_1^{d_1^*}(s\xi)|^2}{|K_1^{d_1^*}(s\xi)|^2 + |K_2^{d_1^*}(s\xi)|^2 + C^2} \right]^{1/2}. \tag{58}$$

$$= s \cdot |c-1| d_1^* \cdot \left[\frac{1}{n} \sum_{\eta=s}^{sn} A(\eta/s) \cdot \frac{|{K_1}'^{d_1^*}(\eta) K_2^{d_1^*}(\eta) - {K_2}'^{d_1^*}(\eta) K_1^{d_1^*}(\eta)|^2}{|K_1^{d_1^*}(\eta)|^2 + |K_2^{d_1^*}(\eta)|^2 + C^2} \right]^{1/2} \tag{59}$$

$$= s \cdot |c-1| d_1^* \cdot \left[\frac{1}{n \cdot s^2} \sum_{\eta=1}^{n} A(\eta/s) \cdot \frac{|{K_1}'^{d_1^*}(\eta) K_2^{d_1^*}(\eta) - {K_2}'^{d_1^*}(\eta) K_1^{d_1^*}(\eta)|^2}{|K_1^{d_1^*}(\eta)|^2 + |K_2^{d_1^*}(\eta)|^2 + C^2} \right]^{1/2}, \tag{60}$$

$$= |c-1| d_1^* \cdot \left[\frac{1}{n} \sum_{\eta=1}^{n} A(\eta/s) \cdot \frac{|{K_1}'^{d_1^*}(\eta) K_2^{d_1^*}(\eta) - {K_2}'^{d_1^*}(\eta) K_1^{d_1^*}(\eta)|^2}{|K_1^{d_1^*}(\eta)|^2 + |K_2^{d_1^*}(\eta)|^2 + C^2} \right]^{1/2}, \tag{61}$$

$$\tag{62}$$

where $\eta = s\xi$.

According to the $1/f$ law (Van der Schaaf and Van Hateren, 1996), the prior power spectra of natural image $A(\xi)$ statistically takes a form of $D \exp 1/\xi^2$, where the power order may vary slightly around 2 with scenes and D is a normalization factor. This spectra function can be roughly approximated as $A(\xi) = D\frac{1}{\xi^\alpha}$ with a proper α, especially when this prior function is applied to finite-resolution images. Then, $A(\eta/s) \approx A(\eta) \cdot s^\alpha$. Therefore, we have

$$M(K_1, K_2, c \cdot d_2^*, d_2^*) \tag{63}$$

$$\approx |c-1| d_1^* \cdot \left[\frac{1}{n} \sum_{\eta=1}^{n} A(\eta) \cdot s^\alpha \cdot \frac{|{K_1}'^{d_1^*}(\eta) K_2^{d_1^*}(\eta) - {K_2}'^{d_1^*}(\eta) K_1^{d_1^*}(\eta)|^2}{|K_1^{d_1^*}(\eta)|^2 + |K_2^{d_1^*}(\eta)|^2 + C^2} \right]^{1/2}, \tag{64}$$

$$= M(K_1, K_2, c \cdot d_1^*, d_1^*) \cdot s^{\alpha/2}. \tag{65}$$

REFERENCES

Caroli E, Stephen J, Cocco G, Natalucci L, Spizzichino A (1987) Coded aperture imaging in X- and Gamma-ray astronomy. Space Science Reviews pp 349–403

Dowski E (1993) Passive ranging with an incoherent optical system. Ph D Thesis, Colorado Univ, Boulder, CO

Dowski E, Johnson G (1999) Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems. In: Proc. SPIE, Citeseer, vol 3779, pp 137–145

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

Favaro P, Soatto S (2005) A Geometric Approach to Shape from Defocus. IEEE Transactions on Pattern Analysis and Machine Intelligence pp 406–417

Gottesman S, Fenimore E (1989) New family of binary arrays for coded aperture imaging. Applied Optics (20):4344–4352

Greengard A, Schechner Y, Piestun R (2006) Depth from diffracted rotation. Optics letters 31(2):181–183

Hasinoff S, Kutulakos K (2009) Confocal stereo. IJCV 81(1):82–104

Hausler G (1972) A method to increase the depth of focus by two step image processing. Optics Communications 6:38–42

Hiura S, Matsuyama T (1998) Depth measurement by the multi-focus camera. In: CVPR, pp 953–959

Klarquist W, Geisler W, Bovik A (1995) Maximum-likelihood depth-from-defocus for active vision. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol 3, pp 374–379

Levin A, Fergus R, Durand F, Freeman W (2007) Image and depth from a conventional camera with a coded aperture. Proc ACM SIGGRAPH 26(3):70

Levin A, Hasinoff S, Green P, Durand F, Freeman W (2009) 4D frequency analysis of computational cameras for depth of field extension p 97

Liang C, Lin T, Wong B, Liu C, Chen H (2008) Programmable aperture photography: multiplexed light field acquisition. Proc ACM SIGGRAPH 27(3):1–10

Mino M, Okano Y (1971) Improvement in the OTF of a defocused optical system through the use of shaded apertures. Applied Optics (10):2219–2225

Nagahara H, Kuthirummal S, Zhou C, Nayar S (2008) Flexible depth of field photography. In: ECCV, vol 3

Nayar S, Watanabe M, Noguchi M (1996) Real-time focus range sensor. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(12):1186–1198

Pentland A (1987) A New Sense for Depth of Field. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(4):423–430

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

Rajagopalan A, Chaudhuri S (1997) Optimal selection of camera parameters for recovery of depth from defocused images. In: CVPR, Citeseer, pp 219–224

Rav-Acha A, Peleg S (2005) Two motion-blurred images are better than one. Pattern Recognition Letters 26(3):311–318

Van der Schaaf A, Van Hateren J (1996) Modelling the power spectra of natural images: statistics and information. Vision Research 36(17):2759–2770

Schechner Y, Kiryati N (1993) The optimal axial interval in estimating depth from defocus. In: ICCV, pp 843–848

Schechner Y, Kiryati N (1998) Depth from defocus vs. stereo: How different really are they? IJCV pp 1784–1786

Subbarao M, Gurumoorthy N (1988) Depth recovery from blurred edges. In: CVPR, pp 498–503

Subbarao M, Surya G (1994) Depth from defocus: A spatial domain approach. IJCV 13(3):271–294

Subbarao M, Tyan J (1997) Noise sensitivity analysis of depth-from-defocus by a spatial-domain approach. In: Proceedings of SPIE, vol 3174, p 174

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

Watanabe M, Nayar S (1998) Rational Filters for Passive

Depth from Defocus. IJCV 27(3):203–225
Weiss Y, Freeman W (2007) What makes a good model of natural images? In: CVPR, pp 1–8
Welford W (1960) Use of annular apertures to increase focal depth. Journal of the Optical Society of America A (8):749–753
Zhou C, Nayar S (2009) What are good apertures for defocus deblurring? In: International Conference of Computational Photography, San Francisco, U.S.