



Research paper

Depth Estimation and Deblurring from a Single Image Using an Optimized-Throughput Coded Aperture

M. Masoudifar^{1,*}, H. R. Pourreza²

¹Department of Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran.

²Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.

Article Info

Article History:

Received 22 January 2022

Reviewed 17 March 2022

Revised 28 April 2022

Accepted 28 May 2022

Keywords:

Coded apertures

Depth from defocus

Defocus deblurring

*Corresponding Author's Email
Address: masoudi@hsu.ac.ir

Abstract

Background and Objectives: Depth from defocus and defocus deblurring from a single image are two challenging problems caused by the finite depth of field in conventional cameras. Coded aperture imaging is a branch of computational imaging, which is used to overcome these two problems. Up to now, different methods have been proposed for improving the results of either defocus deblurring or depth estimation. In this paper, an asymmetric coded aperture is proposed which improves results of depth estimation and defocus deblurring from a single input image.

Methods: To this aim, a multi-objective optimization function taking into consideration both deblurring results and depth discrimination ability is proposed. Since aperture throughput affects on image quality, our optimization function is defined based on illumination conditions and camera specifications which yields an optimized throughput aperture. Because the designed pattern is asymmetric, defocused objects on two sides of the focal plane can be distinguished. Depth estimation is performed using a new algorithm, which is based on perceptual image quality assessment criteria and can discern blurred objects lying in front or behind the focal plane. **Results:** Extensive simulations as well as experiments on a variety of real scenes are conducted to compare the performance of our aperture with previously proposed ones.

Conclusion: Our aperture has been designed for indoor illumination setting. However, the proposed method can be utilized for designing and evaluating appropriate aperture patterns for different imaging conditions.

This work is distributed under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)



Introduction

When a scene is captured by a limited depth of field camera, objects lying at different depths are registered with varying degree of defocus blur. Depth from defocus (DFD) is a method that recovers depth information by estimating the amount of blur in different areas of a captured image. The concept of DFD was first introduced in [1], [2] and then various techniques were proposed, which are briefly reviewed in Sec. 2.

Despite the desirable results of DFD techniques in

conventional apertures, there are some drawbacks rooted in the inherent limitation of circular apertures. For example, single image DFD methods and even some of multiple image DFD methods are unable to distinguish between defocused objects placed before and after the focal plane. In addition, in single image DFD methods, the lower depth of field, which provides enhanced depth discrimination ability, is obtained at the cost of losing image quality. In larger blur scales, most of image frequencies are lost. It makes the estimation of depth and

deblurring more ambiguous and vulnerable to image-noise [3].

Coded aperture photography is a method used for modifying the defocus pattern generated by lens. The shape of PSF (Point Spread Function) can be changed by using a coded mask on lens. So far, a variety of mask patterns have been proposed for improving the results of depth estimation [3]-[5] or defocus deblurring [6]-[8]. However, there are a few number of techniques for extracting both depth and high quality deblurred image [9], [10]. These methods use multiple images captured by a single aperture [9] or multiple aperture patterns [10].

In this paper, we propose an asymmetric single pattern, which is used for capturing a single image. This image is processed to achieve a depth map and an all-focus high quality deblurred image.

To find the proposed optimal aperture pattern, a new multi-objective function containing two evaluation functions is defined. The first function determines the expected value of deblurring error using the correct PSF. The second function computes the expected value of deblurring error using the incorrect PSFs. Both functions are defined in the frequency domain. A non-dominated sorting-based multi-objective evolutionary algorithm [11] is used to find a Pareto-optimal solution. An optimal pattern is chosen in a way that it can also distinguish between defocused objects placed before and after the focal plane. As a result, an asymmetric pattern is proposed which is appropriate for depth estimation and deblurring in a single captured image.

According to [12], [13], illumination conditions and camera specification affect the performance of coded aperture cameras. Therefore, our objective functions are formulated by considering the imaging circumstances. In this way, the designed mask acquires a reasonable throughput that ensures the acceptable signal-to-noise ratio (SNR) of the captured image.

The proposed mask is compared with circular aperture, and a number of state-of-the-art coded aperture patterns. The performance comparison includes depth estimation accuracy and the quality of deblurring results.

In accordance with the proposed multi-objective function, a depth estimation algorithm is introduced in which a blurred image is deblurred by a set of PSF scales. Then, a PSF with the best quality deblurring result is considered as the correct blurring kernel. The quality of deblurred images is measured by an aggregate measure of no-reference image quality assessment criteria.

A. Key Contributions

1) A new multi-objective function is proposed for defining a single pattern, which yields the minimum deblurring error with correct PSF and the maximum deblurring error with incorrect PSFs.

2) The blurring problem is redefined with respect to

the aperture throughput and imaging system conditions. Hence, in the design of coded aperture, the amount of additive noise and image brightness are taken into account.

3) An aggregate no-reference image quality assessment measure is used for depth estimation. The quality of images deblurred by different PSFs is measured and the PSF that yields a deblurred image of the highest quality is chosen as the true PSF.

4) The results of simulation on a dataset show less variance in correct depth/kernel estimation across the entire range of depths/kernel sizes compared to previous aperture patterns.

B. Scope and Limitations

1) The image formation model is assumed to be linear.

2) An affine noise model is used to describe the combined effects of signal-dependent and signal-independent noise. Signal dependent Poisson noise is approximated using a Gaussian noise model. Signal independent noise is assumed only read-noise.

3) The aperture pattern is designed based on the assumption that the exposure time and lighting condition is fixed.

4) The proposed aperture pattern and depth estimation algorithm can be used for both grey-level and color imaging systems.

The rest of this paper is organized as follows: In Sec. 2, related works are briefly reviewed. In Sec. 3, the blurring problem is formulated and pattern evaluation functions are introduced. Section 4 describes the optimization method used to find the optimal pattern. The proposed aperture is analyzed in accordance with spectral properties and depth sensitivity in Sec. 5. Our depth estimation algorithm is presented in Sec.6. Experimental results in both synthetic and real scenes are presented in Sec. 7. Finally, conclusions are drawn in Sec. 8.

Previous Works

The concept of DFD was first introduced in [1], [2] and then a variety of techniques were proposed that used a single image [14]-[19] or multiple images [20]-[24].

Single image DFDs usually estimate the blur scale either by assuming some prior information about PSF [14], [18], texture [16], color information [17] or by using learning methods [19]. However, multiple image DFDs are more variant and use various techniques to extract depth information. Some methods capture two or more images from a single viewpoint under different focus settings or various sizes of aperture [1], [2], [22], [24], [25]. Other methods use two or more images from different viewpoints such as stereo vision with identical focus setting [26] or different focal settings [9].

As mentioned in Sec.1, DFD with conventional apertures suffers from some drawbacks. In the past

decades, coded aperture photography has been used to resolve these problems. Here, some of the proposed apertures and DFD methods are briefly reviewed.

Hiura et al. [27] use multiple images taken by a single aperture pattern from a single viewpoint under different focus settings. Zhou et al. [10] propose a pair of aperture masks. Two blurred images are taken from a single viewpoint with a similar focus setting and two different asymmetric aperture patterns. In real applications, a programmable aperture is needed to ensure that the viewpoint of the two captured images remain unchanged. Otherwise, images should be first registered, and then depth estimation algorithm be applied. Takeda et al. [9] use stereo imaging with a single aperture pattern, yet different focal settings to improve the results of depth estimation presented in [10].

Levin et al. [4] design a single symmetric pattern with the aim of increasing the depth discrimination ability. Kullback-Leibler divergence between different sizes of blur is used to rank aperture patterns. The optimal symmetric pattern is achieved through a full-search of all binary masks. An efficient deblurring algorithm is also used to create high quality deblurred results. Since the proposed mask is symmetric, before and after focal plane cannot be differentiated.

Sellent et al. [5] define a function in the spatial domain for the aperture pattern evaluation. A parametric maximization problem is defined to find a pattern that produce the most possible difference among images blurred of different PSF scales. By solving this problem, non-binary patterns are obtained that can be pruned to binary forms. This technique is then used to find asymmetric patterns suitable to discriminate the front and back of the focal plane [3].

Aperture Evaluation

In this section, first the blurring problem is briefly reviewed and then our criteria for evaluating aperture patterns are introduced. Based on the proposed criteria, a multi-objective function is defined, which is capable to compare aperture patterns with varying throughputs.

A. Problem Formulation

Image degradation due to out of focus blurring and noise can be modeled by convolution of a PSF or kernel function (k_d) with the sharp image (f_s) and then adding noise (ω):

$$f = k_d \otimes f_s + \omega, \quad \sum_i k_d^i = 1 \quad (1)$$

the subscript d indicates that kernel size is a function of depth of scene. The sum of kernel elements (i.e. k_d^i) equals 1, meaning that the image brightness does not change by blurring.

When we use a binary-coded aperture, the shape and

throughput of the aperture are determined by this mask. As noted in [12], [13] an aperture pattern must be evaluated by consideration of both shape and throughput. Therefore, we redefine the well-known defocus problem in terms of these factors.

A binary coded mask with n open cells can be considered as a grid of size $N \times N$, where n holes distributed over the grid are kept open [5], [12]. The pattern of open holes determines the shape of PSF, and their number specifies the mask throughput.

For a simple fronto-parallel object at depth d , defocusing is redefined as the convolution of a defocus kernel (k_d) with a sharp image (f_n) that generates spatial invariant blur:

$$f = k_d \otimes f_n + \omega_n, \quad \omega_n \sim N(0, \sigma_n^2), \quad \sum_i k_d^i = 1 \quad (2)$$

The subscript n shows that the brightness of sharp image (f_n) and the amount of added noise (ω_n) depend on the aperture throughput (n). Due to the additive properties of light, in a constant definite exposure time, the brightness of sharp image (f_n) is increased linearly with an increase in the number of open holes. The value of ω_n also changes with the number of holes. In this study, the growth of ω_n is investigated by considering the number of holes, imaging system's specifications and scene illumination. As mentioned earlier, the sum of kernel elements (i.e. k_d^i) equals 1, meaning that the image brightness does not change by blurring. As we see in Sec. 3. B, the added noise is modeled by normal distribution, which its variance depends of the aperture throughput.

Equivalently, if the Fourier transforms of each variable is shown by its corresponding capital letter, the spatially invariant blur in the frequency domain is defined as follows:

$$F = K_d \cdot F_n + \Omega_n \quad (3)$$

where the convolution operation in the Fourier domain is changed to a simple point-by-point multiplication. The subscripts d and n indicate the depth of scene and aperture throughput, respectively.

B. Noise Model

The imaging noise can be modeled as the sum of two distinct factors: read noise and photon noise [12]. Read noise, which is independent of the measured signal, is commonly modeled by a zero mean Gaussian random variable r with variance σ_r^2 . Photon noise is a signal dependent noise with Poisson distribution. When the mean value of photon noise is large enough, it can be approximated by a random Gaussian variable with variance $\sigma_p^2 = J_n$ [12], [28]. J_n refers to the mean number of photons received by a single pixel in a camera with an

n open-hole aperture.

As noted in [12], the total noise variance is computed as follows:

$$\sigma_n^2 = \sigma_r^2 + \sigma_p^2 = \sigma_r^2 + J_n = \sigma_r^2 + n.J \quad (4)$$

In this study, the mean signal value in photoelectrons (J) of a single-hole aperture is computed by [12]:

$$J = 10^{15} \cdot \frac{1}{F\#^2} \cdot R \cdot I \cdot q \cdot \Delta^2 \cdot t \quad (5)$$

where $F\#$, R , and I refer to camera f-number, average scene reflectivity that varies in range 0 to 1, and amount of scene illumination (measured in lux), respectively. q is the quantum efficacy of the image sensor, which measures the effectiveness of an imaging device to convert incident photons into photoelectrons. Δ is the size of a pixel in an image sensor and t refers to the exposure time. In our experiments, the assumption about scene and imaging system parameters, which represent the typical settings in consumer photography, are as follows:

$q = 0.5$ (typically CMOS sensors)
$R = 0.5$, $t = 10^{ms}$, $F\# = 18$
$\Delta^2 = 5.1 \times 5.1 \mu m^2$ (SLR camera, typically Canon 1100D)
$I = 300^{lux}$ (typically office light level)

In the following section, first our criteria regarding the intensity level of images are proposed. Then, the proposed formula in terms of photoelectron are redefined so that masks with different throughputs can be compared.

C. Mask Search Criteria

Suppose an image F_n is blurred with an unknown Kernel K_1 (3). If it is deblurred with a typical kernel K_2 and Wiener filter is used for deconvolution, then the total error of deblurring (e_n) is computed as (6):

$$\begin{aligned} e_n &= F_n - \hat{F}_n = F_n - \frac{K_2^* F}{|K_2|^2 + |C_n|^2} \\ &= F_n - \frac{K_2^* (K_1 F_n + \Omega_n)}{|K_2|^2 + |C_n|^2} \\ &= \frac{F_n K_2^* (K_2 - K_1)}{|K_2|^2 + |C_n|^2} + \frac{F_n |C_n|^2 - K_2^* \Omega_n}{|K_2|^2 + |C_n|^2} \\ &= \underbrace{\frac{F_n K_2^* (K_2 - K_1)}{|K_2|^2 + |C_n|^2}}_{e_n^{(1)}} + \underbrace{\frac{F_n |C_n|^2 - K_2^* \Omega_n}{|K_2|^2 + |C_n|^2}}_{e_n^{(2)}} \end{aligned} \quad (6)$$

where $|C_n|^2$ is defined as the matrix of expected value for noise to signal power ratios (NSR) of natural images. (i.e. $|C|^2 = \frac{\sigma^2}{A}$ where A is the expected power spectrum of natural images and σ^2 is the variance of additive noise [7].) According to (6), the total error consists of two parts:

$$e_n^{(1)} = \frac{F_n K_2^* (K_2 - K_1)}{|K_2|^2 + |C_n|^2} \quad \text{error of wrong kernel estimation} \quad (7)$$

$$e_n^{(2)} = \frac{F_n |C_n|^2 - K_2^* \Omega_n}{|K_2|^2 + |C_n|^2} \quad \text{deblurring error} \quad (8)$$

If an accurate PSF is used for deblurring (i.e. $K_1 = K_2$), then the only term that determines the total error of deblurring will be $e_n^{(2)}$ (i.e. $e_n^{(1)} = 0$). On the other hand, if a wrong kernel is used as PSF ($K_1 \neq K_2$), both $e_n^{(1)}$ and $e_n^{(2)}$ will generate errors in the deblurring result. As will shown in sec. 4.A, the values of $e_n^{(1)}$ are much greater than $e_n^{(2)}$ (See Fig. 2). Therefore, when $K_1 \neq K_2$, the main determinant of the total error will be $e_n^{(1)}$. Hence, consistent with our objective, a suitable pattern is defined as a pattern that minimizes the norm of $e_n^{(2)}$ and maximizes the norm of $e_n^{(1)}$. The norm of $e_n^{(2)}$ is computed as follows:

$$\begin{aligned} \|e_n^{(1)}\|_2^2 &= \left(\frac{F_n K_2^* (K_2 - K_1)}{|K_2|^2 + |C_n|^2} \right)^* \left(\frac{F_n K_2^* (K_2 - K_1)}{|K_2|^2 + |C_n|^2} \right) \\ &= |F_n|^2 |K_2|^2 \frac{|K_2 - K_1|^2}{\left(|K_2|^2 + |C_n|^2 \right)^2} \end{aligned} \quad (9)$$

Since the power spectra of all natural images follow a certain distribution, the expectation of $\|e_n^{(1)}\|_2^2$ can be computed with respect to F_n . According to $1/f$ law of natural images [29], the expectation of $|F_n|^2$ is computed as $A_n(\xi) = \int_{F_n} |F_n(\xi)|^2 d\mu(F_n)$ where ξ is the frequency and $\mu(F_n)$ is the measure of sample F_n in the image space [7]. Accordingly, the expectation of $\|e_n^{(1)}\|_2^2$ is computed as (10):

$$\begin{aligned} D_n(K_2, K_1) &= \mathbb{E}_{F_n} \{ \|e_n^{(1)}\|_2^2 \} \\ &= \sum_{\xi} \frac{A_{n\xi} |K_2|_{\xi}^2}{(|K_2|_{\xi}^2 + |C_n|_{\xi}^2)^2} |K_2 - K_1|_{\xi}^2 \end{aligned} \quad (10)$$

This measure can be considered as a distance criterion between two kernels. It can also help distinguish between defocus points lying in front or back of the focal plane. It should be noted that the defocus PSF in front of the focal plane is the flipped version of the defocus PSF at the back of the focal plane (See Fig. 1. a), meaning that these PSFs have an identical spectral response but different phase properties. Equation (10) includes the term $K_2 - K_1$, which can compute both spectral and phase differences of two kernels. Hence, by having an asymmetric aperture, the deblurring with the flipped version of a PSF generates $e_n^{(1)}$ error and helps distinguish sides of the focal plane (See Fig. 1. b)

The expected value of $\|e_n^{(2)}\|_2^2$ can be computed in a similar manner. (Details are found in [7]):

$$R_n(K_1) = \|e_n^{(2)}\|_2^2 = \sum_{\xi} \frac{\sigma_n^2}{|K_1|_{\xi}^2 + |C_n|_{\xi}^2} \quad (11)$$

This value has been used by Zhou et al. [7] as a criterion to find aperture patterns with least errors in deblurring results. However, it has been redefined here to allow

functions in the Pareto-front. The values of proposed objective functions are also computed for some other apertures and then added to the figure.

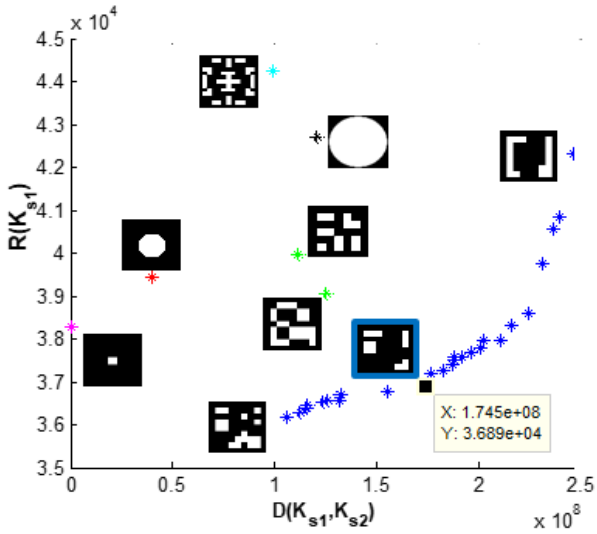


Fig. 2: D values vs. R values of final patterns in the Pareto optimal solution (blue), Open circular aperture (black), conventional aperture (red), pinhole aperture (magenta), patterns proposed in [3] (green) and [4] (cyan). Final selected pattern has been highlighted by the blue border.

According to Fig. 2, in the Pareto optimal solution, with an increase in the symmetry of patterns, the deblurring error (R) and the error of using a wrong scale kernel (D) rises. However, it does not mean that any symmetric pattern outperforms all other asymmetric patterns in terms of discrimination ability (D). For example, objective functions were also computed for the pinhole aperture, open circular aperture and circular aperture with a throughput equal to the selected coded pattern (highlighted by the blue border)¹ as well as the symmetric pattern proposed by Levin et al. [4]. Although these patterns are symmetric, the provided D values are not essentially greater than all asymmetric patterns. On the other hand, R values provided by asymmetric patterns are not essentially smaller than any symmetric ones. In fact, R and D values depend on several factors such as mask throughput and spectral properties.

As noted earlier, NSGA-II provides a set of solutions. Since just one pattern has to be selected, we compute $D_r = D(K_{s1}, rot(K_{s1}, 180))$ for all patterns derived from the Pareto optimal solution. In a similar manner, this value is computed for asymmetric patterns proposed in [3]. Fig. 3 shows the computed values.

As shown in Fig. 3, with an increase in symmetry, D_r declines. Given the significance of criterion D_r , the pattern highlighted by the blue border is selected as a sample of the derived patterns.

¹ In the rest of text, the circular aperture with the same throughput of selected coded pattern is called conventional aperture.

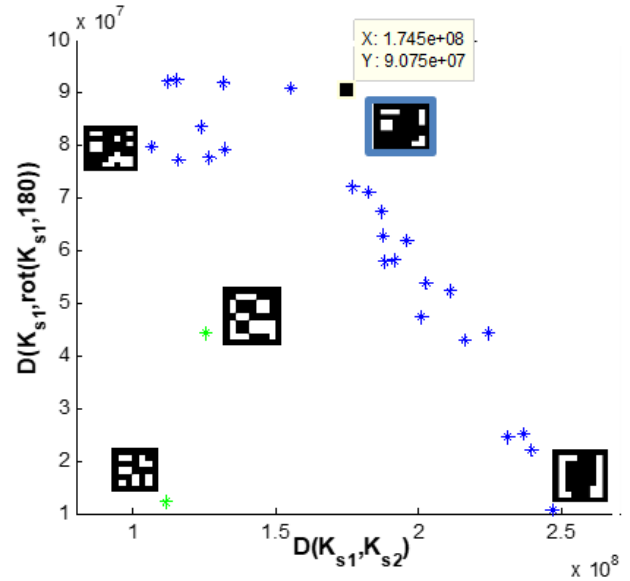


Fig. 3: D value of wrong scale kernels vs. D_r value of the flipped correct scale for the patterns obtained by NSGA-II (blue) and asymmetric patterns proposed in [3] (green).

It must be mentioned that the selected pattern is not the best option under all conditions. However, since it provides appropriate values of D , R and D_r , it is selected as the final pattern. Indeed, the final pattern should provide a minimum value for the weighted sum of all criteria, which each weight representing the importance of the associated criterion. This study adopts NSGA-II, which does not use the weighted sum for optimization.

Aperture Pattern Analysis

In this section, a brief analysis of the proposed pattern is presented. The transmission rate (compared to the open circular aperture) of our optimized aperture is 0.265, which is almost equal to the Levin’s pattern [4]. Hence, the SNR of images captured by this aperture is about 14.4dB², which is in the range of [10..40], meaning that the captured images have an acceptable (not ideal) SNR [34]. In the following; the aperture pattern is examined with respect to its spectral properties and depth sensitivity.

A. Spectral analysis

At the first step, an analogy is drawn between the spectral properties of the selected pattern and the conventional aperture. It should be noted that both apertures have similar throughput so under different imaging conditions; the same amount of additive noise is added to the captured images. In this situation, the spectral properties of apertures determine the results. Fig. 4 shows 1D slices of spectral response for each aperture at five different blur scales. According to [4],

² $SNR_{capture} = 10 \log_{10}(J_n/\sigma_n)$

when a pattern has various frequency responses in each scale, it is more convenient to distinguish blur scales. As shown in Fig. 4, in the conventional aperture, the zero amplitude obtained from different scales overlaps in some frequencies, making it difficult to distinguish between blur scales. However, the coded pattern has diverse spectral responses in different scales.

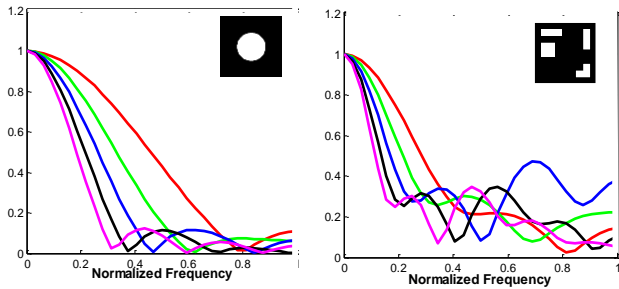


Fig. 4: The 1D slice of spectral response at 5 different blur scales for conventional and coded aperture.

The spectral response of these two apertures is also compared at 4 different scales. As shown in Fig. 5, the minimum spectral response of our pattern is greater than the conventional aperture, especially in larger blur scales. Therefore, in the proposed pattern, attenuation of frequencies in the captured image is reduced and thus deblurring results are improved.

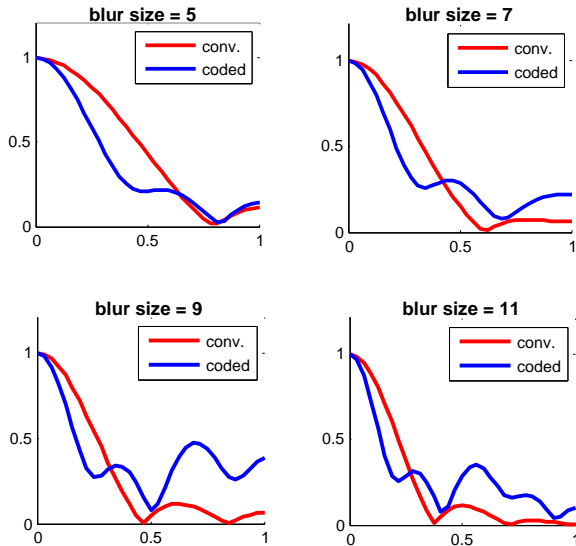


Fig. 5: 1D slices of Fourier transforms for conventional aperture (red) and the proposed pattern (blue) at 4 different scales.

B. Depth Sensitivity

Another advantage of the proposed pattern is its high sensitivity to the depth variation. It is known that DoF declines with an increase in the aperture diameter. In the proposed pattern, open holes are located in the margin of the mask. Hence, this aperture pattern is more sensitive to depth variations than the conventional aperture. To

examine the depth sensitivity difference in these apertures, the blur size is computed in a limited range of depth (before and after the focal point) for a typical lens (EF 50mm f/1.8 II).

The blur size (s) is computed based on thin lens formula [35]:

$$s = \frac{D_a(v - v_0)}{v_0}, \quad v_0 = \frac{Fu_0}{u_0 - F}, \quad v = \frac{Fu}{u - F} \quad (15)$$

The parameters used in (15) were introduced in Fig. 1(a). The aperture diameter (D_a) is assumed 20^{mm} and 8.21^{mm} for coded and conventional patterns respectively. As shown in Fig. 6, the proposed pattern is more sensitive to depth variation. Therefore, depth estimation is easier in images captured by the coded pattern. On the other hand, according to Fig. 5, coded mask gives a higher spectral response, and is thus expected to obtain better results in both deblurring and depth estimation in real imaging.

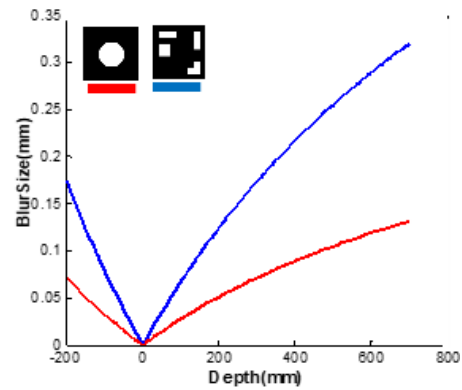


Fig. 6: Blur size vs. depth for conventional (red) and coded (blue) apertures (focus length (u) = 1200^{mm}, v = 50^{mm}). Code aperture is more sensitive to depth variation.

Depth Estimation

Depth estimation is performed using an algorithm described here. The method is based on the proposed objective function (13) and can be used for detecting both the scale and the orientation of PSF. The main idea is that deblurring with inaccurate kernels, whether in scale or direction, produces low-quality images while deblurring with correct kernel yields high quality images (See Fig. 7). For depth estimation, the blurred image is deblurred with a limited set of blurring kernels. The quality of each deblurred image is measured using an aggregate no-reference image quality measure. A PSF, which generates a deblurred image of the highest quality, is selected as the true kernel. As stated earlier, if the aperture pattern is asymmetric, this method can be used for detecting both size and direction of the PSF (Fig. 1).

Several no-reference image quality measures have been proposed in the literature. In one of the most comprehensive studies [36] a weighted sum of 8 different criteria is used for evaluating the image quality (Recent

studies show that using an aggregate measure of image quality assessment criteria is more precise [8], [36], [37] Although this measure can be used for depth estimation, it is more complicated than it is necessary. In our application, a comparison is drawn between the qualities of deblurred versions of the same image.

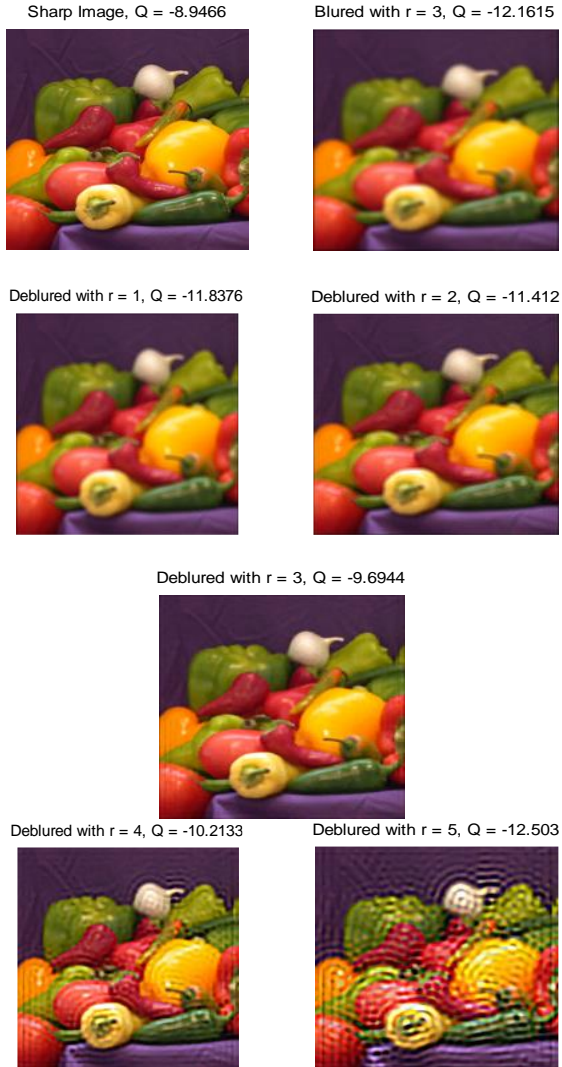


Fig. 7: Deblurring results with different radii of the kernel ($r=1..5$) in imaging with conventional aperture. Deblurring with smaller kernels results in blurry images and deblurring with larger PSFs yields images with artifacts. The quality of each image is evaluated by the no-reference quality assessment measure proposed in [36]. A larger Q-value indicates higher quality.

In fact, here the quality measure is more of a relative measure not a strict one. Therefore, measures of lower complexity can be applied for quality assessment. The speed of depth estimation algorithm is improved by reducing the number of criteria. In this study, the quality of deblurred images is evaluated by an aggregated measure containing four criteria: *Norm-Sparsity-Measure* [38], *Sparsity-Prior* [4], *Sharpness-Index* [39] and *Pyramid-Ring* [36], which are well-suited for our application. These

criteria are sensitive to blur or artifact or both of them. The no-reference aggregate image quality measure is defined in (16), where higher values indicate greater quality. The process of computing this measure has been described in our previous work [40].

$$Quality = -12.65 * normSps + 0.073 * sharpIndex - 0.289 * sparsity - 9.86 * pyrRing \quad (16)$$

A similar measure has been used in [3] to find only the direction of PSF. Sellent et al. [3] use a depth estimation algorithm [35] to determine the scale of PSF. Then, a quality assessment measure is used to find the direction of PSF. Our proposed method is almost similar to [3], but no prepared database is used for PSF estimation here. We use the proposed measure to evaluate the quality of deblurred images (or patch of images) derived by different PSFs. A PSF, which yields a deblurred image with the best quality, is chosen as true PSF. This method is used for detecting both size and direction of PSF.

A. Handling Depth Variations

In real world scenes, there are depth variations. Therefore, each part of an image might be blurred with a different kernel. A common method of depth estimation in these images involves using fairly small patches in which the depth is assumed to be constant. The blur kernel is estimated for the patch, and this estimation is assigned to its central pixel. By repeating this stage for all pixels of the image, a raw depth map is obtained. Then, a coherent map labeling is performed using the raw depth map, image derivative information and some smoothness priors [4], [17].

In this study, first two blur scales that generate deblurred patches of the highest quality are considered as the possible true scales of the central pixel. The probability of each scale is computed based on its relative quality. Higher quality increases probability and the sum of two probability values are equal to 1. At the end of this stage, a three-dimensional matrix is obtained. In other words, for a $H \times W$ image and S possible depths, matrix $D_R \in \mathbb{R}^{H \times W \times S}$ includes the raw depth map in which $D_R(h, w, s)$ represents the probability of depth $s \in S$ in pixel (h, w) .

There may be some errors in the depth estimation of the raw depth map, especially in depth discontinuities. Therefore, in the second step, a coherent blur map is obtained by minimizing an energy function defined as follows [17]:

$$Min E(D_c) = \sum_p D_p(s_p) + \sum_{(p,q) \in N} \lambda_{p,q} V(s_p, s_q) \quad (17)$$

where p and q refer to image pixels. The first term $D_p(s_p)$ indicates fidelity to the previous probability blur scale (s) estimation at position p . The second term $V(s_p, s_q)$ is a

smoothness term, which guarantees that neighbor pixels of similar gray levels have identical blur scales. D_c denotes a solution for coherent data map with minimum energy (E). A coherent map with $\min(D_c)$ is estimated by a method proposed in [17].

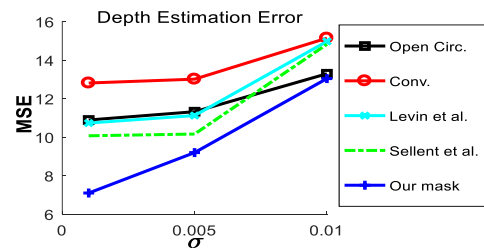
To assign a penalty to depth change in D_p , the early probabilities of blur scale ($p_p(s)$) are convolved with a Gaussian filter ($N(0,0.1)$) to reach the smoothed probabilities ($\hat{p}_p(s)$). Then $-\log(\hat{p}_p(s))$ is used as $D_p(s)$. (See [17]). This function could also be used for cases in which one or more probabilities are assigned to the initial blur scale.

The smoothness term $V(s_p, s_q)$ examines depth discontinuity in neighboring pixels. For each pixel p , depth similarity is investigated with its eight surrounding pixels with $V(s_p, s_q) = |s_p - s_q|$. The relative significance of the difference between depths of two adjacent pixels is determined by the difference of their gray level (g_p and g_q). Hence, $\lambda_{p,q}$ is defined as $\lambda_{p,q} = \lambda_0 e^{-\frac{\|g_p - g_q\|^2}{\sigma_\lambda^2}}$ [17].

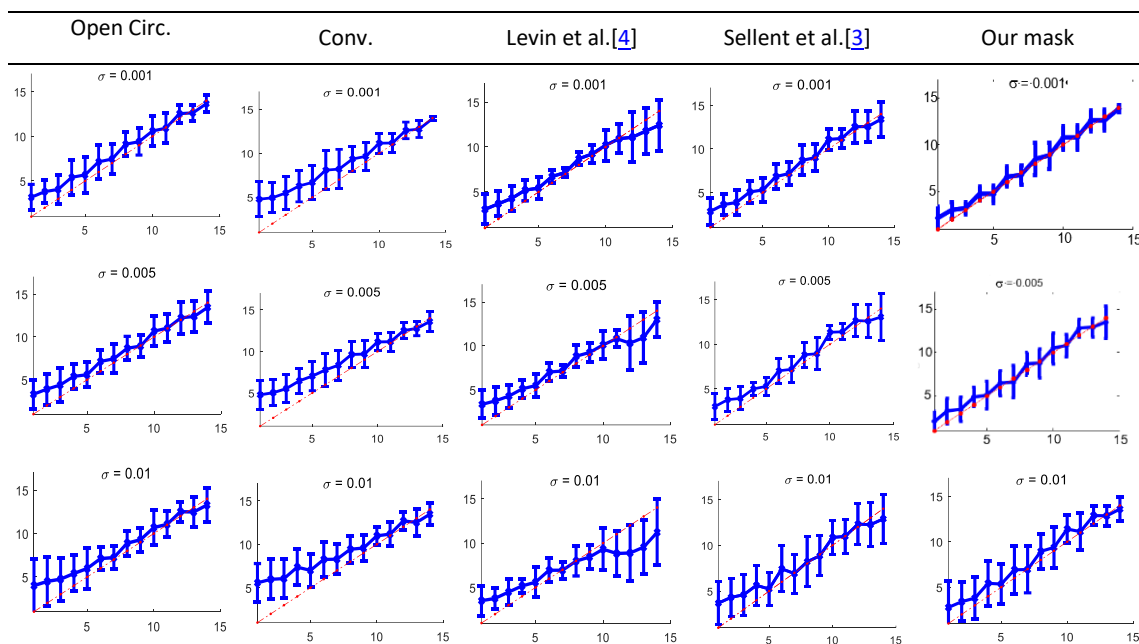
In our experiment, parameters are set to $\lambda_0=1000$ and $\sigma_\lambda=0.006$. Finally, α -expansion is used to minimize the energy function [41].

Experiment

The proposed mask and depth estimation method are validated in several experiments. The mask is compared with circular aperture, conventional aperture and two other masks designed for depth estimation [3], [4] (It must be mentioned that our study does not include aperture patterns proposed for deblurring, which assume to have sufficient information about blurring kernel and only focus on deblurring results). Among the masks proposed by Sellent et al. [3], we choose the 7x7 mask, which is the best according to our evaluating criteria (see Fig. 2 and 3). Our study contains synthetic and real experiments. It is expected that the designed mask increases the accuracy of PSF estimation and provides desirable deblurring results.



(a) A few number of patches used in the experiments. (b) Average of depth estimation error of blur scales ($s = 1:14$).



(c) The average and variance of estimated blur scale (vertical axis) in comparison with ground truth scale (horizontal axis). Red diagonal represents the ideal estimation.

Fig. 8: Results of depth estimation for five apertures at 3 noise levels ($\sigma=0.001, 0.005, 0.01$) and 14 blur sizes ($s = 1:14$).

A. Synthetic Experiments

1) Depth Estimation Accuracy

In the first experiment, a number of various images are blurred uniformly with various blur scales ($s=1:14$). Then, 50 patches of these images are randomly selected and their depth is estimated by the method described in Sec.6. Fig. 8(a) shows some of the selected patches. In each scale, the mean and variance of estimated size of PSFs are computed over all patches.

This experiment is repeated for different aperture patterns at three levels of noise ($\sigma = 0.001, 0.005, 0.01$). Based on the results shown in Fig. 8(c), the depth estimation accuracy is reduced by increasing noise. However, results are satisfactory especially in our mask and the mask proposed by Sellent et al. [3]. It must be mentioned that since both symmetric and asymmetric patterns are studied in this experiment, only one side of the focal plane is considered.

For better comparison of studied aperture patterns, in each scale, the norm of difference between the ground truth blur scale (s_{gt}) and the estimated blur scale (s_{es}) is computed over all patches (i.e. $\sum_{p=1}^{50} (s_{gt}^p - s_{es}^p)^2$). Then, this value is averaged over all studied blur scales. Fig. 8(b) shows the mean square error (MSE) of depth estimation for different apertures at three noise levels. It shows that under equal circumstances, where all imaging conditions (including throughput) are the same, coded pattern has greater performance than its corresponding conventional aperture.

The depth estimation experiment is repeated for asymmetric patterns with blur sizes in the range of -12:12 pixel. Since a blur size of 0 is meaningless and ± 1 indicates a sharp image, 23 different sizes of blur are indeed examined. According to Fig. 9, our method provides favorable results at $\sigma = (0.001, 0.005)$ with the depth estimation error (MSE) of the proposed aperture being less than the pattern in [3].

II) Deblurring Results

In the second experiment, deblurring results of aperture patterns are examined. For different scales of blur, each blurred patch is deblurred with a correct scale of PSF. Then, the Root Mean Square Error (RMSE) of the difference between original sharp image and its deblurred version is computed. The average of RMSE is calculated over all patches.

As shown in Fig. 10, our pattern provides the least error, especially in large blur scales, while the conventional aperture is the best aperture in lower blur scales.

A sample of deblurring result for Circular Zone Plate (CZP) chart is shown in Fig. 11. In all experiments, images are deblurred by the sparse deconvolution algorithm proposed by Levin et al. [4].

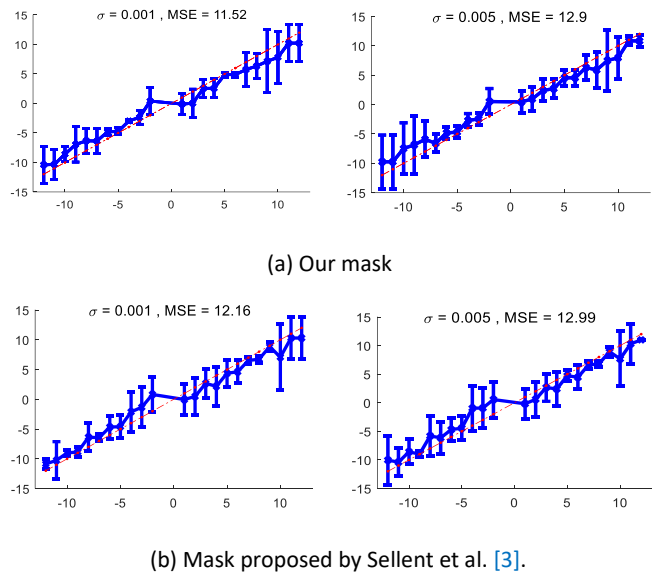


Fig. 9: The average and variance of estimated blur scale (vertical axis) compared to ground truth scale (horizontal axis) at 2 noise levels ($\sigma=0.001, 0.005$) in the depth range of -12:12. Red diagonal represents the ideal estimation.

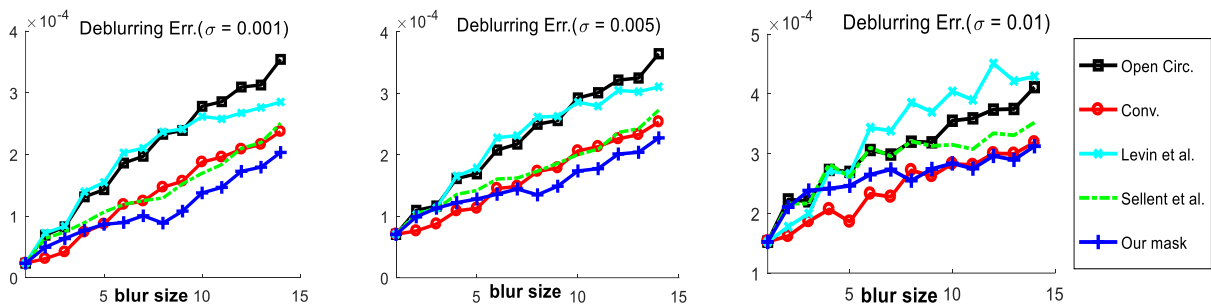


Fig. 10: Deblurring error of five apertures at 3 noise levels ($\sigma=0.001, 0.005, \text{ and } 0.01$) for 14 blur sizes ($s = 1:14$).

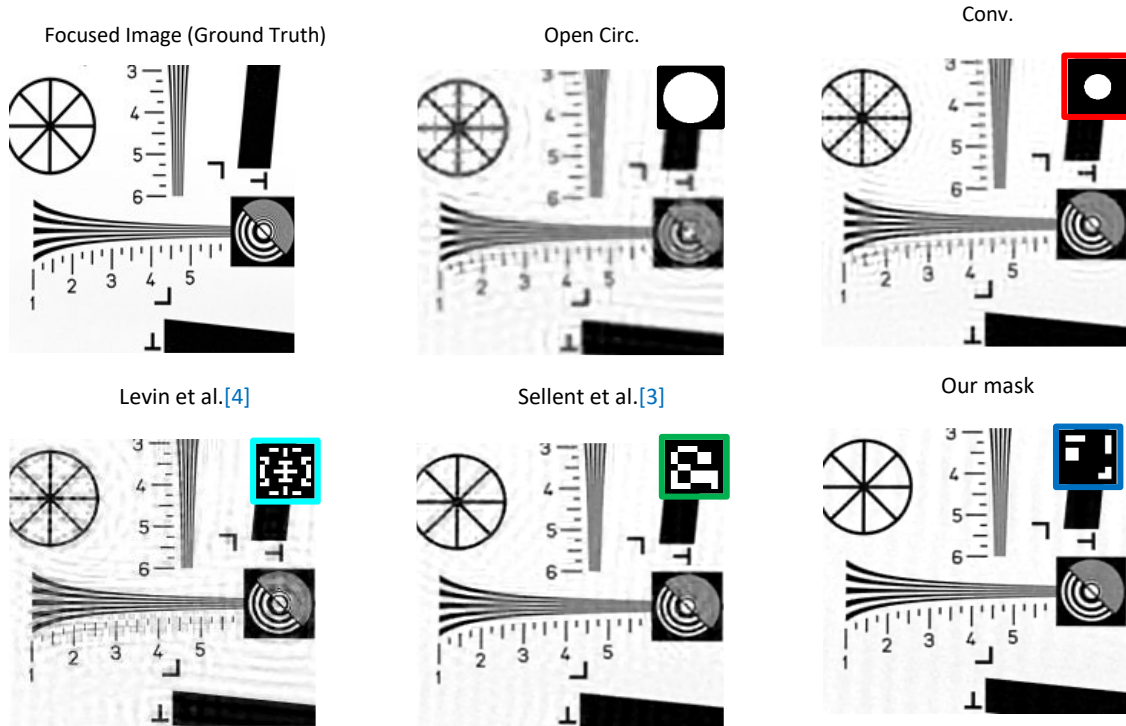


Fig. 11: Comparison of deblurring results derived from different aperture patterns (blur size = 13, $\sigma=0.005$).

B. Real Scene

For real experiments, the proposed pattern is printed on a single photomask sheet. It is cut out of the photomask sheet and inserted into a camera lens. In our experiment, a Canon EOS 1100D camera with an EF 50mm f/1.8 II lens is used. The disassembled lens and the one assembled with the proposed mask are shown in Fig. 12(a, b).

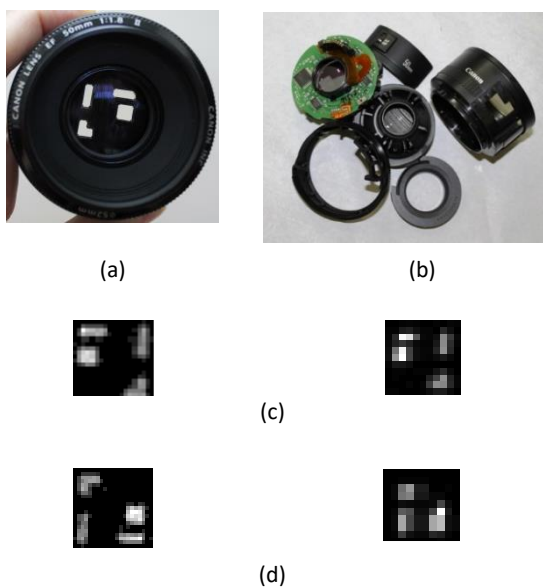


Fig.12: (a) lens assembled with the proposed mask, (b) disassembled lens. (c), (d) calibrated PSFs of evaluated pattern.

A very thin LED is used to calibrate the true PSF. The LED is mounted behind a pierced black cardboard to make a point light source. Since the position of focal point may be changed in each experiment, the camera focus is set to a sample point. Then, the camera is moved back and forth up to 60cm in 5cm increments and an image is captured at each depth. Each image is cropped according to the surface in which the point light spreads. Afterward, using some threshold values, the residual light is cleared and the result is normalized. In some rare cases, there is a jump in the PSF scale in consecutive measured PSFs. Under these conditions, other PSF scales are generated synthetically from the obtained PSFs. In this way, a bank of PSFs is generated that covers all possible sizes of PSF in the range [-19:+19]. The camera is set to $F\# = 2$ and the illumination is set to office room lighting condition (i.e. 300 lux). Fig. 12(c, d) shows some calibrated PSFs in forward and backward points of focus.

In the first experiment, the focal point is set to the farthest point and all objects are placed in front of it. The captured images and results are shown in Fig. 13(a). The index number in the color-bar shows relative distance to the camera so that in each figure, the closer object is colored with smaller index.

Although the results are acceptable, there are some errors of depth estimation on the floor of the scene that should be corrected by the user or other segmentation techniques, which may not be so sensitive to intensity similarity.

In the second experiment, three objects are placed in the back of, over and in front of the focus point.

Fig. 13(b) shows the captured image along with the depth map. In the third experiment, the focal point is set to the nearest object with all other objects being placed behind that.

According to Fig. 13(c) our method can achieve acceptable results in this case.

Each depth-map is slightly corrected and then deblurring [4] is performed with the modified depth map. Fig. 13(c) shows all-focus images derived from deblurring.

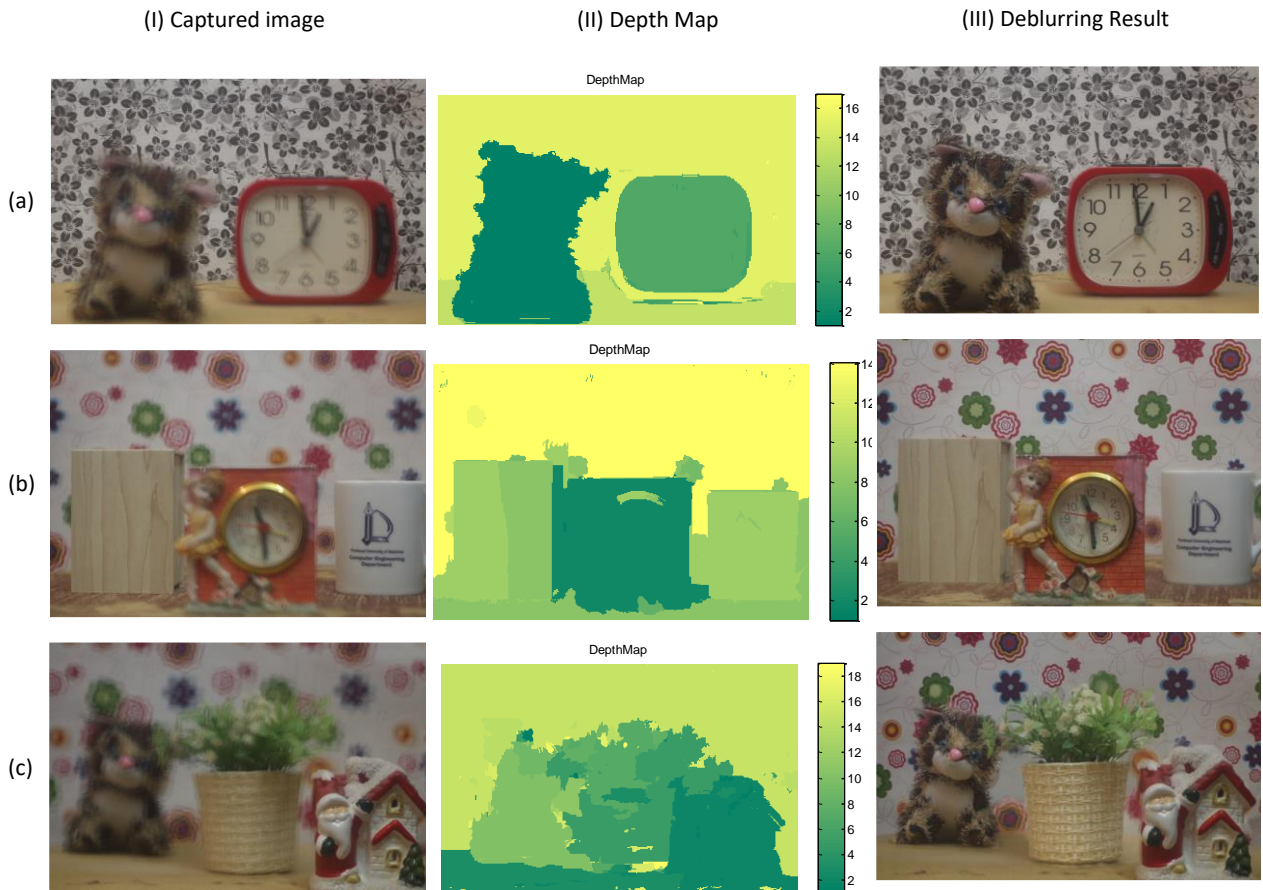


Fig. 13: Depth map estimation of depth varying scenes: (a) in front of the focal plane, (b) both sides of the focal plane, (c) at the back of the focal plane.

Conclusion

In this paper, a new method of aperture mask evaluation was proposed, which could reduce estimation error in both depth map and deblurring results. Asymmetric apertures make different PSFs in the back and front of the focal point. This feature could help discriminate blurred objects on two sides of the focal plane. The aperture pattern was designed for a specific imaging condition. Our future work will be concerned with defining an objective function in which the exposure time is also considered as an unknown variable of the problem and the SNR of captured images determines the lower bound of the mask throughput. Our proposed mask was intended for indoor illumination setting.

by considering the aperture throughput and imaging conditions, an exact evaluation of masks with different throughput could be done. Analytical and experimental results showed that our proposed mask could estimate an appropriate depth map of objects captured in one image regardless of the side of the focal plane. This was achieved with the help of a new depth estimation algorithm proposed in this article. According to the proposed algorithm, the deblurring result of correct PSF has the highest quality, which helps PSF estimation. Although the proposed no-reference quality measure yielded desirable results in depth estimation, more studies are required to obtain better measures which can reduce depth estimation error in both conventional and coded aperture imaging.

Author Contributions

M. Masoudifar designed and implemented the experiments, carried out the data analysis, and wrote the manuscript. H. Pourreza interpreted the results and revised the manuscript.

Acknowledgment

This work is completely self-supporting, thereby no financial agency's role is available. The authors gratefully thank the anonymous reviewers and the editor of JECEI.

Conflict of Interests

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

Abbreviations

<i>DFD</i>	Depth from defocus
<i>PSF</i>	Point Spread Function
<i>NSGA</i>	Non-dominated Sorting Genetic Algorithm
<i>SNR</i>	Signal to Noise Ratio
<i>DoF</i>	Depth of Field
<i>MSE</i>	Mean Square Error
<i>RMSE</i>	Root Mean Square Error

References

- [1] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(4): 523-531, 1987.
- [2] M. Subbarao, N. Gurumoorthy, "Depth recovery from blurred edges," in *Proc. Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'88*: 498-503, 1988.
- [3] A. Sellent, P. Favaro, "Which side of the focal plane are you on?," in *Proc. 2014 IEEE International Conference on Computational Photography (ICCP)*: 1-8, 2014.
- [4] A. Levin, R. Fergus, F. Durand, W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graphics*, 26(3): 70, 2007.
- [5] A. Sellent, P. Favaro, "Optimized aperture shapes for depth estimation," *Pattern Recognit. Lett.*, 40: 96-103, 2014.
- [6] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graphics*, 26(3): 1-12, 2007.
- [7] C. Zhou, S. Nayar, "What are good apertures for defocus deblurring?" in *Proc. 2009 IEEE International Conference on Computational Photography (ICCP)*: 1-8, 2009.
- [8] B. Masia, L. Presa, A. Corrales, D. Gutierrez, "Perceptually optimized coded apertures for defocus deblurring," *Comput. Graphics Forum*, 31(6): 1867-1879, 2012.
- [9] Y. Takeda, S. Hiura, K. Sato, "Fusing depth from defocus and stereo with coded apertures," in *Proc 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 209-216, 2013.
- [10] C. Zhou, S. Lin, S. K. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," *Int. J. Comput. Vision*, 93(1): 53-72, 2011.
- [11] A. Konak, D. W. Coit, A. E. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliab. Eng. Syst. Saf.*, 91(9): 992-1007, 2006.
- [12] K. Mitra, O. S. Cossairt, A. Veeraraghavan, "A framework for analysis of computational imaging systems: Role of signal prior, sensor noise and multiplexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(10): 1909-1921, 2014.
- [13] V. Paramonov, I. Panchenko, V. Bucha, A. Drogolyub, S. Zagoruyko, "Depth camera based on color-coded aperture," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*: 1-9, 2016.
- [14] V. Aslantas, "A depth estimation algorithm with a single image," *Opt. express*, 15(8): 5024-5029, 2007.
- [15] S. Zhuo, T. Sim, "Defocus map estimation from a single image," *Pattern Recognit.*, 44(9): 1852-1858, 2011.
- [16] J. Lin, X. Ji, W. Xu, Q. Dai, "Absolute depth estimation from a single defocused image," *IEEE Trans. Image Process.*, 22(11): 4545-4550, 2013.
- [17] X. Zhu, S. Cohen, S. Schiller, P. Milanfar, "Estimating spatially varying defocus blur from a single image," *IEEE Trans. Image Process.*, 22(12): 4879-4891, 2013.
- [18] S. Gur, L. Wolf, "Single image depth estimation trained via depth from defocus cues," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 7683-7692, 2019.
- [19] P. Hambarde, S. Murala, "S2DNet: Depth estimation from single image and sparse samples," *IEEE Trans. Comput. Imaging*, 6: 806-817, 2020.
- [20] A. N. Rajagopalan, S. Chaudhuri, "Optimal selection of camera parameters for recovery of depth from defocused images," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 219-224, 1997.
- [21] M. Watanabe, S. K. Nayar, "Rational filters for passive depth from defocus," *Int. J. Comput. Vision*, 27(3): 203-225, 1998.
- [22] P. Favaro, S. Soatto, "A geometric approach to shape from defocus," *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3): 406-417, 2005.
- [23] S. Matsui, H. Nagahara, R.I. Taniguchi, "Half-sweep imaging for depth from defocus," *Image Vision Comput.*, 32(11): 954-964, 2014.
- [24] M. Ye, X. Chen, Q. Li, J. Zeng, S. Yu, "Depth from defocus measurement method based on liquid crystal lens," *Opt. Express*, 26(2): 28413-28420, 2018.
- [25] S. W. Hasinoff, K. N. Kutulakos, "Confocal stereo," *Int. J. comput. vision*, 81(1): 82-104, 2009.
- [26] A. Rajagopalan, S. Chaudhuri, U. Mudenagudi, "Depth estimation and image restoration using defocused stereo pairs," *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(11): 1521-1525, 2004.
- [27] S. Hiura, T. Matsuyama, "Depth measurement by the multi-focus camera," in *Proc. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 953-959, 1998.
- [28] O. Cossairt, M. Gupta, S. K. Nayar, "When does computational imaging improve performance?," *IEEE Trans. Image Process.*, 22(2): 447-458, 2013.
- [29] Y. Weiss, W. T. Freeman, "What makes a good model of natural images?," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*: 1-8, 2007.
- [30] P. E. Debevec, J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*: 1-10, 2008.

- [31] O. Cossairt, "Tradeoffs and limits in computational imaging," Columbia University, 2011.
- [32] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, 6(2): 182-197, 2002.
- [33] Y. Gao, "Population size and sampling complexity in genetic algorithms," in *Proc. the Bird of a Feather Workshops*: 178-181, 2003.
- [34] S. Theodoridis, R. Chellappa, *Academic Press Library in Signal Processing: Image, Video Processing and Analysis, Hardware, Audio, Acoustic and Speech Processing*: Elsevier Science, 2013.
- [35] M. Martinello, "Coded aperture imaging," Heriot-Watt University, Edinburgh, Scotland, 2012.
- [36] Y. Liu, J. Wang, S. Cho, A. Finkelstein, S. Rusinkiewicz, "A no-reference metric for evaluating the quality of motion deblurring," *ACM Trans. Graph.*, 32(6): 175, 2013.
- [37] B. Hu, L. Li, J. Qian, "Perceptual quality evaluation for motion deblurring," *IET Comput. Vision*, 12(6): 796-805, 2018.
- [38] D. Krishnan, T. Tay, R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 233-240, 2011.
- [39] G. Blanchet, L. Moisan, "An explicit sharpness index related to global phase coherence," in *Proc. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*: 1065-1068, 2012.
- [40] M. Masoudifar, H. R. Pourreza, "Depth estimation from a single defocused image using no reference image quality assessment metrics," presented at the Fifth International Conference on Technology Development in Iranian Electrical Engineering, 2021.
- [41] A. DeLong, A. Osokin, H. N. Isack, Y. Boykov, "Fast approximate energy minimization with label costs," *Int. J. comput. vision*, 96(1): 1-27, 2012.

Biographies



Mina Masoudifar received the B.Sc. degree in computer engineering from Sharif University, Tehran, Iran, and the M.Sc. and Ph.D. degrees from Ferdowsi University of Mashhad, Iran, in 1996, 1999, and 2017, respectively. She is an Assistant Professor in computer engineering at the Department of Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran. Her main research interests are machine vision, computational photography, image quality assessment, and deep learning.

- Email: masoudi@hsu.ac.ir
- ORCID ID: 0000-0002-9609-1853
- Web of Science Researcher ID: NA
- Scopus Author ID: NA
- Homepage: <http://staff.hsu.ac.ir/persons/?perid=100252>



HamidReza Pourreza is currently Professor of Computer Science and Engineering at Ferdowsi University of Mashhad (FUM). He received his B.S. degree in Electrical Engineering from FUM in 1989, and received his M.S. and Ph.D. degrees in Electrical Engineering and Computer Engineering from Amirkabir University of Technology in 1993 and 2003, respectively. His research interests are in the areas of Deep Learning, Computer Vision, Hardware Design, and Intelligent Transportation Systems (ITS).

- Email: hpourreza@um.ac.ir
- ORCID ID: 0000-0002-3560-8070
- Web of Science Researcher ID: B-8754-2015
- Scopus Author ID: 23968187500
- Homepage: <https://hpourreza.profcms.um.ac.ir/>

How to cite this paper:

M. Masoudifar, H. R. Pourreza, "Depth estimation and deblurring from a single image using an optimized-throughput coded aperture," *J. Electr. Comput. Eng. Innovations*, 11(1): 51-64, 2023.

DOI: [10.22061/JECEI.2022.8630.537](https://doi.org/10.22061/JECEI.2022.8630.537)

URL: https://jecei.sru.ac.ir/article_1718.html

