

ROTATING CODED APERTURE FOR DEPTH FROM DEFOCUS

Jingyu Yang, Jinlong Ma, Bin Jiang

School of Electronic Information Engineering, Tianjin University
Building 26-D, No. 92, Weijin Road, Nankai District, Tianjin 300072, China

ABSTRACT

This paper proposes a rotating coded aperture we called RCA for depth recovery, complementing the previous depth from defocus methods in both recovery accuracy and friendliness for hardware implementation. In the proposed RCA, an optimized coded aperture is rotated around the optical axis, and multiple images are taken at a few locations of the rotated aperture. We derive a criterion to evaluate the depth recovery performance of the RCA. With this criterion, the coded pattern of rotating aperture and the angles of rotations are jointly optimized by a genetic algorithm combined with a multiscale refinement strategy. The proposed RCA is demonstrated to have better depth from defocus performance over other multiple coded aperture schemes.

Index Terms— computational photography, rotating coded aperture, depth from defocus

1. INTRODUCTION

Generally, 3D information of a scene consists of texture information and depth information. However, depth information is more difficult to acquire than the photometric information that can be captured with various matured sensors. There are mainly two categories of depth acquisition methods: 1) active methods, including laser scanning, time-of-flight depth measurement, and structured light based sensing, and 2) passive methods, including stereo matching and depth from defocus (DFD) [1, 2, 3]. While each approach has its own merits and drawbacks, depth from defocus techniques are more robust to occlusion and correspondence difficulty, and are more easy to be integrated into imaging devices with only moderate modifications [4]. When captured through an optical lens, objects of different depths to the camera are recorded with different levels of sharpness due to the out-of-focus blurring. Most of the modifications to recover depth from defocus require multiple images or active methods with extra apparatus. It is challenging to precisely estimate the amount of blur [5, 6]. Coded apertures have shown superiority in out-of-focus deblurring over traditional nearly circular apertures [7, 8, 9]. The depth

recovery performance of DFD has also been improved significantly by using coded apertures instead of circular ones [10]. Levin *et al.* [1] designed a coded aperture that is more discriminative for depth sensing, and proposed a depth estimation approach with sparse regularization. In early DFD work, depth is estimated from a pair of two images captured with two circular apertures of difference sizes [11, 12, 13]. However, the circular shape of aperture pairs makes it difficult to discriminate depth from blurring. Instead of circular apertures, Zhou *et al.* [14, 15] propose optimizing coded aperture pairs for DFD to maximize depth discrimination, significantly enhancing depth recovery performance. However, for practical implementation, it would be problematic since coded aperture pairs require enough room to switch apertures.

To remedy these issues, this paper proposes a rotating coded aperture for DFD. In our approach, a coded aperture is rotated to a few locations around the optical axis, and one image is taken at each location. Then, depth information is estimated from the multiple captured images. At the coarsest scale, the code pattern is first initialized at the solution optimized by a genetic algorithm; then, for each scale, the coded pattern from the last scale is interpolated and refined by coordinate decent. The angles of rotations are optimized by the genetic algorithm with coded pattern. Simulation results demonstrate the superiority of the proposed rotating aperture over the previous coded aperture or aperture pairs. The main contributions of our work are twofold: 1) We propose the RCA scheme for DFD, which does not only have better depth recovery performance, but is also more convenient to integrated in commodity camera. The aperture can be rotated by bridging the aperture to the in-lens electronic focus drive motor with several small driving gears. 2) We derive the optimization of RCA in terms of depth recovery performance, extending the optimization framework in [14].

2. PROPOSED METHODS

2.1. Criterion for RCA Evaluation

When the rotations are more than three times, the theoretical derivation process will become very complex and the performance of coded aperture cannot get an obvious promotion. In this paper, we utilize three rotated versions of an aperture and correspondingly capture three images F_i ($i=1,2,3$) of the same

This work was supported by the Normal Project under Grant 61372084 from the National Natural Science Foundation of China (NSFC), and by the Reserved Peiyang Scholar Program of Tianjin University, Tianjin, China.

scene with different defocus characteristics or PSFs. This process is denoted as

$$F_i = F_0 \cdot K_i^d + N_i, \quad (1)$$

where K_i^d represents the Fourier transform of the corresponding PSF with blur size d . F_0 means the observed blurred image in frequency domain, and N_i is Gaussian white noise. Inspired by the method of computing the optimum aperture in [14], we propose to solve a maximum a posteriori (MAP) problem to obtain the size d and then get an estimate of the depth map. The problem can be expressed as the minimization of an energy function

$$E(\hat{d}|F_1, F_2, F_3, \sigma) = \min_{F_0} \sum_{i=1,2,3} \|\hat{F}_0 \cdot K_i^{\hat{d}} - F_i\|^2 + \|C \cdot \hat{F}_0\|^2 \quad (2)$$

There are two unknown variables, \hat{F}_0 and \hat{d} , in Eq. (2). We first estimate \hat{F}_0 and then use it to calculate \hat{d} using Wiener deconvolution. For an initialized \hat{d} , solving $\partial E / \partial \hat{F}_0 = 0$ yields

$$\hat{F}_0 = \frac{\sum_i F_i \cdot \bar{K}_i^{\hat{d}}}{\sum_i |K_i^{\hat{d}}|^2 + |C|^2}, i = 1, 2, 3 \quad (3)$$

where \bar{K} is the complex conjugate of K . The noise to signal ratio parameter $|C|^2$ is approximated by σ^2/A , where A is defined as the power distribution of natural images according to the $1/f$ law [9]: $A = \int |F_0(\xi)|^2 \mu(F_0)$ and parameter ξ means the frequency. Then we substituting Eqs. (1) and (3) into (2). Using the $1/f$ law of natural images, Eq. (2) can be transformed to

$$E(\hat{d}|K_1^{d^*}, K_2^{d^*}, K_3^{d^*}, \sigma) = \int_{F_0} E(\hat{d}|K_1^{d^*}, K_2^{d^*}, K_3^{d^*}, \sigma, F_0) \mu(F_0) \quad (4)$$

And we arrange and simplify it as:

$$\begin{aligned} & E(\hat{d}|K_1^{d^*}, K_2^{d^*}, K_3^{d^*}, \sigma) \\ &= \sum_{\xi} A \frac{|K_1^d K_2^{d^*} - K_1^{d^*} K_2^d|^2 + |K_1^d K_3^{d^*} - K_1^{d^*} K_3^d|^2 + |K_2^d K_3^{d^*} - K_2^{d^*} K_3^d|^2}{|K_1^d|^2 + |K_2^d|^2 + |K_3^d|^2 + C^2} \\ &+ \sigma^2 \cdot \sum_{\xi} \left[\frac{C^2}{|K_1^d|^2 + |K_2^d|^2 + |K_3^d|^2 + C^2} + 1 \right] \end{aligned} \quad (5)$$

Therefore, we evaluate RCA at d^* when the noise level is σ using

$$\begin{aligned} & R(K_1, K_2, K_3|d^*, \sigma) \\ &= \min_{d \in D/d^*} E(d|K_1^{d^*}, K_2^{d^*}, K_3^{d^*}, \sigma) - E(d^*|K_1^{d^*}, K_2^{d^*}, K_3^{d^*}, \sigma) \end{aligned} \quad (6)$$

where $D = \{c_1 d^*, c_2 d^*, \dots, c_l d^*\}$ is a set of blur size samples. In our implementation, $\{c_i\}$ is set to $\{0.1, 0.15, \dots, 1.5\}$ at step 0.05. We then normalize Eq. (6) and get

$$\begin{aligned} & M(K_1, K_2, K_3, d, d^*) \\ &= \left[\frac{1}{n} \sum_{\xi} A \frac{|K_1^d K_2^{d^*} - K_1^{d^*} K_2^d|^2 + |K_1^d K_3^{d^*} - K_1^{d^*} K_3^d|^2 + |K_2^d K_3^{d^*} - K_2^{d^*} K_3^d|^2}{|K_1^d|^2 + |K_2^d|^2 + |K_3^d|^2 + C^2} \right]^{\frac{1}{2}} \end{aligned} \quad (7)$$

We have

$$R = \min_{d \in D/d^*} M(K_1, K_2, K_3, d, d^*) \quad (8)$$

where n is the pixel number of the PSF. 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.

2.2. RCA Optimization

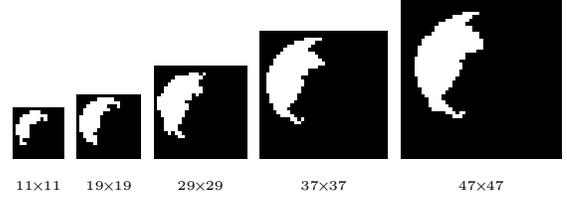


Fig. 1. Resolution enhancement of RCA by up-sampling and coordinate descent.

2.2.1. Generating Rotating Coded Aperture by Genetic Algorithm

We adopt the genetic algorithm (Table 1) to generate the optimal RCA. In Table 1, the aperture pattern k of size $N \times N$ can be computed from a sequence b which contains only 0 or 1 of the length N^2 . The first N elements from b were put into the first row of k and the next N were chosen corresponding to the next row, and so on. So the k can be constructed to represent the aperture pattern in the spatial domain. Using this method, we obtain the optimal RCA pattern with low resolution shown in Fig. 1(a). The optimal rotating angles with the optimal coded aperture patten θ_i are 0° , 110° , and 235° .

2.2.2. Improving the Aperture Resolution by Coordinate Descent

In order to obtain a high resolution optimal aperture pattern for practical application, we refine the RCA pattern obtained in Section 2.2.1. We first amplify the aperture resolution from 11×11 to 13×13 by bicubic interpolation and then refine the resolution through coordinate descent optimization. We repeat this process until reaching the resolution of 47×47 at the step of 2×2 when the performance cannot be improved anymore. Fig. 1 shows the refinement process with resolution varying from 11×11 to 47×47 . The far right aperture is our final optimized RCA pattern for depth estimation.

2.3. RCA-based Depth Estimation

First, we estimate depth information of the captured scene. We reconstruct the sharp images \hat{F}_0^d using Eq. (3) with a set

Table 1. Genetic Algorithm for RCA

Step 1: Initialization:

$D = \{c_1 d^*, c_2 d^*, \dots, c_l d^*\}; \theta_i^0 = [0^\circ, 120^\circ, 240^\circ];$

$g = 0;$ generate S random binary sequences of length $L = N * N;$

Step 2: Calculation (from $g = 1 : G$):

2.1. *Selection* : For each sequence b , we computed the kernel function k . And all of the kernel function $k_{i,(i=1,2,3)}$ are obtained by rotating k with the optimal rotation angel $\theta_{i,(i=1,2,3)}^{g-1}$ in the anti-clockwise direction. Then we calculate the corresponding constraint values by using Eq. (8) to choose the best Z out of S sequences. And the constraint values are computed through Eq. (8) again by using angle set H with the chosen Z sequences. Only the best angle combination is selected to be the optimal rotating angles θ_i^g .

2.2. Repeat until the number of sequences increases to S from Z .

Crossover : Copy two sequences from the Z sequences of Step 2.1 randomly, and then exchange each pair of corresponding bits with a probability of p_1 in turn, to obtain two new sequences.

Mutation : For each new sequence, flip each bit with a probability p_2 .

Step 3: Calculate all of the remaining sequences by Equation 8 and output the optimal result.

$d^*=7, N = 11, S = 4000, Z = 400, p_1 = 0.2, p_2 = 0.05, G = 60$ and H means a angle set, we set the first rotation angle of aperture as 0° , then the other two angles combined from 0° to 355° at the step of 5° .

of PSFs under different depths. Then the corresponding error $e(x, y)$ can be computed through the sharp images,

$$e(x, y) = \min_{d \in D} \sum_{i=1,2,3} \left| f_i - IFT(\hat{F}_0^d \cdot K_i^d) \right| \quad (9)$$

where IFT is the 2-D inverse Fourier transform. The depth corresponding to the minimal residual at each local region is the desired depth information. By averaging the reconstruction error over a small local window, we can get a depth map of the captured scene.

Second, the in-focus image can be reconstructed by using the calculated depth map. We use the Wiener deconvolution similar to Eq. (3) to deblur the defocus image by using corresponding kernel under estimated depth. Finally we can recover the all-focus image \hat{f}_0 as,

$$\hat{f}_0 = IFT(\hat{F}_0^{e(x,y)}(x, y)) \quad (10)$$

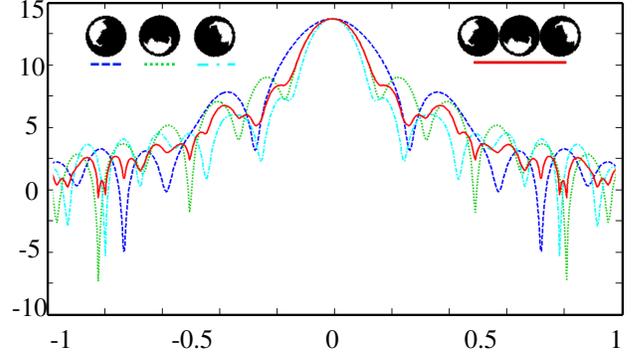


Fig. 2. The joint of power spectra (normalized frequency) of three different rotating aperture patterns: the red solid curve is the average of three rotating coded apertures, the other curves are corresponding to each single rotating aperture.

3. EXPERIMENTAL RESULTS

3.1. Setup

3.1.1. Simple Scenes

We use two simple models which are attached with rich and weak textures half-by-half as shown in Fig. 3(a) and 3(b) to evaluate the RCA patterns.

Ladder: As the distance between two adjacent steps is 10 cm, *Ladder cube* has many granular depth layers distributing from 0.5 m to 2.0 m. The focal plane is set at 1.0 m.

Cone: Depth in this scene is continuously between 1.0 m and 4.0 m, and the focal plane is at 1.0 m.

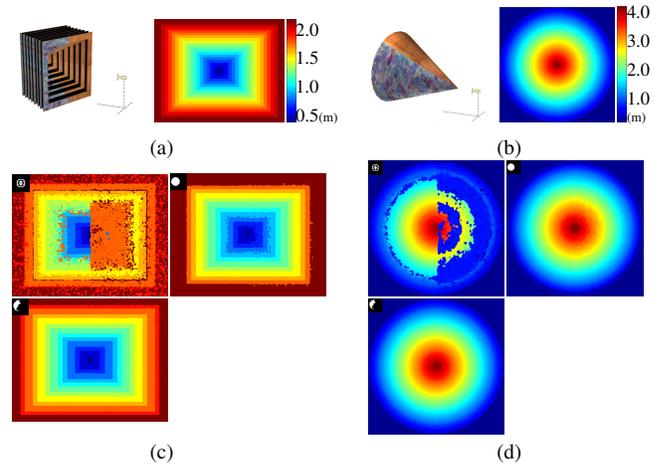


Fig. 3. DFD results on two synthetic 3D scenes: (a) (b) 3D scenes and the corresponding ground-truth depth maps, the left group is *Ladder* and the right group is *Cone*; (c) estimated depth maps of *Ladder* recovered by method in [1], rotating circular aperture, and our rotating aperture; (d) estimated depth maps of *Cone*.

	Rich texture		Weak texture	
	<i>Ladder</i>	<i>Cone</i>	<i>Ladder</i>	<i>Cone</i>
Levin’s aperture [1]	0.6199	0.5036	1.7092	3.1856
Circular aperture	0.5141	0.2464	0.5045	0.2595
Proposed aperture	0.2875	0.1211	0.2889	0.1267

Table 2. Quantitative DFD results on simple models in terms of root mean square error (RMSE).

The distances mentioned above are relative to the associated image planes of the virtual camera with a focal length of 50 mm. All the images taken from the 3D scenes are normalized into $[0, 1]$, and are polluted with white Gaussian noise with standard deviation $\sigma = 0.005$ to verify the error resilience of the evaluated methods.

3.1.2. Complex Scenes

To further evaluate our method in a more realistic scenario, we create more realistic indoor and outdoor scenes, named *Living_room* in Fig. 4(a) and 4(b), and *Garden* in Fig. 4(c) and 4(d). The *Living_room* is an indoor scene. The depth of the scene is between 0.5 m and 3.0 m, and the focal plane is at 1.5 m. The *Garden* is an outdoor scene. The depth of the scene is between 0.8 m and 6.8 m approximately, and the focal plane is at 3.0 m.

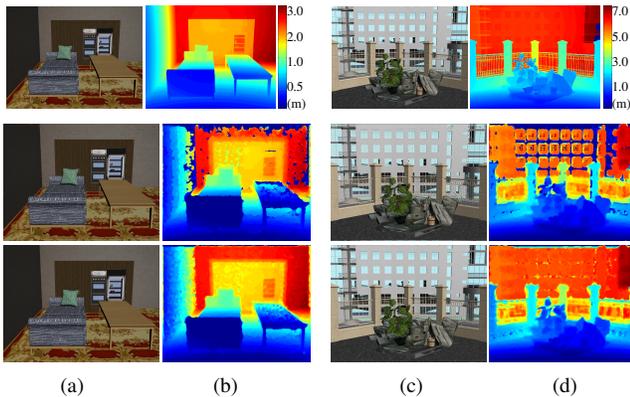


Fig. 4. DFD results on *Living_room* and *Garden*: the top row means the ground truth sharp image and corresponding depth map with two scenes, the middle row shows the recovered all-focus image and corresponding estimated depth map by using the coded aperture pairs [14], and the bottom row shows the recovered image and estimated depth map by using our RCA; (a) sharp images about *Living_room*, (b) depth maps about *Living_room*, (c) (d) sharp images and corresponding depth maps about *Garden*.

	RMSE		PSNR(dB)	
	<i>Living_room</i>	<i>Garden</i>	<i>Living_room</i>	<i>Garden</i>
Zhou’s pairs [14]	3.3045	3.8770	34.37	31.36
Proposed aperture	1.2440	1.7230	39.05	35.38

Table 3. Quantitative DFD results on three realistic 3D scenes in terms of RMSE and PSNR.

3.2. Results and Discussions

Quantitative results of the recovered depth maps are presented in Table 2, and are consistent with the visual results shown in Fig. 3. Our RCA has the lowest root-mean-squared errors (RSME) for the estimated depth maps against the ground truth. The results have shown the superiority of RCA over other aperture patterns. For the realistic indoor and outdoor scenes, we recover the depth estimation and obtain the sharp image from defocus shown in Fig. 4. In Table 3, the RMSE results between the true depth map and estimated depth map and the peak signal-to-noise ratio (PSNR) between the ground truth image and recovered sharp image results show that the RCA is able to recover more accurate depth information and better all-focus images than coded aperture pairs in [14].

Aperture patterns with more zero-crossings in the frequency domain shown in Fig. 2, are more powerful in distinguishing the blurring kernels with respect to the associated depth. In our RCA, the zero-crossings of the aperture are spanned across the 2D frequency domain through rotation. The pattern of the aperture is optimized to maximize the depth distinguishability under the rotation. In this way, the zero-crossings of the blurring kernels at different scales do not coincide so that the associated depth layers can be identified.

4. CONCLUSION

In this work, we represent rotating coded aperture to recover depth information from defocus. In our RCA, we need only an optimized coded mask, and multiple images are captured at a set of optimal rotated locations of the aperture. We first derive a criterion to measure the goodness of a RCA in terms of depth from defocus. The coded pattern is optimized in a multiscale framework: the shape is first searched at the lowest resolution, and then refined progressively via coordinate descent with a designed cost function. Simulation results show that our RCA significantly improve DFD performance over the previous coded apertures. The performance could be improved by selecting a more powerful deconvolution framework, and the key is to trade off the complexity of mathematical manipulations.

5. REFERENCES

- [1] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and depth from a conventional camera with a coded aperture,” *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 70, 2007. 1, 3, 4
- [2] A. Rajagopalan and S. Chaudhuri, “Depth estimation and image restoration using defocused stereo pairs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1521–1525, 2004. 1
- [3] A. Chakrabarti and T. Zickler, “Depth and deblurring from a spectrally-varying depth-of-field,” in *European Conference on Computer Vision (ECCV)*, 2012, pp. 648–661. 1
- [4] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen, “Programmable aperture photography: multiplexed light field acquisition,” *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 55, 2008. 1
- [5] A. P. Pentland, “A new sense for depth of field,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, 1987. 1
- [6] Y. Takeda, S. Hiura, and K. Sato, “Fusing depth from defocus and stereo with coded apertures,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 209–216. 1
- [7] J. Lin, X. Ji, W. Xu, and Q. Dai, “Absolute depth estimation from a single defocused image,” *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4545–4550, 2013. 1
- [8] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, “Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 69, 2007. 1
- [9] C. Zhou and S. Nayar, “What are good apertures for defocus deblurring?,” in *International Conference of Computational Photography (ICCP)*, 2009, pp. 1–8. 1, 2
- [10] C. Ma, J. Suo, Q. Dai, R. Raskar, and G. Wetzstein, “High-rank coded aperture projection for extended depth of field,” in *International Conference of Computational Photography (ICCP)*, 2013, pp. 1–9. 1
- [11] M. Subbarao and N. Gurumoorthy, “Depth recovery from blurred edges,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1988, pp. 498–503. 1
- [12] H. Farid and E. P. Simoncelli, “Range estimation by optical differentiation,” *Journal of the Optical Society of America*, vol. 15, no. 7, pp. 1777–1786, 1998. 1
- [13] S. Hiura and T. Matsuyama, “Depth measurement by the multi-focus camera,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998, pp. 953–959. 1
- [14] C. Zhou, S. Lin, and S. Nayar, “Coded aperture pairs for depth from defocus and defocus deblurring,” *International Journal of Computer Vision*, vol. 93, no. 1, pp. 53–72, 2011. 1, 2, 4
- [15] C. Zhou, S. Lin, and S. Nayar, “Coded aperture pairs for depth from defocus,” in *International Conference on Computer Vision (ICCV)*, 2009, pp. 325–332. 1