End-to-end optimization of coded aperture for extended depth of field

Zhexin Zhao
Electrical Engineering
Stanford University
zhexin@stanford.edu

Lingling Fan
Electrical Engineering
Stanford University
llfan@stanford.edu

Abstract

We apply the end-to-end optimization pipeline that enables extended aperture with the coded aperture and image processing pipeline (ISP). This overcomes the trade-off that in standard optics design or purely image processing towards extended depth of field, where the wide apertures are important to reduce imaging noise, but they also increase defocus blur. By evaluating the end-to-end result with standard designs on a benchmark image, we demonstrate that our coded aperture from the design pipeline can achieve the extended-depth-of-field image reconstruction as the previously much more complicated refractive index distribution design of lens or wave-front control, in a much easier and flexible manner.

1. Introduction

Deep optics by combining optics and post-processing methods provides a new pathway towards optimizing images, compared with the traditional purely image post-processing pipelines or only optics optimization. Such end-to-end optimization can be applied to a variety of applications including extended depth of view, de-blur, and super-resolution imaging. In this study, we apply the end-to-end framework to optimize the coded aperture for extended depth of field.

The study by V. Sitzmann et al [7] achieves the extended depth of field by designing the refractive index distribution. They optimize the geometry of the optical lens along with the image processing, where the tuning of point spread function of the optical system is achieved by changing the phase response of the lens. However, the optimized lens geometries, which require diamond turning or photolithography for fabrication, are costly and challenging for massive manufacturing.

On the other hand, point spread function tuning can also be achieved through changing the amplitude response, i.e. tuning the transparency of the aperture. Therefore, in principle, we can optimize the transmission of the aperture, basically a gray scale, or a binary, coded aperture, to achieve the same objectives. In this way, we only need to insert such a pattern layer in a known camera system. Such a pattern layer should be easier to fabricate. Moreover, the chromatic aberration using such coded apertures may be easier to solve.

2. Related work

Wave-front coding for extended depth of field. Bessel beam is known to achieve neglected diffraction over a large distance, resulting in large depth of field. In [3], they built a phase mask for an extended- depth-of-field optical system based on the ambiguity function and the stationary-phase method, in the near-diffraction-limited regime. The researchers use a single annulus aperture, such that the focusing beam after the optical system resembles a Bessel beam and has longer non-diffraction length compared to the Gaussian beam. Furthermore, comparing using a single annulus with shrinking the area of a circular aperture, which both provide extend depth of field, the former increases the diffraction limited resolution, while the latter decreases the diffraction limited resolution. Therefore, the single annulus aperture has been used to extend the depth of field in some microscopy systems. However, the limit of this work is that it only uses single annulus. Thus, the performance of the EDOF from the designed aperture may be outperformed by aperture with higher freedom optimization such as more anuli.

Focal sweep for extended depth of field. in. In [6], the motion of object or focal plane creates instantaneous invariant point spread function (PSF) at each depth. After integrating over time, the depth is depth-invariant. The limitation of this method is that it involves complicated motion of the optical, which is not easy to implement.

Diffusion coded photography for extended depth of field. In [2], the EDOF is achieved with diffuser as coded aperture, but has better inversion characteristics than lattice focal lens. Therefore the defocus deblurring is easier.

Lattice lens for extended depth of field. [5] superimposes array of lenses with different focal lengths. This optical

overlay creates the depth-invariant point spread function.

Accommodation-invariant (AI) near-eye displays with extended-depth-of-field cameras. In [4] another application of improving accommodation in current-generation VR/AR displays (near-eye displays) with extended-depth-of-field cameras is demonstrated.

End-to-end optimization design. In optical image recording perspective, [7] is using optimized refractive optical distribution to change the phase of the obtained image. In image processing pipeline perspective, this paper is using the Wiener deconvolution to reconstruct and deblur the image. By gradient descent optimization of both refractive profile and image reconstruction parameters, this optical design eventually achieves both depth and chromatic invariance. The limitation of this work is that the optics has more degree of freedom for optimization, not only the refractive profile, which controls the phase distribution of lens, but amplitude, transparency, polarization and so on. With exploring more degrees of freedom, one can expect to push the lens performance, further the whole pipeline to a higher level of extended depth of field.

Deep optics for monocular depth estimation. In [1], an application for monocular depth estimation with encoded blur with the end-to-end optimization framework is demonstrated.

3. Methodology

In this section, we first discuss why it is possible to decrease the depth dependence of the point spread function by tuning the coded aperture. We then introduce our end-to-end optimization pipeline.

The point spread function of the optical system, when the object is located at depth s_1, can be obtained from the diffraction optics:

$$p_{s_1} \propto \left|\mathcal{F}\left[A(x',y')e^{\frac{ik}{2}(\frac{1}{s_1}+\frac{1}{z}-\frac{1}{f})(x'^2+y'^2)}\right]\right|^2, \quad (1)$$

where $\mathcal{F}$ stands for the Fourier Transform, $A(x',y')$ is the amplitude transmission function of the coded aperture, $k = 2\pi/\lambda_0$ is the wave vector at the central wavelength $\lambda_0 = 550$ nm, s_1 is the depth of the object, z is the distance between the lens and the sensor, and f is the focal length of the lens system. In general, the point spread function depends on the depth. However, by designing the coded aperture, i.e. the amplitude transmission function $A(x',y')$, we can make the variation of the point spread function much smaller over a large range of depth.

We assume that the lens system is ideal, i.e. without any aberration, and the focal plane is at depth $s_0 = 1$ m, such that

$$\frac{1}{s_0}+\frac{1}{z}=\frac{1}{f}, \quad (2)$$

where we choose $z = 10$ mm. When the object is at the focal plane, the point spread function is just the absolute square of the Fourier Transform of the coded aperture. Using the lens equation (2), we can simplify the point spread function Eq. (1) as

$$p_{s_1} \propto \left|\mathcal{F}\left[A(x',y')\exp\left(-\frac{ik\delta}{2s_0}\right)\right]\right|^2, \quad (3)$$

where $\delta = (1/s_0 - 1/s_1)s_0$. We notice that the point spread function is the same for δ and $-\delta$. Thus, we focus on $\delta \geq 0$ and minimize the variance of the point spread function for $\delta \in [0,1]$, which corresponds to depth $s_1 \in [s_0, \infty]$. Since the point spread function is independent of the sign of δ, the depth of field is $s_1 \in [s_0/2, \infty]$.

In the image reconstruction, since we are unaware of the depth of the object, we use an averaged point spread function for the image de-convolution. The averaged point spread function is constructed as following:

$$p_{avg} = \frac{1}{N_s}\sum_{i=1}^{N_s} p_{s_1^{(i)}}. \quad (4)$$

To get the point spread function at different depth, we linearly sample $\delta \in [0,1]$ with spacing 0.1, i.e. $\delta = 0, 0.1, 0.2, \cdots, 1$.

It is counter intuitive that optimizing the coded aperture alone can produce a depth invariant point spread function. Without diffraction, it is indeed impossible. However, with diffraction, it is possible to create the second "focal plane" using constructive interference from the coded aperture. Therefore, the depth of field can be significantly extended.

For instance, we can assume that the coded aperture is made of several concentric rings. Suppose that these rings have constructive interference at another depth s_{n1}, then, the central radius r_n of the nth ring should satisfy the following relation:

$$\frac{kr_n^2\delta_{n1}}{2s_0} = \phi_0 + 2\pi n, \quad (5)$$

where $\delta_{n1} = (1/s_0 - 1/s_{n1})s_0$, ϕ_0 is a fixed phase, which is chosen as $\phi_0 = 0$ for simplicity, and n is an integer. These transparent rings are separated by a group of opaque rings, whose central radius $\bar{r}_n$ satisfies

$$\frac{k\bar{r}_n^2\delta_{n1}}{2s_0} = \phi_0 + (2n+1)\pi, \quad (6)$$

In our first attempt, we chose $s_{n1} = 5s_0$, i.e. $\delta_{n1} = 0.8$.

The end-to-end optimization pipeline is shown in Fig. 1. To construct the image on the sensor, we place each image from the dataset at a certain depth and convolve the image with the point spread function for that depth. Then, we add a gaussian noise $\mathcal{N}(0,\sigma_0)$ to obtain the image on sensor. We set the noise $\sigma_0 \propto 1/\sqrt{A}$, where A is the area of the

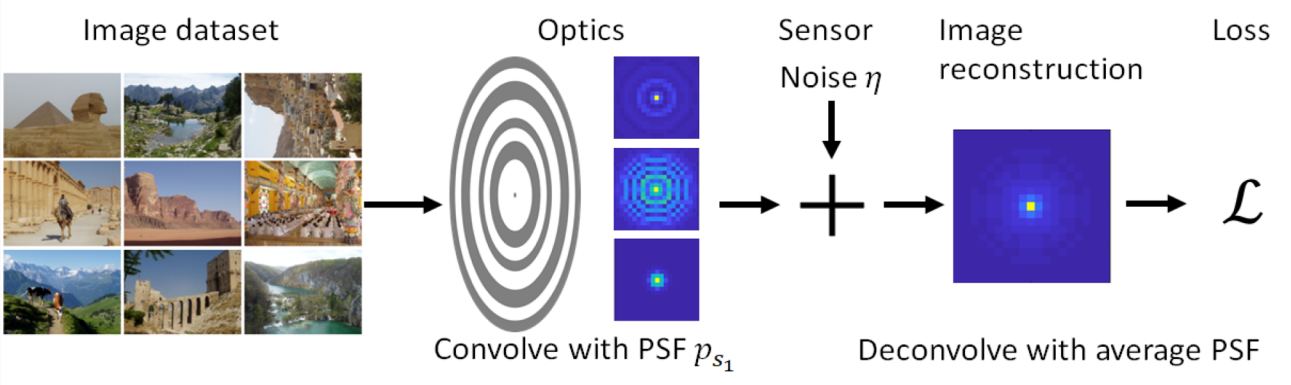

Figure 1. End-to-End optimization pipeline. Each image is placed at certain depth in front of the lens and convolve with the depth invariant point-spread function. The point spread function is obtained from diffraction optics. Adding a Gaussian noise, we get an image on the sensor. In the image reconstruction part, with the depth unknown, we deconvolve with the average point spread function with a Wiener filter. The combined optimization of the optics parameter s_{n1} and deconvolution factor σ. The final loss function is chosen as the distance of reconstructed image from the original one (Eq. 8).

aperture, and let $\sigma_0 = 0.001$ for the full aperture. In the image reconstruction, We de-convolve the image on sensor with the averaged point spread function (Eq. 4) by Wiener filter, i.e.

$$\tilde{I} = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}^*[p_{\mathrm{avg}}]}{\mathcal{F}^*[p_{\mathrm{avg}}]\mathcal{F}[p_{\mathrm{avg}}] + \sigma}\mathcal{F}[I]\Big], \tag{7}$$

where I and $\tilde{I}$ are image on the sensor and the reconstructed image respectively. The parameter σ is a deconvolution parameter of the Wiener filter.

We define the loss for each image as

$$\mathcal{L}^{(j)} = \frac{1}{N_s}\sum_{i=1}^{N_s}\|I^{(j)} - \tilde{I}_i^{(j)}\|^2, \tag{8}$$

where $I^{(j)}$ is the jth image in the image dataset, and $\tilde{I}_i^{(j)}$ is its reconstructed image if it is placed at depth $s_1^{(i)}$. The image reconstruction loss on a dataset is the sum of the loss for each image. In the end-to-end optimization, we minimize the total loss with free parameters s_{n1}, which controls the coded aperture, and σ, which influence the deconvolution.

4. Results

4.1. The point spread function

We studied three types of apertures: a full aperture, an empirical design coded aperture, and the end-to-end optimized one, which are shown in Fig. 2. The maximal diameter of all these apertures is fixed to be 5 mm. For the empirically designed aperture, Eqs. (5) and (6) are used and $\delta_{n1} = 0.8$. The end-to-end optimization gives $\delta_{n1} = 0.69$. To ensure that the interference on axis is never zero, we adjust the first opaque ring to transparent in the coded aperture. The depth dependant point spread functions of these three apertures are shown in the three rows of Fig. 2 respectively, at $\delta = 0, 0.3, 0.5, 0.8, 1$. We find that the point spread function is indeed small for the empirical design at $\delta = 0.8$, where the diffraction optics predict the constructive interference. We therefore claim that the coded aperture designed based on diffraction optics does create a second focal plane. We find that both coded aperture demonstrate smaller depth dependence of the point spread function compared to the full aperture. The average point spread function that is used in the deconvolution is shown in the last column of Fig. 2.

4.2. Qualitative result

As shown in Fig. 3, we compare the image processing results based on (1) a full aperture (row 1), (2) an empirical design (row 2) with $\delta_{n1} = 0.8$ and $\sigma = \sigma_0$, (3) an end-to-end optimized image processing result (row 3). The depth of the images from left to right is from the focal plane $s_1 = 1$ m to infinity, with a linearly inverse depth sampling, i.e. $s_0/s_1 = 1, 0.9, \cdots, 0.1, 0$. Comparing to the full aperture, which gives blur images quickly away from the focal plane, both the empirical coded aperture and the optimized aperture demonstrate the extended depth of field. Moreover, the end-to-end optimized system also provides better image close to the focal plane, where the empirically designed systems gives image sharper than the original image.

We show that the end-to-end optimized parameter is relatively better than the full aperture or empirically designed result: Over a range of objective distance, the image contains relatively low blur, which verifies that the end-to-end optimized result is effective in rendering the extended depth of field.

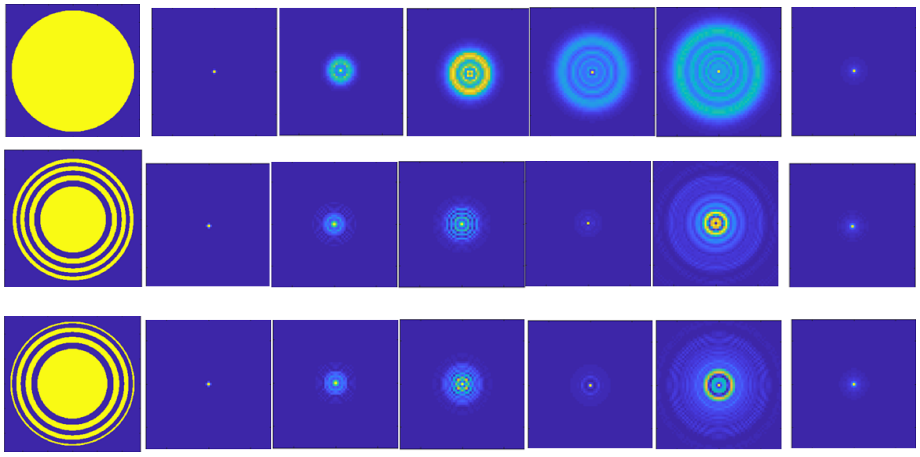

Figure 2. A point spread function (PSF) illustrates the blur introduced for a single point of light at each distance. Aperture profile (column 1) and Point spread function (PSF) (columns 2 - 7) of the full aperture (row 1), empirically designed aperture (row 2), and inverse-designed aperture (row 3). For columns 2 - 6, we evaluate the PSF aperture at distances $\delta = 0,\ 0.3,\ 0.5,\ 0.8,\ 1$ (Here $\delta = (1/s_0 - 1/s_1)s_0$, s_1 is the position of objective.). Column7 represents the averaged PSF over different distances.

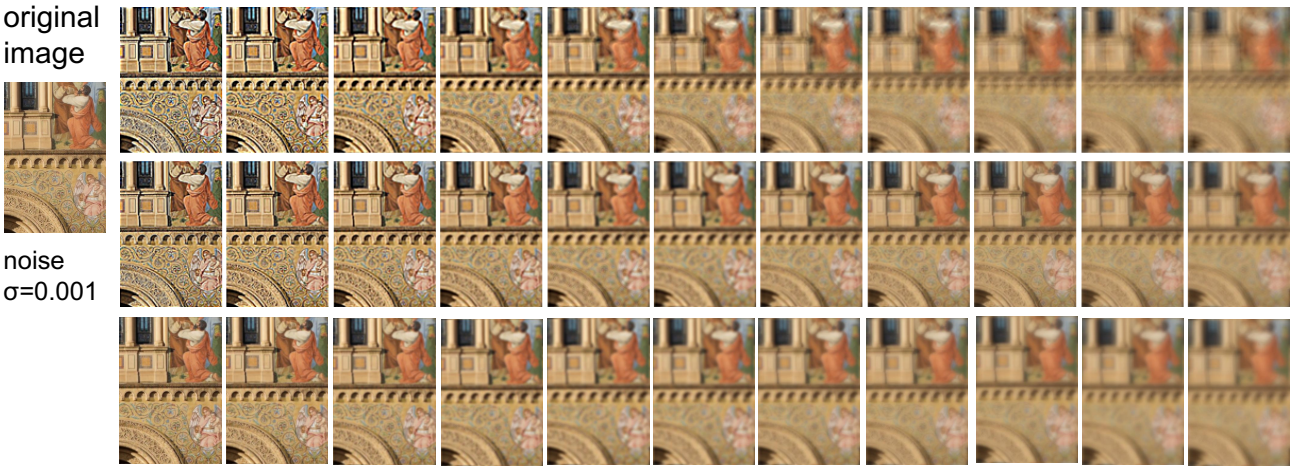

Figure 3. The qualitative comparison of the reconstructed images using (1) a full aperture with deconvolution parameter $\sigma = \sigma_0$ (row 1), (2) an empirical design (row 2) with $\delta_n = 0.8$ and $\sigma = \sigma_0$, or (3) an end-to-end optimized image processing result (row 3). The depth from left to right is from the focal plane $s_1 = 1$ m to ∞, with a linearly inverse depth sampling, i.e. $s_0/s_1 = 1, 0.9, \cdots, 0.1, 0$. We show that the end-to-end optimized parameter is relatively better than the full aperture or empirically designed result.

4.3. Quantitative result

As shown in Fig. 4, we compare the Peak Signal-to-Noise Ratio (PSNR) comparison for the full aperture, diffraction coded aperture and the end-to-end optimized aperture. The solid and dashed curves represent the PSNR of image on the sensor and after reconstruction, respectively. We find that the full aperture gives a highest quality image on the focal plane, but the image quality decreases rapidly away from the focal plane. The empirically designed coded aperture gives highest image quality when the distance from the focal plane is large. However, due to the relatively small σ chosen in the deconvolution, the reconstructed image quality is poor near the focal plane. Nevertheless, the diffractive aperture and the deconvolution parameter after end-to-end optimization render the optimal PSNR over a wide range of objective distance, denoted by

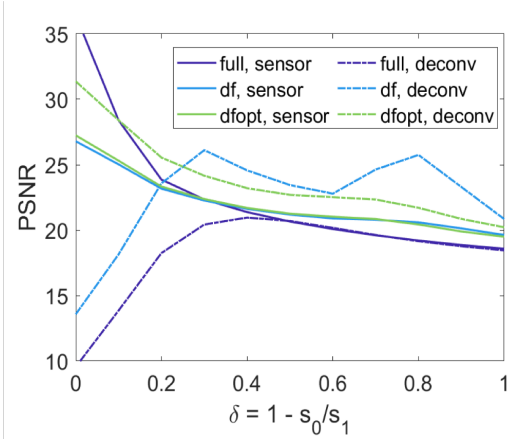

Figure 4. Peak Signal-to-Noise Ratio (PSNR) comparison for the full aperture, diffraction coded aperture and the end-to-end optimized aperture. The green dashed line denoting the diffractive aperture after end-to-end optimization renders the optimal PSNR over a wide range of objective distance, compared with other two designed apertures. This verifies the end-to-end optimized system can indeed improve the depth of field.

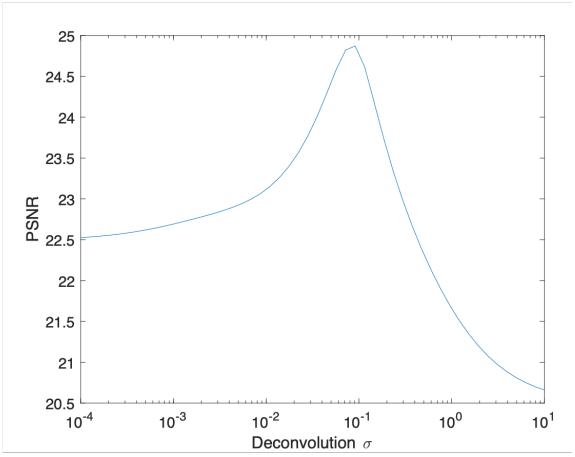

Figure 5. Parameter sensitivity to deconvolution parameter σ. Here we fix the aperture as the first diffractive focus located at $s_{\mathrm{n1}} = 3.3s_0$. The optimal parameter for $\sigma = 0.1$ from the end-to-end pipeline. Here we verify the parameter sensitivity to σ is relatively robust: For σ =10^{-4} to about 0.05, the PSNR is maintained about 10% of the optimal PSNR.

the green dashed line, compared with other two designed apertures. This verifies the end-to-end optimized system can indeed improve the depth of field.

5. Analysis and Evaluation

We produce the depth invariant point-spread function based on the coded aperture by taking into account of the diffraction. The transparent part of the coded aperture enables constructive interference that creates the second focus.

In the rest of this section, we evaluate the PSNR performance of different design and the sensitivity to different parameters of deconvolution σ. As shown in Fig. 5, we fix the aperture as the first diffractive focus located at $s_{\mathrm{n1}} = 3.3s_0$. The optimal parameter for $\sigma = 0.1$ from the end-to-end pipeline. Here we verify the parameter sensitivity to σ is relatively robust: For σ =10^{-4} to about 0.05, the PSNR is maintained about 10% of the optimal PSNR.

6. Discussion

We compared the end-to-end optimization process with other image processing techniques: diffraction aperture, single-annulus aperture, double-annulus aperture, full aperture. There are several points to discuss

- The number of parameter we choose to optimize is two, which is small and they are dominating factors to aperture optical property and deconvolution effect result. This enables fast search of optimal parameter when combining the two processes in image processing.
- By comparing Fig. 3 and Fig. 4, though the end-to-end optimized optical system renders higher PSNR after Wiener deconvolution, the visual resolution is not apparently better than the empirically chosen parameters over several out-of-focus object positions. The metric for optimization should be more carefully chosen in order to reflect the image processing result quality.
- There is degeneracy in optimization result. Namely different optimization combinations can give almost the same PSNR performance. However, we should give preference to image post-processing parameters as the computational difficulty and the cost are both relatively lower than manufacture of diffractive lens. In the sensitivity analysis we should be more careful for the paramter of optical part, since there is higher risk of fabrication error in realistic lens manufacturing.

6.1. Limitations

There are several limitations of the work:

- 1. As we implement our gradient descent optimization algorithm in MATLAB fmincon function, the optimal parameters we obtained may be trapped by the local optimum. Therefore, we may have further space of optimization to improve the performance of the system.
- 2. The image for training this framework is too small. By increasing the processing speed by using GPU or

parallel computing methods, we expect to process the data in a faster way, therefore, we have a more comprehensice training set. The optimization result can then be generalized to other types of image processing.

- 3. The degrees of freedom for optimization is small. Here in optics we use the first-order diffraction focus position as parameter one, and another parameter to be optimized as the deconvolution damping factor. Though achieving relatively remarkable performance in extended depth of field optimization, we can expect to push the performance to a higher livel by exploring the pixel-by-pixel optimization of the two-dimensional aperture profile design, as well as by implementing ADMM or other optimization image reconstruction algorithms to have larger scope of parameter optimization.

6.2. Future work

Corresponding to the limitations discussed in the last part, we have several future plans on this work:

- 1. Since the optimization itself has many parameters to choose, and the standard gradient descent optimizaiton may be often trapped in local optima, we should employ the method by NLopt (Steven G. Johnson, The NLopt nonlinear-optimization package, http://github.com/stevengj/nlopt). This is a library with a combination of mixed optimization solvers, choosing the optimization method depending the context of usage. We expect to have further space of optimization to improve the performance of the system.

- 2. We should use GPU or parallel computing methods to increase the speed of optimizing the structure in a pipeline. We expect to process the data in a faster way. Therefore, we have a more comprehensice training set.

- 3. We should expand the scope fo the degrees of freedom for optimization by exploring the pixel-by-pixel optimization, or more complex periodic unit of the two-dimensional aperture profile design. We also plan to implement ADMM or other optimization image reconstruction algorithms, and convolutional neural network to have larger scope of parameter optimization.

7. Conclusion

To conclude, in this paper, we have shown that the coded aperture is capable of extend the depth of field. The coded aperture designed based on diffraction optics successfully creates the second focal plane. With end-to-end optimization, we obtain the optimized aperture together with the parameter σ in the Wiener filter. This systematic design of a coded aperture and reconstruction algorithm enables a lensless imaging system.

References

[1] J. Chang and G. Wetzstein. Deep optics for monocular depth estimation and 3d object detection. In *Proc. IEEE ICCV*, 2019.

[2] O. Cossairt, C. Zhou, and S. Nayar. Diffusion coded photography for extended depth of field. In *ACM SIGGRAPH 2010 Papers*, SIGGRAPH '10, New York, NY, USA, 2010. Association for Computing Machinery.

[3] E. R. Dowski and W. T. Cathey. Extended depth of field through wave-front coding. *Appl. Opt.*, 34(11):1859–1866, Apr 1995.

[4] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein. Accommodation-invariant Computational Near-eye Displays. *ACM Trans. Graph. (SIGGRAPH)*, (4), 2017.

[5] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4d frequency analysis of computational cameras for depth of field extension. *ACM Trans. Graph.*, 28(3), July 2009.

[6] H. Nagahara, S. Kuthirummal, C. Zhou, and S. K. Nayar. Flexible depth of field photography. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008*, pages 60–73, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[7] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph. (SIGGRAPH)*, 2018.