

End-to-end optimization of a diffractive optical element and aberration correction for integral imaging

Xiangyu Pei (裴翔宇), Xunbo Yu (于迅博)*, Xin Gao (高鑫), Xinhui Xie (谢馨慧), Yuedi Wang (王越笛), Xinzhu Sang (桑新柱), and Binbin Yan (颜玢玢)

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China

*Corresponding author: yuxunbo@126.com

Received April 18, 2022 | Accepted June 27, 2022 | Posted Online August 5, 2022

In the integral imaging light field display, the introduction of a diffractive optical element (DOE) can solve the problem of limited depth of field of the traditional lens. However, the strong aberration of the DOE significantly reduces the final display quality. Thus, herein, an end-to-end joint optimization method for optimizing DOE and aberration correction is proposed. The DOE model is established using thickness as the variable, and a deep learning network is built to preprocess the composite image loaded on the display panel. The simulation results show that the peak signal to noise ratio value of the optimized image increases by 8 dB, which confirms that the end-to-end joint optimization method can effectively reduce the aberration problem.

Keywords: integral imaging; diffractive optical element; end-to-end optimization; deep learning network; aberration correction.

DOI: [10.3788/COL202220.121101](https://doi.org/10.3788/COL202220.121101)

1. Introduction

Three-dimensional (3D) display technology is capable of providing real and natural 3D perception, and, consequently, various efforts have been made to advance this technology further^[1–5]. Light field display, which is considered as the most promising 3D display technology, can reconstruct the real light field distribution of a real 3D scene^[6–9]. Integral imaging (II), a 3D display method that can realize full parallax and true color, and provide depth information, was first proposed, to the best of our knowledge, by Lippmann in 1908^[10]. The II technology is a kind of 3D light field display technology with broad commercial prospects. The currently used II technology consists of two processes: acquisition and reproduction. In the acquisition stage, a lens array (LA) and an electronic sensor capture the information from different angles of the 3D scene and form an element image array (EIA). In the optical reproduction process, the LA with the same parameters is used to project the EIA onto a holographic function screen (HFS) to reproduce the original 3D scene.

Despite numerous advantages, the widespread application of the II method is limited because of several factors such as low viewing resolution, narrow viewing angle, and shallow depth of field (DOF)^[11–16]. To improve the quality of 3D scene reconstruction, various studies on improving the DOF^[12] as well as the lateral resolution^[17–20] have been conducted to date. Accordingly, the previous studies can be divided into two parts: (i) design optimization of the LA to improve its optical

performance and (ii) parallax images optimization algorithm to enable the synthetic image (SI) to reconstruct the original 3D scene with a higher accuracy. However, traditional lenses exhibit small DOF owing to their inherent optical characteristics, which significantly limit the content of the 3D scene that can be rendered. By introducing diffractive optical elements (DOEs)^[21–25], which exhibit the advantages of compact shapes, large and flexible design spaces, and relatively good off-axis imaging performance, the problem of limited DOF can be solved, and the complexity of the optics can be simplified. Although DOEs have numerous advantages, they exhibit strong aberration, resulting in a final display of very poor quality. The resounding success of deep learning in image processing, along with the advent of detailed computational imaging studies, in recent years has allowed a complete transfer of the aberration correction from the optics stage to the post-processing algorithms^[26–31].

In this paper, an end-to-end joint optimization method for optical components and image input sources is proposed. In this method, first, a DOE model is established in the form of thickness variables and combined with the wave optics imaging model to derive the light intensity distribution of the point light source passing through the element at the imaging plane. Second, the deep learning network is employed to pre-correct the input image, and the derived light intensity distribution is subsequently used to simulate the display image of the pre-corrected image on the imaging surface, through the DOE.

Finally, the element images (EIs) in the SI are used as the training dataset, and the stochastic gradient method is used to jointly optimize the aforementioned two parts. In summary, when designing the surface parameters of the DOE, the deep learning network is used to pre-correct the EIA to overcome the aberration problem. Because the light intensity distribution formed at different positions on the input plane is different, the processing method of dividing multiple fields of view—the images in different fields of view are convoluted with the corresponding point spread function (PSF) to simulate the display image of the pre-corrected image at the imaging plane after passing through the DOE—was adopted in this study. The experimental results showed that by using this method the surface parameters of the DOE can be optimized, which effectively mitigates the aberration problem and aids in achieving a highly accurate display result of the original image.

2. End-to-End Optimization Method

The proposed end-to-end optimization of the DOE and aberration correction consists of three steps: (i) input image pre-correction stage, (ii) pre-corrected image subfield convolution stage, and (iii) loss calculation stage, as shown in Fig. 1(b). Figure 1(a) shows the information acquisition stage in the II display method. The camera array collects information from different angles of the 3D scene and encodes it into an SI, which is composed of an EIA. The EIA is used as the training dataset, and each EI is used as the input source to enter the closed-loop process of the method shown in Fig. 1(b). Then, the trained DOE is repeatedly arranged to form a diffractive LA, and the trained pre-correction EI (PCEI) is loaded on the display panel, as shown in Fig. 1(c). In the input image pre-correction stage, the input EI (IEI) passed through the deep learning pre-correction network to obtain the PCEI. Next, the PCEI enters as the input to the pre-corrected image subfield convolution stage, where

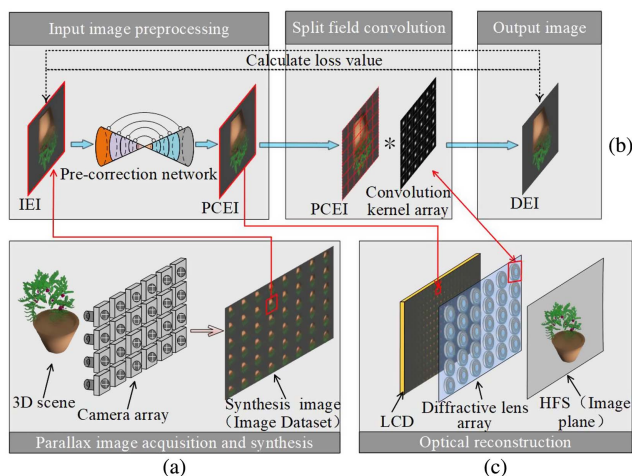


Fig. 1. Overall workflow diagram of this experiment: (a) parallax image acquisition and synthesis; (b) proposed end-to-end optimization method workflow; and (c) optical reconstruction.

the input image is divided into multiple regions, and each region image is convoluted with its corresponding PSF convolution kernel to simulate the final display EI (DEI). Finally, by calculating the loss value between the DEI and the IEI, the surface parameters of the DOE and PCEI are obtained via back propagation.

2.1. Image display model

The ideal optical system differs significantly from the actual optical system. After passing through the actual optical system, the light emitted from an object point in the object space does not converge to a point in the image space; instead, it forms a diffused spot, as shown in Fig. 2. The size of the diffused spot is governed by the aberration of the system. In wave optics theory, ideally, the spherical wave emitted by an object point remains a spherical wave after passing through the optical system. Because of diffraction, the ideal image of an object point is a complicated Airy disk. However, in the actual optical system, because of aberration, the wave surface formed by the optical system does not remain a spherical surface, but exhibits a certain wave phase difference. As a result of aberration, the original input image (OII), after passing through the optical system, becomes distorted, resulting in a different display image. Mathematically, the point light source can be represented by the δ function (point pulse). For an optical system, the light field intensity distribution of the point imaged by the point light source is called the PSF. The image formed by the optical system is a result of the convolution of the object image and the PSF of each point. To simulate the display results, we derive a wave-based image display model, which incorporates the diffraction-related effects on the display. The derived image display model is based on Fourier optics, and we have effectively integrated it into the workflow of the deep learning tools.

To derive the PSF of an optical element, we model the spherical wave emitted by the point light source on a liquid crystal display (LCD) plane and then calculate the wave field expression when it passes through the DOE and subsequently reaches the HFS.

As shown in Fig. 3, the plane where the LCD located is described by a 2D coordinate system XO_1Y , on which a point light source, with an initial amplitude and initial phase of A_0 and φ_0 , respectively, reaches the left surface of the DOE

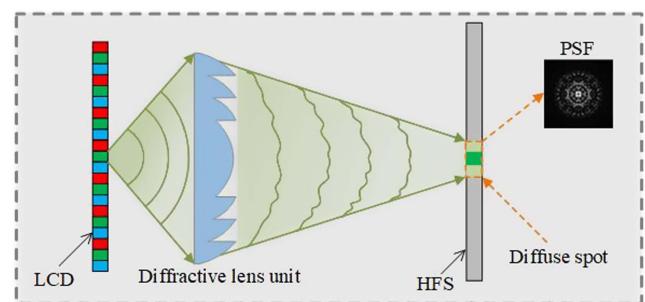


Fig. 2. Imaging process of an actual optical system.

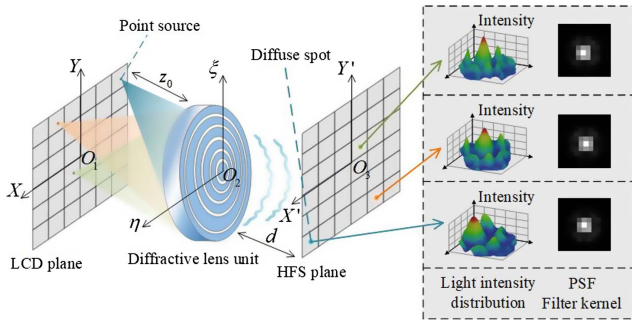


Fig. 3. Schematic diagram of the light intensity distribution at different field points on the image plane.

(the plane described by the 2D coordinate system $\eta O_2 \xi$). Based on wave optics theory, the wave expression of the point light source on the left surface of the DOE is given by (assuming $\varphi_0 = 0$)

$$E(\eta, \xi) = A(\eta, \xi) e^{i\varphi(\eta, \xi)} = \frac{A_0}{[(\eta - x)^2 + (\xi - y)^2 + z_0^2]^{1/2}} \times \exp\{ik[(\eta - x)^2 + (\xi - y)^2 + z_0^2]^{1/2}\}, \quad (1)$$

where $k = 2\pi/\lambda$ (λ is the wavelength of the light wave) is the wave number.

For a single refractive or DOE such as a thin lens, the delay of the wave field phase is proportional to the thickness t at the corresponding position, and the amplitude of the wave field shows negligible changes, i.e.,

$$\Delta\varphi = \frac{2\pi\Delta n}{\lambda} t(\eta, \xi), \quad (2)$$

where λ is the wavelength, and Δn is the refractive index difference between air and the optical element material.

The wave field $E(\eta, \xi)$, with an amplitude of $A(\eta, \xi)$ and a phase of $\varphi(\eta, \xi)$, is incident on the optical element, and, after passing through the element, it changes to

$$E'(\eta, \xi) = A'(\eta, \xi) e^{i\varphi'(\eta, \xi)} = A(\eta, \xi) \exp[i(\varphi(\eta, \xi) + \Delta\varphi)]. \quad (3)$$

As shown in Fig. 3, the wave field $E'(\eta, \xi)$ reaches the HFS (the plane is described by the 2D coordinate system $X'O_3Y'$) after propagating a distance d in free space. The wave field expression at this plane can be expressed as

$$D(x', y') = \frac{e^{ikd}}{i\lambda d} \iint E'(\eta, \xi) \exp\left\{\frac{ik}{2d}[(x' - \eta)^2 + (y' - \xi)^2]\right\} d\eta d\xi. \quad (4)$$

This formula uses the Fresnel propagation operator; when $\lambda \ll d$, Eq. (4) yields an accurate model for near and far distances. The PSF $[p(x', y')]$ is given by the square of the magnitude of the complex-valued wave field, i.e., $|D(x', y')|^2$, as

$$p(x', y') \propto |D(x', y')|^2. \quad (5)$$

Equations (1) to (5) show that the PSF value is related to wavelength. Considering that the pixels on the display panel are composed of subpixels of red, green, and blue colors, we input the PSF corresponding to these three colors into the end-to-end network. In network optimization learning, on the one hand, the surface shape of the DOE can be optimized; on the other hand, the input image can be pre-corrected to adapt to the surface shape of the DOE so that both chromatic aberration and geometric aberration can be taken into account.

In addition, from the derivation process of the formula, it can be seen that the light intensity distribution formed by the point sources at different positions on the object surface on the HFS is different. As shown in Fig. 3, the PSF convolution kernels corresponding to the three point sources at different positions are different. For obtaining a more realistic simulation display result, we evenly divide the input image into $W \times H$ regions, and each subregion is denoted by S_{wh} . The light intensity distribution formed by each point in the subregion, after passing through the optical element, can be approximated by its central point, whose PSF is denoted by P_{wh} . The image formed by the subregion S_{wh} , through the optical system, is denoted by D_{wh} , which can be described by the convolution of S_{wh} and P_{wh} as

$$D_{wh} = S_{wh} * P_{wh}. \quad (6)$$

Using this formula, the display result, CI, of the input image, after passing through the DOE, can be obtained as

$$CI = \sum_{w=1}^W \sum_{h=1}^H D_{wh} = \sum_{w=1}^W \sum_{h=1}^H S_{wh} * P_{wh}. \quad (7)$$

2.2. Pre-correction network

In this study, a network structure was built for the pre-correcting input image and subsequently used to preprocess the original image loaded on the display panel. Further, this developed network structure was used to calculate the parameters of the deep learning network and the surface parameters of the DOE jointly. The proposed network structure consists of two parts: the recognition network and the generation network, as shown in Fig. 4. The recognition network reduces the image size and

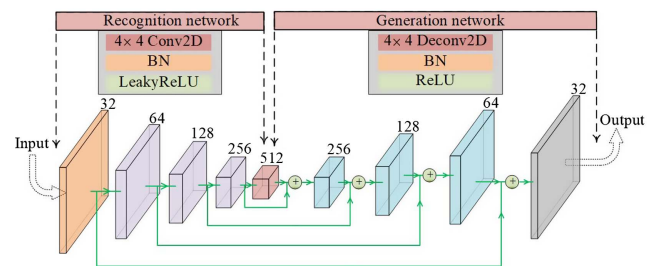


Fig. 4. Pre-correction network structure.

extracts some simple features through convolution and down sampling, while the generation network obtains some deep features through convolution and up sampling. In addition, to make full use of the feature information and retain the richer scene details, we superimpose the feature information on the recognition network and generated network path.

Specifically, each layer of the recognition network is composed of a convolutional layer for down sampling (stride: 2; size: 4×4), a batch normalization (BN) layer for preprocessing the data, and a leaky rectified linear unit (LeakyReLU) activation function. In each down sampling convolution, we reduce the size of the feature map by half and double the number of feature channels. Due to the data diversity of the 3D scene, the BN layer can ensure that the data of each layer is passed down within an effective range, and the LeakyReLU activation function allows the loss signal to be propagated back to the previous layer. Each layer of the generation network consists of a transposed convolutional layer (step size: 2; size: 4×4) for up sampling, a BN layer for preprocessing the data, and a ReLU activation function. In each up sampling convolution, we double the size of the feature map and reduce the number of feature channels by half.

2.3. Overall loss function

Initially, the OII is obtained through the pre-corrected image network proposed and described in section 2.2 to extract the pre-corrected image. Subsequently, the final display image can be calculated using the image display model derived in Section 2.1. The traditional neural network usually uses the mean square error (MSE)^[32] function as the loss function during the experiment. Typically, in practical applications, the image to be displayed should be consistent with the OII to the maximum possible extent. In this study, the experimental results showed that solely using the MSE function does not yield accurate display results, and jointly training the surface parameters of the pre-correction neural network and DOE is not possible. Consequently, we introduced the image quality loss function—structural similarity (SSIM)^[33], which measures the similarity between the reconstructed image and the original image. Additionally, we combined the MSE and SSIM functions, with a certain weight, as the loss function. The loss value between the OII and the calculated display image (CI) is calculated using Eq. (8) as

$$\text{Loss}(\text{OII}, \text{CI}) = \alpha \times \text{MSE}(\text{OII}, \text{CI}) + \beta \times \text{SSIM}(\text{OII}, \text{CI}), \quad (8)$$

where $\alpha + \beta = 1$ and $0 < \alpha < 1$, $0 < \beta < 1$. In this experiment, the lowest loss values were obtained for $\alpha = 0.6$ and $\beta = 0.4$.

3. Experiment and Simulation Results

The aperture diameter, focal length, and F -number of the optical element designed in the experiment are 5 mm, 19.23 mm, and 3.85, respectively. The optical element is discretized with a $12.63 \mu\text{m}$ feature size on a 396×396 grid. Rotational symmetry

is considered in the design of diffractive elements. Although our method can be generalized to rotational asymmetric shapes, rotational symmetry is helpful for manufacturing using turning machines. In order to more accurately construct 3D display spatial voxels and improve light energy utilization, positive first-order diffraction light was selected.

The experiment was carried out according to the workflow shown in Fig. 1, and the experimental parameters are shown in Table 1. First, a virtual camera array is used to capture information from different directions of the 3D scene to develop the 3D model. The LA in the 3D reproduction stage consists of 64×36 lens units. According to the designed light field display system parameters, the SI of the 3D model can be obtained, i.e., the SI contains 64×36 EIs. These EIs become the training dataset in this experiment. We choose the Momentum optimizer with momentum = 0.5, which exhibits robustness to our image dataset. Considering the parameters updating speed and preventing the parameters from hovering near the optimal value, we choose the learning rate of polynomial decay. When the starting learning rate and end learning rate were 10^{-7} and 10^{-9} , respectively, the training effect of network on our image dataset was better.

In the image subfield convolution stage, we divide the pre-corrected EI into 6×6 subregions (as shown in Fig. 5), and the corresponding PSF convolution core array also contains 6×6 PSF convolution cores. The sampling point of each PSF

Table 1. Optical Element and Image Dataset Parameters.

Parameters	Values
Aperture diameter	5 mm
Light source distance (z_0)	20 mm
Propagation distance (d)	500 mm
Focal length	19.23 mm
F number	3.85
Number of training dataset images	64×36 (2304)
Number of image subregions	6×6 (36)

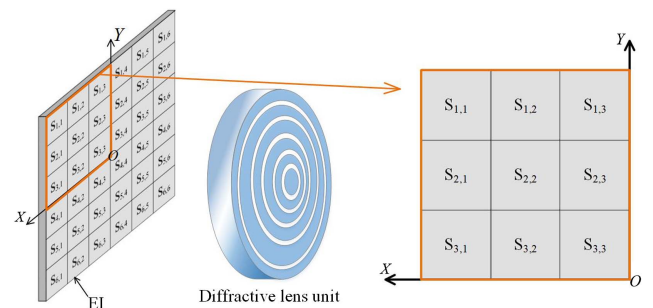


Fig. 5. All 36 subsections of an EI and its corresponding diffractive lens unit.

convolution core is consistent with the resolution of the subregion image. The experimental DOE and loss curve obtained during the training using the proposed end-to-end framework are shown in Figs. 6(a) and 6(b), respectively. The experimental DOE and loss curve obtained during the training without using the pre-correction network are shown in Figs. 7(a) and 7(b), respectively. Due to the rotational symmetry of the lens unit, Figs. 6(c) and 7(c) show the logarithm of nine of the 6×6 PSF convolution kernels (corresponding to the nine subregions marked in Fig. 5).

To verify the effectiveness of the proposed end-to-end joint optimization framework, we conducted optical simulation experiments. As shown in Fig. 8, the simulation experiments were carried out from five different observation positions: central position, above, below, left, and right. Figures 8(a)–8(c) display the image of the original scene, simulation results without using the pre-correction network, and simulation results with end-to-end optimization, respectively. As evident from the enlarged images shown in Figs. 8(b) and 8(c), the experimental results optimized by the end-to-end method as well as the display quality are significantly improved. In addition, the peak signal to noise ratio (PSNR) value is significantly improved, which proves the effectiveness of the framework.

To demonstrate the effectiveness of dividing the field of view to employ multiple PSFs, we supplemented a comparative simulation experiment using a single PSF, and the simulation results are shown in Fig. 9. When using a single PSF approximation, it can be seen from Fig. 9 that only the simulation results from the center viewpoint are better, and this difference will gradually

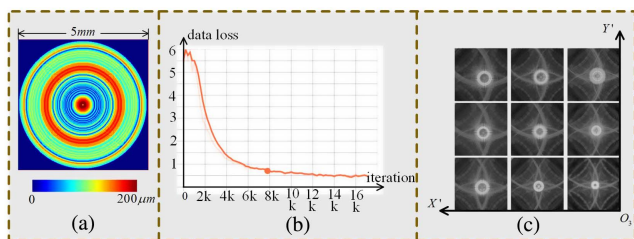


Fig. 6. Experimental results with end-to-end optimization: (a) diffractive element profile; (b) loss curve during training; and (c) PSF of nine fields of view, in logarithmic scale.

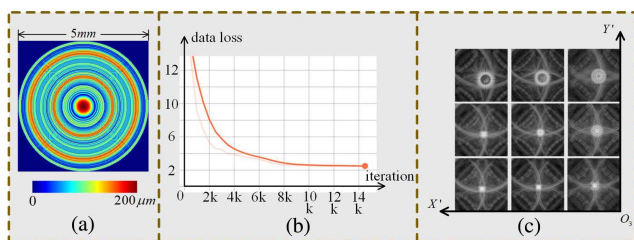


Fig. 7. Experimental results without pre-correction network: (a) diffractive element profile; (b) loss curve during training; and (c) PSF of nine fields of view, in logarithmic scale.

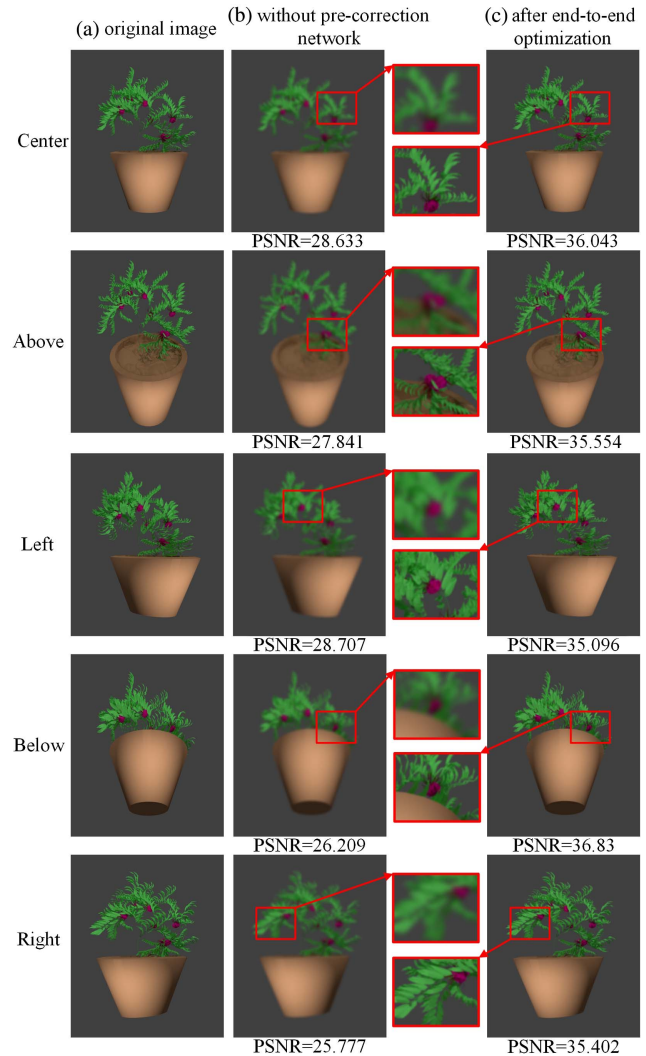


Fig. 8. Simulation results for different viewing positions (with multiple PSFs): (a) original scene; (b) simulation results without pre-correction network; and (c) simulation results with end-to-end optimization.

accumulate as the viewing angle increases. Furthermore, the experimental results shown in Fig. 9 are better than those shown in Fig. 8 with or without the use of the pre-correction network, which shows that dividing the field of view to employ multiple PSFs is the correct choice.

The optical aberrations of the proposed design have the following properties. The chromatic variation is small because the deep diffraction lens surface [as shown in Fig. 6(a)] obtained by using end-to-end frame optimization results in only small focal length differences in the visible wavelength region. In addition, the diffraction lenses also interact with each other, which is taken into account in our design. Similar to the traditional optical LA, the 3D optical field display system based on the diffraction LA will form multiple view areas in space. We form the region of interaction between diffraction lenses outside the visible region to ensure the purity of spatial voxels in the visible region.

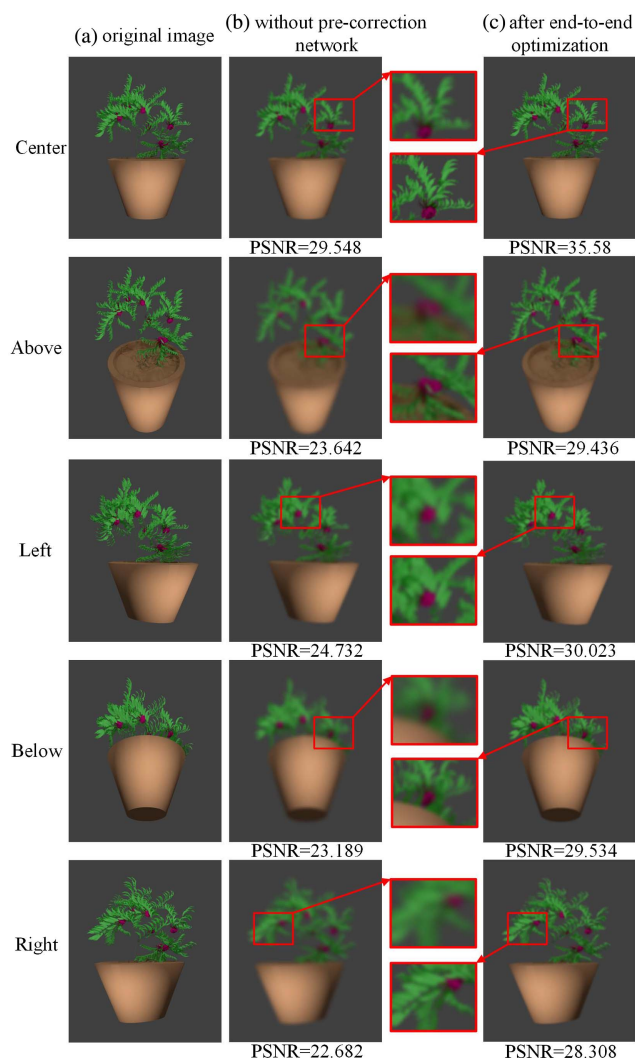


Fig. 9. Simulation results for different viewing positions (with a single PSF): (a) original scene; (b) simulation results without pre-correction network; and (c) simulation results with end-to-end optimization.

One possible solution similar to diffraction lenses is metalenses, which have recently enabled the fabrication of single ultrathin optical elements that encode wavelength-dependent phase patterns onto the incoming light. There is a recent paper^[34] that proposed neural nano-optics, a high-quality, polarization-insensitive nano-optic imager. They jointly optimized the metasurface and deconvolution algorithm with an end-to-end differentiable image formation model. We have similarities in the idea of using end-to-end framework optimization design, but there are great differences in each step. The first is the difference of physical process: the paper mentioned above is about the imaging process, and the ultimate goal is how to capture the picture of the real world and record it clearly. Our work is used for 3D display, and the ultimate goal is to clearly show the content loaded on the display panel to the viewer. Secondly, the modeling process of optical elements is different: they optimize the metasurface in phase function basis as opposed to in a pixel-by-pixel manner to avoid local minima.

We derive a wave-based image formation model, which takes into account the diffraction and wavelength-dependent effects when imaging natural scenes and expresses the phase delay by the variable of height. Finally, there are the differences in the network and its usefulness: the network they built is used to process the images loaded on the sensor and combined with the deconvolution operation to make the sensor image clear. The network we built is not only used to pre-correct the input image loaded on the display panel, but also to extract the characteristics of the input image and jointly optimize the design of optical elements so that the preprocessed image can better match the optical characteristics of the designed optical elements. In actual production, the metalens proposed in the above-mentioned paper has the advantages of ultrathin thickness at the wavelength level, which is comparable to the performance and polarization operation of traditional lenses, but its processing is expensive and time-consuming, difficult to make large aperture, and unsuitable for mass production. However, the diffractive lens we proposed can be mass produced by gray-scale lithography, which is commonly used in the microelectronics industry, and can be processed with centimeter large aperture. In addition, combined with the proposed end-to-end optimization framework, the diffractive lens we designed has good optical performance in 3D display. But, the work of this paper mentioned above is advanced and has potential revolutionary value in redefining the optical industry. At the same time, in the design of optical elements for large viewing angle display, this paper gives us great inspiration for our future work.

4. Conclusion

In this paper, we propose and introduce an end-to-end joint optimization method for DOE and for preprocessing input images. The method mainly includes two steps. In the first step, we establish a DOE model, derive the light intensity distribution of point light sources, from different fields of view, at the imaging surface based on wave optics theory, and obtain the corresponding PSF convolution kernel array. In the second step, we build a pre-corrected image network to preprocess the input image and load the preprocessed image onto the display panel. Finally, combining the MSE and SSIM loss functions and employing the EIs that form the SIs as the training dataset, we jointly calculate the surface parameters of the DOE, the parameters of the pre-corrected image network, and the preprocessed image. The surface parameters of the DOE as well as the PSF of the sampling points, extracted from different fields of view, are obtained through this method. The simulation results show that the end-to-end joint optimization method effectively reduces the aberration problem and significantly improves the quality of the display image.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 62175015, 61905019, and

62075016) and the Fundamental Research Funds for the Central Universities (No. 2021RC13).

References

- G. J. Lv, Q. H. Wang, W. X. Zhao, and J. Wang, "3D display based on parallax barrier with multiview zones," *Appl. Opt.* **53**, 1339 (2014).
- N. Okaichi, M. Miura, J. Arai, M. Kawakita, and T. Mishina, "Integral 3D display using multiple LCD panels and multi-image combining optical system," *Opt. Express* **25**, 2805 (2017).
- S. Xing, X. Sang, X. Yu, C. Duo, B. Pang, X. Gao, S. Yang, Y. Guan, B. Yan, J. Yuan, and K. Wang, "High-efficient computer-generated integral imaging based on the backward ray-tracing technique and optical reconstruction," *Opt. Express* **25**, 330 (2017).
- J. Y. Son, C. H. Lee, O. O. Chernyshov, B. R. Lee, and S. K. Kim, "A floating type holographic display," *Opt. Express* **21**, 20441 (2013).
- W. Song, Q. Zhu, Y. Liu, and Y. Wang, "Volumetric display based on multiple mini projectors and a rotating screen," *Opt. Eng.* **54**, 013103 (2015).
- X. Sang, F. C. Fan, C. C. Jiang, S. Choi, W. Dou, C. Yu, and D. Xu, "Demonstration of a large-size real-time full-color three-dimensional display," *Opt. Lett.* **34**, 3803 (2009).
- N. Balram and I. Tosic, "Light-field imaging and display systems," *Inf. Disp.* **32**, 2 (2016).
- R. Bregović, P. T. Kovács, and A. Gotchev, "Optimization of light field display-camera configuration based on display properties in spectral domain," *Opt. Express* **24**, 3067 (2016).
- H. Huang and H. Hua, "Systematic characterization and optimization of 3D light field displays," *Opt. Express* **25**, 18508 (2017).
- G. Lippmann, "La photographie integrale," *C. R. Acad. Sci.* **146**, 446 (1908).
- J. H. Park, Y. Kim, J. Kim, S. W. Min, and B. Lee, "Three-dimensional display scheme based on integral imaging with three dimensional information processing," *Opt. Express* **12**, 6020 (2004).
- J.-S. Jang, F. Jin, and B. Javidi, "Three-dimensional integral imaging with large depth of focus by use of real and virtual image fields," *Opt. Lett.* **28**, 1421 (2003).
- J.-S. Jang and B. Javidi, "Depth and lateral size control of three-dimensional images in projection integral imaging," *Opt. Express* **12**, 3778 (2004).
- J.-S. Jang and B. Javidi, "Large depth of focus time-multiplexed three-dimensional integral imaging by use of lenslets with nonuniform focal lens and aperture sizes," *Opt. Lett.* **28**, 1924 (2003).
- R. Martínez-Cuenca, G. Saavedra, M. Martínez-Corral, and B. Javidi, "Extended depth-of-field 3-D display and visualization by combination of amplitude-modulated microlenses and deconvolution tools," *J. Display Technol.* **1**, 321 (2005).
- R. Martínez-Cuenca, G. Saavedra, and M. Martínez-Corral, "Enhanced depth of field integral imaging with sensor resolution constraints," *Opt. Express* **12**, 5237 (2004).
- L. Erdmann and K. J. Gabriel, "High resolution digital integral photography by use of a scanning microlens array," *Appl. Opt.* **40**, 5592 (2001).
- S. Kishk and B. Javidi, "Improved resolution 3D object sensing and recognition using time multiplexed computational integral imaging," *Opt. Express* **11**, 3528 (2003).
- J. S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," *Opt. Lett.* **27**, 1144 (2002).
- J. S. Jang and B. Javidi, "Improved viewing resolution of three-dimensional integral imaging by use of nonstationary micro-optics," *Opt. Lett.* **27**, 324 (2002).
- S. Banerji and B. Sensale-Rodriguez, "A computational design framework for efficient, fabrication error-tolerant, planar THz diffractive optical elements," *Sci. Rep.* **9**, 5801 (2019).
- S. Banerji, M. Meem, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, "Imaging with flat optics: metalenses or diffractive lenses?" *Optica* **6**, 805 (2019).
- M. Meem, S. Banerji, C. Pies, T. Oberbiermann, A. Majumder, B. Sensale-Rodriguez, and R. Menon, "Large-area, high-numerical-aperture multi-level diffractive lens via inverse design," *Optica* **7**, 252 (2020).
- S. Banerji, M. Meem, A. Majumder, B. Sensale-Rodriguez, and R. Menon, "Extreme-depth-of-focus imaging with a flat lens," *Optica* **7**, 214 (2020).
- P. Wang, N. Mohammad, and R. Menon, "Chromatic aberration corrected diffractive lenses for ultra broadband focusing," *Sci. Rep.* **6**, 21545 (2016).
- N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," *Optica* **5**, 1 (2018).
- M. S. Asif, A. Ayremlou, A. Veeraraghavan, R. Baraniuk, and A. Sankaranarayanan, "Flatcam: replacing lenses with masks and computation," in *IEEE International Conference on Computer Vision (ICCV)* (2015), p. 663.
- Y. Peng, Q. Sun, X. Dun, G. Wetzstein, and W. Heidrich, "Learned large field-of-view imaging with thin-plate optics," *ACM Trans. Graph.* **38**, 219 (2019).
- K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express* **27**, 28075 (2019).
- 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.* **37**, 114 (2018).
- A. Nikonorov, V. Evdokimova, M. Petrov, P. Yakimov, S. Bibikov, Y. Yuzifovich, R. Skidanov, and N. Kazanskiy, "Deep learning-based imaging using single-lens and multi-aperture diffractive optical systems," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (2019), p. 3969.
- E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting, and variants," *Mach. Learn.* **36**, 105 (1999).
- Z. Wang, A. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600 (2004).
- E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," *Nat. Commun.* **12**, 6493 (2021).