Extended Depth-of-Field Projector using Learned Diffractive Optics

Yuqi Li* — Ningbo University
Qiang Fu† — KAUST
Wolfgang Heidrich‡ — KAUST

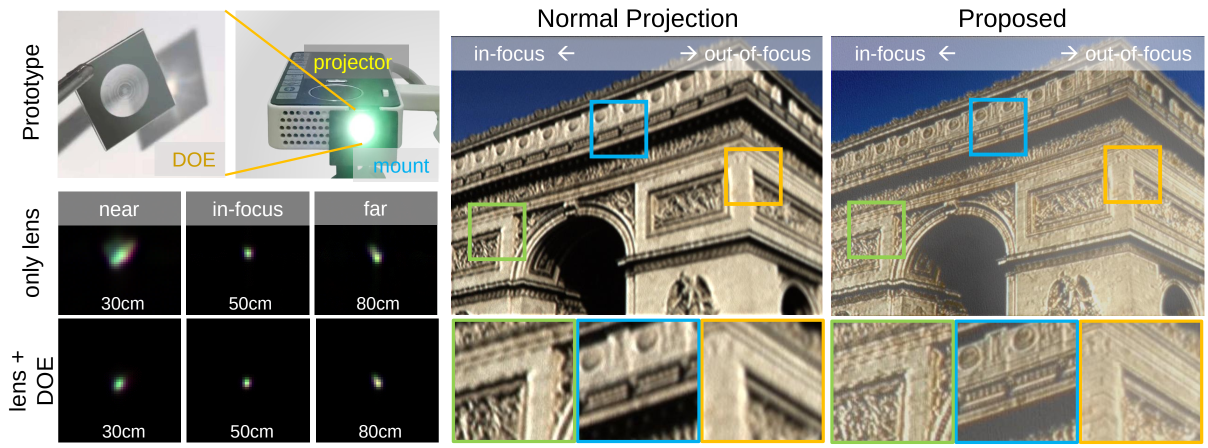

Figure 1: We propose an end-to-end hardware-software joint optimization technique to extend the depth of field (DOF) of projectors. Instead of just algorithmically deconvolving the out-of-focus blur, we learned a custom diffractive optical element (DOE) placed in front of the projector lens. The learned DOE results in a point spread function (PSF) with higher energy concentration over a wide range of projection distances compared to normal projection using a normal lens. With jointly optimized hardware and a deep compensation network, our method can create an all-in-focus image display with sharp details on projection planes at different depths. Here are the PSFs (bottom left) and the results (right) displayed on a tilted projection screen at 50 degrees to the projection direction.

Abstract

Projector Depth-of-Field (DOF) refers to the projection range of projector images in focus. It is a crucial property of projectors in spatial augmented reality (SAR) applications since wide projector DOF can increase the effective projection area on the projection surfaces with large depth variances and thus reduce the number of projectors required. Existing state-of-the-art methods attempt to create all-in-focus displays by adopting either a deep deblurring network or light modulation. Unlike previous work that considers the optimization of the deblurring model and physic modulation separately, in this paper, we propose an end-to-end joint optimization method to learn a diffractive optical element (DOE) placed in front of a projector lens and a compensation network for deblurring. Using the desired image and the captured projection result image, the compensation network can directly output the compensated image for display. We evaluate the proposed method in physical simulation and with a real experimental prototype, showing that the proposed method can extend the projector DOF by a minor modification to the projector and thus superior to the normal projection with a shallow DOF. The compensation method is also compared with the state-of-the-art methods and shows the advance in radiometric compensation in terms of computational efficiency and image quality.

Index Terms: Computing methodologies—Computer graphics—Image manipulation—Image processing

*e-mail: liyuqi1@nbu.edu.cn. Yuqi Li and Qiang Fu contributed equally to this work.
†e-mail: qiang.fu@kaust.edu.sa
‡e-mail: wolfgang.heidrich@kaust.edu.sa

1 Introduction

Projection-based spatial augmented reality (SAR) is a technology that turns 3D objects into display surfaces and changes the appearance of the objects to generate mixed-reality experiences. Benefiting from the low cost of projectors and the convenience of not wearing additional devices, the technology has become popular in a variety of applications, such as virtual exhibition [3], product design [32], serious games [23], and industry manufacturing [46]. With the advancement of geometric calibration and radiometric compensation methods based on projector-camera systems (Procams) [13], precise geometric alignment and color reproduction become possible in 3D projection mapping, greatly enhancing the realism of SAR.

However, projectors are originally invented for display on 2D planar screens, and they usually adopt large apertures for high light throughput. Therefore the degree of focus (DOF) is relatively narrow and not suitable for projection mapping on 3D objects with large depth variances. A straightforward solution is to apply multiple conventional projectors to construct the display. Each projector is responsible for a separate area with a substantially constant depth [4, 34]. However, this will increase the cost of the display as the number of projectors used has increased significantly.

An alternative solution is to extend the DOF of projectors and make a single projector cover as many areas as possible. Many approaches have been proposed to improve the DOF of projectors from the perspectives of optical hardware modulation and projection deconvolution algorithms. Existing optical modulation approaches either spatially modulate the light field or temporally modulate the focal length to achieve wide DOF. However, the light efficiency or the frame rate of the display is sacrificed with the modified projectors. Most of the projector deconvolution approaches compensate for the out-of-focus blur with known point spread function (PSF) kernels. The PSFs have to be accurately measured before compensation as their non-blind compensation algorithm requires. Furthermore, most of the previous approaches treat optical hardware modulation and compensation algorithms as separate tasks, resulting in the loss of some performance optimization freedom.

In this paper, we focus on the joint end-to-end optimization of optical design and compensation algorithms for extending the projector's in-focus display range. We consider optical hardware modulation and compensation algorithm as two interdependent tasks, thus putting forward an end-to-end neural network that can learn a DOE placed in front of a projector lens, and a compensation method for projector deblurring without requiring the PSF measurement before compensation. We fabricate the learned DOE and equip it with a conventional projector as a prototype system. The light efficiency of the prototype is higher than that of previous EDOF projector designs. Both simulations and experiments on the prototype show that our network effectively enlarges the projector DOF, and outperforms state-of-the-art projector compensation methods. Our contributions can be summarized as follows:

- We present the *first* work to jointly design a DOE for light phase modulation and a convolutional neural network for projector compensation. The designed extended depth of field (EDOF) computational projector can achieve high light throughput and real-time performance;

- We demonstrate that this learned optics compares favorably to baselines with conventional projectors, and the learned compensation network outperforms previous state-of-the-art compensation methods in terms of both computational efficiency and compensation quality;

- We implement a laboratory prototype of the computational EDOF projector equipped with the learned DOE, and evaluate it with real display experiments on depth-varying and tilted projection surfaces.

2 Related Works

As mentioned above, many efforts have been made to improve the DOF of a single projector, including non-blind deconvolution algorithms and hardware modulation methods. We also introduce two techniques related to our approach in this section: radiometric compensation and end-to-end optics learning.

2.1 Deconvolution

Previous deconvolution methods treat the depth-dependent PSFs as known degradation models and optimize the input projector images for display on a plane placed at out-of-focus distances [5]. The defocus PSF kernels for each projector pixel can be recovered from the captured images of projected dot patterns [50], and can also be estimated from the captured images of projected natural images [36]. Knowing the PSFs, in order to compensate for the spatially-varying projection blur, a straightforward optimization that minimizes the reprojection error can be solved via iterative gradient descent algorithms [50]. Considering further improving the convergence rate of the deconvolution, Wiener deconvolution in the frequency domain is another practical solution [6]. However, all the deconvolution algorithms attempt to minimize the color numerical difference (e.g., Mean Square Error) while not considering the perceptual difference between the real display and the desired images. We argue that the Mean Absolute Error (MAE) and Mean Square Error (MSE) are not suitable for the objective function in image deblurring task, since they are insensitive to blurring [7], that is, blurring may cause large perceptual differences but only small MSE and MAE changes. Therefore, deep image quality perceptual metrics are more suitable as objective functions for the compensation in projector display.

2.2 Hardware modulation

The optical modulation techniques for projector EDOF can be divided into two categories: coded aperture and focal sweep projector. In the projector design with a coded aperture [11], a high-frequency random blocking mask is placed at the location of the aperture of a camera to generate high-frequency PSFs. Other methods [29] study further to enhance the display quality of a single image display by exploiting a spatial light modulator (SLM) to modulate the light fields in multiple frames. Such projectors with coded apertures can be used in display and surface reconstruction [22]. However, the mask or the SLM blocks at least half of the light energy, leading to relatively low light throughput. As another modulation-based solution, the focal sweep projectors equipped with an electrical focus tunable lens (ETL) to periodically adjust the focal length of the projector and projected content to adapt to the surfaces with pixel-dependent depths [19]. The combination of ETL and high-speed projectors demonstrates its advance in the dynamic display on depth-varying planes [49] and 3D moving objects [47]. Despite the high image quality of the focal sweep projector, such projectors rely on high-speed focal length adjustment, and sacrifice frame rate and a certain amount of light throughput to generate sharp pixel-dependent PSFs.

2.3 Radiometric compensation

Radiometric compensation is a technique to eliminate the indirect light transport effects in projector displays. Indirect light transport effects, such as inter-reflection, refraction, and defocus, heavily degrade the display quality. Previous works consider representing full light transport in projector display by a 4D large-scale tensor. It utilizes low-rank assumption to measure the tensor, and apply the inverse of the tensor to remove the unwanted effects [37, 48]. Some recent works focus on either compensating the defocus blur [20, 21] or textured projection surfaces and geometric distortion [16–18]. Taking full advantage of the power of the convolutional neural networks, these models are trained in a supervised manner, and can deal with complicated scenes with local light transport effects. Most of the methods do not rely on acquiring accurate degradation models, thus no need to project sampling patterns to projection surfaces, resulting in no interruption during the continuous projection procedure. Instead, they project natural images and capture the degraded images by a well-registered camera, then directly feed the desired images and the captured degraded display image into the compensation networks to train the network parameters. This inspires us to construct a deep convolutional neural network for blur compensation without acquiring the out-of-focus degradation model.

2.4 End-to-end optics learning

As the inverse processing of projector display, camera imaging also concerns the properties such as DOF and light throughput. Recently, end-to-end optics learning attracts extensive attention in the computational imaging community. Combining the differentiable optics simulation and various downstream reconstruction neural networks, end-to-end optics learning shows its automaticity and flexibility in optics design, compared with traditional hand-crafted design. Unlike hand-crafted designs that focus on optical aberrations only in the optical system, end-to-end optics learning takes more into account the performance of the whole system on specific tasks, and jointly optimizes the optics and reconstruction algorithms, resulting in higher optimization freedom. The learned optics modulates the optical phase to encode implicit image information into the PSFs, and decode the information using the reconstruction algorithms. Existing end-to-end methods have been widely used in super-resolution and EDOF [43], high-dynamic range imaging [44], hyperspectral imaging [2], large field-of-view imaging [39], seeing through obstructions [42] and so on. Compared with the modulation method using a blocking mask, learned optics usually have much higher light throughput, and can also reconstruct the images in a snapshot.

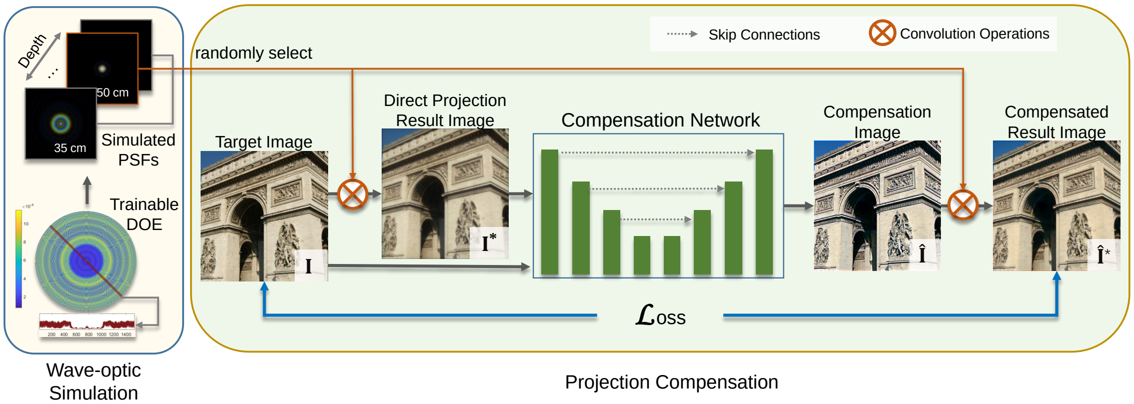

Figure 2: Framework of our end-to-end optimization for both the DOE and the compensation network. We simulate the PSF of a DOE at each depth by substituting its height map into a diffraction propagation model, and randomly choose a PSF as the convolutional kernel to generate the projection result $\mathbf{I}^*$ of the original images $\mathbf{I}$. The compensation network takes the sharp original images $\mathbf{I}$ and the corresponding blur result images $\mathbf{I}^*$ as inputs and outputs the compensation images $\hat{\mathbf{I}}$ for the projector display. Then we simulate the projection result $\hat{\mathbf{I}}^*$ of the compensation images $\hat{\mathbf{I}}$ by convolution with the selected PSF, and treat the difference of $\mathbf{I}$ and $\hat{\mathbf{I}}^*$ as the loss function $\mathcal{L}$ of the network. As our entire framework is differentiable, the DOE height map and the compensation network can be learned by minimizing the loss function $\mathcal{L}$.

To ensure both high frame rate and high light throughput, motivated by the development of end-to-end optics learning in the computational imaging domain, we propose to use learned diffraction optics elements instead of blocking masks to modulate the phase of the light emitted by projectors, and utilize a deep convolutional neural network to compensate for the projected images. The transmitted light by DOEs is much higher than blocking masks, indicating little energy loss through the DOE. Furthermore, convolutional neural networks can be trained with perceptual losses to efficiently handle compensation and achieve high perceptual display quality.

3 Method

The goal of our proposed method is to build a framework that jointly optimizes the projector optics and the compensation algorithms for projection DOF extension as a whole. We present a unified data-driven projector EDOF method to learn a DOE and a reconstruction network in an end-to-end manner. Compared with individual optimizations, our method can achieve better image display quality over a larger projection range. As illustrated in Fig. 2, the framework mainly consists of a wave-optics simulation component and an image compensation component. The wave-optics propagation model is applied to simulate the PSFs and the projection results of input projector images at a specific projection distance, and the image compensation network can generate the compensated image for the projector display at that distance.

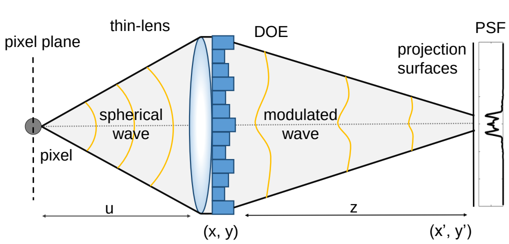

Figure 3: The light propagation from a point on the pixel plane to the projection surface. The phase of the spherical light wave coming from a pixel point is modulated by an ideal thin-lens with focal length f and the optimized DOE. The PSFs along varying depths can be simulated by using the Fresnel diffraction model.

3.1 PSF Simulation

The proposed projector optical system consists of a light source, and a DMD chip followed by a refractive lens and an optimized DOE. In order to simulate the PSFs of the system, we apply the wave optics model [10] to approximate the physical propagation of light and calculate the intensity from a point light source to a plane with a specific projection distance. Specially, we treat a pixel on the pixel plane as a point light source and consider the light wave emitted from the pixel as a spherical wave. As shown in Fig.3, let the position of the pixel be the origin of coordinates, and the distance of the lens principal plane from the pixel plane be u. The light from the pixel and arrival at the lens plane satisfies the spherical wave propagation model, and the phase of the point at position (x,y,u) can be expressed as:

$$\phi_s = \frac{2\pi}{\lambda}\sqrt{x^2+y^2+u^2}, \tag{1}$$

where λ is the wavelength of the incident light. The refractive thin-lens then delay the phase of the light as:

$$\phi_f = \frac{-2\pi}{\lambda}\frac{(x^2+y^2)}{2f}, \tag{2}$$

where f is the focal length of the lens. Right after the lens, the DOE then modulates the phase of the light. We neglect the gap between the lens plane and the DOE plane, and simulate the phase delay caused by the DOE based on the height map h and the wavelength-dependent refractive index n_λ of the DOE:

$$\phi_{DOE} = \frac{2\pi(n_\lambda - 1)}{\lambda} h(x,y). \tag{3}$$

Then, the accumulated phase of the outgoing light after being modulated by the DOE can be expressed as:

$$\phi_m = \phi_s + \phi_f + \phi_{DOE}. \tag{4}$$

Combining the consideration of the ideal thin-lens equation that

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{5}$$

where v denotes the focusing distance of the lens under the focal length f and the distance u, the accumulated phase ϕ_m can be expressed as:

$$\phi_m = \frac{2\pi}{\lambda}\left[(n_\lambda - 1)h(x,y) - \frac{x^2+y^2}{2v}\right], \tag{6}$$

in which the assumption $\sqrt{x^2+y^2} \ll v$ is made to simplify the expression, and variables u and f are eliminated. Note that the phase after the modulation of DOE is only related to the wavelength λ, the refractive index n_λ, spatial location (x,y), the DOE height h, and the original focusing distance v. Finally, the wave field at the DOE plane propagates to the projection plane with the projection distance z, and the PSF kernel $\mathbf{p}$ at the projection plane can be calculated by using the squared magnitude of the complex wave field:

$$\mathbf{p} \propto \left|\mathcal{F}^{-1}\{\mathcal{F}\{A\cdot\mathbf{U}_m\}\cdot\mathcal{H}_z\}\right|^2, \tag{7}$$

where $\mathbf{U}_m$ is the complex-valued light wave after modulation, A denotes a two-dimensional binary pupil function whose spatial resolution is the same as that of $\mathbf{U}_m$, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote Fourier transform and inverse Fourier transform respectively, and the convolution kernel of Fresnel propagation can be expressed as:

$$\mathcal{H} = e^{i\frac{2\pi}{\lambda}(z+\frac{x^2+y^2}{2z})}. \tag{8}$$

Note that the elements in binary pupil function A with value 1 indicate the light is fully transmitted at the point, while those with value 0 indicate the light is blocked. In this paper, we use a circular pupil function that only allows the light to pass through the circle region. Obviously, the PSFs are both depth-dependent and wavelength-dependent. In order to eliminate the effects of the spectral variable λ, the spectral power distribution (SPD) function of the projector's primary needs to be known. Fortunately, the SPD function of the primaries can be obtained by referring to the product specifications or measuring with a spectrometer (e.g., X-Rite ColorMunki) easily. Sampling each wavelength in Eq. 7 and calculating the integral of each channel by substituting the obtained SPD functions separately, the depth-dependent three-channel PSFs can be simulated by giving projection distances z. We randomly set the depth z and generate the corresponding PSF for network training. To force the consistency of the input and output energy, the simulated PSFs are scaled so that the sum of each channel equals one.

3.2 Projection Results Synthesis

Given a PSF generated at a random projection distances z, and an input color image $\mathbf{I}$, we simulate the corresponding projection result $\mathbf{I}^*$ based on the convolutional PSF model:

$$\mathbf{I}^* = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{I}\}\cdot\mathcal{F}\{\mathbf{p}\}\}. \tag{9}$$

Note that we use the fast Fourier transform(FFT) in the frequency domain to get the convolutional result instead of using filtering because the kernel size of $\mathbf{p}$ is usually hundreds due to the diffraction propagation simulation. The computational complexity of filtering is $O(N^2)$, resulting in unacceptable computational cost as the kernel size increases. In contrast, the computational complexity of FFT is $O(Nlog(N))$, resulting in a reasonable computational time. In addition, to handle the image boundaries accurately, we pad the image and kernel with zeros for linear convolution. The PSFs can be used to simulate not only the projection results of target images, but also the projection results of the compensated images.

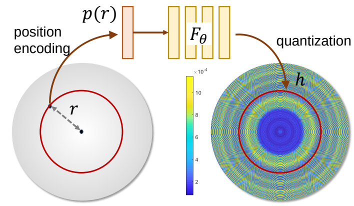

Figure 4: The multiple layer perceptron (MLP) for height representation. In order to constrain the optimization freedom degree of the height map, we force all points that are the same distance from the center of the DOE to have the same height, and the radial distribution height map is represented by an MLP F_θ with trainable parameters θ. Note that we apply position encoding $p()$ to the radius r before sending it to the MLP, and the height map is quantized to meet the manufacturing requirement.

3.3 DOE Height Representation

The height map of the DOE is learned to modulate the incident light wave after being focused by the lens. Our joint optimization framework seeks for optimal combination of the DOE height map and the compensation network.

The most straightforward representation of a $m\times m$ height map is to treat the height of every point on the DOE as learnable parameters. However, this may lead to local minima solutions due to the high degree of optimization freedom [31,44]. Also, in the SAR-oriented projector display problem, rotation-invariant PSFs are much easier for compensation since the normal direction of the projection surface is usually arbitrary in SAR applications. Therefore, we apply a constraint on the height map that all points that are the same distance from the center of the DOE have the same height. This radial distribution constraint drastically reduces the degree of optimization freedom of the height map. Furthermore, as shown in Fig. 4, we adopt the position encoding technique and a multiple layer perceptron (MLP) to implicitly generate the height of the points at radius r. The representation of our height map function is formulated as:

$$h(x,y) = F_\theta\left(p(\sqrt{x^2+y^2})\right), \tag{10}$$

where (x,y) denotes the point position, $p(\cdot)$ denotes the positional encoding operation, $F_\theta(\cdot)$ denotes the MLP that maps the encoded vector of a point to a height value, and θ is the set of trainable parameters in the MLP. A popular trigonometric-functions-based form is applied in the position encoding:

$$p(r) = \left(r, sin(2^0\pi r), cos(2^0\pi r), \cdots, sin(2^{L-1}\pi r), cos(2^{L-1}\pi r)\right), \tag{11}$$

where L denotes the order of frequency of the vector. The frequency of the height map can be inflexibly adjusted by changing the value L. We set $L=9$ with the height map resolution of 2048×2048. It is worth noticing that there are precision limits in manufacturing, so the actual height map of the DOE is the quantization result of the numerical height values.

3.4 Compensation Network

In order to generate the compensated image for practical display, we design a compensation network that takes the target image $\mathbf{I}$ and the projection result $\mathbf{I}^*$ as input, and outputs a compensated image $\hat{\mathbf{I}}$ as the input image of the projector for display. The compensation network aims at generating a display in which the projected result is numerically and perceptually close to the desired image. The goal of taking only $\mathbf{I}$ and $\mathbf{I}^*$ as inputs is to avoid the time-consuming processing of projecting patterns and PSFs estimation.

In the training stage, the input projection result images are simulated by using the convolutional model in Eq.9 and the PSF simulated by the MLP; while in the testing stage, we use a camera to capture the projection result images directly and apply the compensation network for compensated image generation.

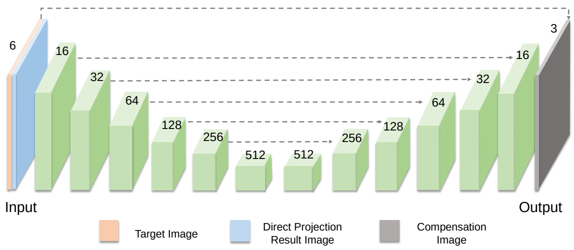

Figure 5: The U-net architecture of the projection compensation network. The network contains seven scales with six downsampling and upsampling blocks. The channel amount of each block is shown in the figure. Note that the input target image is directly added to the output image for residual learning.

The architecture of the compensation network is illustrated in Fig. 5. We adopt a U-net consisting of seven scales with six downsampling and upsampling blocks respectively. In each block, there are two convolutional layers with 3×3 filters, each followed by a Leaky ReLU activation function. To reduce the training burden of the network and accelerate the training, we add a long skip connection from the input target image to the output image and initialize the weights of the network to zeros. This residual-learning mechanism allows the network to focus more on learning how to restore the high-frequency residual details.

As the MLP and the compensation network are both differentiable, the whole framework is trained to minimize the following loss function via backpropagation:

$$\mathcal{L} = \frac{\alpha}{N} ||\hat{\mathbf{I}}^* - \mathbf{I}||_2^2 + \beta D(\hat{\mathbf{I}}^*, \mathbf{I}), \tag{12}$$

where $||\cdot||_2^2$ denotes the squared error term, N denotes the resolution of the projector images, $D(\cdot)$ denotes the DISTSs based perceptual metric [8] which is widely used in image quality assessment, and variables α and β denote the weights for the two terms. We set $\alpha = 1$ and $\beta = 0.05$ as empirical values. The loss function is optimized to jointly learn the parameters of optics and the compensation, in an end-to-end manner.

4 Simulation

4.1 Implementation

To evaluate our method, we train the proposed network and compare our simulation results with previous projector deblurring methods. The dataset used in the simulation is the DIV2K dataset [1], in which 500 high definition high-resolution images are for training while the other 200 images are for testing. We randomly crop image patches, each with the default resolution 1024×1024, to generate the target images. We simulate the PSFs by fixing the focusing distance $v = 500mm$ and substituting the arbitrary depths ranging from 300 mm to 800 mm into the wave optics propagation model. We use a projector Philips PPX4350 in both simulations and experiments. The used SPDs of the projector is measured and sampled from 400 nm to 700 nm with an interval of 30 nm. We generate the projection results of the input projector images by convolution with the PSFs, and crop the results images to the same resolution as the input projector images for concatenation.

Our framework is implemented in PyTorch and is trained on a workstation with Nvidia Tesla V100 GPU with the optimizer Adam. The initial learning rate is set to 1×10^{-3} and decayed by a factor of 0.99. We train the model for 150 iterations with a batch size of 1, which takes about six hours to complete. Note that in the test stage, we do not require PSF kernels, and only exploit the compensation network to generate compensated images from the desired images and its projection results.

4.2 Results

We present the evaluation of the proposed method by comparison with the state-of-the-art methods. Our proposed method is compared with three projector display methods, including directly projecting the desired images without compensation, non-blind deconvolution with known PSF [6], and compensation using OnlineProDeb [21]. Non-blind deconvolution utilizes the known blur kernel and directly solves a least square problem in the frequency domain by applying Fourier transform. OnlineProDeb and ours compensate for the projector's out-of-focus blur without known the PSFs. Since the compared method does not involve any modification of the projector optics system, we simulate the PSFs using only the projector lens for the compared methods; while in our method, the PSFs are simulated using the lens and the optimized DOE. The comparison of the methods is shown in Fig.6.

In the projection results without any compensation, the images are quite blurry since the planes at depths $z = 320mm, 630mm$, and $800mm$ are far away from the focusing plane ($v = 500mm$). Knowing the PSFs at different depths, the non-blind deconvolution method solves a least square problem to minimize the projection error, thereby maintaining the desired image color and contrast in the result image, although a large amount of high-frequency detail is lost. The loss of high-frequency detail is due to the fact that the objective function of the non-blind deconvolution method does not take into account perceptual errors. OnlineProDeb generates the compensated images for out-of-defocus projector display without the prior knowledge of PSF, instead, it learns the compensation for projector deblurring in a data-driven manner. Since the OnlineProdeb can only deal with gray images, we directly apply the pretrained OnlineProDeb for each color channel to compensate for the single lens projector display at varying depths. The results show that the improvement brought by OnlineProDeb is unnoticeable. This is because that OnlineProDeb can only handle slight blur and the quality of compensation is highly sensitive to the PSFs. Our method learns both DOE and compensation, and achieves the best performance compared with others. Furthermore, the high-frequency details are preserved in our results.

We also evaluate the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) [51] metrics of the results for each method in Fig.6. PSNR and SSIM are treated as conventional metrics to evaluate the numerical differences between two images while LPIPS is the perceptual metric to judge the perceptual similarity between two images. We observe that the non-blind deconvolution method performs the best on the PSNR metric with known PSFs. This is reasonable since PSNR is determined by mean square error (MSE), and the goal of the non-blind deconvolution method is exactly to find the solution that minimizes MSE of the display images. Unlike the non-blind deconvolution, OnlineProDeb and our method do not need to know the PSFs. The advantages of our method on SSIM and LPIPS metrics are significant, which is also consistent with our visual perception of comparing these images. Note that although we do not use SSIM and LPIPS in the loss function for training, our method shows superior results on the two metrics and provides more high-frequency image details than other methods.

Finally, we compare the computational time of the three methods with varying resolutions. The results are shown in Tab.1. The computation time of the non-blind method depends on the resolution of the test images with linear complexity. Compensating for an image with a resolution of 256×256, the non-blind method takes

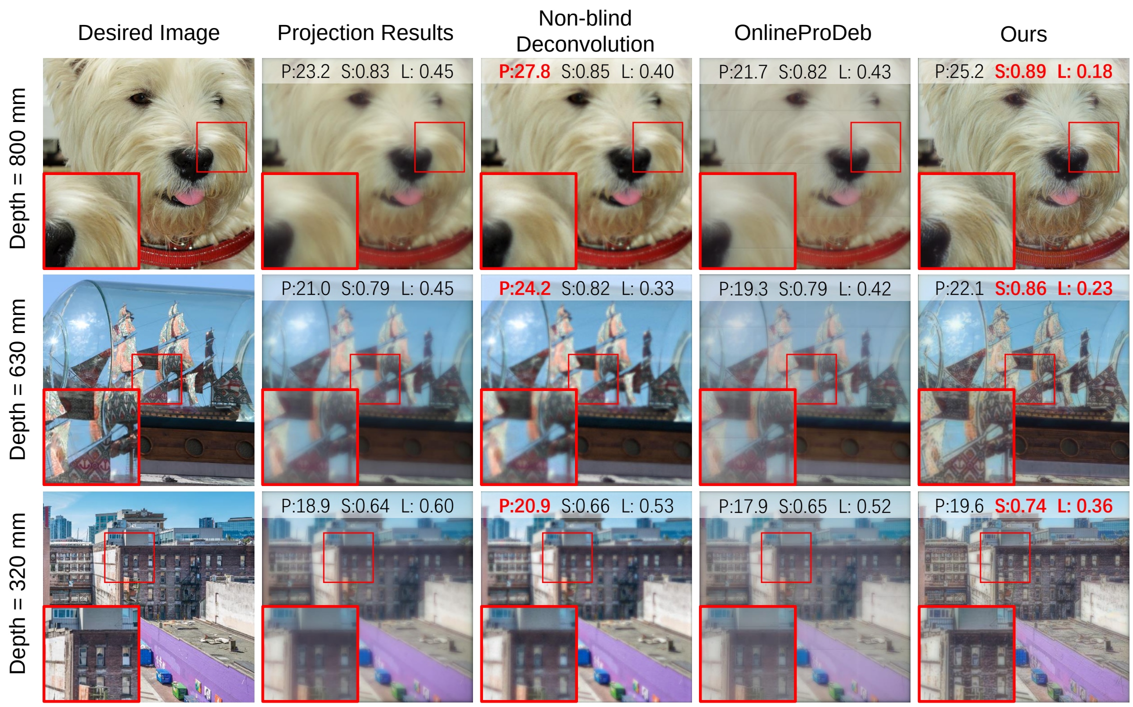

Figure 6: The comparison of simulated display results of the four methods on a projection plane at the projection distance (depth) $z = 320$ mm, 630 mm, and 800 mm. The used kernels are also simulated by either only using the original projector lens (Projection Results, Non-blind Deconvolution, OnlineProDeb) or using both the lens and the learned DOE (Ours). Here we set the focusing distance of the original projector lens $v = 500$ mm. The PSNR (P)↑, SSIM (S)↑, and the LPIPS (L)↓ of each method are given. The best result of each metric of each example is shown in bold red fonts. Our method gives the best image display perceptual quality (LPIPS) and structured similarity (SSIM) among the four methods. Please zoom in to see more details.

Table 1: The computational time of the three methods for deblurring a single projector image with three channels.

Resolution	Computation Time		
	Non-blind [6]	OnlineProDeb [21]	Proposed
256x256	**1.5 ms**	90 ms	3.6 ms
600x800	24 ms	N/A	**4.5 ms**
768x1024	37 ms	N/A	**5.6 ms**

less than half of our computation time. However, it takes five times ours to compensate for a single image with a resolution of 600×800. This is because our convolution network works in parallel and the computational time is less affected by the image resolution. Besides, OnlineProDeb currently only supports the compensation for the images with a resolution of 256×256, and the computational complexity is much higher than ours due to its more complicated network architecture. Our compensation network can achieve real-time efficiency for even higher-resolution images.

5 Experiments

To fully investigate the performance of our method in practice, a prototype of a modified projector is designed and implemented. In the following sections, we will present the built prototype and give the experimental comparisons of real projection results with other methods in detail.

5.1 Prototype

Our prototype consists of a modified projector equipped with the optimized DOE. The used projector is Philips PPX4350 with a spatial resolution of 600×800. We print a mount and fix it closely in front of the projection lens with a lens diameter of 7 mm. The fabricated DOE is placed inside of the mount. The focusing distance v used for training is set to 600 mm.

Fabrication DOE can be fabricated with either subtractive or additive lithography techniques [9]. Since the feature sizes distribute uniformly in a rotational symmetric manner for the optimized height map, we choose to fabricate the DOE with repeated photolithography (PL) and reactive-ion etching (RIE) steps [14, 38]. Before the fabrication, we prepare a double-side polished 4-inch fused silica wafer (thickness of 0.5 mm) that is cleaned by Piranha solution at 110 °C for 20 min. In the PL step, we deposit a 200 nm-thick layer of Chromium (Cr) on the substrate by sputtering, and then spin-coat the photoresist (AZ1505, 0.6 μm) with 1 min of pre-bake at 100 °C. The coated wafer is closely contacted with a pre-defined mask a contact aligner (EVG 6200∞), and UV light (i-line) exposure is applied to the photoresist (9 mJ/cm^2). After UV exposure, the pattern on the mask is transferred to the photoresist, which forms the desired pattern after 18 sec of development in AZ726. A wet etching step is necessary to transfer the pattern again from the photoresist to the Cr film to a create a hard mask. The residual photoresist is then removed by acetone. In the RIE step, we use a combination of gases (CHF_3, 15 sccm and O_2, 5 sccm) at 10 °C to remove the materials on the substrate under the open areas of the Cr mask. The etched depth depends linearly on the etching time. Once the target depth is achieved, the residual Cr is removed by the Cr etchant. This basic PL-RIE fabrication cycle is repeated 4 times to fabricate 16-level micro-structures, with the etched depth doubled each time. The etched depths in the 4 iterations are 75 nm, 150 nm, 300 nm and 600 nm respectively. A final circular Cr aperture is deposited to block the light outside of the effective region.

PSFs To evaluate the optics part of the proposed method, without applying any projector deblurring methods, we compare the

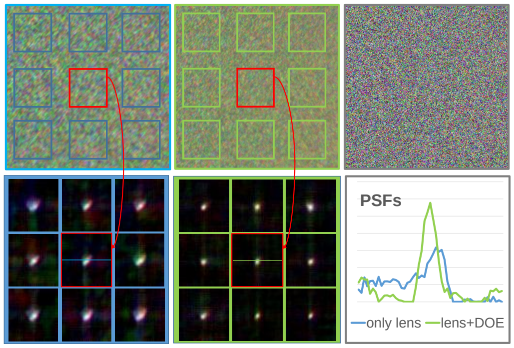

Figure 7: To estimate PSFs of the original and the modified projector at an out-of-focus projection depth $z = 300$ mm with focusing distance $v = 500$ mm, we project random Bernoulli color noise pattern (top-right) to the flat screen and then captured the patterns of using DOE (top-middle) and using only the lens (top-left). The PSFs are estimated by using Eq.13. The estimated PSFs of using DOE (bottom-middle) and using only the lens (bottom-left) at each spatial location show that our optimized optics effectively improve the EDOF of the projector. The example plots (bottom-right) of the middle row (solid line) of the middle PSF show that our PSF is much sharper than that using only the lens. Here the sizes of the PSF windows are 50.

out-of-focus PSFs of the original projector with only the lens and the modified projector with the lens and the optimized DOE. Motivated by the previous camera intrinsic blur kernel estimation [33], we project a sequence of random Bernoulli color noise patterns to the flat screen using the original and modified projector, and capture the images of the projected patterns using a DSLR camera Canon 750D as shown in Fig.7. In order to ensure that the image blur mainly comes from the projectors rather than the camera, we use the camera with minimum aperture and place the flat screen in the wide range of the camera's DOF as shown in Fig.8.

Suppose the pattern sequence have n images, we denote the sequence of the Bernoulli patterns as $\{\mathbf{B}_i | i = 1, ..., n\}$, and denote the pixel-to-pixel corresponding captured pattern as $\{\mathbf{C}_i | i = 1, ..., n\}$, the PSF $\mathbf{p}$ is estimated as follows:

$$\mathbf{p} = \mathcal{F}^{-1}\left\{\frac{\sum_{i=1}^{n} \bar{\mathcal{F}}\{\mathbf{B}_i\}\mathcal{F}\{\mathbf{C}_i\}}{\sum_{i=1}^{n} \bar{\mathcal{F}}\{\mathbf{B}_i\}\mathcal{F}\{\mathbf{B}_i\}}\right\}, \tag{13}$$

where $\bar{\mathcal{F}}$ is the complex conjugate of the Fourier transform. We use $n = 5$ here to enhance the robustness of the PSF estimation. Note that the color of the camera responses is not consistence with the input color of the projector, we apply nonlinear color calibration to match the two color spaces. The color calibration will be described in detail in the following Sec.5.2.

We crop the images into patches at varying locations and estimate the PSFs of the patches. Fig.7 shows the spatial distribution of the PSFs across the full image. The PSFs of both projectors show fewer variations with the change of spatial location. The PSFs created by the modified projector are superior to those of the original projector. We also show the depth-variant PSFs of the two projectors in Fig. 1. This indicates that our modified projector is able to generate sharp images at out-of-focus depths.

5.2 Calibration

As mentioned above, the projection results of the target images are regarded as the direct convolutional results of the target images with a PSF. However, in the practical display, the projection results of the target images are captured by a DSLR camera and thus have to be transformed from the camera color space to the color space

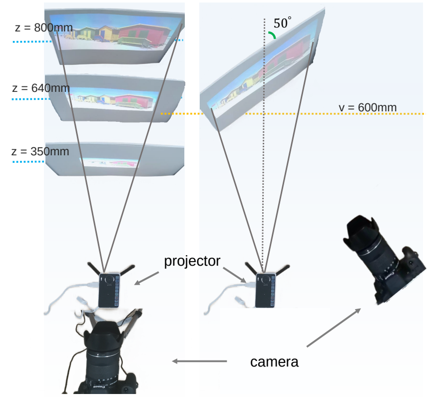

Figure 8: The two projection screen setups of our experiments. The flat screen is placed at different projection depths in the first setup and placed at a certain inclination angle in the second setup. We calibrate a Procams and use it to evaluate the displays.

of the outputs of the projector; the input images of the projector also need to be transformed to the color space of the projector outputs for consistency. The color space alignment in the projector-camera-system (ProCams) usually contains linear and nonlinear component [41]. Therefore, to align the color spaces, it is necessary to calibrate the colors of the ProCams to obtain both the linear color mixing matrix between the projector and the camera and the 3D non-linear response functions of the projector.

We use the projector to project three primary-color (red, green, blue) images to the projection screen and use the camera to capture the images. Although the color-mixing matrix is spectral-reflectance-dependent, we assume that the spectral reflectance of the screen is similar and uniform. Let the outputs of the three single-color images be $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ respectively, the vectors in the color mixing matrix can be constructed using the average pixel intensity of the illuminated regions in the captured images. Applying the inverse of the color-mixing matrix, we can directly transform the colors of the captured image to the colors of the outputs of the projector [35]. In the experiments, we use the Canon 750D camera.

Then we densely sample the colors of the projector on the projection screen and capture the images to obtain the responses. Multiplying the inverse of the known color-mixing matrix, we get the corresponding output colors of the projector input color samples. To construct the 3D non-linear response function of the projector that maps the input colors to the output colors, instead of obtaining an explicit expression of the function, we treat it as a nonlinear interpolation problem and adopt the local linear embedded techniques [26] which interpolate on the non-linear response functions via a linear combination of the nearest k neighbors. The computational complexity of the non-linear interpolation is much lower than the global interpolation [12, 27], and can be accelerated by parallel computation. To achieve geometric calibration, we firstly project a sequence of 16×32 structured dots, then detect the spatial location of each light blob on the captured images, and finally, make pixel-to-pixel correspondence between the projector and the camera to warp the captured images to the coordinate system of the projector images by using bilinear interpolation.

5.3 Results

We implement the methods in the experiments with two types of projection screen setups, as shown in Fig.8. In the first setup, the screen is placed perpendicular to the projection direction at varying depths $z = 350$ mm, 640 mm, and 800 mm while the focusing distance is fixed to $v = 600$ mm; in the second setup, the tilted screen is placed at an angle of 50 degrees to the projection direction,

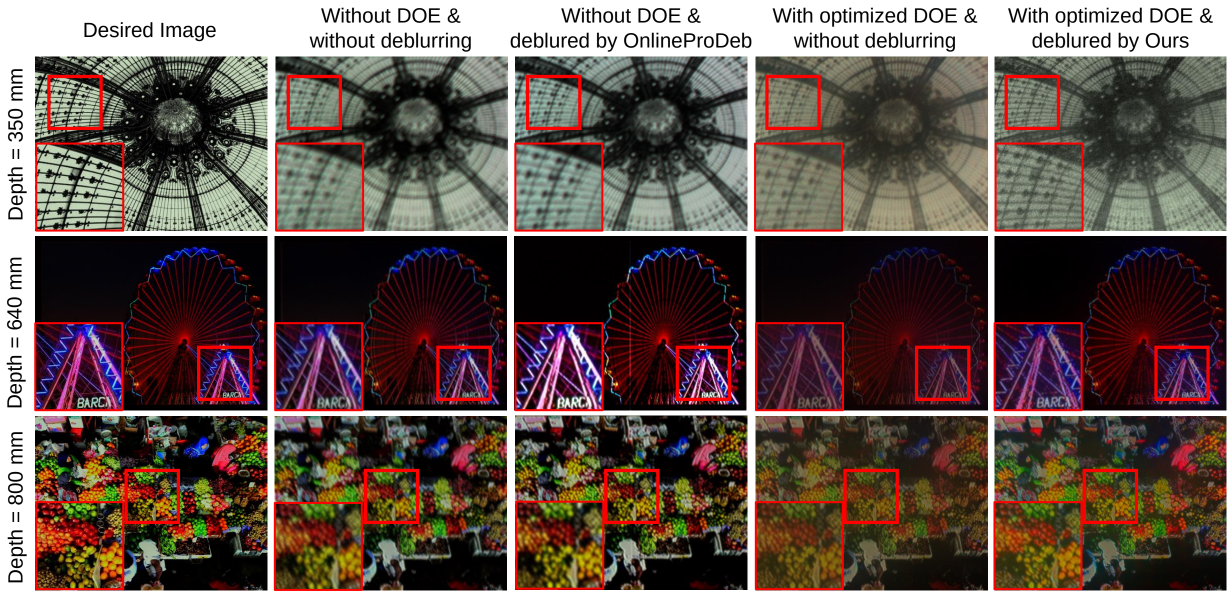

Figure 9: The comparison of captured display results of the four methods on a projection plane at the projection distance (depth) $z = 350$ mm, 640 mm, 800 mm. Here we set the focusing distance of the original projector $v = 600$ mm, and the normal projection (without DOE and compensation) shows a narrow DOF. Note that although the results of solely using learned DOE are hazed, the edge details are preserved. Our method gives the best image display quality among the four methods. Please zoom in to see more details.

and the left region of the displayed images is at the in-focus depth $z = 600$ mm.

To evaluate the generalization ability of our method, the dataset we used for testing is KonIQ-10k IQA Database [15], which is constructed for image quality assessment purposes and is completely different from the training set. The dataset contains 10,073 quality-scored images, all of which are given ratings for the quality of brightness, color, contrast, and sharpness. Because the EDOF task is most concerned with the sharpness of the image, we randomly select 1,000 images from the top 20% sharpest images from KonIQ-10k IQA Database, and randomly crop and resize the images to the patches with a resolution of 600×800 as the test images.

We calibrate the geometry of each setup and the color of each method as presented above. OnlineProDeb can only handle the image with the size 256×256, therefore we crop the testing images into 2×2 sub-images, then resize each sub-image to the resolution 256×256, and finally resize and tile the four compensated sub-images to get the compensated images. Our compensation network is fine-tuned by 20 iterations using the real captured PSFs in the depth range of 300 mm - 900 mm. This is to deal with the inconsistency between the simulated and captured PSFs due to imperfect manufacturing and model approximations.

In the first setup of the real experiments, our results are compared with the uncompensated results (without DOE & without deblurring), the compensated results of OnlineProDeb (without DOE & deblurred by OnlineProDeb), the direct projection results of our modified projector (with optimized DOE & without deblurring), and the compensated results of our modified projector (with optimized DOE & deblurred by ours). Three examples of the results of the first setup are shown in Fig.9. The uncompensated results show the sharpest image at depth $z = 640$ mm since the depth is the closest to the focusing distance $v = 600$ mm. OnlineProDeb improves the image quality but the high-frequency details can not be preserved. It is worth noticing that the projection results with only DOE have sharp edges although the contrast is degraded. This is because the concentration ratio of the PSFs created by the DOE and the lens is higher than that created by only the lens, as shown in Fig.7. Therefore, unlike previous compensation networks, our network can be more focused on the task of contrast enhancement and color adjustment, which is much easier than image deblurring. On the other hand, our compensation network can be lighter and more efficient due to the low computational complexity. With the proposed compensation network, our results show superior image quality to other methods, both at out-of-focus and in-focus distances.

In the second setup of the real experiments, we also compare our method with the uncompensated results (without DOE & without deblurring), and the compensated results of OnlineProDeb (without DOE & deblurred by OnlineProDeb). Two examples of the results of the first setup are shown in Fig.10. The focusing distance is fixed to 600 mm, and the depth of the projection area on the screen ranges from 600 mm to 800 mm. Our method presents sharp details at varying depths in the range, while others can only give in-focus display results on the left region of the projection images.

We also show the quantitative results for the displayed images in Table2. The comparative results in terms of SSIM and DISTS show the advantages of our method.

Table 2: The quantitative results of the displayed images.

Metrics	without DOE& without debluring	without DOE& OnlineProDeb	ours
SSIM↑	0.532	0.579	**0.590**
DISTS↓	0.598	0.484	**0.455**

6 Discussion

The effectiveness of the proposed method is verified in the experiments. However, the prototype still remains three major technical limitations in practical use.

First, the haze effect in the direct output images is noticeable due to the imperfections in the fabrication process and the incoherent light used in the prototype, which decreases the image contrast. The haze artifacts are basically caused by far-range inter-influence between pixels, which is an issue that has been well-studied in previous projection inter-reflection removal literature [25, 28, 37, 45].

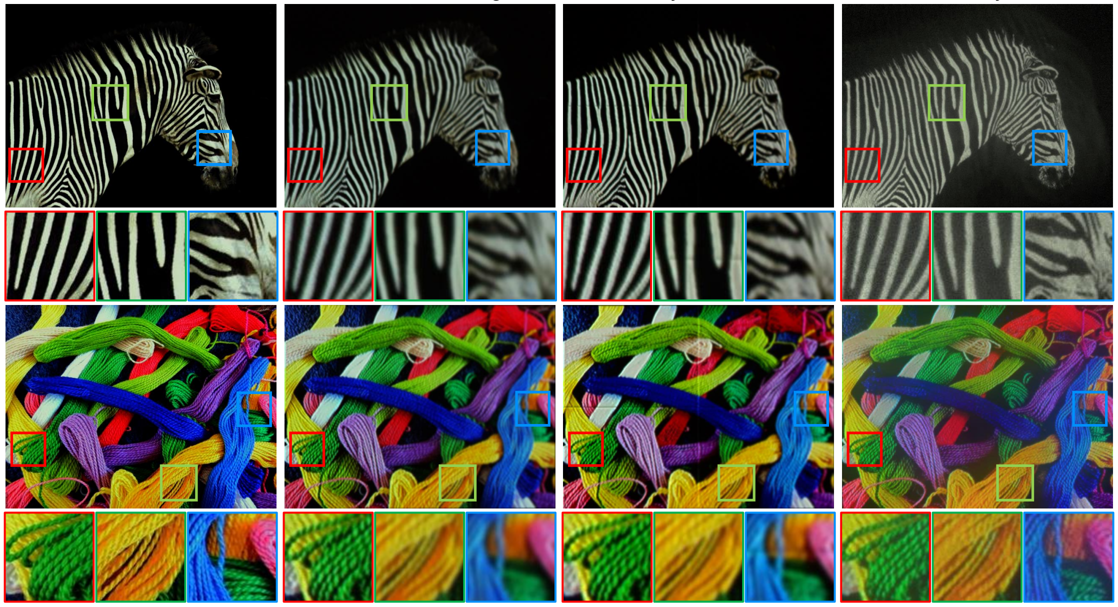

Figure 10: The comparison of captured display results of the three methods on a tilted projection plane inclined 50 degrees from the projection direction. Note that only the left part of the images is in focus when only using the original projector lens. Our method gives the best image display quality among the three methods. Please zoom in to see more details.

Although the multi-scale architecture allows our compensation network to handle the inter-influence of pixels in a large receptive field, the hazing phenomenon can be further improved from four aspects: (1) Increasing the contrast of the projection screen dynamically, the projection surface modulation technique can be helpful to further improve the haze effect [45]; (2) Applying gamut mapping to the desired image in presence of the hazing artifacts to enhance the visual quality [24]; (3) Applying two projectors for a superimposed projection display, one with only the original lens generating low-frequency components of the images and the other with optimized DOE generating high-frequency components (e.g. edges) of the images. Since the high-frequency component in an image is usually given by only a small set of pixels at the edges [40], the haze effect caused by these pixels should be ignorable in the display; (4) Since the learned DOE can produce sharp PSFs with long tails, directly applying the contrast enhancement method with low computational complexity [30] can also greatly improve the haze effect.

Besides, although the light transmittance of the DOE material is relatively high ($\geq 90\%$), we observe a slight color shift in the output images due to the diffraction efficiency variations among different wavelengths. At the design wavelength the DOE achieves maximum diffraction efficiency ($> 95\%$), and drops for other wavelengths, because of the inherent dispersion properties of DOEs. This can be seen as a problem with the primary color shift of the projector. In fact, color shift due to long-term projector lamp usage is quite common, and this problem can be resolved using a linear color mixing matrix correction. In our method, the linear transform is absorbed in the compensation network and thus does not bring extra computational burden. We adopt 16 levels as a comprise of efficiency and feasibility. Ideally a DOE would have the maximum diffraction efficiency with continuous profiles. However, fabricating such continuous DOEs is challenging and costly (using e.g., grayscale lithography). A common trade-off in optics is to approximate the continuous profile with multi-level binary structures as we did in this work. The diffraction efficiency for a binary DOE is only 40.5% in principle, which makes most of the light spreading out in the background. With 16 levels in only 4 steps of fabrication, the diffraction efficiency can reach $> 95\%$ already. Although it is tempting to make more levels, the gain in the improvement is worthless compared with rapidly increased fabrication complexity and cost.

In addition, our compensation network can generate normal resolution compensated images in real-time at frame rates higher than 100 fps, but the efficiency of the method is heavily affected by the latency of the ProCams devices and the readout time of images. Future programmable and configurable computational projectors and cameras embedded with neural processing units may be able to significantly improve the computing efficiency of the system.

7 Conclusions and Future Works

In summary, we propose an end-to-end hardware-software joint optimization framework for projector EDOF. In the framework, we learn both diffraction optics and the deep compensation neural network that can generate sharp image displays in a wide range of DOF. The parameters of diffractive optics are learned in a trainable wave optics model, which is the front-end module of our framework to simulate the projection results; while the compensation network works as a back-end module, and can deal with image blurring and color deviation without known PSFs. We implement a prototype by adding the learned DOE in front of the lens of a projector for phase modulation. Experiments illustrate that the modified projector can effectively create sharper PSFs than the original projector. Our compensation network also shows much higher efficiency and better visual perception compensation effect than previous methods. Our work opens up the possibility of automatically and jointly designing a computational projector for SAR-oriented displays.

In the future, we will explore complex projector lens optimization for dynamic DOF adjustment to adapt to the requirement of dynamic SAR applications. We will seek a more efficient compensation network to deal with both local (e.g., subsurface scattering, and blurring) and global light transports (e.g., inter-reflections). We also would like to work on optimal arrangements for multiple projectors in SAR applications based on their specific field-of-view and DOF.

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[2] S.-H. Baek, H. Ikoma, D. S. Jeon, Y. Li, W. Heidrich, G. Wetzstein, and M. H. Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2651–2660, 2021.

[3] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain. A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2):1–36, 2018.

[4] O. Bimber and A. Emmerling. Multifocal projection: A multiprojector technique for increasing focal depth. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):658–667, 2006.

[5] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer. The visual computing of projector-camera systems. In *ACM SIGGRAPH 2008 classes*, pp. 1–25. 2008.

[6] M. S. Brown, P. Song, and T.-J. Cham. Image pre-conditioning for out-of-focus projector blur. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1956–1963. IEEE, 2006.

[7] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli. Comparison of full-reference image quality models for optimization of image processing systems. *International Journal of Computer Vision*, 129(4):1258–1281, 2021.

[8] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2567–2581, 2022. doi: 10.1109/TPAMI.2020.3045810

[9] Q. Fu, H. Amata, and W. Heidrich. Etch-free additive lithographic fabrication methods for reflective and transmissive micro-optics. *Optics Express*, 29(22):36886–36899, 2021.

[10] J. W. Goodman and P. Sutton. Introduction to fourier optics. *Quantum and Semiclassical Optics-Journal of the European Optical Society Part B*, 8(5):1095, 1996.

[11] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber. Coded aperture projection. *ACM Transactions on Graphics (TOG)*, 29(3):1–12, 2010.

[12] A. Grundhöfer and D. Iwai. Robust, error-tolerant photometric projector compensation. *IEEE Transactions on Image Processing*, 24(12):5086–5099, 2015.

[13] A. Grundhöfer and D. Iwai. Recent advances in projection mapping algorithms, hardware and applications. In *Computer Graphics Forum*, vol. 37, pp. 653–675. Wiley Online Library, 2018.

[14] F. Heide, Q. Fu, Y. Peng, and W. Heidrich. Encoded diffractive optics for full-spectrum computational imaging. *Scientific reports*, 6(1):1–10, 2016.

[15] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020.

[16] B. Huang and H. Ling. Compennet++: End-to-end full projector compensation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7165–7174, 2019.

[17] B. Huang and H. Ling. End-to-end projector photometric compensation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6810–6819, 2019.

[18] B. Huang and H. Ling. Deprocams: Simultaneous relighting, compensation and shape reconstruction for projector-camera systems. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2725–2735, 2021.

[19] D. Iwai, S. Mihara, and K. Sato. Extended depth-of-field projector by fast focal sweep projection. *IEEE transactions on visualization and computer graphics*, 21(4):462–470, 2015.

[20] Y. Kageyama, M. Isogawa, D. Iwai, and K. Sato. Prodebnet: projector deblurring using a convolutional neural network. *Optics Express*, 28(14):20391–20403, 2020.

[21] Y. Kageyama, D. Iwai, and K. Sato. Online projector deblurring using a convolutional neural network. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2223–2233, 2022.

[22] H. Kawasaki, S. Ono, Y. Horita, Y. Shiba, R. Furukawa, and S. Hiura. Active one-shot scan for wide depth range using a light field projector based on coded aperture. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3568–3576, 2015.

[23] J. Lee, S. Jung, J. W. Kim, and F. Biocca. Applying spatial augmented reality to anti-smoking message: focusing on spatial presence, negative emotions, and threat appraisal. *International Journal of Human–Computer Interaction*, 35(9):751–760, 2019.

[24] K.-K. Lee, J.-W. Kim, J.-H. Ryu, and J.-O. Kim. Ambient light robust gamut mapping for optical see-through displays. *Optics Express*, 28(10):15392–15406, 2020.

[25] X. Li, Z. Hu, H. Gu, T. Yang, Q. Lei, H. Chen, P. Cheng, and Z. You. Multiprojector interreflection compensation using a deep convolution network. *Scientific Programming*, 2022, 2022.

[26] Y. Li, A. Majumder, M. Gopi, C. Wang, and J. Zhao. Practical radiometric compensation for projection display on textured surfaces using a multidimensional model. In *Computer Graphics Forum*, vol. 37, pp. 365–375. Wiley Online Library, 2018.

[27] Y. Li, C. Wang, J. Zhao, and Q. Yuan. Efficient spectral reconstruction using a trichromatic camera via sample optimization. *The Visual Computer*, 34(12):1773–1783, 2018.

[28] Y. Li, Q. Yuan, and D. Lu. Perceptual radiometric compensation for inter-reflection in immersive projection environment. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, pp. 201–208, 2013.

[29] C. Ma, J. Suo, Q. Dai, R. Raskar, and G. Wetzstein. High-rank coded aperture projection for extended depth of field. In *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–9. IEEE, 2013.

[30] A. Majumder. Contrast enhancement of multi-displays using human contrast sensitivity. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 377–382. IEEE, 2005.

[31] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1375–1385, 2020.

[32] F. Morosi and G. Caruso. High-fidelity rendering of physical colour references for projected-based spatial augmented reality design applications. 2020.

[33] A. Mosleh, P. Green, E. Onzon, I. Begin, and J. Pierre Langlois. Camera intrinsic blur kernel estimation: A reliable framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4961–4968, 2015.

[34] M. Nagase, D. Iwai, and K. Sato. Dynamic defocus and occlusion compensation of projected imagery by model-based optimal projector selection in multi-projection environment. *Virtual Reality*, 15(2):119–132, 2011.

[35] S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *ICCV workshop on projector-camera systems (PROCAMS)*, vol. 3. Citeseer, 2003.

[36] Y. Oyamada and H. Saito. Focal pre-correction of projected image for deblurring screen image. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, 2007.

[37] M. O'Toole and K. N. Kutulakos. Optical computing for fast light transport analysis. *ACM Trans. Graph*, 29(6):1–164, 2010.

[38] Y. Peng, Q. Fu, F. Heide, and W. Heidrich. The diffractive achromat full spectrum computational imaging with diffractive optics. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, pp. 1–2. 2016.

[39] Y. Peng, Q. Sun, X. Dun, G. Wetzstein, W. Heidrich, and F. Heide. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.*, 38(6):219–1, 2019.

[40] B. Sajadi, M. Gopi, and A. Majumder. Edge-guided resolution enhancement in projectors via optical pixel sharing. *ACM Transactions on Graphics (TOG)*, 31(4):1–122, 2012.

[41] B. Sajadi, M. Lazarov, and A. Majumder. Adict: Accurate direct and

inverse color transformation. In *European Conference on Computer Vision*, pp. 72–86. Springer, 2010.

[42] Z. Shi, Y. Bahat, S.-H. Baek, Q. Fu, H. Amata, X. Li, P. Chakravarthula, W. Heidrich, and F. Heide. Seeing through obstructions with diffractive cloaking. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022.

[43] 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 Transactions on Graphics (TOG)*, 37(4):1–13, 2018.

[44] Q. Sun, E. Tseng, Q. Fu, W. Heidrich, and F. Heide. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1386–1396, 2020.

[45] S. Takeda, D. Iwai, and K. Sato. Inter-reflection compensation of immersive projection display by spatio-temporal screen reflectance modulation. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1424–1431, 2016.

[46] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnulo, G. L. Cascella, and M. Fiorentino. Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *The International Journal of Advanced Manufacturing Technology*, 94(1):509–521, 2018.

[47] L. Wang, H. Xu, S. Tabata, Y. Hu, Y. Watanabe, and M. Ishikawa. High-speed focal tracking projection based on liquid lens. In *ACM SIGGRAPH 2020 Emerging Technologies*, pp. 1–2. 2020.

[48] G. Wetzstein and O. Bimber. Radiometric compensation through inverse light transport. 2007.

[49] H. Xu, L. Wang, S. Tabata, Y. Watanabe, and M. Ishikawa. Extended depth-of-field projection method using a high-speed projector with a synchronized oscillating variable-focus lens. *Applied Optics*, 60(13):3917–3924, 2021.

[50] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. In *ACM SIGGRAPH 2006 Papers*, pp. 907–915. 2006.

[51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.