Post-processing Spatially-Varying Blur Effects with Learned 6D Blur Fields

Zhecheng Wang and Esther Lin

Abstract—The point spread function (PSF) describes a camera's response to a point light source. It describes how a camera renders a 3D scene onto the 2D captured image and explains defocus blurs. Consequently, accurately modeling a camera's PSF is akin to extracting its operational behavior. This projects explores the post-processing of network-represented, spatially-varying blurs, especially in the context of cinematic applications. By controlling the blur of an image, filmmakers can convey a sense of depth, draw attention to the main subject, and add a sense of motion and action. Accurately modeling and controlling spatially varying blur is therefore essential for creating high-quality cinematic images.

Index Terms—Computational Photography, Neural Network

1 Introduction

THis is an ongoing work at the University of Toronto on estimating and representing PSFs. They introduce the novel use of a coordinate-based neural network for representing the PSF. This representation is faithful to the underlying physics: the PSF has a multidimensional dependence on a laundry list of spatial variables, for example, location on the 2D image plane and the distance between the sensor and the lens, and the proposed method is the first to encompass these parameters holistically with the highest accuracy.

While these realistic PSFs can immediately elevate existing depth estimation and deblurring applications, there is another area that remains unexplored: rendering. In this project, we propose the use of these PSFs for two applications:

- Creating photorealistic renderings that capture the aberrations of a real imaging system.
- Synthesizing cinematic effects through realistic synthetic defocusing.

Both of these applications have the potential to be game-changing. First, photorealistic rendering is a battle that many professionals face. Existing rendering engines such as Blender only support the simple thin lens model, which cannot capture the spatially-varying aberrations natural imaging systems inject into an image. Utilizing these PSFs in the rendering pipeline will add realism to the results. Second, lens-adjustment-based techniques such as Dolly zoom and focus adjustment are commonly used in cinematography for dramatic effect or guiding audience attention. However, these effects are expensive in terms of manpower, time, and prior training. Our PSFs can create the same effects digitally through synthetic defocus–at a much lower cost.

In the field of computational photography, it is common to consider images as 2D representations of the visual world.

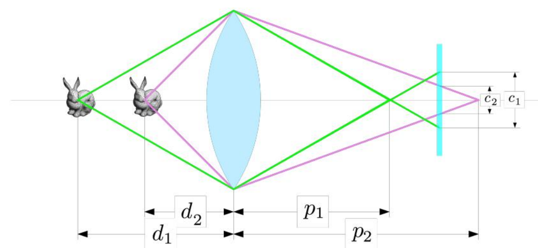

Fig. 1. Figure 1. Thin lens model formulation.

However, in reality, the response of a camera to light can be described as a 4D light field function. This 4D light field passes through the point spread function (PSF) of the imaging system and is projected onto the sensors, where it is captured as a 2D image. By understanding the process by which the camera ”renders” the 4D light field onto the sensor, it is possible to backtrack and reconstruct the original 4D light field from the captured image. Therefore, the first step in this inverse process would be accurately capturing the PSF of the imaging system. The fourth dimension in this context is the angle of the light field.

We summerize our contributions as following:

- Speed up of when rendering images blurred with our spatially-varying PSFs. We use strided patch-based kernel convolutions over a patch of pixels to avoid querying for every pixel.
- A Blender plugin for integrating our PSFs. Previously, Blender only supported image formation with the thin-lens model. To our best knowledge, this plugin is the first one in all render engines to consider camera specific abbreviation.
- Synthesis cinematic effects using the 6D blur fields via image formation.

- *M. Shell is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. E-mail: see http://www.michaelshell.org/contact.html*
- *J. Doe is with Anonymous University.*

2 Related Work

There have been multiple works trying to post-process spatial blur effects on images/videos. Among different blur effects including defocusing/refocusing/portrait mode (Gao et al. 2020), dolly zoom/vertigo effects (Liang et al. 2020) are the most challenging and artistically pleasing one as it involves constant changes of lens position and FoV. In a recent attempt by Gao et al. 2021, the authors jointly trained a time-invariant static NeRF and a time-varying dynamic NeRF and could achieve dolly zoom effects. However, these works do not capture the spatially-varying nuances a real camera can do in general as the PSF information is lost when capturing the RGB image.

3 Proposed Method

The refocus pipeline could be summarized as applying the learned 6D blur field to all in focus image with pixel-wise convolution to get a blur stack and use a nonlinear image formation model to reconstruct the refocused blurred image.

A 6D blur field can provide a more detailed and accurate representation of the blurring effect caused by the point spread function (PSF) of an imaging system than a thin lens model. This is because a 6D blur field can incorporate a wider range of variables that affect the PSF, such as the position on the image plane, the distance between the lens and the sensor, the focus length, and the viewing angle. In contrast, a thin lens model assumes that the blurring is caused solely by the geometric optics of a thin lens, which is a simplification that may not accurately capture the complexity of the PSF in many imaging systems. As a result, using a 6D blur field can enable the development of more effective image restoration and deblurring techniques.

Our method considers a wider range of variables that affect the PSF, such as the position on the image plane, the distance between the lens and the sensor, the focus length, and the viewing angle. In contrast, a thin lens model assumes that the blurring is caused solely by the geometric optics of a thin lens, which is a simplification that may not accurately capture the complexity of the PSF in many imaging systems. As a result, our PSF estimation and representation method can enable the development of more effective image restoration and deblurring techniques, in addition to stylistic renderings.

3.1 Convolution

Patch-based blurring with strided convolution is a method for blurring images that involves dividing the image into overlapping patches, applying a blurring operation to each patch, and then combining the patches back together to create the final blurred image. One advantage of this method is that it can be implemented using a convolutional neural network, which can be trained to learn the best blurring parameters for a given image. This allows for a more flexible and adaptable approach to blurring than traditional methods, which often use fixed blurring kernels. Additionally, using strided convolutions can help reduce the amount of computation required, making the method more efficient.

One disadvantage of this method is that it can sometimes produce artifacts in the resulting blurred image, particularly around the edges of objects. This is because the patches used for blurring do not always align perfectly with the objects in the image, which can cause discontinuities at the patch boundaries. Another disadvantage is that the method is not always well-suited to blurring large images, as the computational cost can become prohibitively high for very large images.

3.2 Lens Model For Photorealistic Rendering

Having an accurate lens model is important for photorealistic rendering because it allows the rendering system to accurately simulate the behavior of light as it passes through the lens of a camera. This is important because the lens of a camera affects the way that light is focused and bent as it enters the camera, and this can have a significant impact on the final appearance of the image. An accurate lens model can help the rendering system to correctly simulate these effects, which can improve the overall realism and quality of the final image.

Additionally, an accurate lens model can help to improve the efficiency of the rendering process, as it can allow the rendering system to more accurately predict which parts of the image will be in focus and which parts will be out of focus. This can help to reduce the amount of computation that is required to generate the final image, which can make the rendering process faster and more efficient.

Overall, the benefits of having an accurate lens model for photorealistic rendering include improved image quality, greater realism, and more efficient rendering processes.

3.3 Nonlinear Defocus Blurring Model

Studied in multiple works [2], linear model are often used for image formation. It uses a simple linear addtion of masked blurred images,

$$I = \sum_{k=0}^{K-1} PSF_k * l_k \tag{1}$$

where K is the number of the layers in the quantized depth map, and PSF_k being the point spread function of the depth layer l_k.

For the purpose of this project, we chose the-state-of-art nonlinear image formation model from [citation needed] as following:

$$I = \sum_{k=0}^{K-1} \tilde{l}_k \prod_{k=k+1}^{K-1} (1 - \tilde{a}_{k'}) \tag{2}$$

where $E_k = PSF_k * \sum_{k'=0}^{k} a_{k'}$, $\tilde{l}_k = (PSF_k * l_k)/E_k$, and $\tilde{a}_k = (PSF_k * a_k)/E_k$. Although based on findings from [2], Using the nonlinear model didn't improve the blurred image PSNR drastically. Using the nonlinear model for image formation has significant visual improvement compared to linear model.

3.4 Cinematic Application

In cinematography, the lens is one of the most important components of the camera, as it is responsible for controlling the way that light enters the camera and is focused onto

Fig. 2. Figure 2. Top left: all in focus image taken on a DSLR camera. Top right: [1] estimated depth map from monocular image. Bottom left: back view is in focus. Bottom right: front view is in focus. The image has a resolution of 2762 x 3718 pixels.

the film or image sensor. Lens choice plays a critical role in determining the final look and feel of a shot.

Having an accurate 6D blur field is important for accurately modeling the lens because it allows for the simulation of the complex way that the lens focuses light onto the film or image sensor. This is particularly important for creating realistic and high-quality visual effects, as well as for simulating the unique visual characteristics of different lenses.

In addition to its importance for creating realistic visual effects, having an accurate 6D blur field is also important for simulating the unique visual characteristics of different lenses. Different lenses have their own distinct optical characteristics, such as their focal length, aperture, and distortion, which can affect the way that they focus light onto the film or image sensor. By applying the 6D blur

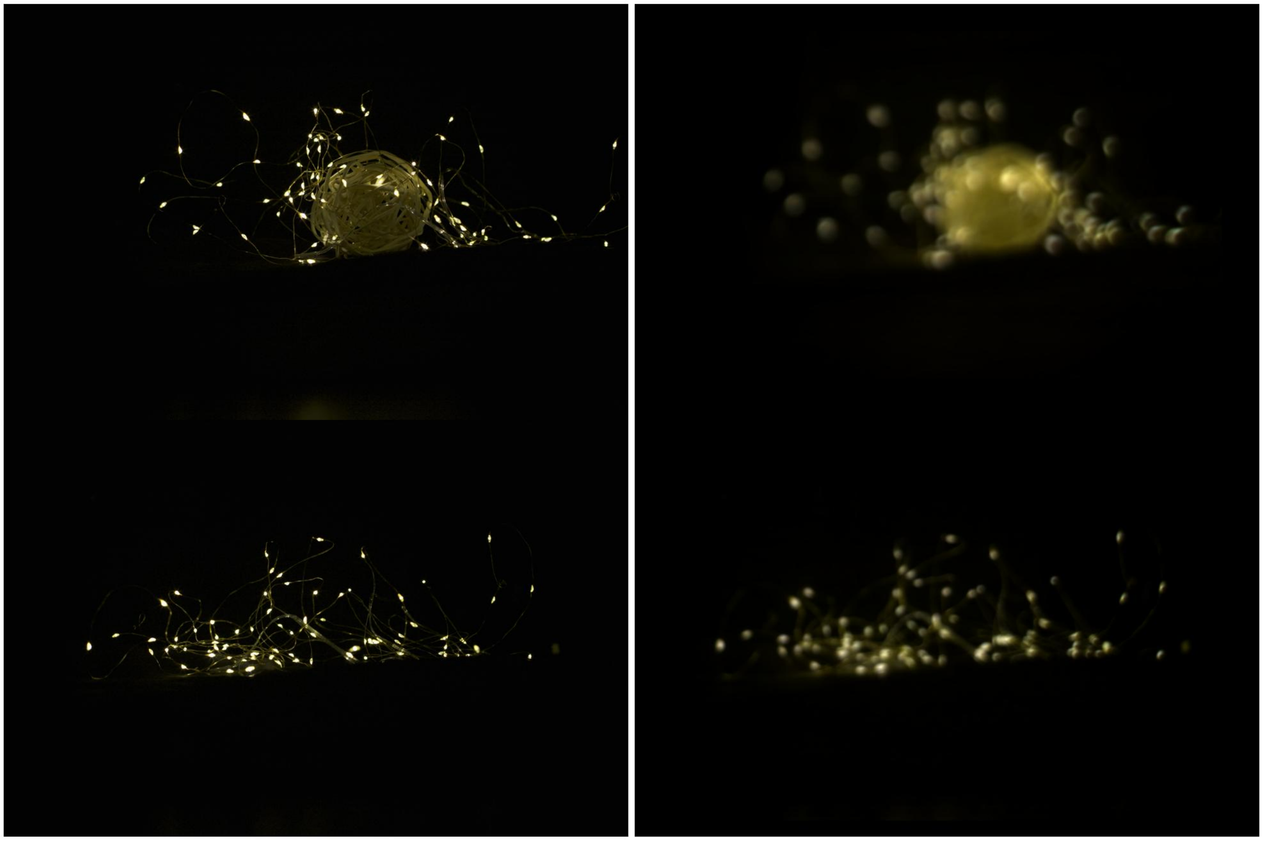

Fig. 3. Figure 2. Left: all in focus images taken on a DSLR. Right: iPhone 12 Pro wide lens 6D blur field applied bokeh.

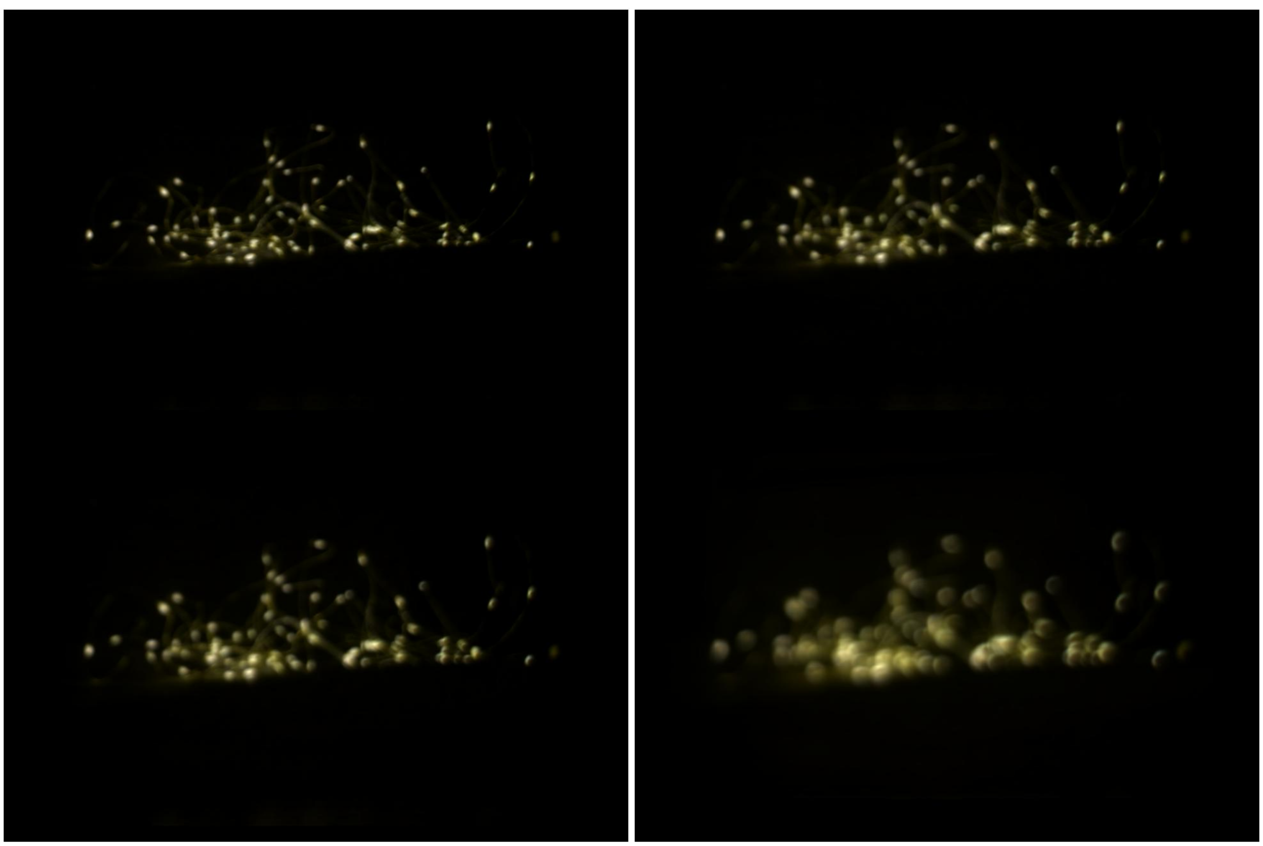

Fig. 4. Figure 4. Top left, top right, bottom left: we extrapolate the network lens positions by 0%, 50%, and 75% respectively. Bottom right: in addition to 75% extrapolation, we also scale up the kernel size by 4 times.

field, filmmakers and visual effects artists can accurately reproduce these characteristics, allowing them to create more faithful and authentic lens simulations.

Our model allows for the simulation of the complex way that the lens focuses light onto the film or image sensor, which is essential for creating realistic and high-quality visual effects, as well as for simulating the unique visual characteristics of different lenses.

4 Implementation

4.1 Convolution

To implement patch-based blurring using strided convolution, the first step would extracing the blur kernels. Like all other blur kerels, the learned PSF kernels are square shaped and normalized to 1. Next, strided convolution would be used to apply the blur kernel to the input image. In strided convolution, the filter is moved across the input image in steps or "strides" rather than being slid pixel by pixel. This allows for a more efficient implementation of convolution and can reduce the computational cost of applying the blur kernel to the input image. This would effectively blur the image by applying the blur kernel in patches, rather than to every single pixel. The output of the strided convolution operation would be a blurred version of the input image.

However, the major draw back of this method is there will be some visual artifacts around the kernel boundaries if the kernel size and/or the stride step is too large. By using overlapped patches, we could reduce this visual effects. It is adviced to use a extremely large stride as single pixel queries takes around 5 seconds on high end GPUs.

4.2 Blender Add-on

We use Blender built-in Python binding to construct the GUI within Blender. We render and cache the rendering and mist pass (normalized depth map automatically generated and clipped to 0-1) to the disk. Then we load it back into the memory and apply forward image formation with depth map (mist pass) and our learned 6D blur fields.

5 Experimental Results

We achieve a range of cinematic effects involving the manipulation of depth and blurring in an image. We use our PSF representations to synthetically create the shallow depth of field effect, add realism and depth with bokeh, and apply unique blurring effects characteristic by optical properties of the lens to a all-in-focus images.

Depth of Field/Refocus(Fig. 2): By manipulating the depth field, we simulate the shallow depth of field effect commonly used in photography and cinematography to focus attention on a specific part of the image and blur out the background.

Bokeh(Fig. 3): Bokeh is the aesthetic quality of the blur produced in the out-of-focus parts of an image. It is often used in cinematography and photography to direct the viewer's attention to a particular part of the image, and is typically created by using a lens with a shallow depth of field. In computational photography, bokeh can be simulated or enhanced using algorithms that manipulate the blur in an image, which can help to create a more pleasing or compelling visual effect. In summary, it is often used to add realism, visual appeal, and draw attention to the main subject (e.g. portrait mode). Our PSF representation can be used to create the visual effect of bokeh. The quality of bokeh is determined by a number of factors, including the shape of the aperture blades in the lens and the amount of light that is allowed to pass through the lens. Lenses with circular aperture blades tend to produce more pleasing bokeh, as the circular shape of the aperture creates circular highlights in the out-of-focus areas of the image. This example 4 also demonstrates how the network is capable of extrapolating blurs beyond what would have originally been physically possible.

6 Conclusion

Acknowledgments

The authors would like to thank the help from Esther Lin, Prof. David Lindell and Prof. Kyros Koutulakos for advice.

References

[1] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 179–12 188.

[2] H. Ikoma, C. M. Nguyen, C. A. Metzler, Y. Peng, and G. Wetzstein, "Depth from defocus with learned optics for imaging and occlusion-aware depth estimation," in *2021 IEEE International Conference on Computational Photography (ICCP)*, 2021, pp. 1–12.