Synthesizing Cinematic Effects with Learned Spatially-Varying Blur Fields

Zhecheng Wang and Esther Lin
{zhecheng, lin}@cs.toronto.edu, University of Toronto

Motivation

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.

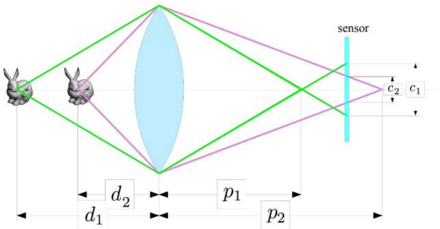

Figure 1: Different object placements result in different defocus blurs sizes.

There is ongoing research at UofT on estimating and representing PSFs using a novel coordinate-based neural network approach. This method is the first to accurately incorporate of the spatial variance of PSFs in a holistic and lightweight manner.

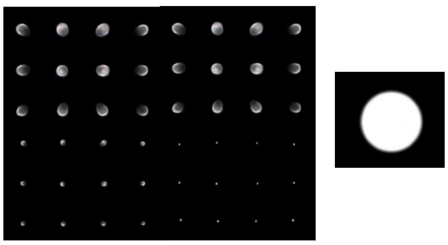

Figure 2: PSFs at different degrees of defocus for an iPhone 12 Pro Wide camera (left), modeled using neural network approach. These calibrated PSFs contain more high frequency detail compared to the pillbox or Gaussian kernels (right), two commonly-used blur models in Computer Vision.

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.

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.

References

[1] Ikoma, Nguyen, Metzler, Peng, & Wetzstein. *Depth from defocus with learned optics for imaging and occlusion-aware depth estimation*. IEEE International Conference on Computational Photography (ICCP), 2021.
[2] Ranftl, Bochkovskiy, & Koltun. *Vision transformers for dense prediction*. IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
[4] Gao, Saraf, Kopf, & Huang, *Dynamic View Synthesis from Dynamic Monocular Video*, IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
[5] Liang, Ranade, Wang, Bai & Lee, *The "Vertigo Effect" on Your Smartphone: Dolly Zoom via Single Shot View Synthesis*, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.
[6] Gao, Shih, Lai, Liang, & Huang,*Portrait Neural Radiance Fields from a Single Image*, arXiv:2012.05903, 2020.

New Technique

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.

This project focuses on building a rendering pipeline that uses our PSF representation. In particular, we developed the following:

1. 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.

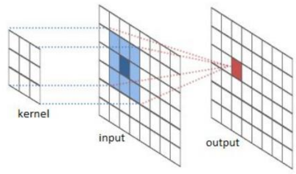

Figure 3: Patch-based convolution with a spatially-varying PSF.

2. 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.

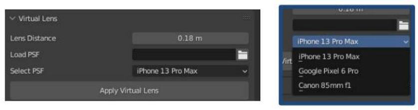

Figure 4: Blender Plugin. The blender plugin supports loading different PSFs and apply spatially varying blur to the rendered image with Blender's mist pass.

3. Implemented a non-linear image formation model [Ikoma et al. 2021] that is capable of capturing occlusions when rendering depth-dependent scenes.

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

Naive Linear Model

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

Ikoma et al. 2021 Nonlinear Model

where $E_k = PSF_k * \sum_{k'=0}^{k} a_{k'}$, $\tilde{l}_k = (PSF_k * l_k)/E_k$, and $\tilde{a}_k = (PSF_k * a_k)/E_k$.

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: 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.

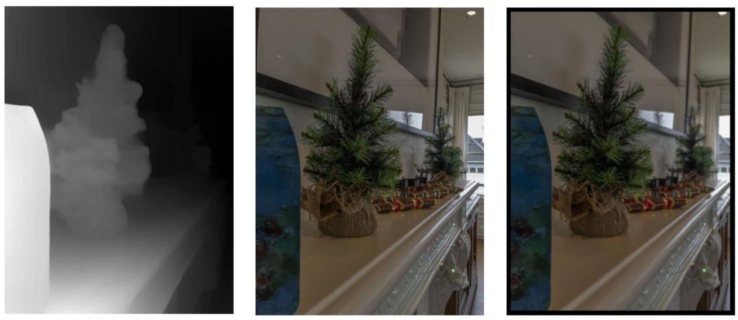

Figure 5: Left: [Ranftl et al. 2021] estimated depth map, Middle: Original in focus image, Right: 6D blur field convolved **novel focus** image. The 6D blur field is trained on iPhone 12 Pro wide lens.

Bokeh: Bokeh is the aesthetic quality of the blur in the out-of-focus areas of an image. 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. This example also demonstrates how the network is capable of extrapolating blurs beyond what would have originally been physically possible.

Figure 6: Left: **75% extrapolated beyond physically possible** network estimated PSF convolved image. Right: Original in focus image. The network used was trained on an iPhone 12 Pro wide lens.