Supplimentary Material for the paper : Camera-Independent Single Image Depth Estimation from Defocus Blur

Lahiru Wijayasingha
Computer Science Department
University of Virginia, USA
lnw8px@virginia.edu

Homa Alemzadeh
Electrical and Computer Engineering Department
University of Virginia, USA
ha4d@virginia.edu

John A. Stankovic
Computer Science Department
University of Virginia, USA
stankovic@cs.virginia.edu

1. Additional details on datasets

1.1. Blender dataset

We create a synthetic dataset by expanding the defocusnet dataset [1]. We create this dataset to show that the models have the ability to generalize its learned knowledge to a new dataset and also it can adapt to different camreas. The original dataset did not have the textures mapped to the 3D objects. We found that this is because the 3D models used are of the STL format (commonly used for 3D printing) which only stores the geometric shape of an object and does not support texture or color information. We convert some of the downloaded STL models into OBJ format after UV mapping to facilitate textures. Different from the defocusnet dataset [1] we create focal stacks taken from several virtual cameras instead of just one camera. Some examples of the defocusnet dataset and our dataset is shown in Figure 3. We vary the virtual cameras in two respects; f-stop and focal length. The f-stops of the simulated cameras were chosen to be 1.0,1.1,1.2,1.5,1.8,2.0,2.2,2.8,3.0,5.0,8.0 and 10.0 to create the $blender_{testN}$ dataset. The focal lengths of the camera were taken to be 3mm, 4mm, 5mm, and 6mm to create the $blender_{testF}$ dataset while keeping the F-number 2. We train our models with the data from the decocusnet dataset and test on the images coming from $blender_{testN}$ and $blender_{testF}$ datasets. $blender_{testN}$ dataset contains 1491 focal stacks. $blender_{testF}$ contains 400 focal stacks. In addition to this we also create another dataset ($blender_{train}$) with a virtual camera with a focal length of $2.9mm$ and F-number of 1. This dataset has 1000 focal stacks. Each focal stack in all the datasets we created contains 6 blurred images and additionally an all-in-focus (AIF) image. The blurred images are focused at distances of 0.1, 0.15, 0.3, 0.7, 1.5 meters and at infinity. We do not use the images focused at infinity in evaluations of this paper. We use the same 20 3D models, 10 textures and a single environment map to create all the datasets. We use the script provided by Maximov et al. (modified) to create our dataset. Note that in the paper we have only used the $blender_{testF}$ dataset.

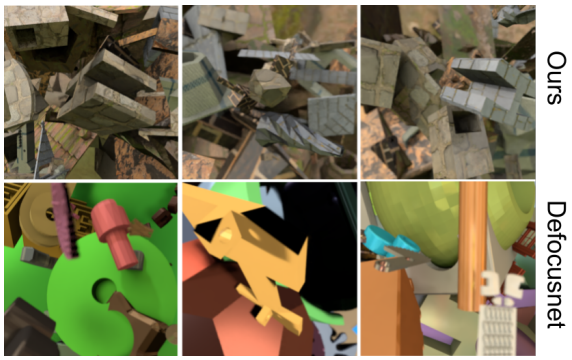

Figure 1. Synthetic dataset samples.

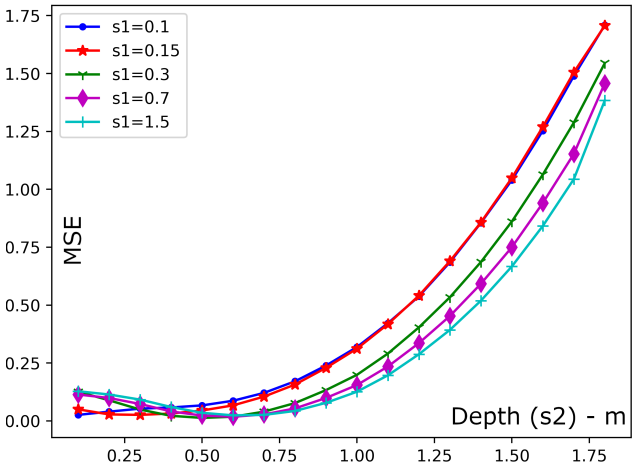

Figure 2. Depth Estimation MSE vs s_1

2. Additional Experiments

2.1. Variation of Depth Estimation Error vs focal distance (s_1)

Figure 2 shows the error of our model under several different focal distances (s_1). We can see that larger focal distances are performing better for larger distances and smaller

focal distances are performing better for smaller distances. Therefore, if we have prior knowledge about the range of distances that we are interested in measuring we can select an appropriate focal distance value to make more accurate depth and blur predictions. Also it can be seen that depth from defocus blur is more accurate for shorter distances. For example, for distances around 0.25m the MSE is around 0.1 and for distances around 1.5m the MSE is around 0.75.

3. Proofs

3.1. Convolution of a 2D Gaussian Function with another 2D Gaussian Function

In the main paper we have shown that the defocus-blurring can be modelled with a convolution of a 2D Gaussian function; the Point Spread Function (PSF) having a standard deviation of σ with the respective image in perfect focus (in-focus image). This Gaussian PSF can be denoted as below.

$$G(x,y) = e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}} \tag{1}$$

The blurred image B can be obtained as follows by convolving the in-focus image F with the blur function (PSF) G.

$$B(x,y) = G(x,y) * F(x,y) \tag{2}$$

Additional blurring of the image occurs due to phenomenon such as pixel binning and post processing which can be modelled as an additional convolution with a Gaussian function with a standard deviation of γ which we can be represented as follows.

$$Q(x,y) = e^{-\frac{1}{2}\frac{x^2+y^2}{\gamma^2}} \tag{3}$$

Let's consider the convolution of the already defocus-blurred image B with the additional blurring function due to pixel binning and post processing Q. The final image we can observe will be denoted by I

$$I = Q(x,y) * B(x,y) \tag{4}$$

$$I = Q(x,y) * (G(x,y) * F(x,y)) \tag{5}$$

Due to the associative nature of convolution

$$I = (Q(x,y) * G(x,y)) * F(x,y) \tag{6}$$

Let's explore the quantity $Q(x,y) * G(x,y)$

$$\begin{aligned}
&Q(x,y) * G(x,y) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(k,m)\cdot G(x-k,y-m)dkdm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{k^2+m^2}{\gamma^2}} \cdot e^{-\frac{1}{2}\frac{(x-k)^2+(y-m)^2}{\sigma^2}} dkdm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{k^2+m^2}{\gamma^2}+\frac{(x-k)^2+(y-m)^2}{\sigma^2})} dkdm \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{\sigma^2(k^2+m^2)+\gamma^2[(x-k)^2+(y-m)^2]}{\sigma^2\gamma^2})} dkdm \\
&= e^{-\frac{1}{2}\frac{\gamma^2(x^2+y^2)}{\sigma^2\gamma^2}} \cdot \\
&\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(\sigma^2+\gamma^2)k^2-2\gamma^2xk}{\sigma^2\gamma^2}} dk \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(\sigma^2+\gamma^2)m^2-2\gamma^2ym}{\sigma^2\gamma^2}} dm \\
&= e^{-\frac{1}{2}\frac{\gamma^2(x^2+y^2)}{\sigma^2\gamma^2}} \cdot \\
&\int_{-\infty}^{\infty} e^{-\frac{(\sigma^2+\gamma^2)}{2\sigma^2\gamma^2}(k-\frac{\gamma^2x}{\sigma^2+\gamma^2})^2+\frac{x^2\gamma^2}{2\sigma^2(\sigma^2+\gamma^2)}} dk \cdot \\
&\int_{-\infty}^{\infty} e^{-\frac{(\sigma^2+\gamma^2)}{2\sigma^2\gamma^2}(m-\frac{\gamma^2y}{\sigma^2+\gamma^2})^2+\frac{y^2\gamma^2}{2\sigma^2(\sigma^2+\gamma^2)}} dm
\end{aligned}$$

The integral of a Gaussian function,
$\int_{-\infty}^{\infty} e^{-a(x+b)^2} dx = \sqrt{\frac{\pi}{a}}$
Therefore,
$\int_{-\infty}^{\infty} e^{-\frac{(\sigma^2+\gamma^2)}{2\sigma^2\gamma^2}(k-\frac{\gamma^2x}{\sigma^2+\gamma^2})^2+\frac{x^2\gamma^2}{2\sigma^2(\sigma^2+\gamma^2)}} dk = \sqrt{\frac{2\sigma^2\gamma^2\pi}{\sigma^2+\gamma^2}}$
and
$\int_{-\infty}^{\infty} e^{-\frac{(\sigma^2+\gamma^2)}{2\sigma^2\gamma^2}(m-\frac{\gamma^2y}{\sigma^2+\gamma^2})^2+\frac{y^2\gamma^2}{2\sigma^2(\sigma^2+\gamma^2)}} dk = \sqrt{\frac{2\sigma^2\gamma^2\pi}{\sigma^2+\gamma^2}}$
The convolution becomes,

$$Q(x,y) * G(x,y) = 2\pi\frac{\sigma^2\gamma^2}{\sigma^2+\gamma^2} \cdot e^{-\frac{1}{2}\frac{(x^2+y^2)}{\sigma^2+\gamma^2}} \tag{7}$$

Therefore the convolution of a 2D Gaussian function with another 2D Gaussian function is also a Gaussian function. Let λ be the standard deviation of the resultant Gaussian function.

$$\lambda^2 = \sigma^2 + \gamma^2 \tag{8}$$

Since we can observe the image I, we can measure its blur (λ). The depth of each pixel on the image I is related the defocus blur (which is described with σ). We can find σ given λ (which we can measure from I) and γ (which we can measure with the calibration process described in the main paper) according to the equation below.

$$\sigma = \sqrt{\lambda^2 - \gamma^2} \tag{9}$$

We use this result in our main paper; in the defous blur calibration section.

References

[1] Maxim Maximov, Kevin Galim, and Laura Leal-Taixé. Focus on defocus: bridging the synthetic to real domain gap for depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1071–1080, 2020. 1