Extracting the Depth and All-In-Focus Image from a Focal Stack

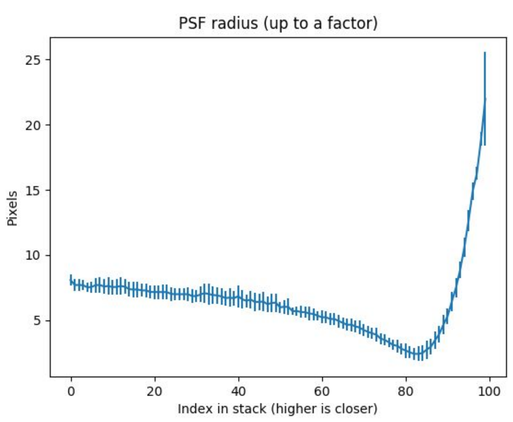
Shane Barratt and Benjamin Hannel
Stanford University
{sbarratt, bhannel}@stanford.edu

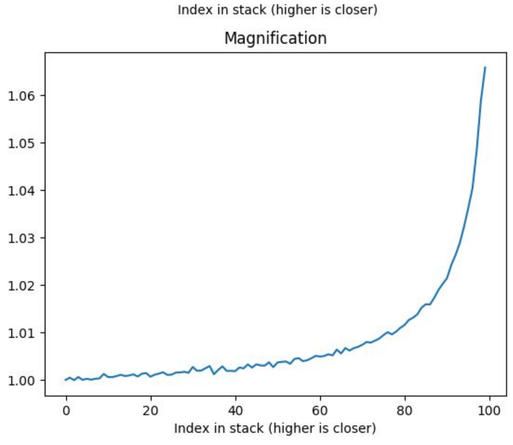
Abstract

- The captured image changes with focus distance.
- We can infer depth from images of a scene at focused at different distances.
- We model image formation, and optimize for the true scene that could have caused our captured images.

Data Capture

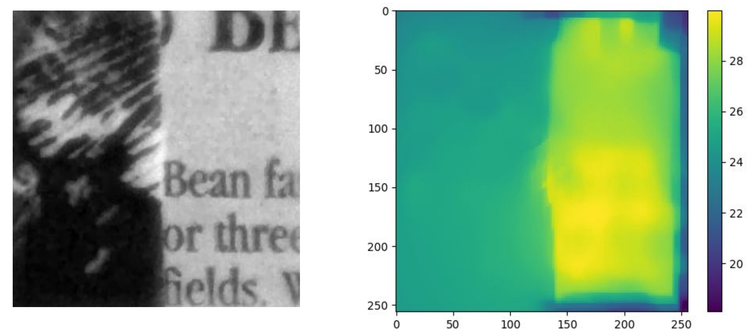
- We used Open Camera on a Samsung Galaxy S8 Active to capture focal stacks.
- We empirically determined the magnification and PSF radius





Method

Start with a guess of the all-in-focus image and the depth map



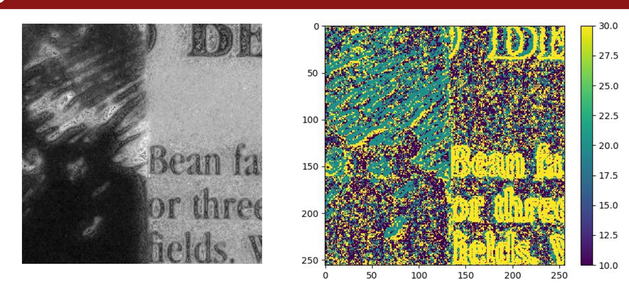
Simulate image formation to render the image at each depth

Compare with observed images at each depth and update our guess

10 cm 20 cm 30 cm

Results Against Baseline

	Depth RMSE (cm)
Our method	4.26
Maximum Gradient	10.20



Output of maximum gradient depth reconstruction

Image Formation

- We model the depth of each pixel as a probability distribution.
- We simulate image formation as a weighted sum of the simulated image at each depth.

$$(\hat{I}_i)_{xy} = \sum_{j=1}^{K} \sum_{u,v} P_{uvj} I_{uv} \mathbb{1}\left[\|(u,v)-(x,y)\|_2 \le C_i(d_j)\right]$$

Inverse Problem

$$\text{minimize} \quad L(I,P) + \lambda_1 r_I(I) + \lambda_2 r_P(P)$$

$$L(I,P) = \sum_{i=1}^{N} \ell(\hat{I}, I_i)$$

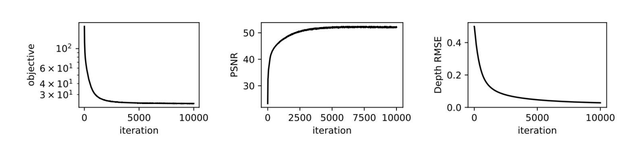
$$\ell(\hat{I}_i, I_i) = |\hat{I}_i - I_i|_2$$

$$r_I(I) = |\nabla_x I|_1 + |\nabla_y I|_1$$

$$r_P(P) = |\nabla_x P|_1 + |\nabla_y P|_1$$

Our loss function has 3 terms:

- L is total reconstruction loss
- ℓ is partial reconstruction loss
- I_i is the ith captured image
- $\hat{I}_i$ is the ith synthesized image
- r_I is the image TV prior
- r_P is the depth TV prior



(a) Objective versus iteration. (b) PSNR versus iteration. (c) Depth MSE versus iteration.

Figure 4: Simulated data optimization results.