

Dynamic Depth of Field Rendering

Junnan Liu

Stanford University

junnanl@stanford.edu

Abstract

Digital refocusing is a technique that generates photographs focused at different depths after a single camera shot and can be treated as replacing the aperture through computing image post processing. This paper will describe a method for refocusing a single image taken by traditional unmodified camera. By obtaining depth information through conventional neural network, the depth map is further used to deconvolve the blurry image and finally blur at another focal plane.

1.Introduction

Digital refocusing is an important artistic tool that can be used to emphasize the subject of a photograph. The classic technique is based on 4D light field image which includes sub-images of the scene but with the views moving 2D dimensions. This can either be done through sequential capture or spatial multiplexing or building camera arrays. However, 4D methods can be expensive and introduce complex set up. Sequential capture requires accurately moving the camera and aperture used for spatial multiplexing is costly. The recent technique of camera arrays also require costly camera devices. Since then, attempts have been made to make digital refocusing a more user friendly tool. Focus and blur a scene from a given viewpoint depends on the depth of the scene. Therefore the problem is broken into obtaining the depth map and performing image deconvolution.

1.1.Camera and Depth of Field

When we visually see an image that contains parts with soft, semi-transparent blurry edges and parts with sharp clear image. We can say there's parts that's out of focus.

The range of depth in a 3D scene that is imaged in sufficient focus through an optics system is called depth of field. In a real camera, depth of field is controlled by four quantities: the distance at which the camera lens is focused at (focal plane), the circle of confusion (c), the f-number and the focal length of the lens (f).

1.2.Depth Map

Depth estimation shows the geometric relations within a scene. Human eyes typical uses multiple visual cues in combination to estimate depth. This includes occlusion, accommodation, retinal blur, shading, texture gradients, convergence etc.

To simulate human eyes visual cues with computing image features, many successful techniques for depth estimation from stereoscopic images have been proposed. With stereoscopic images, depth can be computed from local correspondences and triangularization.

While estimating depth from based on stereo images or motion has been explored extensively, depth estimation from monocular image often arises in practice. This is because multiple images from a scene taken by traditional camera are not practically available. Estimating depths from a single monocular image is a well-known ill-posed problem, since one captured RGB image may correspond to infinite number of real world scenes, and no reliable visual cue can be obtained.

1.3.Image Deconvolution

In a single blurry image restoration problem, if the blur is spatially invariant then the output is related to the input as

$$b = c * x + n \quad (1)$$

where b is the observed blur image, x is the unknown sharp image one attempts to recover, c is the blur kernel, n is the noise and $*$ denotes convolution. The convolution process can typically be done in Fourier domain through multiplication.

When the blur kernel c is known, this issue is called an unkind deconvolution. The algorithm is more mature and includes inverse filtering, wiener filtering which adds a damping factor and ADMM method.

When the blur kernel c is unknown and the one is attempting to get both c and x at the same time, this issue is called a blind deconvolution. Although most of the problems remain challenging, several algorithm have been proposed and have great progress in image recovery.

1.4.Image Blur

Image blurring is the process of average out rapid

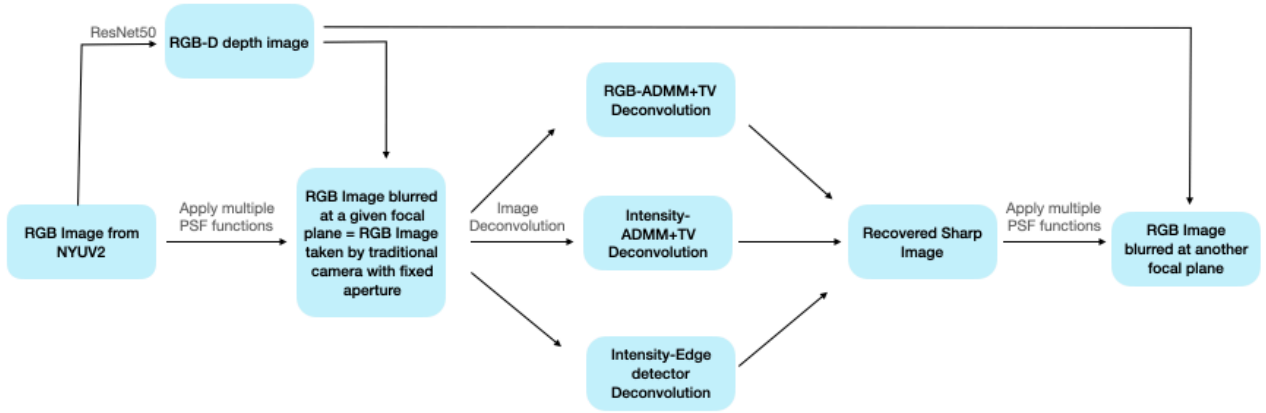


Figure 1: Refocusing Image Processing Pipeline.

changes in pixel intensity. Visually, this can be treated as a smooth color transition from one side of an edge in the image to another side. Image blurry using Gaussian filter is a mature and well-established method. Typically this can be used for denoising, abstracting the majority of the image intact and in our case, simulating the circle of confusion.

2.Related Work

2.1.Refocusing

Most existing methods of refocusing images rely on more information beyond a single 2D input image. The main technique is based on light field rendering, and that a single 2D image is the integral projection of a 4D light field [1]. Francesc, Peter and Shree perform active refocusing using depth information [2]. The defocusing method is powerful can be used on both images and videos, but the depth computing is based on the defocus of a sparse set of dots projected onto the scene which brings set up requirements that cannot be widely used. Fusion-based image enhancement methods can generate images with a set of images obtained under restricted, narrowband, imaging conditions, eg: extended depth-of-field.[3] Yosuke and Tomoyuki propose a local blur estimation method that handles abrupt blur changes at depth discontinuities in a scene. [4] Specifically, the method I'm taking here which utilize depth information and image deblur deconvolution. And some assumptions during the refocusing process is the from Yosuke's method.

2.2.Depth Estimation

For depth estimation, it originally relies on stereo vision

to reconstruct 3D shapes. [5] In single image depth estimation, the classic method is Structure-from-Motion (SfM) which leverages camera motion to estimate depth. [6] From 2015, Convolutional Neural Networks (CNNs) have been widely adopted to learn the implicit relation between RGB and Depth and multiple architecture have been built over the years to improve accuracy. [7,8]

2.3.Image Deconvolution

For image deconvolution, unblind deconvolution methods deal with a known blur kernel, the algorithm is more mature and includes inverse filtering [9], wiener filtering which adds a damping factor and ADMM method [10].

Blind deconvolution method deal with situation when the blur kernel is unknown. In equation 1, there's an infinite set of pairs (c, x) that can explain the image b . This problem is then an ill-posed problem. Although most of the problems remain challenging, several algorithm have been proposed and have great progress in image recovery. [11,12]

3.Method

3.1.Image Processing Pipeline

As shown in Figure 1, this is the image processing pipeline for refocusing a single input image. First one train a model of depth estimation with CNN. After a training model is saved, with an input image, a depth map can be obtained. (Section 3.2)

With the depth map information and single image input, together a blurry image with multiple point spread function (PSF) can be generated. This image can be treated as an image input taken by traditional unmodified camera that already comes with a certain depth of field.

below.

Step 1: Assumption 3 choose the best r in PSFs

for r in rs :

```
b = np.zeros_like(raw_img)
x_imgs = []
```

Step 2: Assumption 1 Choose the best λ ms in TV prior

for λ in λ ms:

```
x_img = np.zeros_like(raw_img)
```

Step 3: Iterate 3 RGB channels

```
for channel in range(raw_img.shape[2]):
    b[:, :, channel] = far_blur_img[:, :, channel]
    bFT = fft2(b[:, :, channel]) # each RGB, calculate bFT
    # ADMM with anisotropic TV prior
    x = np.zeros_like(depth_img.shape) # (513, 513)
    z = np.zeros((2, *depth_img.shape)) # (2, 513, 513)
    u = np.zeros((2, *depth_img.shape))
```

```
residuals = []
psnrs = []
```

Step 4: Iterate and update x , z , u

for it in $tqdm(\text{range}(n_iters))$:

```
# x-update: Fourier Domain Least Squares
# z-update: Soft thresholding
# u-update
```

```
# Residual
```

```
x_img[:, :, channel] = x
```

For intensity method, intensity channel is calculated through the equations below. After that perform the same methods as above.

```
intensity = (20 * far_blur_img[:, :, 0] + 40 * far_blur_img[:, :, 1] +
    far_blur_img[:, :, 2]) / 61
intensity += 1e-20 # important, add a small value so it's not divided
    by 0
```

For edge detection method, detailed layer is abstracted from the intensity layer. And then tuning the parameters that abstract the detailed layer after edge detection.

4.Results

4.1.Depth Estimation



Figure 3: 1) raw image 2) depth output 3) depth output after post processing

4.2.Image Blur with Multiple PSFs

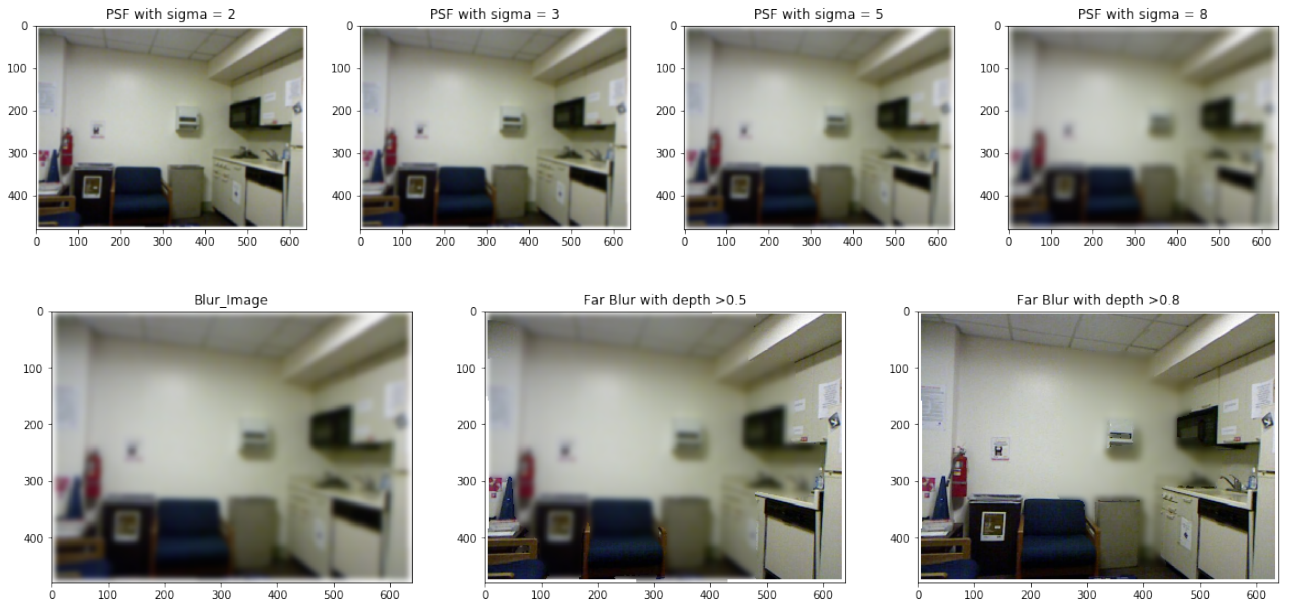


Figure 4: Top - Each figure is applied with single PSF, but multiple strength
Bottom - Each figure is applied with single PSF, but with depth input. 1) Single blur kernel 2) Blur kernel with mask applied with depth information, depth > 0.5. One can see blurred background clear table at the front 3) Blur kernel with mask applied with depth information, depth > 0.8. One can see a clear mark at the sofa.



Figure 5: Multiple blurry PSFs. Compare with Figure 4 (3), one can see that deeper area are more blurred than the edge of the sofa where it's just blurred. Parameters chosen here $\text{far_sigmas} = [2, 3, 5, 8]$, $\text{far_depth_arrays} = [0.8, 0.83, 0.86, 0.97]$

4.3. Image Deconvolution

RGB method with ADMM+TV is shown in Figure 6. Starting from this method, we explore the single radius that provides the best result and we tune the best prior for

RGB ADMM, PSNR = 20.66



Figure 6: RGB method with ADMM+TV, PSNR = 20.66. One can see that the whole image is deconvolved because we took assumption 2 that apply kernel on both blur and sharp area. However, one can see that the whole image is quite patchy. Best recover radius is 5 (Gaussian filter).

the best result as specified by PSNR.

From the output we analyze each channel's performance (Figure 7) and propose intensity channel ADMM+TV method and intensity channel edge detector method. All three methods image and PSNR output are shown in Figure 8. From the sharp image output results shown in Figure 8, method 3 with edge detector has the best PSNR result. When looking close to the top right corner of the ceiling (coordinate [400, 100]) the lines are well recovered. Another benefit is because it tries to maintain the edge information, close sharp area which is not

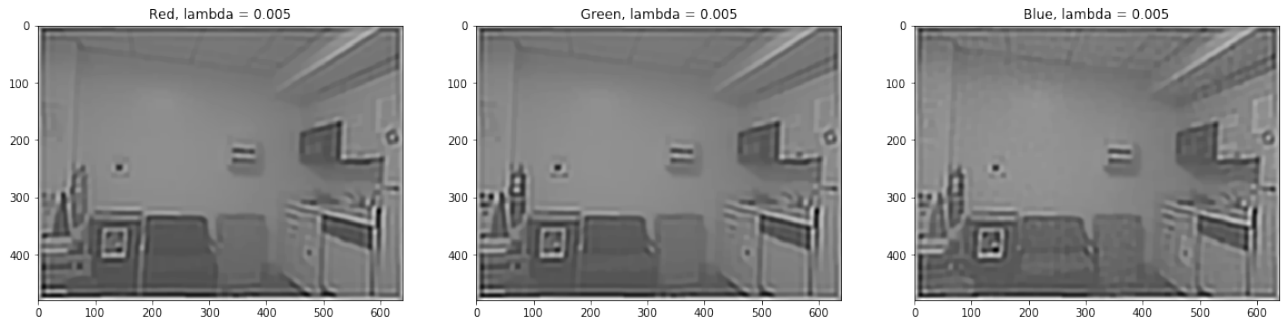


Figure 7: RGB method with ADMM+TV, when drawing 3 channels out together. One can see that red channel and green channels are sharper while blue channel is more noisy. To solve this, consider either get parameters for three channels separately (and will cause color mismatch when putting three channels back) or subtract intensity channel out.

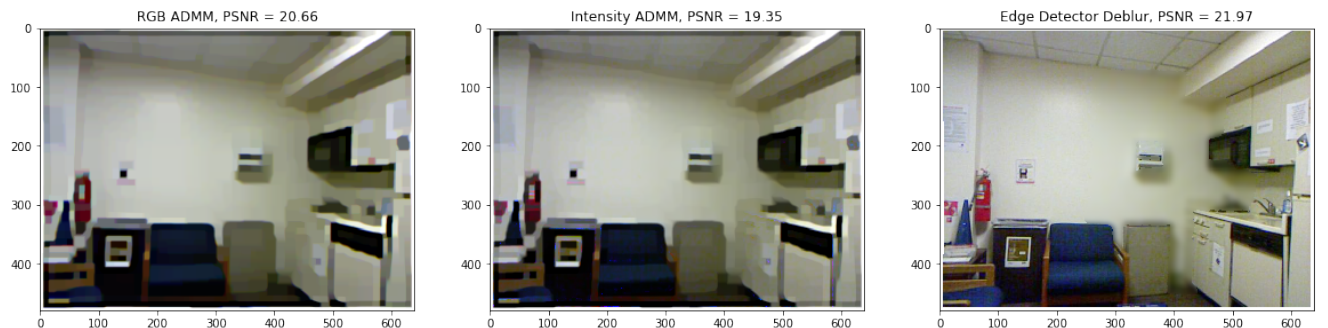


Figure 8: 1) RGB method, PSNR = 20.66; 2) Following the step in Figure 7, apply ADMM+TV method to intensity channel, the PSNR = 19.35; 3) Apply edge detector on intensity channel, PSNR = 21.97

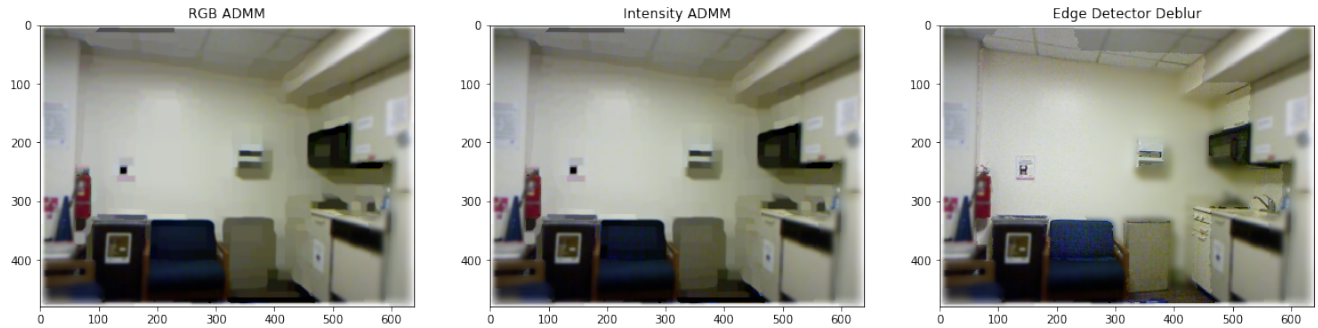


Figure 9: Close multiple PSFs blur on retrieved sharp image.

deblurred is well maintained (coordinate [100, 400]). However around the edge of the sofa where depth-of-field blurry starts, this sharp edge still maintains while the other two method patch and deblur this sofa area (coordinate [250, 350]).

Both RGB and intensity ADMM method attempts to recover the whole image including original sharp area. This can receive good results at area where there's big and less lines like microwave. ADMM method provides a smoother change from the sharp image to blurry image. However, during the iteration, one can see the method tries hard to recover small letter character (coordinate [150, 250], the poster on the wall). However, the results output are larger block of color pads.

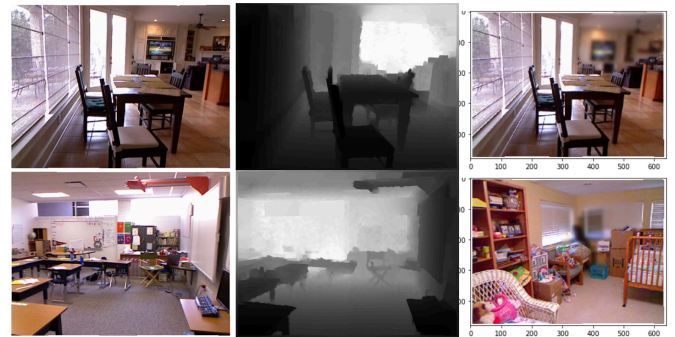


Figure 10: Apply Refocus on more images

4.4. Image Blur on the Recovered Image with Multiple PSFs

Multiple PSFs method is the same as explained in Section 3.3 and Section 4.4. Allowing the user to actively input these blur intensity and depth information enable my proposed method to be able to deal with an input image with either close blur or far blur. And output user expected depth of field. Here the results shown below is an example of employing close blur on the obtained image, shown in Figure 9.

5. Conclusions

5.1. Method Analyze and Comparison

Existing method and papers that discuss single image refocusing have been covered in Section 2.1. This method can be compared with method proposed by Y. Bando. [4] Both of these methods work with single image input and attempts to utilize depth information. Further more, one of the assumption taken in blind deconvolution in my method is employed from Y. Bando's research. The main difference is that Y. Bando's didn't get the depth map in his whole process, instead he actively use multiple radius to recover the image and choose the best result. My method only choose one radius, but utilize the depth

information to recover the image. Future work of my method would be to use both depth map and multiple radius to check if the output can be more satisfying.

5.2. Limitations

Same method is applied to multiple other images from NYUV2. Results are shown as in Figure 10. After trying out on multiple images, this proposed method has the following limitations:

It works the best on an input image from NYUV2. This is due to the limitation of CNN and machine learning. The training model is performed on NYUV2 dataset which is indoor scene. When employing this to a random image, it will show an output but won't be as accurate.

Besides that, due to the assumptions taken to simplify the blind deconvolution, ADMM method will work on the whole image instead of purely the blurry area (explained in Figure 8). Also as a result of all the assumptions, all the parameters (eg: the best radius, the best prior) shall be tuned to achieve the best results. Edge detector method is less sensitive and is the most robust method among these three deconvolutional methods.

5.3. Issues and Future Work

In the future, more areas can be dig deep and improve. This includes utilizing the label information in NYUV2 to

introduce segmentation. As shown in Figure 8, with the same blue sofa, the estimate depth map divide this sofa into two areas, thus there's a sharp circle around the sofa. This can be removed when sofa is labeled as a single object with cable segmentation information existing in NYUV2 dataset.

Also, the depth map generated is relatively distance, far away is 1 and closest one is 0. This is not linked with real image distance.

For blind deconvolution method, this is still a developing area and with more algorithms bringing up, the deconvolution results can be further improved.

References

1. R. Ng. Fourier slice photography. *ACM Trans. Graphics*, 24(3):735–744, 2005.
2. Francesc Moreno-Noguer, Peter N. Belhumeur, and Shree K. Nayar. 2007. Active refocusing of images and videos. *ACM Trans. Graph.* 26, 3 (July 2007), 67–es.
3. P. J. Burt and R. J. Kolczynski. Enhanced image capture through fusion. In *Proc. IEEE Int. Conf. Computer Vision*, pages 173–182, 1993.
4. Y. Bando and T. Nishita, "Towards Digital Refocusing from a Single Photograph," 15th Pacific Conference on Computer Graphics and Applications (PG'07), Maui, HI, USA, 2007, pp. 363-372, doi: 10.1109/PG.2007.22.
5. F. H. Sinz, J. Q. Candela, G. H. Bakır, C. E. Rasmussen, and M. O. Franz. Learning depth from stereo. In *Pattern Recognition*, pages 245–252. Springer, 2004.
6. R. Szeliski. Structure from motion. In *Computer Vision, Texts in Computer Science*, pages 303–334. Springer London, 2011.
7. D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374, 2014.
8. I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab. Deeper depth prediction with fully convolutional residual networks. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 239–248. IEEE, 2016.
9. Alex P Pentland. Depth of scene from depth of field. Technical report, DTIC Document, 1982.
10. BOYD, S., PARIKH, N., CHU, E., PELEATO, B., AND ECKSTEIN, J. 2001. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3, 1, 1–122.
11. P. Patil and R. B. Wagh, "Implementation of restoration of blurred image using blind deconvolution algorithm," 2013 Tenth International Conference on Wireless and Optical Communications Networks (WOCN), Bhopal, India, 2013, pp. 1-5, doi: 10.1109/WOCN.2013.6616234.
12. A. Levin, Y. Weiss, F. Durand and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 1964-1971, doi: 10.1109/CVPR.2009.5206815.
13. Silberman, Nathan, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. "Indoor segmentation and support inference from rgb-d images." In *European conference on computer vision*, pp. 746-760. Springer, Berlin, Heidelberg, 2012.
14. T. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
15. Y.-L. You and M. Kaveh, "Anisotropic blind image restoration," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 1, 1996, pp. 461–464 vol.2.
16. A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, 2011.