What is a Good Model for Depth from Defocus?

Fahim Mannan* and Michael S. Langer†
School of Computer Science
McGill University
Montreal, Quebec H3A 0E9, Canada
{*fmannan, †langer}@cim.mcgill.ca

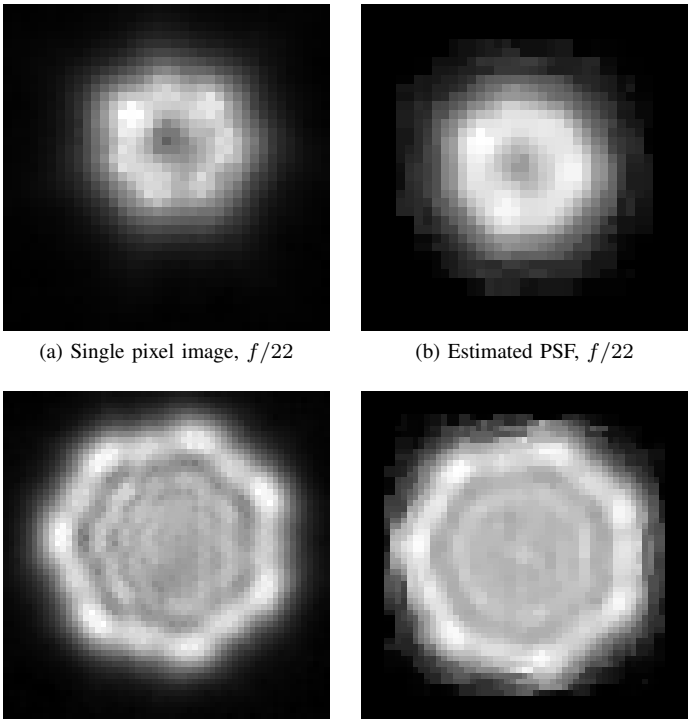
***Abstract*—Different models for estimating depth from defocused images have been proposed over the years. Typically two differently defocused images are used by these models. Many of them work on the principle of transforming one or both of the images so that the transformed images become equivalent. One of the most common models is to estimate the relative blur between a pair of defocused images and compute depth from it. Another model known as the Blur Equalization Technique (BET) works by blurring both images by an appropriate pair of blur kernels. The inverse approach is to deblur both images by an appropriate pair of blur kernels. In this paper we compare the performance of these models to find under what conditions they work best. We show that the common approach of using the Gaussian approximation of the relative blur kernel performs worse than a more general approximation of the relative blur kernel. Furthermore, we show that despite the reduction in signal content in BET, it works well in most circumstances. Finally, the performance of deconvolution based approaches depends on a large part on the shape of the blur kernel and is more appropriate for the coded aperture setup.**

***Keywords*-Depth from Defocus, Point spread functions, Relative Blur, Blur Equalization Technique, Deconvolution**

I. Introduction

Images taken with a finite aperture camera can have regions that are out of focus. These regions correspond to scene points that are outside the focal plane of the lens. The amount of defocus blurs in these regions will vary with depth and by measuring the defocus blur we can measure depth. Different methods have been proposed for estimating depth from defocused images. Most of the methods use a pair of defocused images where two images of the same scene are captured with different camera parameters. All of these methods rely on modelling the relationship between the observed defocused images.

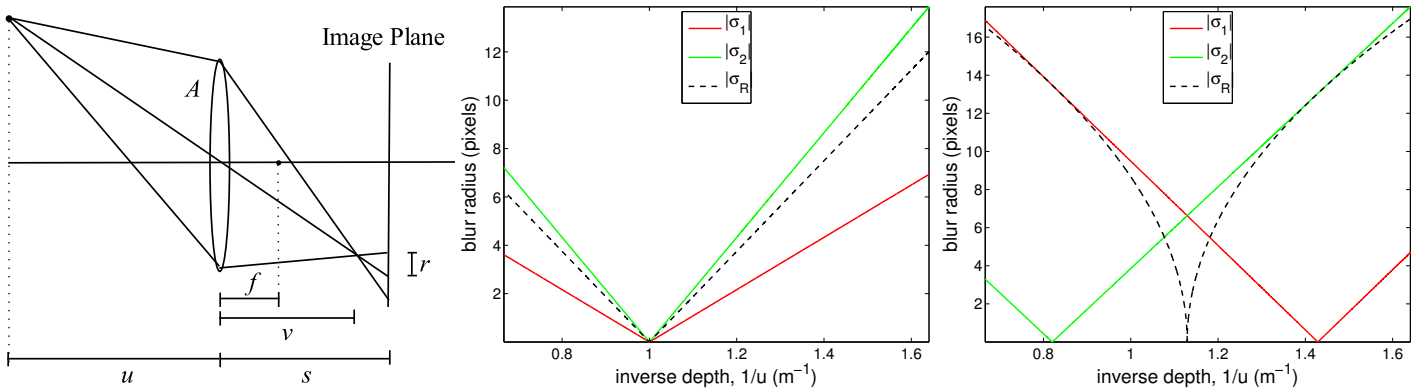
In this paper we evaluate three of these DFD models. The models considered are the widely used relative blur model, blur equalization model, and deblurring based model. Most DFD models including the ones evaluated in this work are based on the assumption that the camera is a linear shift-invariant system (LSI). In such systems the input output relationship can be modelled using the system's impulse response which for optical systems is the point spread function or the image of a point light source as seen by the camera. For real cameras these PSFs are the result of defocus, diffraction, lens aberration, and sensor properties. Defocus affects the scale of the PSF and diffraction, lens aberration, aperture-stop affect the shape and smoothness of



(a) Single pixel image, $f/22$ (b) Estimated PSF, $f/22$

(c) Single pixel image, $f/11$ (d) Estimated PSF, $f/11$

Figure 1: Examples of PSFs extracted from a single pixel (left column) and estimated PSFs (right column) from our calibration process. Object to sensor distance is 0.9 m and focus distance 0.5 m. The depression in the middle of (a, b), the ringing in (c, d) and also the brightening near the corners, and smooth edges are the effects of diffraction. In general the PSF is the result of defocus (i.e. depth dependent scaling from Eq. 2), aperture shape, diffraction, different optical aberrations (e.g. chromatic, coma, etc.), and sensor (e.g. pixel size and color filter array pattern). Our PSF estimation process – described in more details in [1] – can capture these subtle effects.



(a) Defocus Formation (b) Variable Aperture (2A) (c) Variable Focus (2F)

Figure 2: a) Defocus blur formation. b-c) Absolute blur $|\sigma_1|$ (red), $|\sigma_2|$ (green) and relative blur σ_R (dashed) from Eq. (8) versus inverse depth u^{-1} in diopters (D) i.e. m^{-1}. b) Variable Aperture (2A) configuration, where the two images have the same focal length and focused at the same distance but different f-number, and c) Variable Focus (2F) configuration, where the two images have the same focal length and f-number but focused at two different depths.

the PSF. DFD models measure the scale of the PSF which is the only depth dependent component of the PSF (e.g. diffraction creates ringing artifacts which only depends on the aperture size and wavelength).

The most commonly used model assumes that the blurrier image can be obtained by convolving the sharper image with a *relative blur* kernel. This is an exact model in the case of Gaussian point spread functions. For non-Gaussian PSFs this is an approximation. We consider two variations of this model where the relative blur kernel is either estimated using general smoothness constraints on the kernel or approximated as a Gaussian. The blur equalization model makes use of the commutative property of convolution and does not make any assumption about the type of the PSF. Deconvolution based models such as the Generalized Wiener Deconvolution used by Zhou et al. [2], try to estimate the true sharp image and choose depth based on the reconstruction error.

It should be noted that we do not try to model the PSF or evaluate what is a good general shape for the PSF. In fact the shape of the PSF can have large variability and cannot be represented by a single model (see Fig. 1). Instead our main contribution is to evaluate different DFD models under different PSF shapes and camera parameters.

The paper is organized as follows. Sec. II gives some of the necessary background for DFD. In Sec. III we give an overview of the models that are evaluated in this paper and discuss their strengths and limitations. Sec. IV presents the experimental evaluation of these models. Finally, Sec. V concludes the paper with some possible applications of this work.

II. Defocus Blurred Image Formation

In this section we first describe the geometric model for depth dependent scaling and then discuss how that relates to the LSI model of the PSF.

Fig. 2a shows how a scene point at distance u is imaged by a lens of focal length f and aperture diameter A. Light rays emanating from the scene point fall on the lens and converge at distance v on the sensor side of the lens. The relationship between these variables is specified by the thin lens model as:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \ . \tag{1}$$

If the imaging sensor is at distance s from the lens then the imaged scene point creates a circular blur pattern (as shown in Fig. 1, the exact shape is more complex) of radius r as shown in the figure. The thin lens model (Eq. 1) and similar triangles from Fig. 2a give the radius of the blur in pixels:

$$\sigma = \rho r = \rho \frac{fs}{2N}\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right). \tag{2}$$

In the above equation the ratio of focal length (f) and f-number (N) is used instead of the aperture (i.e. $A = f/N$). The variable ρ is used to convert from physical to pixel dimension. In the rest of this paper we will use σ to denote blur radius in pixels.

Note that the blur can be positive or negative depending on which side of the focus plane a scene point resides. For circularly symmetric aperture the sign of the blur has no effect in the blurred image formation process. In practice the blur is not necessarily circularly symmetric (Fig. 1) and can be used to disambiguate the sign of blur (e.g. [3] and Fig. 6b and 6c).

If the scene depth is nearly constant in a local region, then an observed blurred image i, can be modelled as a

convolution of a focused image i_0, with a depth-dependent point spread function (PSF) $h(\sigma)$.

$$i = i_0 * h(\sigma) \tag{3}$$

In the case of a circular aperture the PSF kernel will be a pill-box function (i.e. a cylindrical shaped function) with radius σ. However, it has been shown that due to lens diffraction and other unmodelled characteristics of the optical system the blur kernel is similar to the Bessel function [4]. To simplify the depth estimation model one commonly used shape is a Gaussian. When the PSF is modeled as a Gaussian the spread of the Gaussian is, $\sigma_G = \sigma/\sqrt{2}$. However in practice as shown in Fig. 1 it depends on the the shape and size of the aperture as well as diffraction and lens aberration. The main goal of this paper is to evaluate different DFD models under such varying PSF.

The models evaluated in this paper take a pair of differently defocused images (i.e. captured using different optical parameters) for depth estimation. The most common settings for taking two defocused images are the variable aperture (2A) and variable focus (2F) configurations. These configuration are shown in Fig. 2b and Fig. 2c respectively. In the variable aperture case both camera settings have the same focus distance and the apertures are varied. In the variable focus case both camera settings have the same aperture and the focus distances are varied. There are many other possible settings and the relationship between the settings can give insights into the accuracy of different models. More details on the optimal camera settings can be found in [5].

III. Models used for DFD

The main objective in DFD is to estimate the blur radius at each pixel and in turn estimate depth using Eq. 2. Many different methods were proposed for this purpose using one or more images, ranging from blur estimation from an image of a step edge [6], [7], to using a pair of focused and defocused image [6], to multiple defocused images e.g. [4], [8]. In this work, we address the most common case of using a pair of defocused images for depth estimation. In the rest of this section, we discuss how a pair of defocused images is used by the three models – the relative blur model, BET and deconvolution based model – to estimate depth. We highlight their underlying assumptions, approximation ability, advantages and disadvantages.

A. Relative Blur Model

The relative blur model aims to find the amount by which the sharper image is blurred to obtain the blurrier image. This model has been used both in the frequency and in the spatial domain [6], [9], [10]. The spatial domain relative blur model was first proposed by Ens and Lawrence in [11]. They estimated the relative blur kernel from two observed images. A simplified version of it assumes Gaussian kernel and knowing the camera setting is sufficient for finding the corresponding depth. This approach is used in the diffusion based variational methods [8], [12].

In general, the blur in corresponding regions of two defocused images will vary with the depth and camera parameters. Therefore a region will either be blurrier in one image but sharper in the other or vice versa. For example in the variable focus example shown in Fig. 2c, a point at 1.4D is sharper in one image (red) and another point at 1D is sharper in the other (green). To simplify the explanation of the relative blur model, let us assume a fronto-parallel plane being imaged with two different camera parameters. In this case, one of the images will be sharper (i_S) and the other blurrier (i_B). The observed image can be modeled as convolution of the hypothetical sharp image i_0 with an appropriate PSF h and additive noise. Therefore the observed sharper and blurrier images are:

$$i_S = i_0 * h_S + n_S \tag{4}$$

$$i_B = i_0 * h_B + n_B. \tag{5}$$

where n_S and n_B are additive noise.

Let h_R be the relative blur which is the amount by which the sharper (h_S) PSF is blurred to get the blurrier (h_B) PSF, that is,

$$h_B \approx h_S * h_R. \tag{6}$$

For Gaussian PSFs the above equation holds with equality. The key assumption made in the model is that,

$$\begin{aligned} i_B &\approx i_S * h_R \\ \text{or, } n_B &\approx n_S * h_R \end{aligned} \tag{7}$$

The equation holds with equality when $n_B = 0$ and h_R is large.

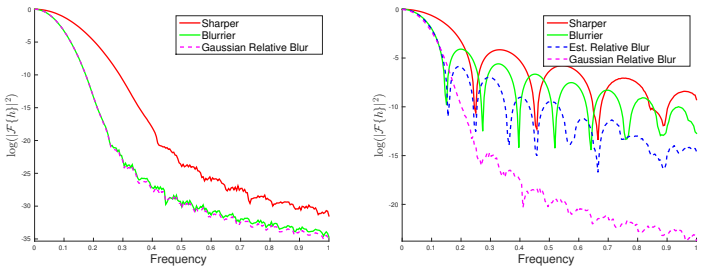
Let the blur radii of the kernels h_B and h_S be σ_B and σ_S respectively. The radius σ_R of the relative blur h_R in Eq.6 is:

$$\sigma_R \equiv \sqrt{\sigma_B^2 - \sigma_S^2}\,. \tag{8}$$

The relative blur estimation problem can be written as the following optimization problem:

$$\underset{h_R}{\operatorname{argmin}} \|i_B - i_S * h_R\|_2^2 \tag{9}$$

Eq. 9 can be solved using least squares. However, this assumes we know in advance which image is the blurrier one at each pixel. In the more general case given two defocused images i_1 and i_2, we would have to decide which one to blur and by what amount (i.e. σ_R). In such cases, it is more convenient to use signed relative blur. We can choose the sign of σ_R to be negative when $|\sigma_1| > |\sigma_2|$ and positive otherwise. Therefore when the sign of relative blur is negative $\sigma_B = \sigma_1$ and $\sigma_S = \sigma_2$ and vice-versa when the sign is positive. This idea is used when fitting a Gaussian relative blur in the real experiments for this paper.



(a) Gaussian PSF (b) Pillbox PSF

Figure 3: Reconstruction accuracy shown using the log power spectrum ($\mathcal{F}\{.\}$ is the Fourier transform) for sharper, blurrier and the reconstructed PSF (dashed). a) Relative blur reconstruction is best when the PSFs are Gaussian. b) For pillbox (or any arbitrary) PSF our estimated kernel gives a better approximation than the Gaussian relative blur kernel.

In this work we estimate the relative blur between two defocused images by solving the following Quadratic Programming (QP) problem.

$$\begin{aligned} \underset{h_R}{\operatorname{argmin}} & \sum_{j=1}^{n} \lambda_j \| f_j * (i_S * h_R - i_B) \|_2^2 \\ & + \lambda_{n+1} \| \nabla h_R \|_2^2 \\ \text{subject to } & \|h_R\|_1 = 1\,,\ h_R \geq 0. \end{aligned} \tag{10}$$

In the above optimization problem, i_B is the observed blurrier image and i_S is the sharper image. The sharper and blurrier images can be the absolute PSFs or texture images or both. h is the PSF kernel that is to be estimated. f_j is a filter that is applied to the images. In the experiments, we use $f_1 = \delta$, $f_2 = G_x$, and $f_3 = G_y$, where G_* is the spatial derivative of a Gaussian in the horizontal and vertical directions. The constraints ensure that the kernel is non-negative and preserves the mean intensity after convolution. The optimization function is more general than the one proposed by Ens and Lawrence in that ours support asymmetric relative blur kernel. Furthermore the additional image derivative, non-negativity and unity constraints results in a robust and more accurate model of relative blur. The QP formulation also allows us to use fast off-the-shelf QP solvers (in our case Matlab's `quadprog`).

For Gaussian approximation of relative blurs, we consider the signed relative blur. For a given signed relative blur corresponding to the spread of the Gaussian, we choose the sharp image based on the sign of the hypothetical relative blur. Then the assumed sharp image is synthetically blurred by the appropriate Gaussian and compared with the assumed blurred image. The signed relative blur that minimizes the sum-of-squared error between the hypothetical blurred and sharp images is chosen to be the spread of the Gaussian.

The relative blur model requires a single convolution operation for every depth hypothesis. This is the least amount of blurring that can be performed to generate the blurred image. Therefore there is minimal loss in the signal component. Fig. 3a and 3b show the reconstructed blurrier PSF using the relative blur method. For Gaussian PSF pair in Fig. 3a the reconstructed blurrier PSF is almost exact. However real PSFs are not Gaussian and can have sharp edges e.g. in coded aperture. While this can be modelled by finding the closest relative blur approximation, using a Gaussian approximation will result in large error. Fig. 3b shows an example of relative blur using pillbox PSF pair. A simple Gaussian approximation does not work as well as the estimated relative blur from Eq. 10.

B. Blur Equalization Technique

The Blur Equalization Technique (BET) was proposed by Xian and Subbarao in [13] for application in the S-transform based relative blur estimation. More recently it has been used for motion deblurring and optical flow [14]. It works on the same principle as blurring one image to get the other, but here both images are blurred so that they produce the same result. The commutative property of convolution is used to equalize the blurs in the observed image pair:

$$\begin{aligned} i_1 * h_2 &= (i_0 * h_1 + n_1) * h_2 \\ &= (i_0 * h_2) * h_1 + n_1 * h_2 \\ &\approx i_2 * h_1 \end{aligned} \tag{11}$$

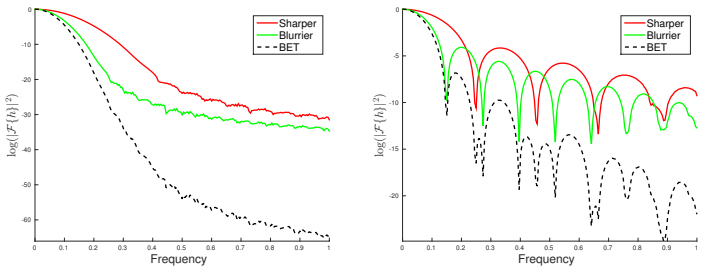
The assumption made here is that,

$$n_1 * h_2 \approx n_2 * h_1 \approx 0 \tag{12}$$

This assumption does not hold when exactly one of the images is in focus. However if we blur both images by a small amount then we can expect the assumption to be valid. The approximation error is small when both images have large blur. But in that case, the signal content of the images is also reduced which in theory will result in larger variance in the depth estimates.

One of the advantages of this model is that it only requires the absolute blur PSFs at every depth for a given camera setting. This makes it more scalable in practice, since we do not have to consider every possible pair of camera settings as in the relative blur model. Furthermore, BET correctly models the blurring process for any type of PSF kernels. The main disadvantage is that BET requires synthetically blurring both images. This reduces the signal content in both images, which in turn increases the variance in depth estimation. Furthermore Eq. 12 suggests that there will be approximation error when a scene point in one of the images is within the depth-of-field.

Finally, BET can have problem for certain type of PSFs. For instance Fig. 4a shows that for Gaussian PSFs, applying BET results in rapid decrease in the higher frequency components. This will result in higher variance in the estimate.



(a) Gaussian BET (b) Pillbox BET

Figure 4: Log power spectrum for applying BET for a) Gaussian and b) pillbox PSFs. For Gaussian PSFs, BET causes faster decay in higher frequency components.

On the other hand for pillbox PSFs (Fig. 4b) the rate of decrease in frequency is less significant.

C. Deconvolution based Model

Deconvolution based models work on the same principle of equalizing two defocused images. Here the goal is to find the latent sharp image. Every depth has a pair of PSFs associated with it. The correct PSF pair will produce a sharp image with the lowest reconstruction error. In this paper we consider the Generalized Wiener Deconvolution reconstruction cost used by Zhou et al. [2]. The cost function used in this model is a general version of Wiener deconvolution. It uses the pair of images to estimate the sharp image. The deconvolution model is specified in the Fourier domain as follows:

$$\underset{d}{\operatorname{argmin}} \sum_{j} ||\hat{I}_0 H_j(d) - I_j||^2 + ||C\hat{I}_0||^2 \qquad (13)$$

where, I_j is the Fourier transform of the j^{th} defocused image and $\hat{I}_0$ is the Fourier transform of the latent sharp image that is being estimated. The goal is to find the depth dependent blur kernel $H_j(d)$ in the Fourier domain that explains the observed blurred images. In the last term, C is a prior derived from the average of multiple natural images. There are other variations of this model that use more complex natural image priors [15]. It should be noted that, in practice, the prior is only used during the latent image estimation step. For depth estimation, only the data cost is used.

Similar to BET, this model requires linear number of PSF calibration steps. The model is applicable for single image depth estimation and can also estimate the latent sharp image. However this method requires solving an inverse problem by finding the depth hypothesis that produces the best sharp image. In the frequency domain deconvolution is equivalent to dividing the observed blurred image with the PSF. For some PSFs there can be zeros in the frequency spectrum (e.g. pillbox in Fig. 3b) of the PSFs or in some cases rapid decrease in the high frequency components (e.g. Gaussian in Fig. 3a). Having zeros in the frequency spectrum can help in recovering depth but not the sharp image (because wrong PSF hypothesis will result in large error) whereas having small but non-zero frequency components can result in large ambiguity because nearby depths will produce the same deblurred image. Therefore deconvolution based models require carefully choosing the shape of the PSF and can also be sensitive to the type of texture in the scene.

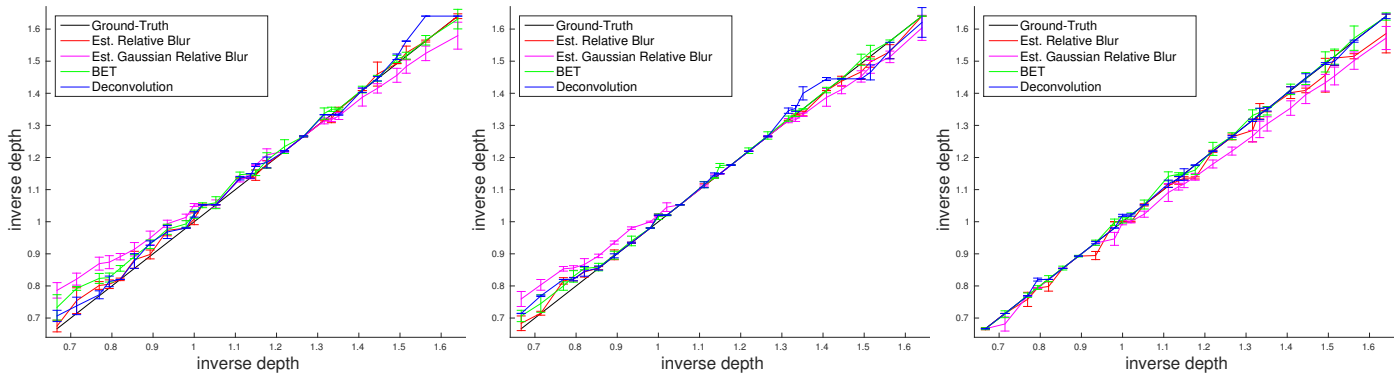
IV. Experimental Evaluation

A. Real Defocus

Calibration: As shown in Fig. 1 real PSFs are not just a scaled version of the apeture where the scaling is determined by Eq. 2. They are the result of defocus blur, diffraction, optical aberrations, and the sensor. Since the objective of this work is to evaluate the DFD models under such widely varying PSFs, for the real defocused experiments, we need to calibrate the absolute and relative blur PSFs for different depths and camera settings. For this we use a grid of disks with known radius and disk center distance as the calibration image. The main idea here is to use Eq. 10 with i_S as the latent sharp image of a disk and i_B the observed defocus blurred image of the disk to infer the blur kernel. We use a 24 inch LED display of resolution 1920×1200 for rendering both the textures and the calibration pattern. The images are captured in raw linear format using a Nikon D90 camera with a 50 mm prime-lens, under varying focus and aperture settings, and placing the camera fronto-parallely to the display at different distances. The processing pipeline includes first radiometrically correcting the raw images, performing magnification correction and alignment, and normalizing the intensity of the images by their mean. Our approach for estimating PSFs from disk images is robust to display pixel size and camera configuration. Further details can be found in [1].

DFD Evaluation: To evaluate the DFD models we first calibrate the PSFs for 27 object-to-sensor distances spaced uniformly (roughly) in inverse depth space ranging between 0.61 m to 1.5 m. We choose uniform subdivision in inverse depth space because blur changes linearly with inverse depth (recall Eq. 2). In practice, the relationship is approximately linear because we are moving the lens instead of the sensor. We capture images of a $1/f$ texture pattern under various defocus conditions. For a pair of input images depth estimation is performed by choosing the appropriate PSFs for every depth hypothesis and evaluating the model cost. The per-pixel cost is then averaged over a finite window and the depth label is chosen to be the one that minimizes the cost at every pixel.

The depth from defocus results are shown in Fig. 5. Figures 5a and 5b correspond to variable focus setting with $f/11$ with focus distances being 0.61 m and 1.5 m,



(a) 0.61 m and 1.5 m, $f/11$ (b) 0.7 m and 1.22 m, $f/11$ (c) Focused at 1.5 m $f/11$ and $f/22$

Figure 5: Results from real defocused images of a texture.

and 0.7 m and 1.22 m. The theoretical blur radii for Fig. 5b corresponds to Fig. 2c. Since Fig. 5a focuses at the scene boundary the maximum blurs are larger than Fig. 2c. Fig. 5c corresponds to variable aperture with $f/22$ and $f/11$ and focus distance set to 1.5 m. In all cases the Gaussian relative blur approximation has the most error and sometimes deviating from the true depth by about two standard deviations. When both PSFs are large e.g. near 1.6 D for variable aperture (Fig. 5c), relative blur estimation has more approximation error and as a result the estimates start to deviate from the true value. The deconvolution based method is more sensitive than the other models in that it sometimes deviates abruptly from the true value (e.g. near 1.6 D in (a) and 1.4 D in (b)). Among all the models BET usually has the least amount of estimation error.

B. *Synthetic Defocus*

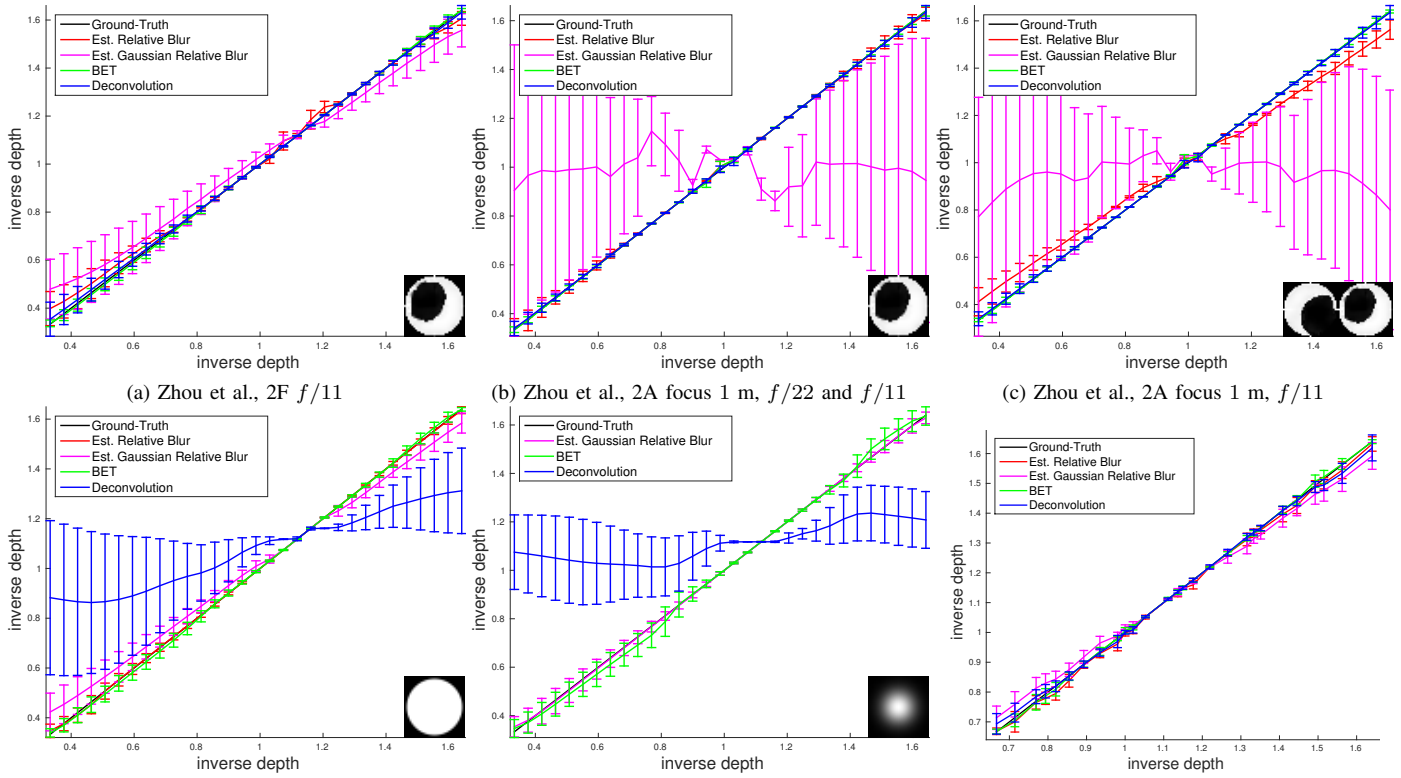
For generating synthetically defocused images, we first generated a discrete set of PSF kernels of different size and shape and their corresponding relative blur kernels or Gaussian approximations of them. The PSF kernels are scaled and rotated based on the chosen camera and scene parameters. In all the experiments we have $f = 50$ mm and $\rho = 180$ pixels-per-mm. The scene is considered to be within 0.61 m to 3 m. A 512×512 image of $1/f$ noise texture is synthetically blurred with the PSF pair that is being used for evaluation. This was followed by adding additive Gaussian noise $\sigma_n = 2\%$ to the blurred images. Finally, depth estimation was done using the relative blur model with estimated relative blur kernel, the Gaussian relative blur approximation, Blur Equalization Technique (BET), and the deconvolution model using the generalized Wiener deconvolution algorithm.

For the synthetic experiments we tried to highlight the limitations of different models. Therefore we choose PSFs of different shapes with different camera configurations. For PSFs we used Zhou et al's coded aperture [2] with both variable focus and variable aperture configuration, pillbox and Gaussian PSFs with variable focus configuration.

The PSFs for the coded aperture was simulated by first computing the signed blur radius at a given depth for our chosen camera settings and then appropriately scaling and flipping the aperture pairs. We then estimate the relative blur using Eq. 10 and also find its Gaussian approximation. For our calibrated PSFs we simply take the observed absolute PSFs at a given depth and estimate their relative blur.

In the variable focus setting the camera is focused at 0.7 m and 1.22 m i.e. within the scene rather than at the boundary similar to Fig. 2c. For the variable aperture setting with the coded aperture we used a slightly unconventional setting in that the camera is focused within the scene (i.e. similar to Fig. 2b) which in our case is at 1 m. Usually with variable aperture the focus is at one end of the scene. In the case of symmetric PSF such as Gaussian or pillbox this circumvents the problem of detecting the sign of the blur in Eq. 2. But with asymmetric coded aperture pair we do not have any sign ambiguity [5].

Results for the synthetic experiments are shown in Fig. 6 with (a)–(e) using simulated PSFs using the apertures shown in the figure inset and (f) using estimated PSFs from real defocused images. The top row contains experiments with Zhou et al's coded aperture. In the first two cases defocused image pairs are simulated assuming a fixed aperture shape. In the 2F case (Fig. 6a), the Gaussian approximation of relative blur (magenta curve) performs worse than estimated relative blur. Fig. 6b shows variable aperture with aperture sizes $f/22$ and $f/11$. When focused within the scene, the Gaussian relative blur approximation fails because it is symmetric and cannot differentiate whether a scene point is in front or behind the focal plane (for asymmetric aperture the image of the PSF will be different in the two cases). In contrast the estimated relative blur can differentiate between the two sides because the estimated relative blur is asymmetric. In Fig. 6c we used two apertures of the same size but



(a) Zhou et al., 2F $f/11$ (b) Zhou et al., 2A focus 1 m, $f/22$ and $f/11$ (c) Zhou et al., 2A focus 1 m, $f/11$

(d) Pillbox, 2F $f/11$ (e) Gaussian, 2F $f/11$ (f) Est. PSF, 2F $f/11$

Figure 6: Synthetically defocused image with different PSFs, camera settings, and noise $\sigma_n = 2\%$. The blur radii for variable focus (a, d–f) and the variable aperture in (b) correspond to Fig. 2c and 2b. Examples of the estimated relative blurs can be found in http://cim.mcgill.ca/~fmannan/relblur.html

different shape. In this case Gaussian approximation fails as expected and estimated relative blur has more approximation error resulting in larger bias in the estimates than the first two cases.

Figs. 6d and 6e show experiments with symmetric aperture with variable focus. In both cases deconvolution based depth estimation (blue curve) fail to accurately estimate depth. Also as expected the Gaussian approximation performs the best for Gaussian PSF. Fig. 6f uses the estimated PSFs from the real experiments. In this case all models perform reasonably well with Gaussian approximation of relative blur having slightly larger error than the other. The result in this case is similar to the real experiments indicating that the simulation process does not produce any significant artifacts. It should be noted that in all cases BET performs consistently well. Also except when the relative blur estimation problem is complex e.g. differently shaped aperture with fixed size, estimated relative blur has acceptable accuracy.

V. Conclusion

In this paper we evaluated three different models for DFD, namely, the relative blur model with estimated relative blur kernel and its Gaussian approximation, BET, and deconvolution. We discussed the advantages and limitations of these models in terms of scalability and accuracy, and experimentally evaluated their performance under different noise and PSF kernels. We showed that BET works well despite blurring both images and is robust when both images are initially blurred by a slight amount. The relative blur model works well for different sizes and shapes of relative blurs. The Gaussian relative blur approximation works well mainly for small relative blurs and when the aperture has a simple shape. We found the general version of Wiener deconvolution to perform similar to the estimated relative blur and BET.

The relative blur model with Gaussian approximation has been widely used in practice and more often than Ens and Lawrence's approach. This is primarily because the Gaussian approximation makes the estimation problem easier and depth estimation can be solved in the continuous domain e.g. using variational methods. However as we showed in this paper, even under standard settings the Gaussian approximation does not always give the best result. The conclusions we can make are as follows. If relative blur model is used then it should be approximated using Eq. 10 or some

similar method. If the Gaussian relative blur approximation is used then one needs to ensure that the camera settings are suitable for such approximation. In general BET is the simplest to use. In fact BET is less affected by different PSF shapes and camera configurations. The simple Generalized Wiener Deconvolution has comparable performance but it is sensitive to the shape of the PSF and can be affected by errors in the PSF estimation, image noise and weak texture. Relative blur based model is not very good when the PSFs have complex shape or when it is difficult to estimate the relative blur. However it is still much better than Gaussian approximation of the relative blur which is widely used in the literature.

For real images there are many factors that can dominate depth estimation accuracy. For instance for sparse or low contrast textures all models will perform poorly. Furthermore, it is also possible to choose the camera parameters so that the modelling errors have less influence. For instance large relative blurs are likely to exacerbate the relative blur approximation error. Therefore in that case we can choose camera parameters that reduces the relative blur between the two defocused images. The goal of this paper was to highlight the modelling errors and experimentally verify using both synthetic and real defocused images. We believe this would help DFD users to pay attention to the different issues and help them choose the appropriate model, accurately and easily estimate PSFs and appropriately choose the camera settings.

ACKNOWLEDGEMENTS

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). Computations were performed on the HPC platform Guillimin from McGill University, managed by Calcul Qubec and Compute Canada. The operation of this compute cluster is funded by the Canada Foundation for Innovation (CFI), NanoQubec, RMGA and the Fonds de recherche du Qubec - Nature et technologies (FRQ-NT).

REFERENCES

[1] F. Mannan and M. S. Langer, "Blur calibration for depth from defocus," in *CRV*, 2016.

[2] C. Zhou, S. Lin, and S. Nayar, "Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring," *IJCV*, vol. 93, no. 1, p. 53, May 2011.

[3] A. Sellent and P. Favaro, "Which side of the focal plane are you on?" in *ICCP*, May 2014, pp. 1–8.

[4] S. Chaudhuri and A. N. Rajagopalan, *Depth from defocus - a real aperture imaging approach*. Springer, 1999.

[5] F. Mannan and M. S. Langer, "Optimal camera parameters for depth from defocus," in *3DV*, Oct 2015, pp. 326–334.

[6] A. P. Pentland, "A new sense for depth of field," *PAMI*, vol. 9, pp. 523–531, July 1987.

[7] T.-L. Hwang, J. Clark, and A. Yuille, "A depth recovery algorithm using defocus information," in *CVPR*, 1989, pp. 476–482.

[8] P. Favaro and S. Soatto, *3-D Shape Estimation and Image Restoration - Exploiting Defocus and Motion Blur*. Springer, 2007.

[9] M. Subbarao, "Parallel depth recovery by changing camera parameters," in *ICCV*, dec 1988, pp. 149–155.

[10] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *IJCV*, vol. 13, no. 3, pp. 271–294, 1994.

[11] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *PAMI*, vol. 15, no. 2, pp. 97–108, 1993.

[12] P. Favaro, S. Soatto, M. Burger, and S. Osher, "Shape from defocus via diffusion," *PAMI*, vol. 30, no. 3, pp. 518–531, March 2008.

[13] T. Xian and M. Subbarao, "Depth-from-defocus: Blur equalization technique," *SPIE*, vol. 6382, 2006.

[14] T. Portz, L. Zhang, and H. Jiang, "Optical flow in the presence of spatially-varying motion blur," in *CVPR*, June 2012, pp. 1752–1759.

[15] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.