On Defocus, Diffusion and Depth Estimation

Vinay P. Namboodiri and Subhasis Chaudhuri

Vision and Image Processing Laboratory
Department of Electrical Engineering.
Indian Institute of Technology-Bombay
Mumbai - 400076. INDIA.
(vinaypn,sc)@ee.iitb.ac.in

Abstract

An intrinsic property of real aperture imaging has been that the observations tend to be defocused. This artifact has been used in an innovative manner by researchers for depth estimation, since the amount of defocus varies with varying depth in the scene. There have been various methods to model the defocus blur. We model the defocus process using the model of diffusion of heat. The diffusion process has been traditionally used in low level vision problems like smoothing, segmentation and edge detection. In this paper a novel application of the diffusion principle is made for generating the defocus space *of the scene. The defocus space is the set of all possible observations for a given scene that can be captured using a physical lens system. Using the notion of defocus space we estimate the depth in the scene and also generate the corresponding fully focused equivalent pin-hole image. The algorithm described here also brings out the equivalence of the two modalities, viz. depth from focus and depth from defocus for structure recovery.*

1 Introduction

Structure recovery from images has been one of the important goals in computer vision. Many techniques have evolved to that effect which make use of cues like stereo, shading and defocus. Here we are concerned with the use of the defocus cue. When an image is in focus, knowledge of the camera parameters can be used to estimate the depth of the object point. When the image is defocused, the structure can be recovered through an estimation of the defocus blur. There have been various ways in which the problem of structure recovery have been solved using the defocus cue. Here we consider using the process of diffusion for modeling the blurring process.

The idea of diffusion has been one of the important methodologies in the field of computer vision. It stems largely from the idea of modeling the image (observation) generation process using the heat equation. The pioneering work was done by Witkin in [1] where he proposed a scale space for images based on smoothing of images using a Gaussian kernel. Koenderink in [2] proved that this was equivalent to solving the heat equation. This approach has subsequently been widely used in low level vision tasks like smoothing, segmentation and edge detection.

In this paper we discuss how the linear diffusion principle can be used for depth estimation based on defocus as the cue. In depth estimation using defocus as the cue, the basic principle is to use the characteristics of the imaging system. There have been two methodologies in the literature, one is to obtain depth from focus [3] and the other to obtain depth from defocus [4].

In the procedure for obtaining depth information from focus, a sequence of images of a scene is obtained by continuously varying the distance between the lens and the image detector [5]. The corresponding fully focused observation is locally estimated from the sequence of images. A measure of image sharpness is used to decide whether the point is in focus or not. From the fully focused image point the distance of the corresponding object point is calculated using the standard lens equation $1/f = 1/u + 1/v$ where f is the focal length, u is the distance of the object from the principal plane and v is the distance of the focused image from the lens plane. For a good ranging accuracy, one is required to sample the observation space densely by changing v slowly.

When a point light source is in focus, all light rays that are radiated by the object point and intercepted by the lens converge at a point on the image plane. When the point light source is not in focus, its image on the image plane is not a point, but a circular patch resulting in a blur as can be seen from fig (1). In depth from defocus, given two images of a scene recorded with different camera settings, one obtains an estimate of the blur at each point [4]. Subsequently, by using the estimate of the blur, one can recover the depth information in the scene with the knowledge of the lens parameters.

In this paper we show that, given two observations obtained by two sets of lens parameters as is commonly employed in depth from defocus technique; we can generate the entire set of images in the *defocus space* of the input images using the diffusion equation. The defocus space of a particular scene refers to the continuous space of all possible observations obtainable by varying the lens parameters in between those two lens settings. This concept is further elucidated in section 3.3. In this method, the defocus blur is never explicitly calculated as it is done in depth from defocus techniques. Instead, by using diffusion, for each pixel we can obtain the corresponding fully focused observation in the defocus space. Then using that observation and the corresponding
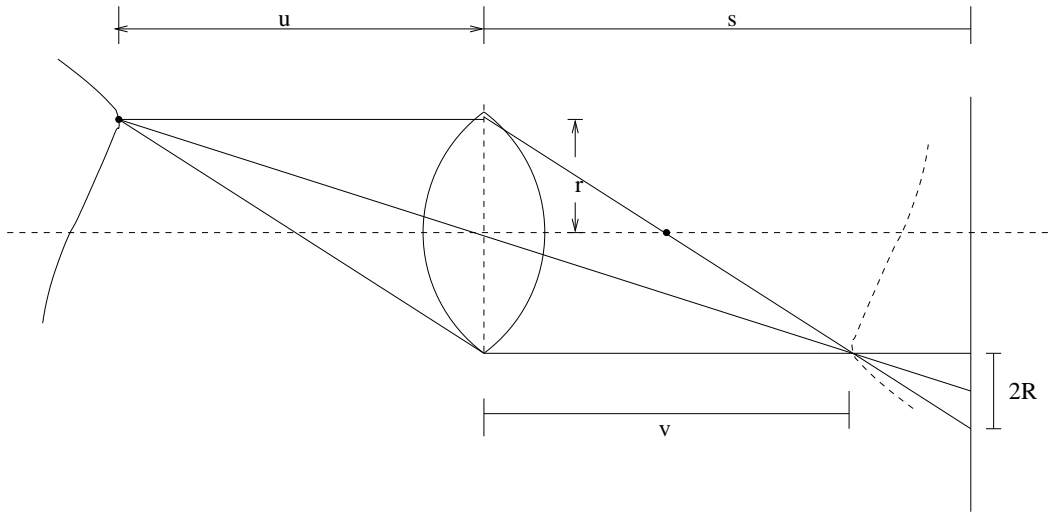
Fig. 1. Illustration of image formation in a convex lens.

virtual lens parameters we can recover the depth information from the lens
equation. As a by-product, we also obtain the fully focused pin-hole image
from these two defocused observations. The diffusion process simulates the
depth from focus technique by generating images in the defocus space of the
observation. Many separate observations as required for the depth from focus
technique are no longer required. In fact, using the diffusion technique, the
two modalities of estimation of depth can be considered to be equivalent. This
is discussed further in section 3.4.

In the next section we give a brief overview of the related work done. In section
3 we outline the theoretical basis for the formation of the defocus image space
of an observation based on the diffusion process. In section 4 we present the
basic algorithm for depth estimation using diffusion. In section 5 we analyze
the procedure and consider the practical issues involved in the implementation
of this method. In section 6 we present the experimental results obtained. We
conclude the paper in section 7.

2   Related Work

2.1   Depth From Focus (DFF)

There are a number of papers in the literature which address the problem of
obtaining depth information from focus. This includes work by Nayar and Nak-
agawa [6], and work by Subbarao and Choi [5]. The basic method followed has
been to obtain different focus levels by adjusting the camera parameters, i.e.
either the lens to image plane distance v, the focal length f or the aperture

3

radius r. The methods involve obtaining many observations for the various camera parameters and estimating the focus measure using various criterion functions. Krotkov [3] has experimentally evaluated several such criteria including the Laplacian and Teningrad operators. In [7] the authors discuss a method in which the blur is evaluated from the intensity change along corresponding pixels in the multi-focus images instead of using window-based blur estimation operators. The fundamental weakness of the DFF method is, however, the time required for image acquisition. In practice about ten or even more images are required to estimate the depth of a scene for a reasonable level of accuracy.

2.2   Depth From Defocus (DFD)

The basic problem addressed in the depth from defocus methodology is the measurement of the relative defocus between observations. Broadly speaking, the approaches have been based on active and passive techniques. The active techniques ([8], [9], [10], [11]) use structured lighting as a cue in obtaining the depth estimate. They do give good results, but they can be used in controlled environments only and are not generally applicable. We are concerned here mainly with the passive techniques. The research in depth from defocus was initially done in the passive domain itself and was introduced by Pentland [12]. He identified the problem of DFD as an estimation of linear space variant blur. The defocus parameter was recovered using the deconvolution in the frequency domain. However, the method depended on the availability of a perfectly focused image of the scene as one of the observations. Subbarao [13] proposed a more general method in which he removed the constraint of one image being formed with the pinhole aperture. There have been other approaches as well in the frequency domain ([14], [15], [16], [17]). The main issue in frequency domain based methods have been the trade off involved in frequency and spatial resolution. In [17], for instance a large number of narrow band filters are used and the depth is estimated in a least square sense. In [18], the authors suggest the use of broadband rational filters. There has also been a substantial amount of work done to estimate depth from defocus where the image analysis is done in the spatial domain itself ([19], [20], [21], [22], [23], [24], [25]). In [19], the authors propose a S-Transform which does deconvolution in the spatial domain itself. In [20], the authors propose a technique based on image decomposition using the Hermite polynomial basis. The difference in blur is then computed by resolving a system of equations. In [21], the authors propose a matrix-based method using regularization. Along similar lines the authors in [23] propose the use of functional singular value decomposition to compute the point spread function. In [22], the authors pose the problem as one of reconstructing the shape and the radiance that minimizes a measure of information divergence between blurred images. Recently, in [25] Favaro *et al.*, have used

the idea of diffusion for estimating depth from defocus which is very similar to our work. However they approach the problem in the traditional manner of casting it into a variational framework. We have introduced a more general idea of generating the entire defocus space of a scene. All these approaches have been deterministic. There has been extensive work done in the statistical framework as well ([4], [26], [27], [28]). There the authors provide various methods for solving this problem by modeling the depth and the image as separate Markov random fields (MRF). Most of the methods in depth from defocus literature assumed that the observations do not suffer from occlusion. However it was brought out in [29] that occlusion can occur in depth from defocus techniques as well. Occlusion effects have also been studied in [30]. The handling of occlusion effects in depth computation was addressed in [28] and [31].

2.3   Heat Equation and Diffusion

In this paper we use the technique of diffusion for synthesizing new virtual observations in the defocus space. The idea of diffusion can be traced to that of scale space filtering by Witkin [1]. Koenderink [2] showed that this is equivalent to solving the heat equation. This scale space approach was extended by Perona and Malik in their landmark paper [32] where they proposed a nonlinear scale space model, aimed at preserving important features such as edges. The model changes its behavior based on the conduction coefficient associated in a region of an image and achieves forward diffusion in the interior region and at the boundaries it acts in the opposite direction. In general the inverse diffusion approach can be thought of as reversing the heat equation in time. This reverse heat equation is however ill-posed and there has been a substantial amount of work done for stabilizing the reverse heat equation. Rudin, Osher and Fatemi in [33] introduced the "shock filter" where they proposed a pseudo-inverse, where the inverse diffusion propagation term is tuned by the sign of the Laplacian. There has been a lot of research done along similar lines where various nonlinear inverse diffusion models have been proposed. In linear scale space theory, recently an interesting work has been done by Lindeberg [34], where he provides a theoretical analysis of the linear scale space theory and also observes that Gaussian and higher orders of the Gaussian kernel are the only admissible kernels based on the admissibility conditions for linear scale space.

3 Defocus as a Diffusion Process

3.1 Diffusion Process

Consider the classical equation for the isotropic diffusion of heat given by the following partial differential equation:

$$\frac{\partial I(x,y;t)}{\partial t} = a \left(\frac{\partial^2 I(x,y;t)}{\partial x^2} + \frac{\partial^2 I(x,y;t)}{\partial y^2} \right) \tag{1}$$

Here the constant a is the thermometric conductivity or diffusivity [35]. The equation above describes how heat diffuses over a surface, given an initial temperature distribution with time. It is assumed here that the diffusion of heat is uniform in all directions. Consider that $I(x, t = 0)$ is an image. The solution of the heat equation can be obtained in terms of convolution of the image with a temporally evolving Gaussian kernel [36]. This is known as the source solution for the heat equation [35] and is given by $\sigma^2 = 2at$ where σ denotes the spread of the Gaussian kernel and is used through out in this paper. As the image is progressively convolved with a Gaussian kernel, it gets increasingly more blurred thereby representing the image information at a different scale. Note that as $t \to \infty$ this corresponds to a fully diffused image. This is the basic idea underlying scale space analysis. Also note that the process is not defined for $t < 0$, a fact that will be utilized later to define the *extended defocus space*.

3.2 Basic Model of Defocus

Consider the image formation process in a real aperture camera employing a thin lens [4]. When a point light source is in focus, all light rays that are radiated from the object point and intercepted by the lens converge at a point on the image plane. When the point is not in focus, its image on the image plane is no longer a point but a circular patch of radius that defines the amount of defocus associated with the depth of the point in the scene. It can be shown that [4]

$$\sigma = \kappa r s \left(\frac{1}{F} - \frac{1}{s} - \frac{1}{u} \right) \tag{2}$$

where r is the radius of the aperture, s is the lens-to-image plane distance, F is the focal length of the lens, u is the depth at that point and κ is a camera constant that depends on the sampling resolution on the image plane.

Let $I(x, y)$ be the pin-hole image of the scene. From the eqn.(2) we note that $C = (r, F, s)$ defines the camera parameters each of which may be changed to effect a different amount of defocus blur for a fixed depth.

The depth related defocus process is linear but not space invariant. Assuming a diffraction-limited lens system and a constant depth in the scene (this assumption will be relaxed at a later stage), the point spread function of the camera system at a point (x, y) may be approximately modeled as a circularly symmetric 2-D Gaussian function [4]:

$$h(x, y) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{3}$$

where the blur parameter σ is obtained from eqn.(2). Assuming the depth to be constant everywhere, the observed defocused image $E(x, y)$ is given by

$$E(x, y) = I(x, y) * h(x, y). \tag{4}$$

This equation can be directly related to the solution of the diffusion equation in terms of the Gaussian kernel as discussed in section 3.1. The real aperture imaging can thus be thought of as providing a real world example of scale space theory. The eqn(4) can be represented by taking its Fourier transform. Denoting the Fourier transform of a function $f(x, y)$ by $\hat{f}(\omega_x, \omega_y)$ we obtain

$$\hat{E}(\omega_x, \omega_y) = \hat{I}(\omega_x, \omega_y)\hat{h}(\omega_x, \omega_y) = \hat{I}(\omega_x, \omega_y)exp(-\frac{\sigma^2(\omega_x^2 + \omega_y^2)}{2})$$

3.3 Defocus Space

For a given scene, one can have two defocused observations E_1 and E_2 corresponding to two different camera parameter settings C_1 and C_2, such that the resulting blur parameters are σ_1 and σ_2, assuming $\sigma_1 > \sigma_2$ without loss of generality. For the two observations E_1 and E_2, a defocus space can be defined.

Definition 1: *Defocus space*

The defocus space is defined to be the set of all possible observations E for a given scene generated by varying the blur σ as a combination of the associated blur parameter σ_1 and σ_2 in the two observations E_1 and E_2 respectively, by the following relation

$$\sigma^2 = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 \tag{5}$$

for all values of $0 \leq \alpha \leq 1$.

This is equivalent to generating $I(x, y, t)$ for $t_1 \leq t \leq t_2$ given the states $I(x, y, t_1)$ and $I(x, y, t_2)$ at two specified time instants t_1 and t_2 in the heat diffusion eqn(1). Substituting eqn(5) in eqn(5) we obtain:

$$\hat{E}(\omega_x, \omega_y) = \hat{I}(\omega_x, \omega_y) \exp\left[-\frac{1}{2}(\alpha\sigma_1^2 + (1-\alpha)\sigma_2^2)(\omega_x^2 + \omega_y^2)\right]$$

$$= \left\{\hat{I}(\omega_x, \omega_y) \exp\left[-\frac{\sigma_1^2(\omega_x^2 + \omega_y^2)}{2}\right]\right\}^\alpha \left\{\hat{I}(\omega_x, \omega_y) \exp\left[-\frac{\sigma_2^2(\omega_x^2 + \omega_y^2)}{2}\right]\right\}^{1-\alpha}$$

or

$$\hat{E}(\omega_x, \omega_y) = \hat{E}_1^\alpha(\omega_x, \omega_y)\hat{E}_2^{(1-\alpha)}(\omega_x, \omega_y). \tag{6}$$

The relation given in eqn(6) is equivalent to the notion of scale space as formed by the diffusion equation. This can be noticed as eqn(6) can be thought of as convolving the image $I(x, y)$ with a time varying Gaussian kernel. This is because convolving a Gaussian function with another Gaussian function always results in a Gaussian function. The eqn(6) effectively reduces to convolving the original image $I(x, y)$ with a Gaussian kernel which varies with time (in this case α) according to the relation given in eqn(5).

The defocus blur σ could be present physically due to any of the following camera parameters : aperture, the lens to image plane distance, the focal length or even a combination of these, as shown in eqn(2). A monotonic variation in any of the lens parameters can generally result in a non-monotonic variation in the blur (for instance as v is changed from an initial value, σ reduces, becomes zero and then subsequently increases), signifying both sides of the defocus cone (see Fig.2 for illustration). The diffusion based defocus space generation process however generates the blur in a monotonic manner, i.e we are restricted to one side of the defocus cone. By continuously varying the parameter α, we can generate any virtual observation for defocus setting lying between the lines AB and CD in Fig.2 using the eqn(6). The defocus space thus consists of all possible observations of the defocus blur $\sigma_1^2 \leq \sigma^2 \leq \sigma_2^2$.

Corresponding to the notion of continuous defocus space as introduced in the previous section, a practical counterpart of this defocus space would be a sampled defocus space. This corresponds to generating the defocus space for discrete values of α between 0 and 1. In Fig.2, the lines corresponding to $A_1B_1, A_2B_2, ...A_nB_n$ may represent one such possible set of sampled defocus space. The sampled defocus space generated for an image is similar to the physically obtained focused image space described in [5].

8

So far we have considered a restricted range of α between $[0, 1]$. Now we relax this condition and something interesting happens. If the values of α beyond the range $[0, 1]$ are considered then the defocus space generated is the extended defocus space.

Definition 2: *Extended defocus space*

The extended defocus space is defined to be the set of all possible observations E for a given scene generated by varying the blur σ as a combination of the associated blur parameter σ_1 and σ_2 in the two observations E_1 and E_2 respectively, by the following relation

$$\sigma^2 = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 \tag{7}$$

for all values of $\beta \leq \alpha \leq \infty$.

Here the value of β is the value of α such that $\sigma^2 = 0$ in eqn(7), resulting in a fully focused observation. This can be obtained from the diffusion equation since corresponding to the image $I(x, y, t)$ with $t \to \infty$ we can obtain an observation $E(x, y)$ with $\alpha \to \infty$. This represents the fully diffused image. Similarly for each point there exists a value $\alpha = \beta < 0$ corresponding to $t = 0$. This corresponds to a fully focused observation, i.e. $\sigma^2 = 0$. Thus the extended defocus space is defined for the range $\alpha \in [\beta, \infty)$. In the range $\alpha = [\beta, 0]$ the process, instead of being a diffusion becomes an inverse diffusion. Beyond this range, the defocus space is undefined since one cannot have the blur $\sigma^2 < 0$. This is illustrated in Fig.2.

Depth from defocus(DFD) methodology estimates the space variant blur whereas depth from focus (DFF) methodology estimates the focused image point. It is possible to use the techniques in DFD methodology to estimate the space variant blur using just two observations, whereas DFF requires many samples to estimate the fully focused point. Here as we have shown, it is possible to generate the extended defocus space for the image using just two observations. Thus both the techniques can be considered fundamentally equivalent, rendering the need for multiple samples to be redundant. The diffusion based process thus provides an equivalent means for estimating the depth from the known lens parameters using either depth from defocus or depth from focus.

4   Algorithm for Depth Estimation

The derivation of eqn(6) is based on the assumption of constant depth. When there is depth variation in the scene, eqn(6) is no longer valid as the blurring
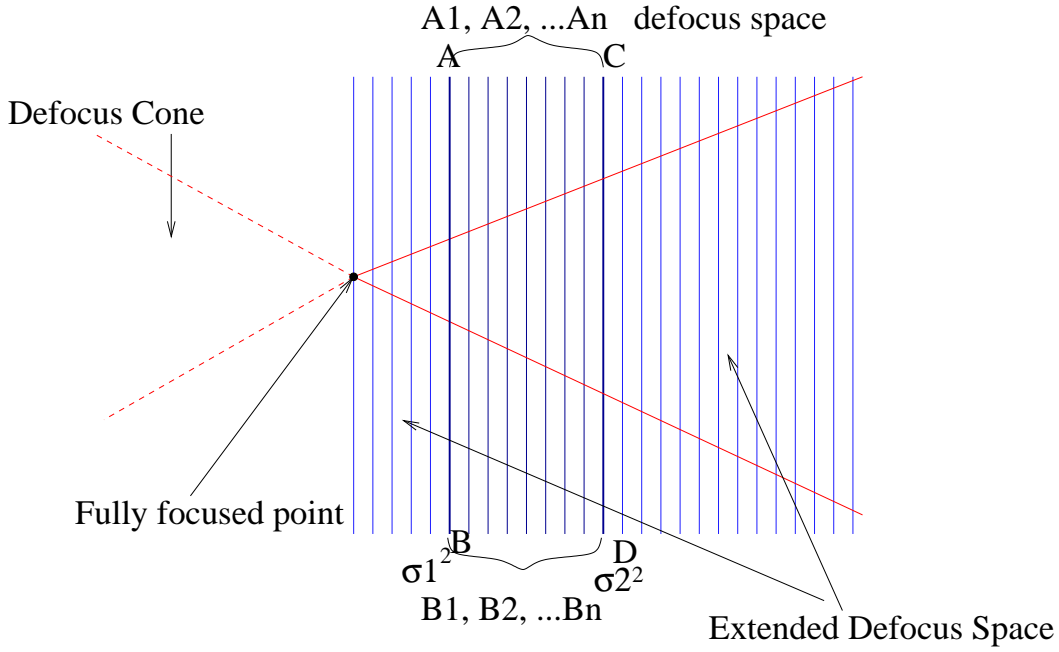
Fig. 2. Illustration of the concept of defocus space for a particular scene

process becomes shift variant, implying a non-homogeneous diffusion process. This corresponds to the following diffusion equation

$$\frac{\partial I(x,y;t)}{\partial t} = a(x,y)(\frac{\partial^2 I(x,y;t)}{\partial x^2} + \frac{\partial^2 I(x,y;t)}{\partial y^2})$$
(8)

Here a is no longer a constant but is now a function $a(x,y)$ and this is handled by forming a small MxM window about a point over which the depth can be assumed to be constant as is done commonly in all literature. Using this modification the defocus space for a scene can be created locally even in the depth varying case. The depth estimation is done by obtaining the fully focused point for each image. The process of creating the defocus space is a monotonic process. As α varies, the characteristics of the process changes from diffusion to inverse diffusion and the deblurring of the defocused observations takes place. In obtaining the fully focused image the value of α is not restricted to lie between 0 and 1, rather we go for values of $\alpha < 0$. The characteristic of the convolution changes from a low pass filter to a high pass filter for $\alpha < \beta$. The defocus process has to be stopped when the fully focused point is reached. This stopping point is estimated empirically from the virtually synthesized observations using a band pass filter, similar to the way it is done in DFF methods [3]

The various steps of the algorithm for depth estimation are as follows:

STEP 1: Divide the observed images E_1 and E_2 into overlapping MxM windowed representations.

STEP 2: Obtain the FFT of the corresponding windows in E_1 and E_2.

STEP 3: Synthesize a sample of the defocus space corresponding to a particular value of $\alpha \in [\beta, 0)$ for each window using eqn(6). Note that β is unknown as the value of β would give us the depth.

STEP 4: Estimate the amount of focus using a sharpness criterion function which is essentially a band pass filter and decide whether a fully focused point is reached. Else change the value of α and go to STEP 3.

STEP 5: Using the corresponding values of the virtual lens parameters, calculate the value of depth at the point. Save the pixel value as the restored one.

This algorithm is sequentially executed for all pixels in the image till the corresponding pin-hole observation of the scene is obtained and a dense depth map is generated.

5   Computational Difficulties

The algorithm uses the windowed Fourier transform. In some cases, especially, where the gray level variance in the window is very low, signifying a textureless scene, there might be a problem as the spectral components are nearly zero. When the value of α goes beyond the 0 to 1 range, potentially a division by zero can occur in eqn(6). This can be avoided by marking such windows out of computation. Mathematically it signifies that the depth cannot be estimated for homogeneous regions.

Another factor which adversely affects the accuracy is its sensitivity to quantization error. Generally, an 8 bit quantization of the scene results in a very noisy virtual observations. This is because the defocus space generating process acts as a high pass filter when we take $\alpha < \beta$, which greatly enhances the quantization error. Further the inverse diffusion process is inherently unstable and the quantization error aggravates the instability. Practical implementation suggests the use of a 16-bit representation of the intensity function.

Generation of virtual observations using eqn(6) locally may demand quite a bit of computation. This is more so due to the fact that a finer sampling of the extended defocus space would lead to a better accuracy in the depth estimate. To obtain better estimates of the fully focused points efficiently, a hierarchical virtual sampling technique is used wherein, using the algorithm defined earlier a value of α is quickly estimated using coarser discrete steps in the range $[\beta, 0]$. Then a further dense sampling is performed in a small neighborhood ϵ around the best current estimate of α, i.e. $\alpha \in [\hat{\alpha} - \epsilon, \hat{\alpha} + \epsilon]$ and the estimate of α is refined in a hierarchical manner.

6 Results

The algorithm has been tested with real as well as simulated data. In the case of real data, there is a substantial amount of noise in the recovered structure. This is mainly because the real world data is in eight bit form and the resultant quantization error is quite significant. However, the overall structure recovered still resembles the true structure in the scene. In a similar way, the corresponding deblurred observation, in general, does not resemble the actual pin-hole image, but the result is definitely better focused and less blurred than the observations given as input to the algorithm. The results obtained with synthetic data can be observed to be of better quality due to the use of 16 bit representation.

In Fig.3, two images of a ball are taken with varying lens-to-image plane distances. In the experimental setup the base was at a distance of 117 cm. from the camera. The point on the ball nearest to the camera was at 121.8 cm. while the points lying on the occluding boundary of the ball were at a distance of 132.3 cm. from the camera. The change in the lens-to-image plane distance introduces a small amount of change in magnification. This was taken into account and corrected. Fig.3.c shows the recovered dense depth map with the darker shading corresponding to a nearer distance. The darkest points (gray level 0) refer to the homogeneous regions for which the depth cannot be estimated as explained in section 5. Fig.3.d shows the corresponding deblurred image obtained.

The second experimental setup was the "blocks world" where three blocks were arranged at different depths (see Fig. 4(a,b)), the nearest one at a distance of 73 cm., another at 82.7 cm. and the farthest block at 96.6 cm. Again images were taken with varying lens-to-image plane distances to obtain different amount of defocus in different observations. Fig.4.c shows the dense depth map estimated in this case and Fig.4.d shows the deblurred image obtained.

The Fig.5 shows a test data where a textured image is synthetically blurred with a continuously varying Gaussian kernel. The variance of the Gaussian kernel was increased in a ramp like manner from left to right. The second observation was simulated using different values of the blur kernel. Fig.5.c shows the corresponding dense depth map and Fig.5.d shows the deblurred image. The left to right variation in depth is clearly visible. Similarly the restored image is much sharper, although it contains dark spots where depth could not be estimated due to reasons mentioned in section 5. Although we could have copied the intensities at these pixels from one of the observations as they correspond to fairly homogeneous regions, we refrain from doing it so that the effect can be highlighted. Fig.6 shows another synthetically generated test data where a textured image is blurred with a continuously varying Gaussian
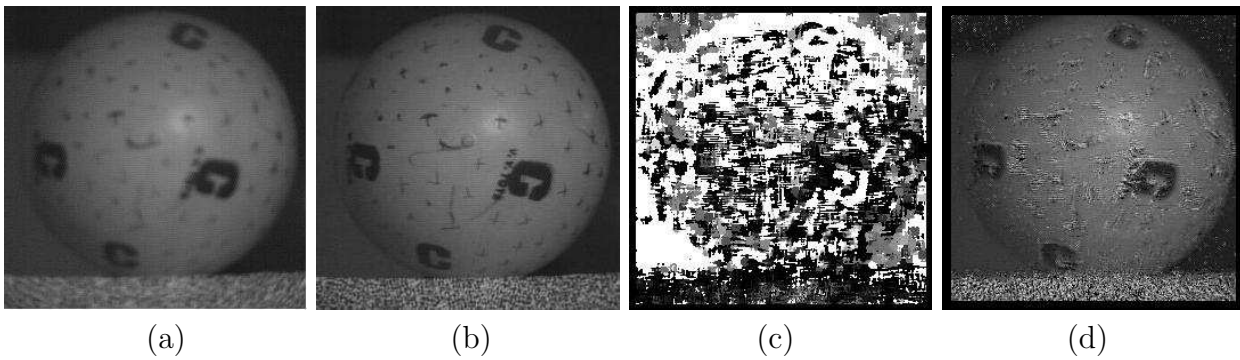
| (a) | (b) | (c) | (d) |

Fig. 3. Ball Image: (a,b) Two observations with the right one being less blurred, (c,d) recovered structure and the deblurred image, respectively.
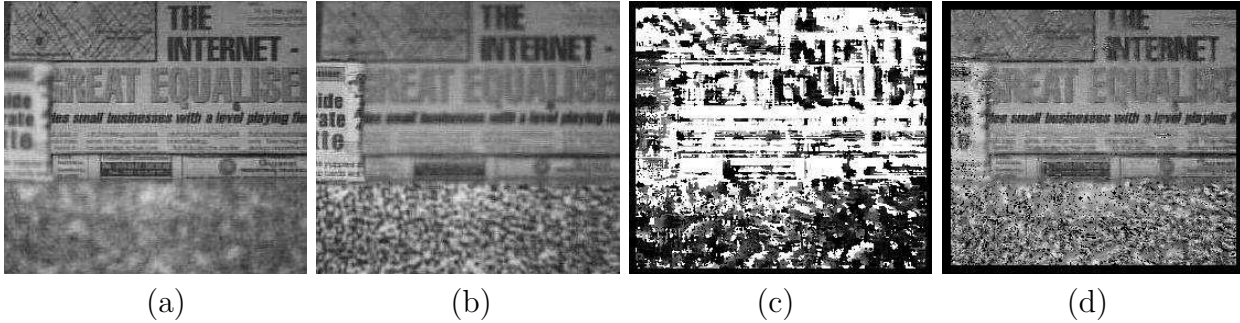

| (a) | (b) | (c) | (d) |

Fig. 4. Two observations of the Block World. (a) The furthest block is in focus, (b) the nearest block is in focus, (c,d) recovered structure and the deblurred image, respectively.

kernel. However, here the variance of the blur was increased in a radially outward manner. Fig.6.c shows the corresponding dense depth map and Fig.6.d shows the corresponding deblurred image. Once again the depth variation is quite clear from the plot.

The results appear to be noisy as the linear diffusion process suffers from instability in the extended defocus space as the process corresponds to inverse diffusion. One does require a suitable regularizing functional to make the problem better posed. This method, however, presents a theoretical basis for understanding depth from focus/defocus in the light of the diffusion equation and is thus important in its own merit.

7   Conclusion

For a given scene in the real world, we have defined a defocus space which is a virtual space of all observations based on the properties of a real aperture imaging system. A method for generating the defocus space based on the diffusion equation has been presented. We have also presented an algorithm

13

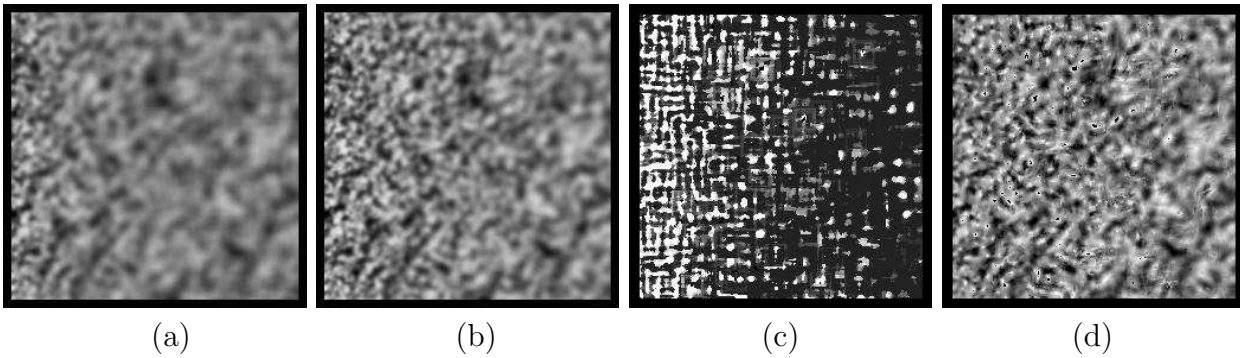<p align="center">(a)   (b)   (c)   (d)</p>

Fig. 5. (a,b) Two synthetically generated blurred observations. Here the blur increases progressively from left to right. (c) Recovered structure, and (d) the deblurred image.

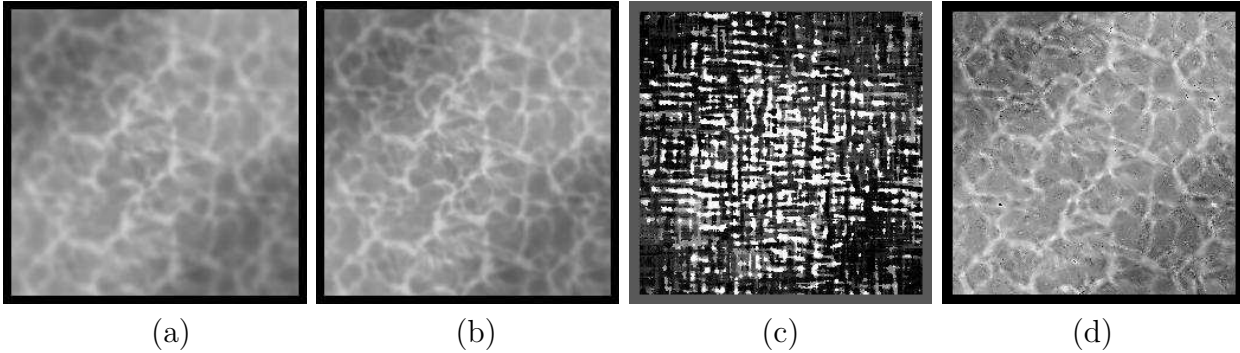<p align="center">(a)   (b)   (c)   (d)</p>

Fig. 6. (a,b) Two simulated observations. Here the blur increases radially outward. (c,d) Recovered structure and the deblurred image, respectively.

for recovering the scene structure based on the defocus space. An interesting outcome of this work is that it brings out the equivalence of the depth from focus and depth from defocus modalities for depth estimation. This algorithm has been tested with real as well as synthetic images. A possible extension of the current algorithm is to incorporate a facet based modeling of the depth of the scene while calculating the diffusion coefficient for improved accuracy. It is also possible to consider multiple exposures of the scene as is commonly done in the DFF process and this is expected to improve the results. We are also exploring the suitability of regularizing the diffusion process in order to obtain a smooth depth map.

References

[1] A. P. Witkin, "Scale-Space Filtering," in *Proc. of the 4th International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, 1983, pp. 1019–1022.

[2] J. J. Koenderink and A. J. van Doorn, "Dynamic shape," *Biological Cybernetics*, vol. 53, pp. 383–396, 1986.

[3] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, pp. 223–237, 1987.

[4] S. Chaudhuri and A. N. Rajagopalan, *Depth From Defocus: A Real Aperture Imaging Approach*, Springer Verlag, New York, 1999.

[5] M. Subbarao and T. Choi, "Accurate Recovery of Three-Dimensional Shape From Image Focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 266–274, march 1995.

[6] S. K. Nayar and Y. Nakagawa, "Shape From Focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[7] Takashi Matsuyama Naoki Asada, Hisanaga Fujiwara, "Edge and depth from focus," *International Journal of Computer Vision*, vol. 28, no. 2, pp. 153–163, 1998.

[8] B. Girod and S. Scherock, "Depth from defocus of structured light," in *Proc. Optics, Illumination, and Image Sensing for Machine Vision IV, SPIE vol. 1194*, 1989, pp. 209–215, held in Philadelphia, Penn., November 1989.

[9] A. P. Pentland, S. Scherock, T. Darrell, and B. Girod, "Simple range cameras based on focal error," *J. Optical Soc. America A*, vol. 11, no. 11, pp. 2925–2934, november 1994.

[10] M. Noguchi and S.K. Nayar, "Microscopic shape from focus using active illumination," in *Proc. of 12th IAPR Int. Conf. on Pattern Rec.*, 1994, vol. 1, pp. 147–152, Jerusalem, Israel.

[11] S.K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1186–1198, Dec. 1996.

[12] A. P. Pentland, "A new sense for depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, July 1987.

[13] M. Subbarao, "Parallel depth recovery by changing camera aperture," in *Proceedings of International Conference on Computer Vision*, 1988, pp. 149–155, held in Florida, USA.

[14] M. Gokstorp, "Computing depth from out-of-focus blur using a local frequency representation," in *Proc. Intl Conf. Pattern Recognition*, 1994, pp. 153–158, Jerusalem, Israel.

[15] A.N. Rajagopalan and S. Chaudhuri, "A block shift-variant blur model for recovering depth from defocused images," in *Proc. Intl Conf. Image Processing*, 1995, pp. 636–639, Washington D.C., USA.

[16] Y. Y. Schechner and N. Kiryati, "The optimal axial interval in estimating depth from defocus," in *Proceedings of the Seventh International Conference on Computer Vision*, September 1999, vol. 2 of *Annual Conference Series*, pp. 843–848, held in Corfu, Greece, September 20-25, 1999.

[17] Y. Xiong and S.A. Shafer, "Moment filters for high precision computation of focus and stereo," in *Proc. Intl Conf. Intelligent Robots and Systems*, August 1995, pp. 108–113, Pittsburgh, Pennsylvania, USA.

[18] M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *International Journal of Computer Vision*, vol. 27, no. 3, pp. 203–225, May 1998.

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

[20] D. Ziou and F. Deschenes, "Depth from defocus estimation in spatial domain," *Computer Vision and Image Understanding*, vol. 81, no. 2, pp. 143–165, February 2001.

[21] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 97–108, 1993.

[22] P. Favaro, A. Mennucci, and S. Soatto, "Observing shape from defocused images," *International Journal of Computer Vision*, vol. 52, no. 1, pp. 25–43, April 2003.

[23] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 406–417, March 2005.

[24] P. Favaro and S. Soatto, "Shape and radiance estimation from the information divergence of blurred images," in *Proc. European Conf. Computer Vision, June 2000*, 2000, vol. 2, pp. 755–768, Dublin, Ireland.

[25] P. Favaro, S. Osher, S. Soatto, and L. Vese, "3d shape from anisotropic diffusion," in *Proceedings of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, 2003, vol. 1, pp. 179–186, Madison, Wisconsin, USA.

[26] A.N. Rajagopalan and S. Chaudhuri, "Optimal recovery of depth from defocused images using an mrf model," in *Proc. Intl Conf. Computer Vision*, 1998, pp. 1047–1052, Bombay, India.

[27] A.N. Rajagopalan and S. Chaudhuri, "Optimal selection of camera parameters for recovery of depth from defocused images," in *Computer Vision and Pattern Recognition*, 1997, pp. 219–224, San Juan, Puerto Rico.

[28] S.S. Bhasin and S. Chaudhuri, "Depth from defocus in presence of partial self occlusion," in *Proc. Eighth IEEE Int'l Conf. on Computer Vision*, July 2001, vol. 1, pp. 488–493, Vancouver, Canada.

[29] Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?," *International Journal of Computer Vision*, vol. 39, no. 2, pp. 141–162, September 2000.

[30] Takashi Matsuyama Naoki Asada, Hisanaga Fujiwara, "Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 155–167, 1998.

[31] P. Favaro and S. Soatto, "Seeing beyond occlusions (and other marvels of a finite lens aperture)learning shape from defocus," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 03)*, 2003, vol. 2, pp. 579–586, Madison, Wisconsin, USA.

[32] P. Perona and J. Malik, "Scale Space and Edge Detection using Anisotropic Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[33] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[34] T. Lindeberg and B. M. ter Haar Romeny, "Linear-Scale -Space," in *Geometry-Driven Diffusion in Computer Vision*, B. M. ter Haar Romeny, Ed., pp. 1–41. Kluwer Academic Publisher, Boston, 1994.

[35] D. V. Widder, *The Heat Equation*, Academic Press Inc., London, UK, 1975.

[36] N. Sochen, R. Kimmel, and A. M. Bruckstein, "Diffusions and confusions in signal and image processing," *J. Math. Imaging and Vision*, vol. 14, pp. 195–209, 2001.