Depth from Diffusion

Changyin Zhou
Columbia University

Oliver Cossairt
Columbia University

Shree Nayar
Columbia University

Abstract

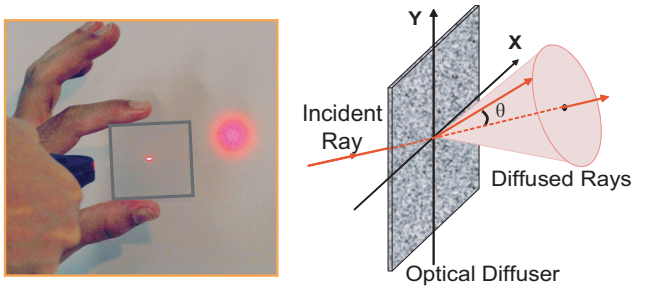
An optical diffuser is an element that scatters light and is commonly used to soften or shape illumination. In this paper, we propose a novel depth estimation method that places a diffuser in the scene prior to image capture. We call this approach depth-from-diffusion (DFDiff).

We show that DFDiff is analogous to conventional depth-from-defocus (DFD), where the scatter angle of the diffuser determines the effective aperture of the system. The main benefit of DFDiff is that while DFD requires very large apertures to improve depth sensitivity, DFDiff only requires an increase in the diffusion angle – a much less expensive proposition. We perform a detailed analysis of the image formation properties of a DFDiff system, and show a variety of examples demonstrating greater precision in depth estimation when using DFDiff.

1. Introduction

A diffuser is an optical element that scatters light and is widely used to soften or shape light in illumination and display applications [1][2]. Optical diffusers are also commonly used in commercial photography. Photographers place diffusers in front of the flash to get rid of harsh light, in front of the lens to soften the image, or at the focal plane to preview the image. Most commercially available diffusers are implemented as a refractive element with a random surface profile. These surfaces can be created using random physical processes such as sandblasting and holographic exposure, or programmatically using a lithographic or direct writing method [3][4][5][6]. Figure 1(a) shows an off-the-shelf diffuser scattering a beam of light.

A diffuser converts an incident ray into a cluster of scattered rays. This behavior is fundamentally different from most conventional optical devices used in imaging, such as mirrors and lenses. Figure 1(b) illustrates the geometry of light scattering in Figure 1(a). The scattering properties of a diffuser can be generally characterized by its diffusion function $\mathcal{D}(\theta_i, \psi_i, \theta_o, \psi_o)$, where $[\theta_i, \psi_i]$ is the incident direction and $[\theta_e, \psi_e]$ is the exitance direction. Since diffusers have usually been designed so that scatter is invariant to incident direction, the diffusion function can be simply written as $\mathcal{D}(\theta, \psi)$, where θ and ψ are the angular coordinates of the exiting ray relative to the incident direction. For most commercial diffusers (e.g., the one shown in Figure 1), the diffusion functions are radially symmetric and can be further simplified to $\mathcal{D}(\theta)$.



(a) An Optical Diffuser (b) Geometry of Optical Diffusion

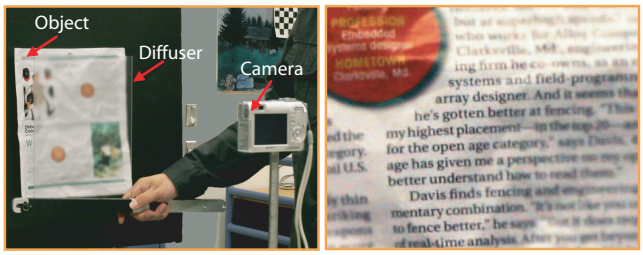
Figure 1. (a) A laser beam is diffused by a holographic diffuser. (b) The geometry of the optical diffusion.

Diffusers have been studied extensively by the optics community; however, researchers in computer vision have paid little attention to them. In this paper, we analyze the image formation properties of an imaging system augmented by a diffuser. When a diffuser is placed in front of an object, we capture a diffused (or blurred) image that appears similar to a defocused image. By assuming a locally constant diffusion angle, a small patch in a diffused image can be formulated as a convolution between a focused image and the diffusion blur kernel. The diffusion blur kernel is determined both by the diffusion function and the object-to-diffuser distance. This effect is quite similar to lens defocus, which is often formulated as the convolution between an in-focus image and a defocus kernel. We perform a detailed comparison between diffusion and lens defocus in Section 5.1. In addition, we analyze image formation in the presence of both diffusion and defocus.

To implement our depth from diffusion (DFDiff) technique, we place an optical diffuser between the scene and the camera as shown in Figure 2(a). Our analysis shows that the diffusion blur size is proportional to the object-to-diffuser distance (see Figure 2(b)). We can therefore infer depth by estimating the diffusion blur size at all points in the image. Since the depth estimation problem for DFDiff is similar to conventional DFD, many existing algorithms can be used to find a solution [7][8][9][10].

2. Related Work

Researchers have used optical diffusers in the aperture of a camera to computationally increase resolution [11] and depth of field (DOF) [12]. For DFDiff, we place a diffuser in the scene to effectively *reduce* DOF. We show in Section 3.2 that this decrease in DOF is a result of the diffuser



(a) Experimental Setup (b) Captured Image (close-up)

Figure 2. (a) An optical diffuser is placed in front of the camera and close to the object (a crinkled magazine). (b) A close-up of the captured image. We can see that the blur of the text is spatially varying as a function of depth.

effectively enlarging the lens aperture.

The DFDiff technique is most similar to conventional DFD, which has been studied extensively by the vision community (e.g., [7][8][9][10]). While DFDiff is similar in principle to DFD, it offers three significant advantages:

- **High-precision depth estimation with a small lens.** For DFDiff, the precision of depth estimation depends only on the mean scattering angle of the diffuser and is independent of lens size. Note that while it is often difficult to make lenses with large apertures, it is relatively easy to make diffusers with large diffusion angles.
- **Depth estimation for distant objects.** By choosing the proper diffuser, DFDiff can achieve high precision depth estimation even for objects at very large distances from the camera. For DFD, depth sensitivity is inversely proportional to the square of object distance [13][14][15]. In many scenarios, it is necessary to place objects far from the camera in order to achieve a reasonable field of view.
- **Less sensitive to lens aberrations.** Lens aberrations cause the shape of the defocus point spread function (PSF) to vary with field position. This effect is strong, particularly in the case of inexpensive lenses, and degrades the precision of depth estimation. In contrast, as we show in Section 6, diffusion PSFs are more invariant to field position.

DFDiff does, however, require the flexibility to place a diffuser in the scene, which is impractical or impossible in some situations.

3. Image Formation with a Diffuser

3.1. Geometry of Diffusion

When an optical diffuser is placed between the scene and the camera, the captured image will be diffused, or blurred. The diffusion varies with camera, scene, and diffuser settings. We first show in Figure 3 the geometry of diffusion in a simple pinhole imaging system. Placed between the pinhole O and the scene point P is a diffuser with a pillbox diffusion function $\sqcap_\theta(x)$:

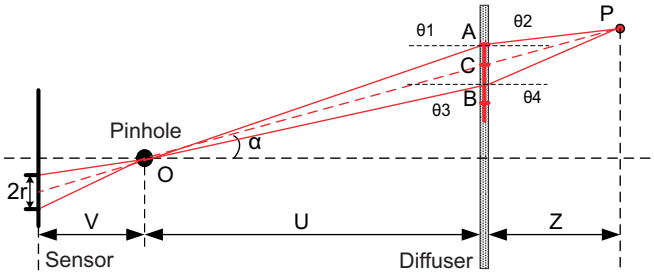


Figure 3. Geometry of diffusion in a pinhole camera. An optical diffuser with a pillbox diffusion function of degree θ is placed in front of a scene point P and perpendicular to the optical axis. From the viewpoint of pinhole, a diffused pattern AB appears on the diffuser plane.

$$\sqcap_\theta(x) = \begin{cases} \frac{1}{\pi\cdot\theta^2} & x < \theta \\ 0 & otherwise, \end{cases}$$

where θ is the diffusion angle of the diffuser.

As shown in Figure 3, the light from an arbitrary scene point P is scattered by the diffuser. Due to the limit of the diffusion angle θ, only the light scattered from a specific region AB can reach the pinhole O. From the viewpoint of the pinhole, a line AB (or pillbox in 2D) appears on the diffuser plane instead of the actual point P.

Proposition 3.1 *When an optical diffuser is placed parallel to the sensor plane (see Figure 3) and the diffusion angle θ is small ($\sin\theta \approx \theta$), we get*

$$\frac{2\tan\theta}{\cos^2\alpha}\cdot\frac{1}{\overline{AB}} = \frac{1}{U} + \frac{1}{Z}, \tag{1}$$

where α is the field angle and $\overline{AB}$ is the diffusion size. The perspective projection of P on the diffuser plane C can be approximated with high precision as the center of AB when α is not too large. (see Appendix I in the supplementary material for the proof.)

It is interesting to observe that this equation has a form that is similar to the Gaussian lens law. It shows that for any given U, the diffusion size $\overline{AB}$ is uniquely determined by the distance Z and the diffusion angle θ. In addition, the perspective projection C and the center of the diffusion pattern AB are the same. Therefore, the diffuser blur does not cause geometric distortions.

Then, the radius r of the PSF can be obtained using Equation 1:

$$r = \frac{V}{U}\cdot\frac{\overline{AB}}{2} = m\cdot\frac{Z}{\cos^2\alpha}\cdot\tan\theta, \tag{2}$$

where $m = V/(Z+U)$ is the image magnification.

In this paper we assume the diffuser is parallel to the sensor plane. The equations governing DFDiff can easily be extended to include tilted planes. Please see Appendix in the supplementary material for details.

3.2. Equi-Diffusion Surfaces and Image Formation

From Equation 2, we can see that the diffusion size r is related to the field angle α. Given r, we can derive a surface using Equation 2:

$$Z = \frac{r \cdot U \cdot \cos^2 \alpha}{\tan\theta \cdot V - r \cdot \cos^2 \alpha}, \tag{3}$$

referred to as an equi-diffusion surface. All scene points on an equi-diffusion surface will be equally blurred by diffusion. Under the paraxial approximation ($\sin\alpha = \alpha$), the surface is planar, since the term $\cos^2\alpha$ approaches 1. For a large field of view, the equi-diffusion surface is no longer planar. But note that the field angle α of each pixel can usually be computed directly from the effective focal length and the pixel position according to the geometry of image formation. For lenses with severe distortions (e.g., fish-eye lens), the mapping between α and pixel position needs to be calibrated. A set of equi-diffusion surfaces in 1D space are shown in Figure 4.

For any equi-diffusion surface with $r = r_0$, the diffused image F can be written as the convolution of the latent clear pinhole image F_0 and the pillbox PSF $\sqcap_{r_0}$: $F = F_0 \otimes \sqcap_{r_0}$. Similarly, when a diffuser with Gaussian diffusion function is used, we will have $F = F_0 \otimes g_{r_0}$, where g_{r_0} is a Gaussian function with standard deviation $\sigma = r_0$. More generally, for a diffuser with an arbitrary diffusion function $\mathcal{D}$, the image formation can be written as the convolution of the image F_0 and the diffusion function $\mathcal{D}$ of size r_0:

$$F = F_0 \otimes \mathcal{D}_{r_0}. \tag{4}$$

3.3. Diffusion + Defocus

It is well known that for a lens camera without a diffuser, the defocused image of a fronto-planar object can be formulated as $F = F_0 \otimes L$, where F_0 is the latent focused image (pinhole image) and L is the defocus PSF. On the other hand, we know from Section 3.2 that for a pinhole camera augmented by a diffuser, the image of an equi-diffusion surface can be written as $F = F_0 \otimes \mathcal{D}$, where $\mathcal{D}$ is the diffusion PSF. But how will the lens blur interact with the diffuser blur when a diffuser is used in a lens camera?

Proposition 3.2 *Suppose a lens camera is focused at an arbitrary distance, and an optical diffuser, which is parallel to the lens, is placed between the lens and a scene point P.*

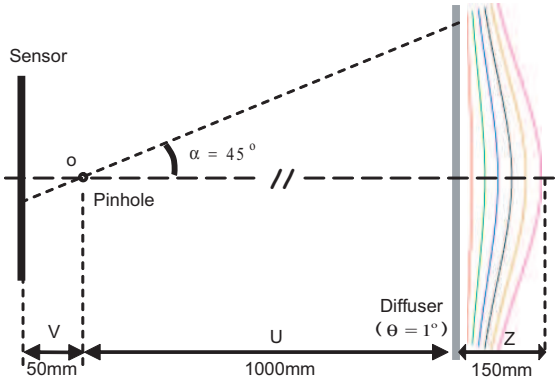


Figure 4. Equi-diffusion surfaces of a simulated pinhole camera with a diffuser. Six equi-diffusion surfaces (1D) are shown in different colors.

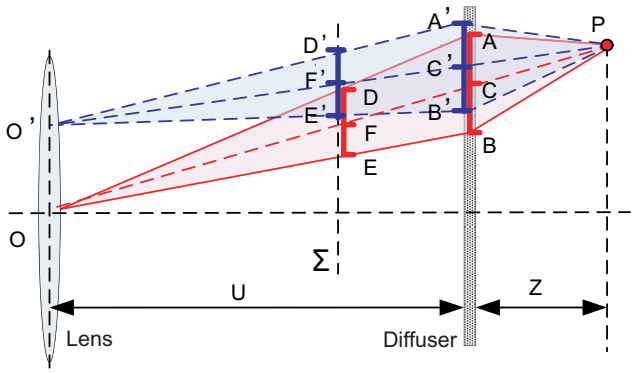


Figure 5. Diffusion in a lens camera. An optical diffuser with a pillbox diffusion function of degree θ is placed in front of a pinhole camera and perpendicular to the optical axis.

When the distance from P to the lens plane is much larger than the aperture size, we have

$$\mathcal{K} = \mathcal{L} \otimes \mathcal{D}, \tag{5}$$

where K is the image of P (the PSF), L is the image of P that would be captured if the diffuser were removed (the defocus PSF), and D is the image of P that would be captured if a pinhole were used instead of the lens (the diffusion PSF).

Proof: As shown in Figure 5, suppose the lens is focused at Plane Σ and the diffuser is placed at a distance U, perpendicular to the optical axis, and a scene point P is located behind the diffuser at a distance Z. From Section 3.1 we know that from the perspective of O, a scene point P appears as AB in the diffuser plane, or as DE on the focus plane Σ. The image of DE on the sensor is $\mathcal{D}$, the diffusion PSF of P if a pinhole camera were used. Similarly, for an arbitrary point O', P appears as $A'B'$ on the diffuser plane and $D'E'$ on the focus plane. Since $U + Z \gg O'O$, the view angles of P with respect to O and O' can be regarded as equal, thus $AB = A'B'$ and $DE = D'E'$. Therefore, the image of $D'E$ on the sensor is a shifted version of $\mathcal{D}$.

For an arbitrary O', the center of the virtual image F' is the projection of P on the focus plane. Note that this effect is independent of the diffuser properties. When all the points on the aperture are considered, each point forms a virtual image of P on the focus plane, whose image on the sensor is the lens defocus pattern L. Hence, the image of P on the sensor K is the sum of a set of shifted $\mathcal{D}$'s whose centers are given by $\mathcal{L}$. That is $\mathcal{K} = \mathcal{L} \otimes \mathcal{D}$. □

Now, suppose we have two images of a scene captured using a normal lens, one without a diffuser and one with a diffuser placed in front of the object, as illustrated in Figure 5. Consider arbitrary corresponding small patches P_1 and P_2 in the two images. By assuming that the diffusion and defocus are locally constant, we have $P_2 = P_1 \otimes \mathcal{D}$, since $P_2 = P_0 \otimes (\mathcal{L} \otimes \mathcal{D})$ and $P_1 = P_0 \otimes \mathcal{L}$, where P_0 is the latent focused patch. According to Equation 1, $\mathcal{D}$ is determined by the diffusion profile of the diffuser and the

distance from the patch to the diffuser plane. Note that according to Proposition 3.2, this relation holds regardless of the lens focus.

4. Depth from Diffusion Algorithm

The basic idea of depth from diffusion (DFDiff) is straightforward. As shown in Figure 2(a), an optical diffuser is placed between the scene and the camera, and a blurred image is captured (shown in Figure 2(b)). The diffusion size is uniquely determined by the distance between objects and the diffuser. By estimating the diffusion size in the image, we can infer the scene depth relative to the diffuser plane.

To estimate the diffusion size, we can take two images F_1 and F_2 with and without a diffuser, respectively. According to Section 3.2, for an arbitrary small patch pair P_1 and P_2 in these images, we have $P_2 = P_1 \otimes \mathcal{D}_{s_0}$, where s_0 is the diffusion size. To estimate depth, we must infer the diffusion size s_0 from the two captured patches P_2 and P_1. Note that this is exactly the same formulation as conventional DFD, which computes depth from two input images, one defocused and one focused. Therefore, most existing DFD algorithms can be applied to estimate the diffusion size s_0. For complicated scene surfaces, different diffusion sizes have to be computed for different pixels. The same problem also exists in DFD and many strategies have been proposed to estimate maps of blur size.

In our implementation, we adapt a straightforward algorithm, similar to those in [16] and [17], to recover the map of diffusion size, S(x, y). For every sampled diffusion size s, a residual map R^s is computed as

$$R^s(x,y) = |F_1(x,y) \otimes \mathcal{D}_s(x,y) - F_2|. \tag{6}$$

Then, for each pixel (x, y), its diffusion size $S(x, y)$ is selected to minimize the corresponding residual:

$$S(x,y) = \arg\min_s R^s(x,y). \tag{7}$$

Based on the estimated diffusion map $S(x, y)$, we can then compute the depth map $Z(x, y)$ according to Equation 1. Note that the field angle α can be computed directly from the pixel position (x, y) and camera parameters, so that it is straightforward to convert between $S(x, y)$ and $Z(x, y)$.

4.1. Reflections from Diffuser Surface

Although the light transmission efficiency of diffusers can be quite high (92% for the Luminit holographic diffusers [18], which will be used in our experiments), some light is still directly reflected by the diffuser surface to the camera. Thanks to its extremely rough surface, light reflected from the diffuser is usually quite uniform. Therefore, its contribution to the captured image can be approximately modeled as $F = a * F' + b$, where F is the actual diffused image captured with reflections, F' is the ideal diffused image captured without any reflection, and a and b are two constants mainly determined by the light transmission efficiency of the diffuser.

Obviously, for the mean brightness $\bar{F}$ and $\bar{F'}$, $\bar{F} = a * \bar{F'} + b$ still holds. $\bar{F'}$ can be estimated using the mean brightness B of the image captured without a diffuser. In addition, note that for a captured RGB image, $[a, b]$ is consistent over the three color channels. Therefore, given one image captured with a diffuser and one image captured without a diffuser, we can easily compute a and b by solving a simple linear equation. Then, the effects of reflectance can be removed by applying $F' = (F - b)/a$.

4.2. Illumination Changes due to the Diffuser

When a diffuser is placed over the object, the illumination will be first diffused by the diffuser before reaching the object. Illumination is usually low-frequency and the diffusion makes it even more uniform. Furthermore, non-specular surfaces are known to low-pass filter incident illumination. Therefore, illumination changes due to the diffuser will only affect low-frequencies in the captured images. To account for this effect, we apply a high-pass filter to Equation 6 and get

$$R^s(x,y) = |\mathbb{H}[F_1(x,y) \otimes D_s(x,y) - F_2]|, \tag{8}$$

where $\mathbb{H}$ is a high-pass filter. We use a Derivative of Gaussians (DOG) filter in our implementation. Note that the depth estimation mainly relies on the high-frequency information, so that applying a high-pass filter has little effect on depth estimation performance.

5. Analysis

5.1. Diffusion vs. Lens Defocus

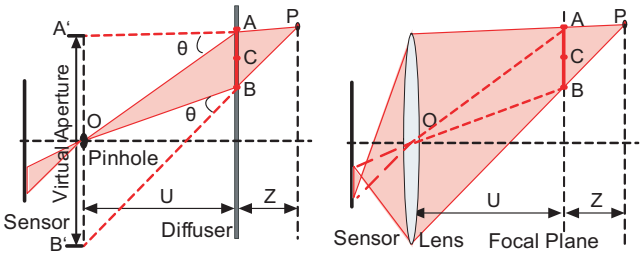
Diffusion caused by a diffuser can be shown to be geometrically equivalent to lens defocus. Figure 6(a) shows a pinhole camera with a diffuser placed in front of the scene point P, perpendicular to the optical axis. From the perspective of P, the pinhole O appears like a large aperture $A'B'$ which collects a cone $A'PB'$ of light from P. It should be noted that if we replace the pinhole with a lens of size $A'B'$, set the focus at the diffuser plane, and remove the diffuser as shown in (b), P will have the same projection AB on the focus plane, mapping to the same PSF on the sensor plane.

From Figure 6(a), we can see the size of the virtual aperture $A'B' = \frac{U+Z}{Z} \cdot AB$. We can compute AB from Equation 4, giving

$$A'B' = 2\tan\theta \cdot U/\cos^2\alpha. \tag{9}$$

When α is small, $A'B' = 2\tan\theta \cdot U$. For instance, a DFDiff system which consists of a pinhole camera and a 5° pillbox diffuser placed $1m$ away is equivalent to a DFD system whose lens has a huge aperture (diameter$= 17.5cm$) and is focused at $1m$.

While it is often expensive or even impossible to manufacture large lenses, it is relatively easy to make large diffusers with large diffusion angles. Several companies now



(a) Diffusion (b) Defocus

Figure 6. Equivalence between diffusion and lens defocus. The diffusion (a) caused by a diffuser in a pinhole camera is equivalent to the defocus (b) in a regular lens camera which has a large lens of size $A'B'$ and is focused at the diffuser plane.

supply off-the-shelf optical diffusers with diffusion angles ranging from $0.2°$ to $80°$ [18][19]. Because a diffuser effectively increases the lens aperture without physically increasing lens size, DFDiff provides an economical alternative for applications that require high precision in depth estimation.

5.2. Depth Sensitivity

In depth from stereo, the disparity r is used to compute the depth Z [15][13][14]. The derivative of r with respect to Z, is often referred to as depth sensitivity $S = \partial r/\partial Z$. Usually, we have $S \approx B \cdot V/U^2 = m \cdot B/U$, where B is the baseline, U is the distance to the object, V is the distance from the lens to the sensor, and m is the image magnification. The higher the depth sensitivity is, the more precise is the depth estimation.

DFD can also be regarded as a triangulation-based method, as shown in [9]. The aperture size in DFD plays the same role as the baseline B in stereo vision. We can thus apply the depth sensitivity analysis used in stereo vision to a DFD system as follows:

$$S \approx m \cdot B/U = m \cdot D/U, \qquad (10)$$

where D is the aperture diameter. For any given magnification m, the sensitivity is proportional to the aperture size D and inversely proportional to the distance U.

A DFDiff system is equivalent to a DFD system with aperture size $D \approx 2U \cdot \tan\theta$ (Equation 9) when α is small. Therefore, we have $S \approx m \cdot 2\tan\theta$, where θ is the diffusion angle of the diffuser. For any given magnification m, the sensitivity only relies on θ.

To increase the depth sensitivity with DFD, one has to either increase the aperture size of the lens, which may be prohibitively expensive, or move the camera closer to the object, which reduces the field of view (FOV). However, for DFDiff, it is easy to achieve high depth precision at a large distance, even with a low-end lens.

5.3. Sensitivity, Distance, and Field of View

Suppose we have a Canon EOS 20D D-SLR camera, whose sensor has a dimension of $22.5mm \times 15mm$ 8 microns pixel size, and we have a target object of size $225mm \times 150mm$. Table 1 shows the required F# or

FOV	U	S	EFL	DFD		DFDiff
mm×mm	*mm*	*pixel/mm*	*mm*	*F#*	*D (mm)*	θ
225×150	500	10	50	**0.125**	**400**	21.80°
225×150	500	1	50	1.25	40	2.29°
225×150	500	0.1	50	12.5	4	0.23°
225×150	1000	10	100	**0.125**	**800**	21.80°
225×150	1000	1	100	**1.25**	**80**	2.29°
225×150	1000	0.1	100	12.5	8	0.23°
225×150	5000	10	500	**0.125**	**4000**	21.80°
225×150	5000	1	500	**1.25**	**400**	2.29°
225×150	5000	0.1	500	12.5	40	0.23°

Table 1. Comparison of DFD and DFDiff for different depth precision requirements and object distances. On the left are FOV, object distance, and depth sensitivity that we want to achieve; on the right are the required EFL, F# or aperture size D in DFD and diffusion angle θ in DFDiff. In bold are lenses required by DFD which are too complicated to manufacture (e.g. a $500mm$ focal length lens with $4m$ diameter aperture).

Aperture diameter, D, in DFD, and the required diffusion angle θ in DFDiff for different depth precision requirements ($10\ pixel/mm, 1\ pixel/mm, 0.1\ pixel/mm$) and object distances ($500mm, 1000mm, 5000mm$). To ensure that the field of view (FOV) covers the whole object, the effective focal length (EFL) is increased with object distance. For example, the first row shows that if a depth precision of $10\ pixel/mm$ is required, for an object placed $500mm$ from the camera, then DFD requires a lens with $EFL = 50mm$ and F# = 0.125 ($D = 400mm$). DFDiff, on the other hand, can estimate depth with the same precision using any lens when a $21.80°$ diffuser is used.

We can see that for high precision and large object distance requirements, DFD demands lenses with unreasonably large apertures (e.g. a $500mm$ focal length lens with $4m$ diameter aperture). These lenses are shown in bold. DFDiff, on the other hand, can estimate high-precision depth maps using lenses with small apertures.

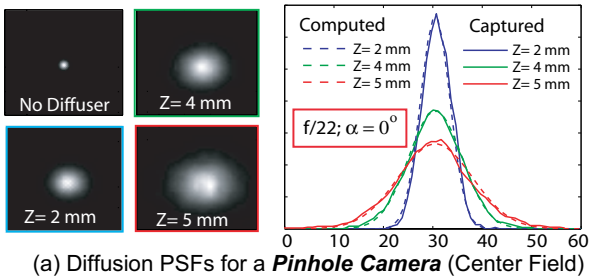
6. Experiments

Today, several companies sell off-the-shelf diffusers reproduced onto glass or plastic sheets up to 36" wide. In our experiments, we use holographic diffusers with Gaussian diffusion functions from Luminit Optics [18]. These diffusers have different diffusion angles, ranging from $0.5°$ to $20°$, and different sizes, ranging from $2'' \times 2''$ to $10'' \times 8''$. Their feature sizes change between 5 and 20 microns depending on the diffusion angles. In each experiment, the proper diffuser was chosen according to the scene and precision requirements.

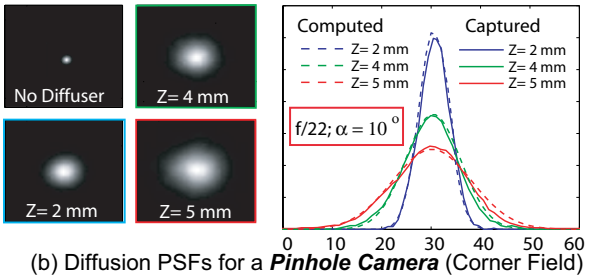
6.1. Model Verification

6.1.1 Pinhole Camera

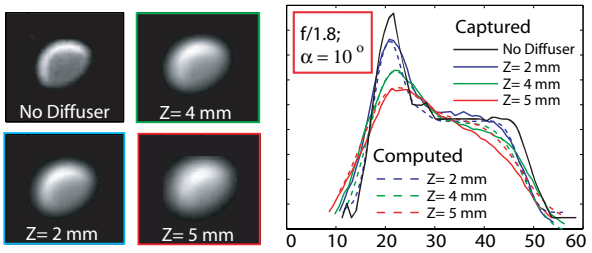
We first conducted experiments to verify the image formation model derived in Section 3. An array of point light sources was placed $1m$ in front of a Canon EOS T1i D-SLR camera with a Canon EF $50mm$ F/1.8 lens, perpendicular



(a) Diffusion PSFs for a ***Pinhole Camera*** (Center Field)



(b) Diffusion PSFs for a ***Pinhole Camera*** (Corner Field)



(c) Defocus+Diffusion PSFs for a ***Lens Camera*** (Corner Field)

Figure 7. Model Verification. (a) Captured and computed diffusion PSFs of a center point source in a pinhole camera. (b) Captured and computed diffusion PSFs of a corner point source ($\alpha = 10°$) in a pinhole camera. (c) Captured and computed diffusion defocus+diffusion PSFs of a corner point source ($\alpha = 10°$). We can see that in all these three cases, the PSFs computed using our derived diffuser model (dashed curves) are fairly consistent with the captured ones (solid curves). Note that the defocus pattern in (c) is asymmetric because of lens aberrations.

to the optical axis. First, to emulate a pinhole camera, we stopped down the aperture size to F/22. We mounted a $10°$ Luminit diffuser to a high-precision positioning stage, placing it just in front of the point light source array. We then captured a set of images while slowly moving the diffuser away from the light source array ($Z = 2mm - 10mm$).

Figure 7(a) left shows a focused image of the center point light source captured without a diffuser. On the right we show three images captured with a diffuser placed at different positions ($2mm, 4mm,$ and $5mm$). These three blurred images should be a convolution between the focused image and the three corresponding diffusion PSFs. Cross sections of the blurred images are plotted in solid curves on the right of Figure 7(a). Since the diffusion function of the diffuser and the distances Z are known, we can compute the diffusion PSFs according to our proposed imaging model. We then compute three diffused images by convolving these computed PSFs with the focused image. These three computed images are plotted in dashed curves. Figure 7(b) shows the captured images of a point light at the corner field ($\alpha = 10°$), as well as a comparison with the computed images.

We can see from both Figure 7 (a) and (b) that the computed images are quite consistent with the captured ones. This indicates the real diffusion PSFs not only fit the designed patterns well, but also are spatially invariant.

6.1.2 Lens Camera

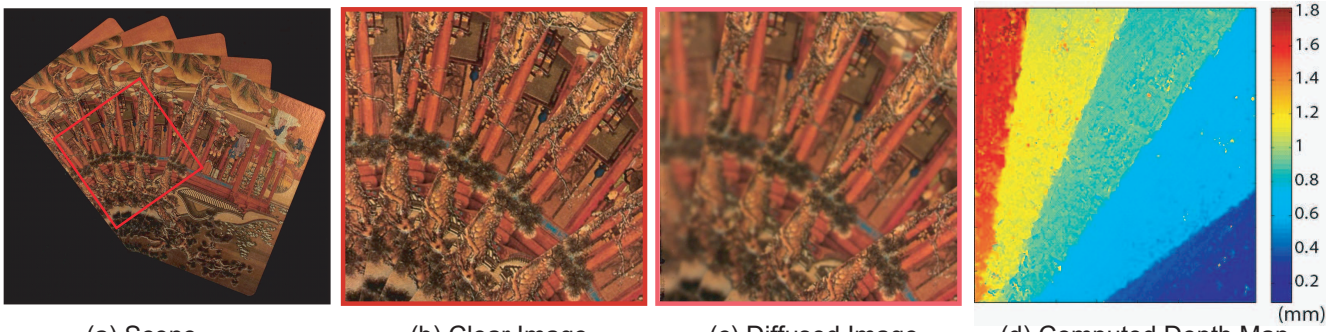
To verify the proposed imaging model in the presence of defocus, we open up the aperture of the lens to F/1.8, focus the camera at a distance of $1.9m$, and repeat the same experiment as in Section 6.1. Figure 7(c) left shows the defocused image of a corner point source ($\alpha = 10°$) captured without a diffuser. On the right we show three diffused and defocused images that were captured with the diffuser placed at different depths. We computed the diffusion PSFs from our diffusion model and convolved them with the defocused image captured without a diffuser. The computed diffused and defocused images are plotted in Figure 7(c) (dashed curves).

In Figure 7(b), note that although the aperture pattern of this Canon lens is circular, the captured defocus pattern is not circular at the periphery of the FOV, due to lens aberrations. The defocus PSF variation with field position will degrade the estimation precision of DFD. Meanwhile, we can see the plots of computed PSFs in Figure 7 are fairly consistent with the captured PSFs (solid curves). This verifies our derived Proposition 3.2 and confirms that the proposed DFDiff does not rely on the shape of defocus PSFs (Equation 8). This property relaxes requirements on the camera lens and enables high precision depth estimation with small, low-end lenses .

6.2. Depth from Diffusion: D-SLR Camera

Figure 8 shows an example where we use the proposed DFDiff method to estimate the depth map of an artificial scene. Five playing cards are arranged as shown in Figure 8(a). Each card is only $0.29mm$ thick. To estimate depths, we captured an image using a Canon EOS 20D D-SLR camera with a Canon EF $50mm$ F/1.8 lens. The distance was set to be $500mm$, which approaches the minimal working range of this camera. The camera was focused at the plane of cards. Note that for this setting, the depth of field is about $6mm$, far larger than the scene depth, and therefore all the cards are in focus. A clear image taken without a diffuser is shown in (b). Then, we placed a $20°$ Luminit Gaussian diffuser just in front of the first card and captured a diffused image, as shown in Figure 8(c). From these two captured images, DFDiff recovers a high-precision depth map, as shown in (d).

According to Equation 9, by using the diffuser, we have effectively created a huge virtual lens with F# = 0.12, 15 times larger than the F# of the actual lens. Note that for a



(a) Scene (b) Clear Image (c) Diffused Image (d) Computed Depth Map

Figure 8. Recovered depth map of five playing cards, each of which is $0.29mm$ thick. (a) An overview of the scene. (b) A captured image without a diffuser. (c) A captured image with a 20° Gaussian diffuser. (d) The recovered depth map which has a precision $\leq 0.1mm$

regular $50mm$ F/1.8 lens, the depth of field is $6mm$, much larger than the required depth precision. Therefore, DFD cannot be used effectively in this setting.

6.3. Depth from Diffusion: Consumer-level Camera

DFDiff imposes fewer restrictions on the camera lens, so that a low-end consumer camera can be used to estimate a high-precision depth map. Figure 9 shows a small sculpture of about $4mm$ thickness. For this experiment, we used a Canon G5 camera with a $28.8mm$ F/4.5 lens and a diffusion of 5° angle. The camera was set up $300mm$ away from the object. The captured focused and diffused images are shown in (b) and (c), respectively. From these two images, we compute the 3D structure of the sculpture of precision $\leq 0.25mm$, as illustrated in (d) and (e). To achieve the same precision in the same scene setting, DFD requires a much larger lens ($F\# \approx 0.5$).

6.4. Large Field of View Depth from Diffusion

For our last example, we use DFDiff to recover a depth map of a larger object ($650mm \times 450mm$), as shown in Figure 10(a). To recover the shape of the stars and stripes on this object, depth precision of at least $1mm$ is required. A Canon EOS 20D D-SLR camera with a $50mm$ lens is mounted on a tripod placed $2000mm$ from the object, so that the FOV covers the whole object.

We use a 10° diffuser of size $250mm \times 200mm$, which is smaller than the object. Multiple images are captured to cover the whole FOV and the diffuser is scanned sequentially over the curved surface. First, from each diffused image, we compute one depth map relative to an unknown diffuser plane. The relative positions and orientations of these diffuser planes can be easily calibrated by fitting the overlapping depth maps. We can then stitch all these depth maps into one, as shown in Figure 10. Three close-ups of the depth map are shown in (c). Note the viewpoint is not changed during the capture process, making the process of stitching straightforward.

If the same camera and lens were used to perform conventional DFD, the camera would have to be moved much closer to the object ($< 100mm$) to achieve a similar precision, which would greatly reduce the FOV. To capture the entire object with a DFD system, one would have to move the camera and capture many more images. However, uncalibrated movements introduce significant difficulties in both capturing (controlling the focus or aperture settings) and processing (aligning captured data), which may also degrade the precision of depth estimation.

7. Conclusion and Discussion

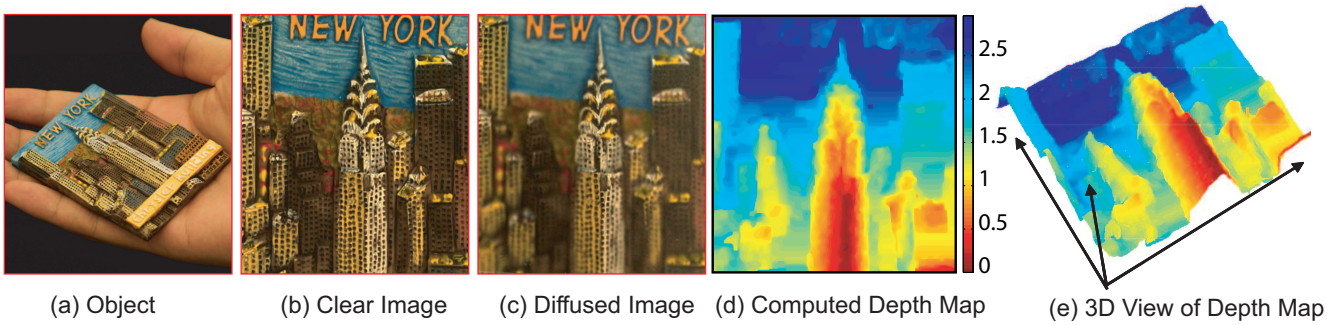
In this paper, we have demonstrated that optical diffusers can be used to perform high-precision depth estimation. In contrast to conventional DFD, which either requires a prohibitively large aperture lens or small lens-to-object distances which restricts the FOV, DFDiff relaxes requirements on the camera lens and requires only larger diffusion angles, which are much cheaper to manufacture. Even a low-end consumer camera, when coupled with the proper diffuser, can be used for high-precision depth estimation.

One of the beneficial properties of the DFDiff technique is that depth estimation is measured relative to a proxy object instead of a camera lens, which introduces more flexibility in the acquisition process. However, this same property is also a major drawback since it requires a diffuser to be placed near objects being photographed, which is not possible in many situations.

In our implementation, we have chosen diffusers with Gaussian diffusion functions for simplicity. Diffusers with a variety of diffusion functions are currently commercially available [18][19]. An interesting question that warrants further investigation is: "What is the optimal diffusion function for depth estimation?". For simplicity, we have used a typical DFD algorithm, which requires two input images. Another interesting topic for further research is how to design diffusers and algorithms that enable depth estimation using only a single image.

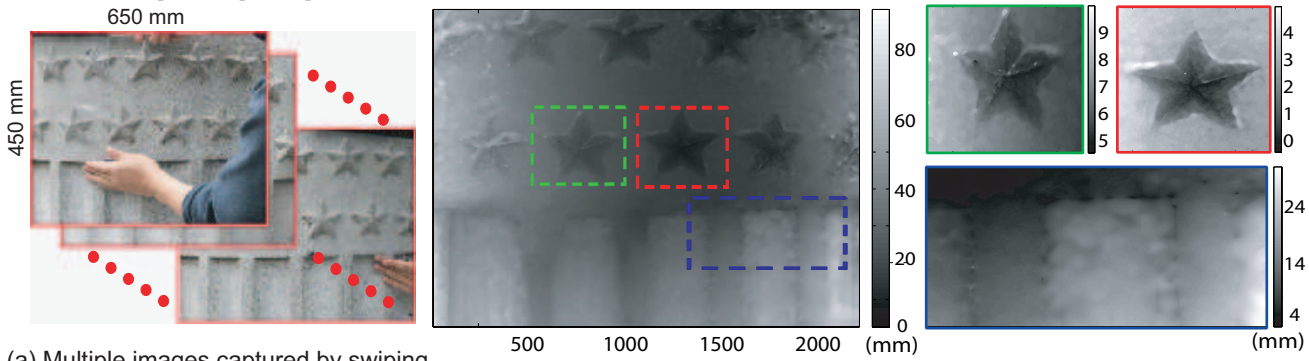
Acknowledgements:

This research was funded in part by ONR awards N00014-08-1-0638 and N00014-08-1-0329.



(a) Object (b) Clear Image (c) Diffused Image (d) Computed Depth Map (e) 3D View of Depth Map

Figure 9. DFDiff results for a thin sculpture captured using a Canon G5 camera. (a) Wide view of the sculpture. (b) A clear image without a diffuser. (c) An image captured using a 5° Gaussian diffuser. (d) The computed depth map which has a precision $\leq 0.25mm$. (e) A 3D view of the computed depth map.



(a) Multiple images captured by swiping the diffuser over the surface (b) Computed large FOV depth map (c) Close-ups of the depth map

Figure 10. A $650mm \times 450mm$ sculpture with stars on a curved surface. (a) Ten diffused images captured by swiping 10° diffuser over the surface. A Canon EOS 20D camera with a $50mm$ lens was placed $2000mm$ from the object; (b) The computed and stitched large FOV high-precision depth map with precision $\leq 1mm$. (d) Three close-ups of the depth map

References

[1] K. Mori. Apparatus for uniform illumination employing light diffuser, July 17 1984. US Patent 4,460,940. 1

[2] G.M. Mari-Roca, L. Vaughn, J.S. King, K.W. Jelley, A.G. Chen, and G.T. Valliath. Light diffuser for a liquid crystal display, February 14 1995. US Patent 5,390,085. 1

[3] P. Beckmann and A. Spizzichino. The scattering of electromagnetic waves from rough surfaces. *New York*, 1963. 1

[4] H.M. Smith. Light scattering in photographic materials for holography. *Applied Optics*, 11(1):26–32, 1972. 1

[5] P.F. Gray. A method of forming optical diffusers of simple known statistical properties. *Journal of Modern Optics*, 25(8):765–775, 1978. 1

[6] S. Chang, J. Yoon, H. Kim, J. Kim, B. Lee, and D. Shin. Microlens array diffuser for a light-emitting diode backlight system. *Optics letters*, 31(20):3016–3018, 2006. 1

[7] A.P. Pentland. A new sense for depth of field. *Radiometry*, page 331, 1992. 1, 2

[8] S. Chaudhuri and A.N. Rajagopalan. *Depth from defocus: a real aperture imaging approach*. Springer, 1999. 1, 2

[9] Y.Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *IJCV*, 39(2):141–162, 2000. 1, 2, 5

[10] G. Surya and M. Subbarao. Depth from defocus by changing camera aperture: a spatial domainapproach. In *CVPR*, pages 61–67, 1993. 1, 2

[11] A. Ashok and M.A. Neifeld. Pseudorandom phase masks for superresolution imaging from subpixel shifting. *Applied optics*, 46(12):2256–2268, 2007. 1

[12] E.E. Garcia-Guerrero, E.R. Méndez, H.M. Escamilla, T.A. Leskova, and A.A. Maradudin. Design and fabrication of random phase diffusers for extending the depth of focus. *Optics Express*, 15(3):910–923, 2007. 1

[13] U.R. Dhond and J.K. Aggarwal. Structure from stereo-a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989. 2, 5

[14] R.S. Allison, B.J. Gillam, and E. Vecellio. Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, 9(1):10, 2009. 2, 5

[15] S.T. Barnard and M.A. Fischler. Computational stereo. *ACM Computing Surveys (CSUR)*, 14(4):553–572, 1982. 2, 5

[16] A. Levin, R. Fergus, F. Durand, and W.T. Freeman. Image and depth from a conventional camera with a coded aperture. In *Proc. ACM SIGGRAPH*, 2007. 4

[17] C. Zhou, S. Lin, and S. Nayar. Coded Aperture Pairs for Depth from Defocus. In *ICCV*, 2009. 4

[18] Luminit optics. http://www.luminitco.com/. 4, 5, 7

[19] RPC photonics. http://www.rpcphotonics.com/. 5, 7