Role of optics in the accuracy of depth-from-defocus systems

Ilya Blayvas, Ron Kimmel, and Ehud Rivlin

Computer Science Department, Technion–Israeli Institute of Technology, Haifa 32000, Israel

Received August 10, 2006; revised October 3, 2006; accepted November 5, 2006;
posted November 22, 2006 (Doc. ID 73824); published March 14, 2007

The depth-from-focus–defocus approach to 3D reconstruction is based on the fact that objects closer to or farther from the object in focus appear blurred, and the amount of blur increases with the distance from the object in focus. An important characteristic of any depth-from-defocus system is the depth reconstruction accuracy. Several 3D reconstruction algorithms have been proposed, and the influence of image noise and image spectrum on the system accuracy has been studied. However, so far the effect of optics on the accuracy has not been fully explored. Here, we derive an expression estimating the system accuracy as a function of its optical parameters. It turns out that optics plays a major role in the accuracy, and tenfold increase of the lens focal length, and the aperture can increase the overall accuracy by a factor of more than 1000. The derived expression allows one to review several results, revealing that the accuracy is defined primarily by the optics. We also provide guidelines for the design of new depth-from-defocus systems in compliance with predefined specifications by choosing the appropriate optics.

OCIS codes: 110.6880, 150.5670, 150.6910.

1. INTRODUCTION

Reconstruction of the geometric structure (3D) from 2D images is an important task in computer vision. Among other approaches for 3D reconstruction—commonly referred to as "depth-from-*X*," where *X* stands for stereo, shading, motion, structured light, etc.—depth-from-defocus and closely related depth-from-focus approaches have attracted substantial attention.[1–12] These approaches are based on the limited depth of view of the lens. When the lens is focused on a certain object, objects located closer to or farther from the object in focus appear blurred. The blur increases with the distance from the object in focus. This phenomenon allows one to estimate the geometry of a scene by measuring the amount of blur in an image. In a depth-from-defocus approach, two images with different defocus levels are used. The depth-from-focus approach relies on multiple images, focused at different distances evenly distributed through the depth range. This gives the depth-from-focus approach some advantage in accuracy over depth from defocus at the expense of a larger number of the acquired images. The derivations we present here are relevant to both approaches.

An important characteristic of depth from defocus is the reconstruction accuracy, which depends on many factors, such as camera settings,[13] number of observations,[14] image noise,[15] and spatial frequencies.[16] Numerous algorithmic ways to improve the reconstruction accuracy have been proposed.[17–23]

Among the papers tackling the role of optics, most relevant to our discussion are Refs. 18 and 24. Subbarao and Surya in Ref. 18 used a thin lens formula to derive a quadratic dependence of the reconstruction error on the distance L to the object. However, they did not estimate the magnitude of the error. Nayar *et al.* in Ref. 24 decomposed an image blur into separate defocus, sensor, and optical transfer function blurs. However, the blurring model was used to design a linear operator for computing the relative blur of the two images and not for the accuracy analysis.

Here, we follow Ref. 24 and decompose the blur into defocus, sensor, and optics blur. We extend the analysis by expressing the blur in terms of basic optical parameters and use the resulting model in order to estimate the system accuracy. The expression for the system accuracy confirms the quadratic dependence of the error on L, derived in Ref. 18.

The main contribution of this paper is an estimation of the system accuracy as a function of its optical parameters. We show that the system optics actually defines the reconstruction accuracy. For example, we show how the change of the lens focal length and the aperture by a factor of 10 can result in a change of depth resolution by a factor of more than 1000. Review of the published results from the perspective of this estimation confirms the fundamental role of the optics on the system accuracy.

2. ROLE OF OPTICS IN DEPTH ACCURACY

A. Spatial Domain Analysis

An image $i(\mathbf{x})$ captured by an image sensor is formed from a sharp preimage $s(\mathbf{x})$, blurred by the optics, the defocus, and the finite size of the pixels. It also includes some noise. Formally, the image can be described via the convolution of $s(\mathbf{x})$ with the point spread function $h(\mathbf{x})$, as

$$i(\mathbf{x}) = h(\mathbf{x}) * s(\mathbf{x}) + \epsilon(\mathbf{x}), \tag{1}$$

where $\mathbf{x}=\{x,y\}$ is a coordinate in the 2D image and $\epsilon(\mathbf{x})$ describes an additive noise.

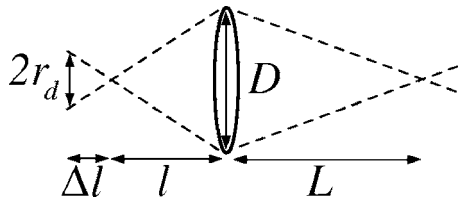

Fig. 1. Defocus spot size.

$h(\mathbf{x})$ can be considered as a convolution of the optical blur $\mu(\mathbf{x})$, the defocus blur $\eta(\mathbf{x})$, and the sampling blur $\rho(\mathbf{x})$ kernels[24]:

$$h(\mathbf{x}) = \int\int \rho(\mathbf{x}-\mathbf{s})\,\eta(\mathbf{s}-\mathbf{t})\,\mu(\mathbf{t})\mathrm{d}\mathbf{s}\mathrm{d}\mathbf{t}. \tag{2}$$

Accurate modeling of $\mu(\mathbf{x})$, $\eta(\mathbf{x})$, $\rho(\mathbf{x})$ is technically difficult and requires a special software and a detailed physical model of the lens and the sensor,[25] which are usually unavailable.

The popular measure of lens quality is the modulation transfer function, which is defined as the ratio of relative image contrast divided by relative object contrast, where the object is the sinusoidally varying brightness at some spatial frequency. The MTF depends on the location on the image plane and decreases with increase of the spatial frequency. Despite the relation of MTF to the lens resolution, the blur kernel $\mu(\mathbf{x})$ cannot be derived from it. Usually, as the F number of the lens increases from low to high, the blur of the lens first decreases due to the improvement of the optical wavefront, and then increases due to increasing diffraction blur.[26] The blur for the lower end of the F-number range varies among different lenses, while the blur for the higher end of the F-number range of most decent lenses approaches the diffraction limit,[26,27]

$$\mu(\mathbf{x}) = \left[\frac{2J_1(\gamma)}{\gamma}\right]^2. \tag{3}$$

Here, $\gamma=(\pi|\mathbf{x}|D)/(\lambda l)$, J_1 is a Bessel function,[28] $\lambda=0.7\times10^{-6}$ m is the wavelength of the light, l is the distance from the lens to the image plane, and D is the lens diameter (see Fig. 1). In practical cases $l\approx f$, where f is the focal length of the lens. Using the definition of the F number, $F\equiv f/D$, we can write $\gamma\approx(\pi|\mathbf{x}|)/(\lambda F)$.

The defocus kernel has a cylindrical shape given by

$$\eta(\mathbf{x}) = \begin{cases} \dfrac{1}{\pi r_d^2}, & |\mathbf{x}| \le r_d \\ 0, & |\mathbf{x}| > r_d \end{cases}, \tag{4}$$

(Refs. 29 and 30) where the radius of the cylinder (see Fig. 1) is given by

$$r_d = \left|\frac{\Delta l}{l}\frac{D}{2}\right| = \left|\frac{\partial l}{\partial L}\frac{\Delta L}{l}\frac{D}{2}\right| = \left|\frac{f^2}{L^2}\frac{\Delta L D}{l}\right| \approx \left|\frac{f^2}{L^2}\frac{\Delta L}{F}\right|. \tag{5}$$

Here, the derivative $\partial l/\partial L$, is obtained from the thin lens formula,[29] connecting the focal length f, the distance L to the object, and the distance l to the focused image of the object:

$$\frac{1}{l}+\frac{1}{L}=\frac{1}{f}. \tag{6}$$

Finally, the sampling kernel describes averaging over a square pixel of size Δx:

$$\rho(\mathbf{x}) = \begin{cases} \dfrac{1}{\Delta x^2}, & \max(|x_1|,|x_2|) \le \dfrac{\Delta x}{2} \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

Now consider an image in focus $I(\mathbf{x})=\rho*\mu*s(\mathbf{x})$ and a defocused image $J(\mathbf{x})=\rho*\eta*\mu*s(\mathbf{x})$. (See Fig. 2.) The minimal depth resolution ΔL is the depth change, at which the difference between the image in focus and the defocused image becomes distinguishable.

Figure 3 presents results of computer simulation of the diffraction and defocus blurs. The sharp preimage $s(x)$ has a steplike profile, shown as a dashed curves. The focused image of this profile, blurred only by the optics diffraction with $\lambda F=1$ is shown by a solid curve $I=\mu(1)*s(x)$. The dotted curves show the defocused images, blurred by defocuses with $r_d=1$ and $r_d=2$; $J_1=\eta(1)*\mu(1)*s(x)$, and $J_2=\eta(2)*\mu(1)*s(x)$.

Arbitrary small defocus blur will result in some difference between the focused and defocused blurs. In practical cases, such difference is obscured by the image quantization and noise. The difference between the images should exceed the quantization and noise level, in order to be reliably detectable. The defocus might be detectable also when the difference between the images is below the noise level. However, rigorous analysis of the minimum distinguishable difference between two noisy images seems to be a somewhat unrelated task. Actually the field of image denoising in image processing introduces many solutions, each one tailored for a specific scenario. Noise analysis requires assumptions about the nature of the noise and the image contents and is beyond the scope of this paper.[13,15]

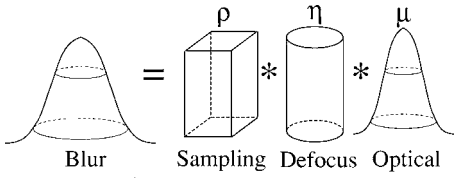

Fig. 2. Overall blur kernel is formed by convolution of the optical, defocus, and sampling kernels.

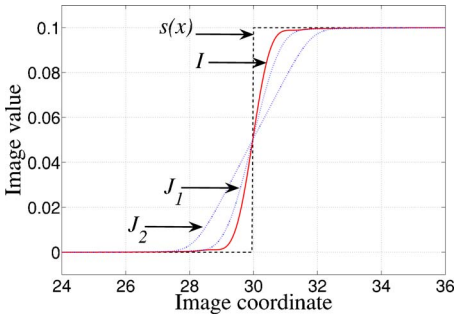

Fig. 3. (Color online) Step on the sharp preimage, smoothed by the optics blur and further smoothed by defocus.

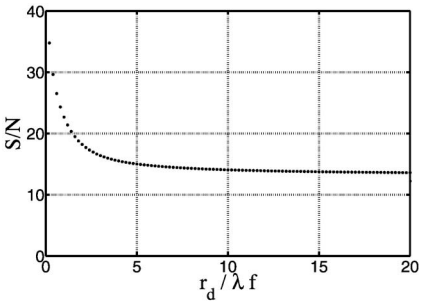

Fig. 4. Ratio between the distinguishable defocus spot r_d and the optics blur λF as a function of the signal-to-noise ratio.

For the sake of simplicity, we assume that the noise exceeds the image quantization level and that the minimal detectable difference between the images must be equal to the noise level. Furthermore, we assume that the sharp preimage has features equivalent to a step with the magnitude 10% of the full intensity range. These assumptions may seem too pessimistic or the contrary, but even if they are false, all the further conclusions would still remain qualitatively the same, although the numbers will be offset by a constant factor.

The maximum difference between the focused and defocused images, shown in Fig. 3 are 0.0113 for an image blurred with $\eta(r_d=1)$ and 0.0264, for an image blurred with $\eta(r_d=1)$. They will become distinguishable from the focused image at the $S/N=15.8$ dB and $S/N=19.5$ dB, respectively.

For any ratio $r_d/(\lambda F)$ there is a corresponding noise level at which the defocus becomes detectable. Figure 4 shows $r_d/(\lambda F)$ as a function of the signal to noise of the image. The simulation was performed for an image with a step of 10% of the intensity range.

B. Fourier Domain Analysis

To take into account the sampling blur (image pixelization) and derive an analytical estimate for the resolution as a function of optical parameters, we proceed to the analysis in the Fourier domain. In the Fourier domain, any blur is a low-pass filter, that cuts off or significantly suppresses high frequencies, starting from some characteristic frequency $\nu \approx 1/\zeta$, where ζ is the scale of the blur kernel in the spatial domain. In the previous section, we have seen that at the reasonable signal-to-noise ratios, the scales of defocus and optical blurs must be comparable. This implies that in the Fourier domain their characteristic frequencies must be comparable.

In the Fourier domain, the diffraction blur is transformed into

$$\mathcal{F}(\mu(\mathbf{x})) = \begin{cases} F^2(\gamma - \sin\gamma), & \sqrt{u^2+v^2} \leqslant \dfrac{2}{\lambda F} \\ 0, & \sqrt{u^2+v^2} > \dfrac{2}{\lambda F} \end{cases}, \tag{8}$$

where u and v denote spatial-frequency parameters in the x and y directions, respectively, and $\gamma = 2\arccos(\lambda F\sqrt{u^2+v^2}/2)$.[24] The defocus blur then reads

$$\mathcal{F}(\eta(\mathbf{x})) = \frac{L^2 F}{2\pi f^2 \Delta L\sqrt{u^2+v^2}} J_1\left(\frac{2\pi f^2 \Delta L}{L^2 F}\sqrt{u^2+v^2}\right) \tag{9}$$

(Refs. 24 and 27). Finally, the sampling kernel blur is given by

$$\mathcal{F}(\rho(\mathbf{x})) = \frac{1}{2\pi}\frac{1}{\Delta x^2}\operatorname{sinc}\frac{u\Delta x}{2}\operatorname{sinc}\frac{v\Delta x}{2}. \tag{10}$$

Now, consider two images I and J. Let us assume that the first image is in focus, and the second one is out of focus. Then, $I(\mathbf{x}) = h(\mathbf{x}, L) * (s(\mathbf{x}) + \epsilon_1(\mathbf{x}))$ and $J(\mathbf{x}) = h(\mathbf{x}, L + \Delta L) * (s(\mathbf{x}) + \epsilon_2(\mathbf{x}))$. In the Fourier domain, we have

$$\begin{aligned} \mathcal{F}(I) &= \mathcal{F}(\rho(\mathbf{x}))\mathcal{F}(\mu(\mathbf{x}))\mathcal{F}(s(\mathbf{x}) + \epsilon_1(\mathbf{x})), \\ \mathcal{F}(J) &= \mathcal{F}(\rho(\mathbf{x}))\mathcal{F}(\eta(\mathbf{x}))\mathcal{F}(\mu(\mathbf{x}))\mathcal{F}(s(\mathbf{x}) + \epsilon_2(\mathbf{x})). \end{aligned} \tag{11}$$

The minimal depth resolution $\Delta L_{\min}$ is the value, at which the difference between $\mathcal{F}(I)$ and $\mathcal{F}(J)$ is detectable. The images I and J are blurred, and therefore they have a band-limited spectrum with characteristic frequencies ν_I and ν_J. We can say that $\nu_I \approx \min\{\nu_\rho, \nu_\mu, \nu_s\}$, $\nu_J \approx \min\{\nu_\eta, \nu_\rho, \nu_\mu, \nu_s\}$, where $\nu_{\eta,\rho,\mu,s}$ denotes the characteristic frequencies of the band-limited kernels η, ρ, μ, and s. For the image in focus I, the defocus kernel η, given by Eq. (4) approaches a delta function, and does not limit the bandwidth of I, which is defined by diffraction and sampling blurs. As the image J gets out of focus, the defocus radius r_d increases, and the defocus blur first becomes comparable and then even exceeds the diffraction and sampling blurs. At this moment, the second image becomes more blurred than the first one. Therefore, the images I and J become distinguishable when the defocus blur exceeds the diffraction and sampling blurs. In the frequency domain, this means that the characteristic frequency of the band-limited defocus kernel becomes lower than characteristic frequencies of diffraction and defocus blurs: $\nu_\eta \approx \min\{\nu_\rho, \nu_\mu, \nu_s\}$.

Here, we assume that the resolution is optically limited and not limited by the absence of high spatial frequencies in the image: $\nu_s > \{\nu_\rho, \nu_\mu\}$. In this case, the defocus blur becomes distinguishable when

$$\nu_\eta \approx \min\{\nu_\rho, \nu_\mu\}, \tag{12}$$

which can be rewritten as

$$\frac{1}{\nu_\eta} \approx \max\left\{\frac{1}{\nu_\rho}, \frac{1}{\nu_\mu}\right\} \quad \Rightarrow \quad \frac{1}{\nu_\eta} \approx \sqrt{\frac{1}{\nu_\rho^2} + \frac{1}{\nu_\mu^2}}. \tag{13}$$

From Eqs. (8)–(10), describing the spectrums of the kernels, we can estimate their characteristic frequencies as $\nu_\eta = L^2 F/(2\pi f^2 \Delta L_{\min})$, $\nu_\rho = 2/(\Delta x)$, and $\nu_\mu = 2/(\lambda F)$. Substituting into Eq. (13), we obtain

$$\frac{2\pi f^2 \Delta L_{\min}}{L^2 F} \approx \sqrt{\left(\frac{\Delta x}{2}\right)^2 + \left(\frac{\lambda F}{2}\right)^2}, \tag{14}$$

and thereby

$$\Delta L_{\min} \approx \frac{L^2 F}{2\pi f^2}\sqrt{\left(\frac{\Delta x}{2}\right)^2+\left(\frac{\lambda F}{2}\right)^2}. \quad (15)$$

Again, we would like to stress that $\Delta L_{\min}$ measures the depth shift at which the defocus blur becomes comparable to the diffraction and sampling blurs. $\Delta L_{\min}$ can serve only as an estimate for the resolution of the system. In the case of low image noise, feature rich image, and efficient processing algorithm, the resolution can, in principle, exceed $\Delta L_{\min}$. In the case of noisy and textureless image, the resolution can be lower than $\Delta L_{\min}$.

Substituting some typical numbers into Eq. (15), such as $\{f=20$ mm, $F=16$, $L=1$ m, $\Delta x=5$ μm$\}$, we obtain $\Delta L_{\min}=39$ mm. Replacing the lens for one with ten times longer focus and opening the aperture by a factor of 10, such as $\{f=200$ mm, $F=1.6$, $L=1$ m, $\Delta x=5$ μm$\}$, we obtain $\Delta L_{\min}=0.016$ mm. Thus, we have shown that the change of the lens can improve the accuracy by a factor of $\sim$2500.

In the literature, the popular figure of merit of an algorithm is the relative error $\Delta L_{\min}/L$. We can see from Eq. (15) that

$$\frac{\Delta L_{\min}}{L} = \frac{LF}{2\pi f^2}\sqrt{\left(\frac{\Delta x}{2}\right)^2+\left(\frac{\lambda F}{2}\right)^2} \quad (16)$$

depends on L, f, F, and Δx—the optical parameters of the system, and therefore the measure $\Delta L_{\min}/L$, often used in publications that explore the depth-from-defocus problem, does not provide an optics-invariant comparison between the algorithms.

C. Depth from Defocus versus Depth from Focus

The above analysis refers to the case where one image is focused, and the second one is defocused, which is a depth-from-focus approach. In depth from defocus, both images are defocused. In that case, the achievable accuracy degrades. Figure 5 shows a step profile, blurred by different defocuses, with $r_d \in [0.2, 20]$. Figure 6 shows the maximum derivative between adjacent profiles, $\max(\eta(r_d+\Delta r_d)\mu(1)s(x)-\eta(r_d)\mu(1)s(x))/\Delta r_d$. One can see that after a prompt increase, near $r_d \approx 0$, the difference decays from its maximum value of 0.02, near $r_d \approx 2$ to $\sim$0.005 and below near $r_d \sim 20$ and beyond. This figure illustrates, that the resolution of the depth-from-defocus approach gradually degades with increasing defocus, but remains proportional to the resolution of the depth-from-focus approach.

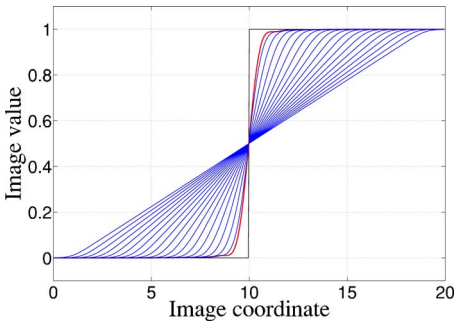

Fig. 5. (Color online) Steplike profile with different defocus blurs.

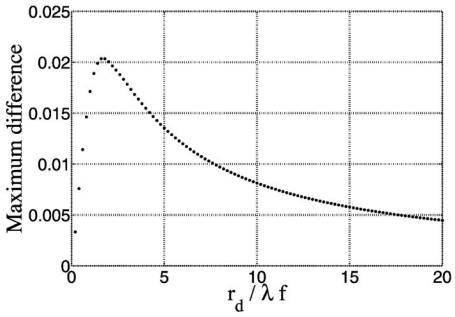

Fig. 6. Difference between adjacent defocused images, as a function of defocus.

3. REVISITING DEPTH RECONSTRUCTION FROM DEFOCUS

Most of the depth-from-defocus publications evaluate their accuracy as an rms error divided by the distance range. As one can see from Eq. (16), this measure depends on the optics. We would like to compare reported results in an optics-invariant way, by dividing each reported accuracy by its corresponding $\Delta L_{\min}$, derived from the reported optical parameters of the system. The ratio between the reported accuracy ΔL_{rep} and its optical limit $\Delta L_{\min}$ will be referred to as the accuracy factor, $\alpha \equiv \Delta L_{\text{rep}}/\Delta L_{\min}$. Estimation of $\Delta L_{\min}$ requires knowledge of L, f, F, and Δx. Unfortunately, some of these parameters are sometimes missing from the reports.

Among the relevant publications that we could not evaluate here due to the missing optical parameters are[24] with microscopic depth from defocus, where L, f, Δx are missing[31]; where L, ΔL, and Δx are missing[18]; and[32] where L, f, and Δx are missing; and [23] where Δx is missing. Actually the pixel size Δx is not reported in most of the papers, but in some cases it can be recovered from the documentation of the camera.

Subbarao and Choi in Ref. 17 proposed to use the focused image surface in order to reconstruct the 3D shape of an object. They derived and experimentally confirmed that $\Delta L \sim L^2$, in compliance with Eq. (15). The optical parameters were $f=35$ mm, $F=4$, $L=1$ m, and $\Delta x=13$ μm, while $\Delta L_{\text{rep}}=3$ cm. Substituting the optical parameters into Eq. (15), we obtain $\Delta L_{\min}=3.5$ mm. Therefore, for this algorithm, the accuracy factor is $\alpha=8.6$.

Subbarao and Surya in Ref. 18 apply a spatial domain transform to extract the depth information from two images, each with different camera parameters. In their system, $f=35$ mm, $F=4$, $L=2$ m, $\Delta x=13$ μm, and $\Delta L_{\text{rep}}=0.1$ m. Substituting these figures into Eq. (15), we obtain $\Delta L_{\min}=13.8$ mm, thereby $\alpha=7.24$.

Ens and Lawrence in Ref. 19 calculate the depth information from two defocused images, acquired with two different F numbers. Defocusing by change of the F number preserves l and allows one to exclude an unwanted scaling of the image. The blur is treated as a convolution of a sharp image $s(x,y)$ with a low-pass filter $h(x,y)$. A less blurred image $i_1(x,y)=s(x,y)*h_1(x,y)$ is acquired with a defocused system with higher F number F_1 and a more blurred image $i_2(x,y)=s(x,y)*h_2(x,y)$ with $F_2<F_1$. Then, the inverse problem is solved to find a blurring function $h_3(x,y)$, which transforms $i_1(x,y)$ into $i_2(x,y)$, $i_2(x,y)$

$=i_1(x,y)*h_3(x,y)$. In the Fourier domain, this deconvolution problem translates into a simple division:

$$F(h_3(x,y)) = \frac{F(i_2(x,y))}{F(i_1(x,y))}. \quad (17)$$

The one-to-one relation between $h_3(x,y)$ and the depth is derived from the geometric optics or found from a look-up table, evaluated on a calibration object.

A Javelin JE2062 CCD camera was used with a focal length of $f=50$ mm. The pixel size Δx was not reported. We could not recover the sensor type used in this camera, but from the camera focal length, field of view, and number of pixels, the pixel size can be estimated to be $\Delta x \approx 50$ μm. The two different F numbers were $F_1=1.3$ and $F_2=2.0$. The distance to the object was in the range of 80–95 cm, where we use an average of $L=87.5$ cm. Substituting these figures into Eq. (15), we obtain $\Delta L_{\min}=1.8$ mm.

The authors present several solutions to the inverse filtering problem (17), obtaining an rms error of 6.8% for consistent inverse filtering, 1.7% for iterative matrix solution, and 1.3% using experimentally measured blurring operators. For the average distance $L=87.5$ cm this translates into $\Delta L_{\text{rep}}=60$, 15, 11 mm, respectively. Thus, for the best case, the accuracy factor is $\alpha=11$ mm/1.8 mm=6.1.

Watanabe and Nayar in Ref. 20 used a near- and far-focused images to extract the depth information. Let us denote by I_{near} and I_{far} a nonzero frequency of the images in the spatial-frequency domain. Then, the expression $(I_{\text{near}}-I_{\text{far}})/(I_{\text{near}}+I_{\text{far}})$ is a monotonic function of distance in between the focused range of near-focused image [L_{near} and the focused range of the far-focused image L_{far}]. To increase the processing speed, a small set of broadband filters derived by precisely modeling an image blur was used.

A Sony XC-77 camera was used with Cosmicar B1214D-2 lens. The optical parameters were $f=25$ mm, $L=70$ cm, $\Delta x=11$ μm, and $F=8.3$. For this optical setup $\Delta L_{\min}=9.2$ mm, while $\Delta L_{\text{rep}}=5$ mm. Thus, the algorithm accuracy factor was $\alpha=5/9.2=0.54$, which means that the the algorithm detects the defocus blur when it is still below the diffraction and sampling blurs. Such good accuracy can be explained by the fact that for each of the two focus settings, 256 images were averaged to reduce noise. Averaging over 256 images before the linear processing of depth reconstruction is equivalent to reconstructing first and then averaging of 256 depth maps, which increases the reconstruction accuracy.

Xiong and Shafer in Ref. 21 used a depth from focus approach, when the distance to the point is estimated by focusing on it. They used a flat black–white step edge as an artificial target to measure the depth resolution. Combination of the Fibbonacci search and curve fitting was used to detect the focus peak. The focal length was $f=130$ mm, the distance to the object $L\approx 1.2$ m, the F number $F=1.7$, and the pixel size $\Delta x=23$ μm. $\Delta L_{\text{rep}}=1.18$ mm, while the accuracy estimated by Eq. (15) for this system is $\Delta L_{\min}=0.22$ mm. Therefore, $\alpha=1.18/0.22=5.36$.

Baba *et al.*[22] estimated the depth accounting for zoom, focus, F number, and the lens center transition. The authors used SONY XC-007 camera with the pixel size $\Delta x=50$ μm and a Canon J16x9.5B4RAS lens. The focal lengths were $f=9.5-152$ mm, used at the 40–130 mm range, the $F=1.8$, $L\in\{1,1.5,2,2.5,3\}$ m. The depth accuracy was $\Delta L_{\text{rep}}\in\{22,45,71,93,112\}$ mm, respectively, while the corresponding accuracies predicted by Eq. (15) for this system, with $F=70$ mm, are $\Delta L_{\min}\in\{1.5,3.3,5.8,9.1,13.2\}$ mm. The corresponding accuracy factors are $\alpha_i=\{15,13.7,12.1,10.2,8.5\}$, respectively.

Table 1 summarizes the results reported in the reviewed papers, with focal length f, distance to the object L, F number F, sensor pixel size Δx, and the reported accuracy of depth estimation ΔL_{rep}, all lengths in millimeters. The two right columns are the corresponding accuracy bound $\Delta L_{\min}$ and the algorithm accuracy factor $\alpha=(\Delta L_{\text{rep}})/(\Delta L_{\min})$.

4. SUMMARY

We have shown that the optical parameters of a depth-from-defocus system—the focal length, the distance to the object, the F number, and the sensor pixel size—have a crucial role in the accuracy. A change of the focal length and the F number by a factor of 10 each can increase the system depth resolution by a factor of more than 1000.

Thus, comparing the published methods by reported accuracy does not allow separating between the role of optics and the algorithm. For example, the absolute resolution of five out of the six published results reviewed in this paper varied by a factor of 95 from 1.18 to 112 mm, their relative accuracies $\Delta L_{\text{rep}}/L$ varied by a factor of 37 from 0.1% to 3.7%, while the resolution relative to the corresponding accuracy estimate varied only by a factor of 1.5. Therefore, normalization of the system resolution by the accuracy estimate $\Delta L_{\min}$, given in Eq. (15), reveals a relatively modest role of the reconstruction algorithm on the system performance. Equation (15) estimating the system accuracy as a function of optical parameters can help to design depth-from-defocus systems in compliance with a given accuracy specification.

Table 1. Summary of Reviewed Depth-from-Defocus Publications (Lengths in Millimeters)

Reference	f	L	F	Δx	ΔL_{rep}	$\Delta L_{\min}$	α
17	35	1000	4	0.013	30	3.5	8.6
18	35	2000	4	0.013	100	13.8	7.24
19	50	875	1.3	0.050	11	1.8	6.1
20	25	700	8.3	0.011	5	9.2	0.54
21	130	1200	1.7	0.023	1.18	0.22	5.36
22	70	3000	1.8	0.050	112	13.2	8.5

I. Blayvas's e-mail address is blayvas/ron/ehudr @cs.technion.ac.il.

REFERENCES

1. P. Favaro, A. Menucci, and S. Soatto, "Observing shape from defocused images," Int. J. Comput. Vis. **52**, 25–43 (2003).
2. P. Favaro and S. Soatto, "A geometric approach to shape from defocus," IEEE Trans. Pattern Anal. Mach. Intell. **27**, 406–415 (2005).
3. M. Gokstorp, "Computing depth from out-of-focus blur using a local frequency representation," in *Proceedings of International Conference on Pattern Recognition* (IEEE, 1994), pp. 153–158.
4. H. Jin and P. Favarao, "A variational approach to shape from defocus," in *Proceedings of European Conference on Computer Vision* (2002), pp. 18–30.
5. S. K. Nayar and Y. Nakagawa, "Shape from focus: an effective approach for rough surfaces," IEEE Trans. Pattern Anal. Mach. Intell. **16**, 824–831 (1994).
6. M. Noguchi and S. K. Nayar, "Microscopic shape from focus using active illumination," in *Proceedings of International Conference on Pattern Recognition* (IEEE, 1994), pp. 147–152.
7. A. P. Pentland, "A new sense for depth of field," IEEE Trans. Pattern Anal. Mach. Intell. **9**, 523–531 (1987).
8. A. P. Pentland, T. Darrel, M. Turk, and W. Huang, "A simple, real time range camera," in *Proceedings of International Conference on Computer Vision and Pattern Recognition* (IEEE, 1989), pp. 256–261.
9. A. N. Rajagopalan and S. Chaudhuri, "A block shift-variant blur model for recovering depth from defocused images," in *Proceedings of International Conference on Image Processing* (IEEE, 1995), pp. 636–639.
10. S. Soatto and P. Favaro, "A geometric approach to blind deconvolution with application to shape from defocus," in *Proceedings of International Conference on Computer Vision and Pattern Recognition* (IEEE, 2000), pp. 10–17.
11. D. Ziou, "Passive depth from defocus using spatial domain approach," in *Proceedings of International Conference on Computer Vision* (IEEE, 1998), pp. 799–804.
12. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: how different really are they," in *Proceedings of International Conference on Pattern Recognition* (IEEE, 1998), pp. 1784–1786.
13. A. N. Rajagopalan and S. Chaudhuri, "Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters," Int. J. Comput. Vis. **30**, 175–190 (1998).
14. A. N. Rajagopalan, S. Chaudhuri, and R. Chellappa, "Quantitative analysis of error bounds in the recovery of depth from defocused images," J. Opt. Soc. Am. A **17**, 1722–1731 (2000).
15. M. Subbarao and J. K. Tyan, "Noise sensitivity analysis of depth-from-defocus by a spatial-domain approach," Proc. SPIE **3174**, 174–187 (1994).
16. Y. Schechner and N. Kiryati, "The optimal axial interval in estimating depth from defocus," in *Proceedings of International Conference on Computer Vision* (IEEE, 1993), pp. 843–848.
17. M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," IEEE Trans. Pattern Anal. Mach. Intell. **17**, 266–274 (1995).
18. M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," Int. J. Comput. Vis. **13**, 271–294 (1994).
19. J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," IEEE Trans. Pattern Anal. Mach. Intell. **15**, 97–108 (1993).
20. M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," Int. J. Comput. Vis. **27**, 203–225 (1998).
21. Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," in *Proceedings of International Conference on Computer Vision and Pattern Recognition* (IEEE, 1993), pp. 68–73.
22. M. Baba, N. Asada, A. Oda, and T. Migita, "A thin lens based camera model for depth estimation from defocus and translation by zooming," in *Proceedings of ICVI* (2002), pp. 274–281.
23. O. Ghita and P. Whelan, "A video-rate sensor based on depth from defocus," Opt. Laser Technol. **33**, 167–176 (2001).
24. S. K. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," IEEE Trans. Pattern Anal. Mach. Intell. **18**, 1186–1198 (1996).
25. A. E. Savakis and H. J. Trussel, "On the accuracy of PSF representation in image restoration," IEEE Trans. Image Process. **2**, 252–259 (1993).
26. W. J. Smith, *Modern Lens Design* (McGraw-Hill, 1992), pp. 43–49.
27. M. Born and E. Wolf, *Principles of Optics* (Pergamon, 1975), pp. 395–397.
28. G. Arfken, *Mathematical Methods for Physicists* (Academic, 1985), pp. 573–636.
29. E. Hecht and A. Zajac, *Optics* (Addison-Wesley, 1974).
30. B. K. P. Horn, *Robot Vision* (MIT Press, 1986).
31. F. Deschenes, D. Ziou, and P. Fuchs, "Simultaneous computation of defocus blur and apparent shifts in spatial domain," in *Actes de 15th International Conference on Vision Interface* (1992), pp. 236–243.
32. A. P. Pentland, S. Scherock, T. Darrel, and B. Girod, "Simple range cameras based on focal error," J. Opt. Soc. Am. A **11**, 2925–2934 (1994).