Error Bounds in Depth from Defocus

A.N. Rajagopalan[1] S. Chaudhuri[2] Rama Chellappa[1]

Center for Automation Research[1]
University of Maryland
College Park, MD - 20742, USA

Department of Electrical Engineering[2]
Indian Institute of Technology
Powai, Bombay - 400 076, India

Abstract

Depth from defocus (DFD) involves estimating the relative blur between a pair of defocused images of a scene captured with different lens settings. When a priori information about the scene is available, it is possible to estimate the depth even from a single image. However, experimental studies indicate that the depth estimate improves with multiple observations. In this paper, we provide a mathematical underpinning to this evidence by deriving and comparing the theoretical bounds for the error in the estimate of blur corresponding to the case of a single image, and a pair of defocused images, respectively. A new theorem is proposed which proves that the Cramér-Rao bound on the variance of the error in the estimate of blur, decreases with an increase in the number of observations. The difference in the bounds turns out to be a function of the relative blurring between the observations. Results on synthetic as well as real data are given to validate the claim.

1 Introduction

Of the several methods used for depth recovery: depth from stereo, structure from motion, shape from shading, structure from focus, and depth from defocus (DFD), only the last two approaches involve the principle of real aperture imaging, while all the other techniques assume an ideal pin-hole model of the camera. The blurring of image regions due to a finite depth of field is usually considered a liability. The degree of defocus, however, is a function of the lens settings and the depth of the scene. Hence, it is possible to recover depth from two defocused versions of a scene obtained with different sets of camera parameters. Precise knowledge of the camera parameters and accurate estimates of blur are necessary to obtain a good depth map of the scene. The accuracy of the DFD method is roughly comparable to that of methods based on stereo disparity or motion parallax. The depth map recovery process is parallel, and requires no search unlike depth from focus [1].

Although in its fullest generality, the blurring process in DFD is space-variant, simplifying assumptions, such as local space-invariance, are often made in the literature for reasons of mathematical tractability. In this paper also, the depth will be assumed to be locally constant. As early as 1982, Pentland [2] showed that if some a priori knowledge about the scene characteristics can be assumed, the depth can be recovered from even a single defocused image of the scene. In particular, he assumed a Gaussian kernel for the point spread function and showed how the relation between the Laplacian of the observed image and the spread of the Gaussian could be used for computing the depth. At about the same time, Grossman's work [3], which was mostly experimental, demonstrated that useful depth information can be obtained from blurred step and ramp edges. In yet another method [4], the depth was recovered from blurred edges for the more general case of rotationally symmetric point spread functions. Lai et al. have proposed a depth estimation algorithm [5] in which the raw image data in the vicinity of an edge is used to estimate the depth from a single image. In all the above approaches, which are based on the processing of a single image, depth estimation is restricted only to feature points such as edges.

Most of the works on DFD are, however, based on comparing two defocused versions of a scene captured with different camera parameters [6] - [12]. The only requirement for this approach is that the scene must have sufficient spectral content. The relative blur between the defocused images is measured to determine the depth of the scene. It may also be mentioned here that, in one of the earliest approaches to DFD, Pentland [6] had derived depth estimates by assuming one of the two images to be focused. Recently, attempts have been made to use an appropriate model for the focused image of the scene and then solve for the depth. For example, Surya and Subbarao [13] propose a spatial domain approach that approximates the original image by a cubic polynomial over local

regions. Rajagopalan and Chaudhuri [14] model the original focused image by an autoregressive (AR) process and the computation of depth from two defocused images is formulated as a maximum-likelihood (ML) estimation problem.

Although it is possible to recover the depth from even a single image, researchers in DFD have observed that an improved estimate of depth can be obtained by using multiple observations [6, 7, 14, 15]. In this paper, we provide a theoretical justification for the improvement in the accuracy of the estimate of depth with multiple defocused images. We give an elegant proof in support of this evidence in terms of the Cramér-Rao lower bound (CRLB), under fairly general conditions. The original focused image is modeled as an AR process, and the CRLBs for the variance of the error in the ML estimate of blur derived from a single blurred image and from two blurred images, are compared. A new theorem is proposed that not only establishes that the CRLB of the variance of the error in the DFD estimate of blur is indeed lower when two defocused images are used instead of one, but also gives the exact relationship between the two error bounds. Apart from the theoretical novelty, the practical significance of the theorem for the DFD problem, in terms of the choice of the camera parameters, is also confirmed by experiments.

2 Real Aperture Imaging

The point spread function (PSF) of the camera describes the image intensity caused by a single point light source. Geometric optics approximates the PSF in the spatial domain by simple ray tracing. An object point is imaged to a circular disk, referred to as the circle of confusion, on the image plane by the ray tracing model [8]. However, in practice, the image of a point object is not a crisp circular patch of constant brightness, as suggested by geometric optics. Instead, due to diffraction and lens aberrations, it will be roughly a circular blob with the brightness falling off gradually at its border [7]. For a diffraction-limited lens system, the point spread function (PSF) of the camera system may be approximately modeled [6, 16, 17] as a circularly symmetric 2-D Gaussian function $h(i,j) = \frac{1}{2\pi\sigma^2} \exp\frac{-(i^2+j^2)}{2\sigma^2}$, where the blur parameter σ is given by $\rho\ r_b$ and is related to the depth of the scene as

$$\sigma = \rho\ r_0\ v_0\ \left(\frac{1}{F_l} - \frac{1}{v_0} - \frac{1}{D}\right) \qquad (1)$$

where r_0 is the lens aperture, v_0 is the distance between the lens and the image plane, F_l is the focal length, and D is the depth. A positive value of the blur parameter indicates that the vertex of the blur cone is in front of the image plane, while a negative value indicates that the vertex is behind the image plane. Here ρ is a constant that depends on the particulars of the optics and the sampling resolution, and must be determined initially by an appropriate calibration procedure. Equation (1) suggests that an estimate of the blur parameter in conjunction with the values of the lens settings can be used to calculate the depth.

If the scene is thought of as a collection of point sources of light, then the intensity at a particular location in the image is the result of contributions from many overlapping defocused point sources. Since the blur parameter is a function of the depth of the scene, the PSF is also a function of depth. Hence, the blurring process is linear but space-variant in DFD. However, from a practical and computational standpoint, most methods assume local space-invariance in calculating the depth; the depth is assumed to be constant over any local region in the image. The observed image $g(i,j)$ can then be modeled as the output of a linear space-invariant system, which is characterized by its PSF $h(i,j)$. Here $h(i,j)$ refers to the PSF corresponding to the local region under consideration.

The noisy and blurred image $g(i,j)$ can then be expressed as

$$g(i,j) = \sum_{m,n\varepsilon\mathcal{S}_h} h(m,n) f(i-m, j-n) + w(i,j)\ , \qquad (2)$$

where $\mathcal{S}_h$ is the support of the PSF. The observation or sensor noise $w(i,j)$ is assumed to be an additive, zero-mean, white Gaussian process. It is more convenient to express the above relation in the frequency domain as

$$\overline{G} = \Lambda_H \overline{F} + \overline{W}\ , \qquad (3)$$

where $\overline{G}$ and $\overline{W}$ are the DFTs of the sequences $g(i,j)$ and $w(i,j)$, respectively. Matrix Λ_H is diagonal with entries that correspond to the DFT $H(k,l)$ of the sequence $h(i,j)$.

3 Depth from a Single Image

It must be noted that the image data are the result of the characteristics of both the scene and the lens system. To disentangle these factors, one possibility is to look for places in the image with known characteristics. For example, at edges, the rate of change in intensity in the image is due primarily to the point spread function. Since it is possible to recognize sharp discontinuities with some degree of confidence, the image data surrounding them can be used to determine

the blur, and hence the depth [6]. A limitation of this method, however, is that its application is restricted to special points such as step discontinuities. Instead, a priori knowledge about the scene may be incorporated by assuming a suitable model for the focused image as in [13, 14]. We use an AR model for this purpose. The focused image corresponding to a given scene can then be treated as a realization of this AR process. Computation of depth is posed as an ML estimation problem. The reasons for choosing the AR model are three-fold: (i) This model is fairly general and has been quite successfully used by researchers in the image processing community. (ii) For a reliable estimate of the depth, the DFD technique requires that the image have sufficient texture, and the AR model is well-suited for modeling texture patterns. (iii) The relationship among the dependent variables in the AR model is linear. This helps to preserve Gaussianity in the mathematical analysis.

Since the focused image $f(i,j)$ is modeled by a 2-D AR process,

$$f(i,j) = \sum_{m,n \varepsilon \mathcal{S}_a} a(m,n) f(i-m, j-n) + v(i,j) \ , \quad (4)$$

where the $a(i,j)$'s are the image model coefficients, and $\mathcal{S}_a$ is the support for the causal image model. The noise process $v(i,j)$ is assumed to be zero-mean, white Gaussian, and independent of $f(i,j)$. Correspondingly, in the frequency domain we get

$$\overline{F} = (I - \Lambda_A)^{-1} \overline{V} \ , \quad (5)$$

where $\overline{F}$ and $\overline{V}$ are the DFTs of the rasterized sequences $f(i,j)$ and $v(i,j)$, respectively. It is also assumed that the observation noise process $w(i,j)$ (in equation(2)) is independent of $f(i,j)$ and $v(i,j)$. Matrix Λ_A is diagonal with entries that correspond to the DFT $A(k,l)$ of the sequence $a(i,j)$.

Because all the transformations are linear, it is clear from equation (3) that $\overline{G}$ is also Gaussian. Let $\mathcal{E}$ denote the expectation, and $\mathcal{H}$ the Hermitian operator. For a sequence of length N, we have

$$p(G) = \frac{1}{(2\pi)^{\frac{N}{2}} (\det P)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} G^{\mathcal{H}} (P)^{-1} G \right) \ , \quad (6)$$

where $p(G)$ is the probability density function (pdf) of $\overline{G}$. The bar over G represents the process while G is a realization. It can be shown that

$$P = \mathcal{E}\left[\overline{G}\ \overline{G}^{\mathcal{H}}\right] = \Lambda_H \mathcal{E}\left[\overline{FF}^{\mathcal{H}}\right] \Lambda_H^{\mathcal{H}} + \mathcal{E}\left[\overline{W}\ \overline{W}^{\mathcal{H}}\right] \ ,$$

$$= \Lambda_H (I - \Lambda_A)^{-1} \mathcal{E}\left[\overline{V}\ \overline{V}^{\mathcal{H}}\right] (I - \Lambda_A)^{-\mathcal{H}} \Lambda_H^{\mathcal{H}} + \sigma_w^2 \ I$$

$$= \sigma_v^2 \ \Lambda_H (I - \Lambda_A)^{-1} (I - \Lambda_A)^{-\mathcal{H}} \Lambda_H^{\mathcal{H}} + \sigma_w^2 \ I \ .$$

Hence, P is a diagonal matrix and

$$P(k,k) = \frac{H(k) H^*(k)}{|1 - A(k)|^2} \sigma_v^2 + \sigma_w^2 \ , \ k = 0, 1, \ldots, N-1,$$

where '$*$' represents complex conjugate. Let $s \stackrel{\triangle}{=} \sigma^2$ where σ is the blur parameter. Since the blur is modeled by a Gaussian function, we have

$$P(k,k) = \frac{\exp\left(-\left(\frac{2\pi k}{N}\right)^2 s \right)}{|1 - A(k)|^2} \ \sigma_v^2 + \sigma_w^2 \ . \quad (7)$$

The unknown parameters to be estimated are given by $\theta = \{a(m), s, \sigma_w^2, \sigma_v^2\}$. From equation (6), the maximum likelihood estimate of θ may be equivalently expressed as

$$\min_{\theta} \quad F_1(\theta)$$
$$\text{where} \quad F_1(\theta) = \log(\det P) + G^{\mathcal{H}}(P)^{-1} G \ .$$

Because the matrix P is diagonal, the likelihood function is given by

$$F_1(\theta) = \sum_{k=0}^{N-1} \log(P(k,k)) + \frac{|G(k)|^2}{P(k,k)} \ . \quad (8)$$

The unknown parameter vector θ is found by minimizing $F_1(\theta)$ using a gradient descent technique. The estimate of s can then be used to calculate the depth from equation (1). It may be noted here that the above relationships are derived for 1-D signals for notational simplicity only. Extension to the 2-D case is straightforward.

3.1 Error Analysis for a Single Defocused Image

The Cramér-Rao bound (CRLB) provides a fundamental lower limit on the variance of the error attainable with an unbiased estimator for an unknown parameter. It expresses the minimum error variance of any estimator $\hat{x}(z)$ of x in terms of the conditional density $p(z|x)$ of the data [18]. Since the latter is usually not difficult to obtain in practical applications, the bound gives a convenient standard against which to judge the performance of any other estimator that one may wish to use.

We now derive the Cramér-Rao bound of the variance of error in s when a single defocused image is given. The CRLB of Var($\tilde{s}$) is given by

$$\mathcal{E}\left[\tilde{s}^2\right] \geq \frac{1}{-\mathcal{E}\left[\frac{\partial^2 \log p(G)}{\partial s^2}\right]} \ ,$$

where $\tilde{s} = s - \hat{s}$ and $\hat{s}$ is an estimate of s. We have from equation (6)

$$\log p(G) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{k=0}^{N-1}\left[\log P(k,k) + \frac{|G(k)|^2}{P(k,k)}\right]$$

Since $\mathcal{E}\left[G(k)G^*(k)\right] = P(k,k)$, it can be shown that

$$-\mathcal{E}\left[\frac{\partial^2 \log p(G)}{\partial s^2}\right] = \frac{1}{2}\sum_{k=0}^{N-1}\left[\frac{1}{P(k,k)}\frac{\partial^2 P(k,k)}{\partial s^2} - \frac{1}{(P(k,k))^2}\left(\frac{\partial P(k,k)}{\partial s}\right)^2 - P(k,k)\left(\frac{1}{(P(k,k))^2}\frac{\partial^2 P(k,k)}{\partial s^2} - \frac{2}{(P(k,k))^3}\left(\frac{\partial P(k,k)}{\partial s}\right)^2\right)\right]$$

Therefore,

$$\frac{1}{\mathcal{E}\left[\tilde{s}^2\right]} \;\leq\; \frac{1}{2}\sum_{k=0}^{N-1}\left[\frac{1}{(P(k,k))^2}\left(\frac{\partial P(k,k)}{\partial s}\right)^2\right]. \quad (9)$$

The partial derivative of $P(k,k)$ with respect to s can be derived from equation (7).

4 Depth from Two Defocused Images

Yet another methodology for estimating the depth is to observe what happens to the image when some aspects of the lens system (such as aperture, focal length or lens-to-image plane distance) are changed. The difference in blurring between two defocused images captured with different lens settings is a source of depth information. For two different lens settings, we have from equation (1)

$$\sigma_m = \rho r_m v_m \left(\frac{1}{F_{l_m}} - \frac{1}{v_m} - \frac{1}{D}\right) \;,\; m = 1, 2. \quad (10)$$

Here σ_m corresponds to the blur parameter in the m^{th} observation. By eliminating D from the above equations, one can obtain a relationship between σ_1 and σ_2 in terms of the lens settings [7].

Given two differently blurred and noisy images of the scene, we have in the frequency domain

$$\overline{G^i} = \Lambda_{H_i}\overline{F} + \overline{W^i} \;,\; i = 1, 2 \;.$$

In many of the techniques for DFD, the contribution of the scene to the two images is factored out. However, this assumes a noise-free situation, which is seldom true in practice. To account for the presence of observation/sensor noise, an ML-based method was proposed in [14] for estimating depth. In this section, we give a brief description of this method to enable performance comparison with the ML-based method for the single-image case.

Let the matrix $P^{i,j} = \mathcal{E}\left[\overline{G^i}\;\overline{G^j}^H\right]$. Let $\sigma_2 = \alpha\sigma_1$ where α depends on the lens settings and can be determined. We define $s \triangleq \sigma_1^2$, where σ_1 corresponds to the blur in the first defocused image. Using well-known matrix properties, the ML problem of estimating θ, given two defocused images of the scene, can be posed as [14]

$$\min_{\theta} \quad F_2(\theta)$$

$$\text{where} \quad F_2(\theta) = \sum_{k=0}^{N-1}\Big[\log P^{1,1}(k,k) - \log C_2(k,k) + \left|G^1(k)\right|^2 A_2(k,k) + 2Re\{G^{1^*}(k)B_2(k,k)G^2(k)\} + \left|G^2(k)\right|^2 C_2(k,k)\Big] \;. \quad (11)$$

Note that the unknown parameter vector θ remains the same as in the case of the single image because the relation between the blur parameters is known.

In the above expression, the quantities $A_2(k,k)$, $B_2(k,k)$, and $C_2(k,k)$ are given by

$$\begin{aligned} A_2(k,k) &= \frac{1}{P^{1,1}(k,k)} + \frac{|P^{1,2}(k,k)|^2 C_2(k,k)}{(P^{1,1}(k,k))^2} \;, \\ B_2(k,k) &= -\frac{P^{1,2}(k,k)C_2(k,k)}{P^{1,1}(k,k)} \;, \\ C_2(k,k) &= \frac{P^{1,1}(k,k)}{P^{1,1}(k,k)P^{2,2}(k,k) - |P^{1,2}(k,k)|^2} \end{aligned} \quad (12)$$

Since the PSF of the camera system is modeled by a Gaussian function, we have

$$P^{1,1}(k,k) = \frac{\exp\left(-\left(\frac{2\pi k}{N}\right)^2 s\right)}{|1 - A(k)|^2}\;\sigma_v^2 + \sigma_w^2 \;,$$

$$P^{1,2}(k,k) = P^{2,1}(k,k) = \frac{\exp\left(-\left(\frac{2\pi k}{N}\right)^2 \left(\frac{1+\alpha^2}{2}\right) s\right)}{|1 - A(k)|^2}\;\sigma_v^2 \;,$$

$$P^{2,2}(k,k) = \frac{\exp\left(-\left(\frac{2\pi k}{N}\right)^2 \alpha^2 s\right)}{|1 - A(k)|^2}\;\sigma_v^2 + \sigma_w^2 \quad (13)$$

The unknown parameters can be estimated by minimizing $F_2(\theta)$ using a gradient descent algorithm. The estimated value of s and the lens settings can then be used to infer the depth from equation (10).

4.1 Error Analysis for Two Images

In this section, we discuss briefly the Cramér-Rao lower bound on the variance of the error in s when two differently blurred images of the scene are given. Details of the derivation may be found in [19]. The CRLB of $\mathrm{Var}(\tilde{s})$ is given by

$$\mathcal{E}\left[\tilde{s}^2\right] \geq \frac{1}{-\mathcal{E}\left[\frac{\partial^2 \log p(G_2)}{\partial s^2}\right]} ,$$

$$\text{where} \quad \log p(G_2) = -N \log 2\pi - \frac{1}{2} F_2(\theta) ,$$

and $F_2(\theta)$ is given by equation (11).

After some mathematical exercise [19], one can show that

$$\frac{1}{\mathcal{E}\left[\tilde{s}^2\right]} \leq \frac{1}{2} \sum_{k=0}^{N-1} \Bigg[\frac{1}{P^{1,1}(k,k)} \frac{\partial^2 P^{1,1}(k,k)}{\partial s^2} - \frac{1}{(P^{1,1}(k,k))^2} \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 - \frac{1}{C_2(k,k)} \frac{\partial^2 C_2(k,k)}{\partial s^2} + \frac{1}{C_2^2(k,k)} \left(\frac{\partial C_2(k,k)}{\partial s}\right)^2 + P^{1,1}(k,k) \frac{\partial^2 A_2(k,k)}{\partial s^2} + 2P^{1,2}(k,k) \frac{\partial^2 B_2(k,k)}{\partial s^2} + P^{2,2}(k,k) \frac{\partial^2 C_2(k,k)}{\partial s^2} \Bigg] \quad (14)$$

5 Comparison of the Cramér-Rao Lower Bounds

We now address the following important issue: how does the CRLB of the variance of the error in the estimate of the blur parameter s of a single defocused image compare with that of the corresponding CRLB when two defocused images are used?

Theorem 1 *If $CRLB_1$ and $CRLB_2$ denote the Cramér-Rao bound on the variance of the error in the estimate of the blur given a single defocused image and two defocused images of a scene, respectively, then $CRLB_2 \leq CRLB_1$.*

Proof: Using the expression for $C_2(k,k)$ from equation (12), and combining the terms containing the second-order partial derivative of $C_2(k,k)$ in equation (14), the CRLB for the two-image case can be rewritten as

$$\frac{1}{E\left[\tilde{s}^2\right]} \leq \frac{1}{2} \sum_{k=0}^{N-1} \Bigg[\frac{1}{P^{1,1}(k,k)} \frac{\partial^2 P^{1,1}(k,k)}{\partial s^2} - \frac{1}{(P^{1,1}(k,k))^2} \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 + \frac{1}{C_2^2(k,k)} \left(\frac{\partial C_2(k,k)}{\partial s}\right)^2 + P^{1,1}(k,k) \frac{\partial^2 A_2(k,k)}{\partial s^2} + 2P^{1,2}(k,k) \frac{\partial^2 B_2(k,k)}{\partial s^2} + \frac{(P^{1,2}(k,k))^2}{P^{1,1}(k,k)} \frac{\partial C_2(k,k)}{\partial s^2} \Bigg] \quad (15)$$

From the relations in equation (12), the second-order partial derivatives of $A_2(k,k)$, $B_2(k,k)$, and $C_2(k,k)$ can be derived. After some mathematical exercise, it can be shown that

$$P^{1,1}(k,k) \frac{\partial^2 A_2(k,k)}{\partial s^2} + 2P^{1,2}(k,k) \frac{\partial^2 B_2(k,k)}{\partial s^2} = -\frac{1}{P^{1,1}(k,k)} \frac{\partial^2 P^{1,1}(k,k)}{\partial s^2} + 2 \frac{1}{(P^{1,1}(k,k))^2} \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 + 2 \frac{C_2(k,k)}{P^{1,1}(k,k)} \left(\frac{\partial P^{1,2}(k,k)}{\partial s}\right)^2 - \frac{(P^{1,2}(k,k))^2}{P^{1,1}(k,k)} \frac{\partial^2 C_2(k,k)}{\partial s^2} - 4 \frac{P^{1,2}(k,k)}{(P^{1,1}(k,k))^2} C_2(k,k) \frac{\partial P^{1,1}(k,k)}{\partial s} \frac{\partial P^{1,2}(k,k)}{\partial s} + 2 \frac{(P^{1,2}(k,k))^2}{(P^{1,1}(k,k))^3} C_2(k,k) \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2$$

Substituting the above expression for $P^{1,1}(k,k) \frac{\partial^2 A_2(k,k)}{\partial s^2} + 2P^{1,2}(k,k) \frac{\partial^2 B_2(k,k)}{\partial s^2}$ into equation (15) and cancelling common terms, we get

$$\frac{1}{E\left[\tilde{s}^2\right]} \leq \frac{1}{2} \sum_{k=0}^{N-1} \Bigg[\frac{1}{(P^{1,1}(k,k))^2} \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 + \frac{1}{C_2^2(k,k)} \left(\frac{\partial C_2(k,k)}{\partial s}\right)^2 + 2 \frac{C_2(k,k)}{P^{1,1}(k,k)} \left(\frac{\partial P^{1,2}(k,k)}{\partial s} - \frac{P^{1,2}(k,k)}{P^{1,1}(k,k)} \frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 \Bigg]$$

Hence, the variance of the error in the estimate of the blur parameter, given two defocused images of a scene, is lower bounded by

$$\mathcal{E}\left[\tilde{s}^2\right] \geq \frac{2}{\sum_{k=0}^{N-1} \left[\frac{1}{(P^{1,1}(k,k))^2} \left(\frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 + A'' \right]} , \quad (16)$$

where

$$A'' = \frac{1}{C_2^2(k,k)} \left(\frac{\partial C_2(k,k)}{\partial s}\right)^2 + \frac{2\ C_2(k,k)}{P^{1,1}(k,k)} \left(\frac{\partial P^{1,2}(k,k)}{\partial s} - \frac{P^{1,2}(k,k)}{P^{1,1}(k,k)} \frac{\partial P^{1,1}(k,k)}{\partial s}\right)^2 . \quad (17)$$

As derived earlier, for the single-image case, we have from equation (9)

$$\mathcal{E}\left[\tilde{s}^2\right] \;\geq\; \frac{2}{\sum_{k=0}^{N-1}\left[\frac{1}{(P(k,k))^2}\left(\frac{\partial P(k,k)}{\partial s}\right)^2\right]}\,, \qquad (18)$$

where $P = P^{1,1}$.

In the expression for A'', $P^{1,1}(k,k)$ is positive-valued (see 13). Also, $C_2(k,k)$ (which is given by (12)) is positive-valued because $P_2 = \begin{bmatrix} P^{1,1} & P^{1,2} \\ P^{2,1} & P^{2,2} \end{bmatrix}$ is a covariance matrix. Since all the other terms in A'' are squared quantities, it is easily verified that A'' is positive-valued. By comparing equations (16) and (18), and using the fact that A'' is positive, it is clear that

$$CRLB_2 \;\leq\; CRLB_1$$

as asserted in the theorem.

From Theorem 1, it is now clear that one can, indeed, get improved estimates of depth from multiple defocused images of a scene. Because $P^{1,2}(k,k)$ and $P^{2,2}(k,k)$ (as given by equation (13)) are functions of α, so is A''. Hence, the difference in the error bounds CRLB_1 and CRLB_2 is a function of the degree of the relative blur between the two defocused images.

Corrolary 1 *If $CRLB_M$ denotes the Cramér-Rao bound of the variance of the error in the estimate of the blur given M number of defocused images of a scene, then*

$$CRLB_M \;\leq\; CRLB_{M-1} \quad for \quad M \geq 2.$$

6 Experimental Results

In this section, we consider some examples and compare the Cramér-Rao lower bounds in the single-image and the two-image cases. Synthetic as well as real examples are used for experimentation. The importance of the result of Theorem 1 and its practical significance are further exemplified by actually computing and comparing the ML estimates of the blur for the single-image and the two-image cases.

We first consider the example of a synthetic image. A set of known AR parameters were chosen for the focused image process. The image size was taken to be 64×64 pixels and a second-order AR model was assumed. The value of the blurring parameter σ_1 was chosen as 2.5. The SNR for the noisy defocused image was 20 dB. Using the above values for the various parameters, the CRLB of $\text{Var}(\tilde{s})$ was calculated for the case of the single defocused image using equation (18). The value was found to be 0.228; it is shown dotted in Fig. 1. The CRLB corresponding to the case of two defocused images was next calculated for a range of values of α using equation (16); these values are plotted as a solid curve in Fig. 1. A comparison of the CRLBs corresponding to the single-image case and the two-image case clearly reveals that the CRLB for the two-image case is lower for all values of α. This observation is perfectly in accordance with the claim made in Theorem 1.

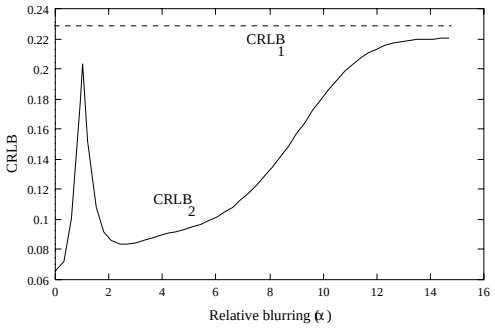

Figure 1: **An illustration of the CRLB of $\text{Var}(\tilde{s})$ for the single and the two-image case. The dotted and the continuous lines correspond to CRLB_1 and CRLB_2, respectively.**

Furthermore, it is interesting to note in Fig. 1 that when the two defocused images are not very different (i.e., $\alpha \approx 1$) or when the second defocused image is severely blurred (i.e., $\alpha >> 1$), the value of the CRLB for the two-image case is very close to that for the single-image case. It must be mentioned here that Ghiglia [20] had made a similar observation in the context of image restoration from multiple images. His experimental studies indicated that there was little to gain from multiple images that looked very similar to one another. The more different the images were, the better the restoration result was. It is very interesting to note that our quantitative analysis of the comparison of the CRLBs suggests a similar behavior for the depth from defocus problem. For the two-image case, it may also be noted that the CRLB is quite small for α close to zero. This is to be expected as it corresponds to the case when one of the two observations is focused. It is well-known that a high spectral content in the scene results in a good estimate of the depth.

Our next experiment illustrates the practical significance of the CRLB analysis. For this purpose, a focused image of the scene was synthetically generated with the same AR parameters that were used in the first experiment. The focused image thus obtained is shown in Fig. 2(a). A noisy defocused image was then generated by blurring this image with $\sigma_1 = 2.5$. The

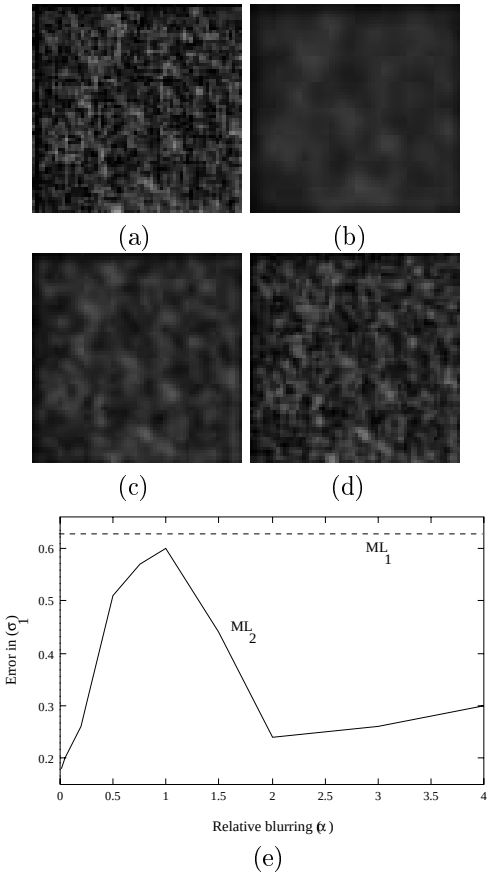

Figure 2: **(a) A synthetically generated focused AR image. (b,c,d) Defocused and noisy versions of the original image for different values of α. (e) Magnitude of the error in the ML estimate of σ_1. The dotted and the continuous lines correspond to the single and the two-image case, respectively.**

SNR was again chosen to be 20 dB, as in the previous experiment. The corresponding defocused observation is shown in Fig. 2(b). The ML estimate of σ_1 was obtained from the noisy defocused image by minimizing the likelihood function (equation (8)) corresponding to the single-image case. The magnitude of the error in the estimate of σ_1 was found to be 0.62 and is shown dotted in Fig. 2(e). The estimate is not very satisfactory because the blurring is severe. The blur parameter σ_1 was next estimated using a second defocused image of the scene. The second observation was generated for different values of α; some of these blurred and noisy images are shown in Fig. 2(c,d). The ML estimate of σ_1 was next computed by minimizing the likelihood function corresponding to the two-image case (equation(11)). The magnitude of the error in the estimate of the blur parameter is plotted in Fig. 2(e) for different values of α. It is very interesting to note that the plot in Fig. 2(e) corresponds well to what was predicted by the theoretical CRLB curve (Fig. 2) in the first experiment. Moreover, the CRLB plot reveals the very interesting fact that when capturing images in an experimental setup, one may want to use that value of α which yields a sufficiently large difference between the two CRLBs. Defocused images thus captured can result in a very good estimate of depth because the corresponding value of CRLB_2 is quite low for such an α. Thus, apart from the theoretical aspect, the practical importance of the CRLB analysis for the DFD problem is amply evident from this example.

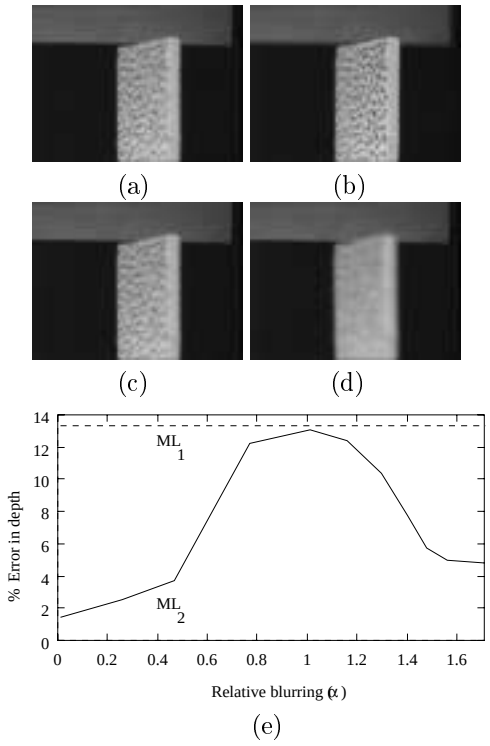

Figure 3: **(a,b,c,d) Defocused images corresponding to a real experimental setup for different focusing ranges of the camera. (e) Plot of the percentage error in the ML estimate of the depth for the nearest end of the textured planar object. The dotted and the continuous lines correspond to the single and the two-image case, respectively.**

Experiments were also conducted on real images. A Pulnix CCD camera was used for this purpose. The lens aperture was kept constant at an f-number of 4 and the focal length was 2.5 cm. A planar textured

object was placed in front of the camera. The resulting variation in depth was linear, as shown in Fig. 3. Note that for a good estimate of the depth, there must be sufficient texture (spectral content) in the scene. The nearest point of the object was about 80 cm from the camera while the farthest point was at 100 cm. The camera was coarsely calibrated with a different object at a known depth. A defocused image of the scene was captured with respect to a focusing range of 140 cm (Fig. 3(a)). Since, the variation in the depth of the scene was gradual, the blur was assumed to be locally space-invariant. The depth corresponding to the nearest end of the object was estimated using a local window. The percentage error in the ML estimate of the depth was found to be 13.4 and this is shown dotted in Fig. 3(e). A second defocused image was next captured for different values of the focusing range, which varied from 80 cm to 3 m. Some of the defocused images of the object are shown in Fig. 3(b,c,d). The ML estimate of the depth for the nearest point of the object was again computed, but now with two defocused images by again minimizing the corresponding likelihood function. The magnitude of the error in the estimate of the depth is plotted in Fig. 3(e) for different values of α. Again, it is interesting to note that the nature of the plot in Fig. 3(e) is as expected from the result of Theorem 1. A second observation gives a better estimate of depth as long as the relative blurring between the first and the second observation is reasonably large.

7 Conclusions

We have discussed maximum-likelihood estimation of depth from a single image and two defocused images of a scene. The error bounds on the estimates of depth corresponding to these two cases were compared and it was proved, under fairly general conditions, that the Cramér-Rao bound on the variance of the error in the estimate of blur becomes smaller as additional observations are used. The actual difference in the error bounds was computed and it turned out to be a function of the relative blur of the defocused images. The experimental results were found to conform quite well with theoretical expectations.

References

[1] E.P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*, (Springer-Verlag, NY, 1989).

[2] A.P. Pentland, "Depth of scene from depth of field", *Proc. Image Understanding Workshop* (Palo Alto, CA), 1982, pp. 253-259.

[3] P. Grossman, "Depth from focus", *Pattern Recognition Letters*, vol. 5, pp. 63-69, 1987.

[4] M. Subbarao and N. Gurumoorthy, "Depth recovery from blurred edges", *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* (Ann Arbor, MI), 1988, pp. 498-503.

[5] S. Lai, C. Fu and S. Chang, "A generalized depth estimation algorithm with a single image", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 405-411, 1992.

[6] A.P. Pentland, "A new sense for depth of field", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 9, pp. 523-531, 1987.

[7] M. Subbarao, "Parallel depth recovery by changing camera parameters", *Proc. IEEE Intl. Conf. on Computer Vision* (Florida), 1988, pp. 149-155.

[8] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 97-107, 1993.

[9] M. Gökstorp, "Computing depth from out-of-focus blur using a local frequency representation", *Proc. IEEE Intl. Conf. on Pattern Recognition*, 1994, pp. 153-158.

[10] M. Watanabe and S.K. Nayar, "Minimal operator set for passive DFD", *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* (San Francisco, CA), 1996, pp. 431-438.

[11] Y.Y. Schecner and N. Kiryati, "Depth from defocus vs Stereo: How different really are they?", *Technical report EE PUB No. 1155*, Dept. of Electrical Engg., Technion - Israel Institute of Technology, 1998.

[12] D. Ziou, "Passive depth from defocus using a spatial domain approach", *Proc. IEEE Intl. Conf. on Computer Vision* (Bombay, India), 1998, pp. 799-804.

[13] G. Surya and M. Subbarao, "Depth from defocus by changing camera aperture: A spatial domain approach", *Proc. IEEE Intl. Conf. on Comp uter Vision and Pattern Recognition* (New York), 1993, pp. 61-67.

[14] A.N. Rajagopalan and S. Chaudhuri, "Optimum camera parameter settings for recovery of depth from defocused images", *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition* (San Juan, PR), 1997, pp. 219-224.

[15] W.N. Klarquist, W.S. Geisler and A.C. Bovik, "Maximum-likelihood depth-from-defocus for active vision", *Proc. IEEE Intl. Conf. on Intelligent Robots and Systems* (Pittsburgh, PA), 1995.

[16] M. Born and E. Wolf, *Principles of Optics*, Pergamon, London, 1965.

[17] W.F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer-Verlag, 1986.

[18] J.M. Mendel, *Lessons in Digital Estimation Theory*, Prentice-Hall, Englewood Cliffs, 1987.

[19] 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", *Intl. J. Comput. Vision*, vol. 30, pp. 175-190, 1998.

[20] D.C. Ghiglia, "Space-invariant deblurring given N independently blurred images of a common object", *J. Opt. Soc. Amer.*, vol. 1, pp. 398-402, 1984.