

An Investigation of Methods for Determining Depth from Focus

John Ens and Peter Lawrence

Abstract—The concept of *depth from focus* involves calculating distances to points in an observed scene by modeling the effect that the camera's focal parameters have on images acquired with a small depth of field. This technique is passive and requires only a single camera. The most difficult segment of calculating depth from focus is deconvolving the defocus operator from the scene and modeling it. Most current methods for determining the defocus operator employ inverse filtering. This paper will reveal some fundamental problems with inverse filtering: inaccuracies in the finding the frequency domain representation, windowing effects, and border effects. A general, matrix-based method using regularization will be presented, which eliminates these problems. The new method will also be confirmed experimentally, with the results showing an rms error of 1.3%.

Index Terms—Depth from focus, range detection.

I. INTRODUCTION

A. Depth From Focus

THE CONCEPT of *depth from focus* involves calculating distances to points in an observed scene by modeling the effect that the camera's focal parameters have on images acquired with a small depth of field. This technique avoids the *correspondence* and *occlusion problems* [8]. It is passive and requires only a single camera.

The problem addressed by this paper is as follows: Given a blocks scene with featured objects an unknown distance from the camera but within known range bounds, derive the distance to features in the scene using the concept of depth from focus. This will be attempted in a setting that is as generalized as possible.

B. Criteria for the Most General Solution

A key motivation of this paper is to find a general solution to the depth from focus problem, rather than a solution that takes advantage of special features that may be found in a particular application. The characteristics of the most general solution are outlined below.

1) *Least Constraints on Scene*: It is desirable to place as few constraints on the scene as possible. Therefore, as an

example, it is *not* desirable to do any of the following:

- Introduce any structured illumination into the scene since this illumination may be confused with features in the scene.
- Assume that the edges in the scene are perfect black-to-white transitions since this may not be the case.
- Require one image to be taken with a pin-hole aperture since this in turn requires a very sensitive camera or high scene illumination.

It will be necessary, however, to require that the minimum and maximum distances from the camera to objects in the scene be known and that the scene contain visual features such as edges.¹

2) Method Independent of Defocus Operator Model:

Although the defocus operator has to be modeled, it is desirable that the method of solving the depth-from-focus problem be *independent* of any particular model. Some researchers have simplified the problem of depth from focus by restricting their solutions to a particular defocus operator. The Gaussian function advocated by Pentland [20]–[22], [19], seems to be the most common in the literature. As observed by Subbarao [30], however, the Gaussian may not be an appropriate model. Therefore, a general method for depth from focus should be independent of the model used for the defocus operator.

3) *Selective Control of Tradeoffs*: It is inevitable that the solution to the problem will involve tradeoffs, which change the nature of the solution. To keep the solution general, however, it is desirable that the tradeoffs remain explicit and adjustable by the user.

C. Theory of Defocus

1) *Geometric Optics*: The *geometric optics* model for defocusing relies on ray tracing and results in a first-order approximation. It is explained by Horn [15], [16].

From Fig. 1, similar triangles yield a formula for the radius R of the blur circle

$$R = \frac{L\delta}{2D_{fi}} \quad (1)$$

where

- L is the diameter of the lens or the aperture
- D_{fi} is the distance from the lens to a sharply focused image of a particular object

¹ These requirements are not unusual for a passive system of depth perception.

Manuscript received October 1, 1990; revised November 15, 1991. This work was supported by the Natural Sciences and Engineering Research Council of Canada Grant A-4924 and the Science Council of British Columbia G.R.E.A.T. Award 36(GC-10). Recommended for acceptance by Associate Editor E. Grimson.

J. Ens is with Intense Technologies, Richmond, Canada V6X 3C8.

P. Lawrence is with the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada V6T 1W5.

IEEE Log Number 9203758.

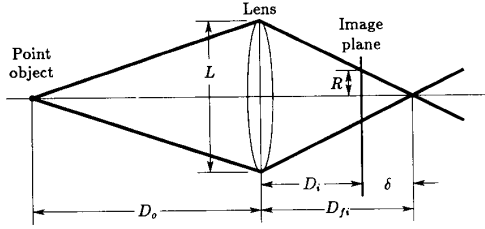


Fig. 1. Geometric optics model of defocus.

- δ is the displacement of the image plane from sharp focus.²

To a first-order approximation, the brightness within the blur circle is uniform. The defocus operator $h(x, y)$, where x and y are coordinates in the image plane, is defined as a pillbox of radius R

$$h(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{if } x^2 + y^2 > R^2 \end{cases} \quad (2)$$

Blurring due to defocus can be modeled as convolution with this pillbox [16].

A few manipulations can transform (1) into an expression that relates the radius of the blur circle with the distance to an object. From Fig. 1, the relation

$$\delta = D_{fi} - D_i \quad (3)$$

is also obtained, where D_i is the distance from the lens to the defocused image. In an optical system, the diameter of the aperture L is often given as [16], [12]

$$L = \frac{F}{f} \quad (4)$$

where F is the focal length, and f is the *f number*. Substituting (3) and (4) into (1) yields

$$R = \frac{F D_{fi} - F D_i}{2 f D_{fi}} \quad (5)$$

Then, using the lens law

$$\frac{1}{D_{fi}} + \frac{1}{D_o} = \frac{1}{F} \quad (6)$$

where D_o is the distance from the lens to a point object, D_{fi} is eliminated from (5) to form

$$R = \frac{F D_i + F D_o - D_i D_o}{2 f D_o} \quad (7)$$

or, solved for D_o , (7) can be rewritten as

$$D_o = \frac{F D_i}{2 f R + D_i - F} \quad (8)$$

Equations (7) and (8) are simple expressions that relate the radius of the blur circle R to the object distance D_o . Equation (8) shows that defocus (designated by R), is a unique indicator of depth from a point source object.

² It is assumed that objects are placed in front of the point of best focus.

Since the spatial domain model for defocus is circularly symmetric, the optical transfer function of defocus can be obtained by finding the Hankel transform of (2) [5], [16].

To find the Hankel transform, polar coordinates in both the spatial and spatial frequency domains must be defined:

$$x = r \cos \theta \text{ and } y = r \sin \theta \quad (9)$$

$$f_x = \rho \cos \phi \text{ and } f_y = \rho \sin \phi \quad (10)$$

where

- r is the radial component in the spatial domain
- θ is the angular component in the spatial domain
- f_x and f_y are the spatial frequency coordinates expressed in cycles/unit and *not* in radians
- ρ is the radial component in the spatial frequency domain
- ϕ is the angular component in the spatial frequency domain.

Then, if a circularly symmetric $f(x, y) = f(r)$, Bracewell gives the formula for the Hankel transform of $f(r)$ as [5]

$$F(\rho) = 2\pi \int_0^\infty r f(r) J_0(2\pi r \rho) dr \quad (11)$$

where $J_0(x)$ is the zeroth-order Bessel function.

Using (11), the Hankel transform of (2) is

$$H(\rho) = 2\pi \int_0^R \frac{r}{\pi R^2} J_0(2\pi r \rho) dr \text{ and} \quad (12)$$

$$H(\rho) = 2 \frac{J_1(2\pi R \rho)}{(2\pi R \rho)} \quad (13)$$

using the identity

$$\frac{d}{dz} z J_1(z) = z J_0(z). \quad (14)$$

Equation (13) is an expression for the optical transfer function of defocus, assuming geometric optics.

Since the experimental results will be obtained in distances mensurated in pixels, it is helpful to also express R in pixel units.³ Therefore, if P is the digitization ratio (i.e., pixels/mm) of the camera and R_p is R in pixel units, then

$$R_p = R \cdot P. \quad (15)$$

The radial frequency component ρ can, with the help of (15), be expressed in terms of pixels as ρ_p [23]

$$\rho = \frac{\rho_p P}{N} \quad (16)$$

where N is the number of points in the transform.

2) *Diffraction Optics*: The *diffraction optics* model of defocus uses the wave theory of light, and its results are exact.⁴ The *frequency transfer function* for a defocused system using *diffraction optics* is given by Hopkins [14] and is also covered in the standard reference on optics by Born and Wolf [3]. However, because of the aberrations present in most low-cost optical systems, the geometric optics model should be an adequate first approximation.⁵

³ Currently, R takes on the same units as F , D_i , and D_o .

⁴ That is, assuming the optics are ideal.

⁵ It will be seen later that for the best accuracy, the point-spread function of the lens system should be experimentally modeled.

D. Measuring Defocus

The most recent work in depth from focus has attempted to measure the amount of *defocus* throughout an image and, hence, infer distances to objects in the scene. This concept was first reported by Helmholtz [13]; however, the first experimental work was done by Pentland in 1982 [20], which showed that two images formed with different apertures are a source of depth information.

The amount of defocus in an image is measured by isolating the defocus operator from the scene. Unfortunately, one unfocused image of an unknown scene does not contain enough information to recover the defocus operator. Soft edges in an image may be either unfocused hard edges in the scene *or* focused soft edges in the scene. Therefore, at least two different images of the same scene are usually obtained with two different defocus operators.⁶ The defocus operator $h(x, y)$ can be changed by varying one or more of three camera parameters

- position of image plane D_i
- focal length F
- aperture f

as shown by (7) and (2).

The most common method of isolating the defocus operator from the scene is inverse filtering. Pentland acquired one image with a pinhole aperture and the other with a larger aperture [20]–[22]. The defocus operator, which was assumed to be a Gaussian, was recovered by a division in the spatial frequency domain.

In his latest paper [19], Pentland's technique was extended. He estimated the power in one or more frequency bandwidths for one image taken with a pinhole aperture and one image taken with another aperture. These two values were compared with a lookup table to produce an estimate of range. Parseval's theorem was used to perform the calculations in the spatial domain.

Pentland also extended his data acquisition equipment to include a two-camera system that acquired two images at different apertures simultaneously. By performing the calculations in hardware, he was able to process four images per second with an rms accuracy of 2.5%.

Using a similar two-camera system and acquiring one image with a pinhole aperture, Bove [4] described a slightly different technique. He took the 2-D Fourier transform of windowed regions in each image, collapsed them to the radial dimension, and divided the two spectra. Bove used a higher order regression solution than Pentland to model the division result to a polynomial approximation of the geometric optics model of defocus. No accuracy results were given.

Assuming that the point-spread function is a Gaussian, Subbarao derived the equations that relate changes in each of the camera parameters to the Gaussian spread parameter and, hence, the distance to the scene. This was done for both *infinitesimal* changes [28], [27] and *finite* changes [29]. Subbarao also performed the theoretical work for a more general solution [31], [26], [30]. He removed Pentland's

⁶That is, unless other cues like color or texture are used to identify sharp features in the scene.

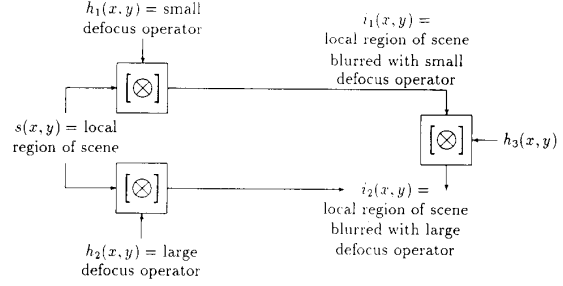


Fig. 2. Block diagram of the depth from focus problem. The computational goal is to find h_3 .

constraint of one image being formed by a pinhole camera and gave a closed-form solution assuming a Gaussian defocus operator. Subbarao also extended this technique to other models of the defocus by generalizing the Gaussian spread parameter σ to be the *square root of the second central moment* of the defocus operator. Like Pentland, he advocated inverse filtering; however, he gave no experimental results to confirm these methods.

Another approach, which has not previously been applied to the depth from focus problem, is a matrix-based method. The approach taken here seeks to deconvolve the defocus operator from the images by characterizing the problem as a system of linear equations.

II. ANALYSIS

In the analysis of the depth from focus problem, it will be argued that a matrix-based approach is more accurate than inverse filtering. A block diagram of the problem is shown in Fig. 2.

If $s(x, y)$ is a local region of the scene and $h_1(x, y)$ is the point spread function caused by defocus with the smaller aperture,⁷ then a local region of the acquired image $i_1(x, y)$ is

$$i_1(x, y) = s(x, y) [\otimes] h_1(x, y). \quad (17)$$

The operator $[\otimes]$ designates restricted convolution, where the borders of the kernel $h_1(x, y)$ are *not* convolved past the borders of the scene $s(x, y)$. This restricted definition is necessary because the result of convolving the kernel past the ends of the local region of the scene is *unknown*.

Similarly, if $h_2(x, y)$ is the point spread function caused by defocus with the larger aperture, then a local region of the acquired image $i_2(x, y)$ is

$$i_2(x, y) = s(x, y) [\otimes] h_2(x, y). \quad (18)$$

The computational goal will be to find a function $h_3(x, y)$, which will transform $i_1(x, y)$ into $i_2(x, y)$ according to

$$i_1(x, y) [\otimes] h_3(x, y) = i_2(x, y). \quad (19)$$

⁷For this analysis, the defocus operator will be varied by changing the aperture, although this is not a limitation of the technique. The defocus operator can be changed by adjusting any of the variables given in (7); however, changing the distance between the image plane and the lens means that the magnification of the images is different [30].

For any local region, $h_3(x, y)$ is then uniquely related to the distance from the camera to that region in the scene. More details will be given later.

Equation (19) can also be expressed in terms of $h_1(x, y)$ and $h_2(x, y)$. By substituting (17) and (18) into (19)

$$s(x, y) [\otimes] h_1(x, y) [\otimes] h_3(x, y) = s(x, y) [\otimes] h_2(x, y), \quad (20)$$

$$h_1(x, y) [\otimes] h_3(x, y) = h_2(x, y). \quad (21)$$

Since $h_1(x, y)$ has a limited spatial extent, (21) can be generalized to

$$h_1(x, y) \otimes h_3(x, y) = h_2(x, y). \quad (22)$$

In (22), $h_3(x, y)$ will be called the *convolution ratio* of the two defocus operators $h_2(x, y)$ and $h_1(x, y)$. The convolution ratio $h_3(x, y)$ can be isolated by matrix-based methods or inverse filtering. However, it will be shown that inverse filtering incurs a number of errors.

A. Inverse Filtering Solution for the Convolution Ratio

From a theoretical standpoint, inverse filtering in the spatial frequency domain is convenient because convolution in the spatial domain now becomes multiplication. To enable transformation to the spatial frequency domain, (17) is windowed [11] with $w_1(x, y)$

$$i_1(x, y) \cdot w_1(x, y) = (s(x, y) [\otimes] h_1(x, y)) \cdot w_1(x, y). \quad (23)$$

Taking the Fourier transform of both sides

$$\begin{aligned} I_1(f_x, f_y) \otimes W_1(f_x, f_y) \\ = (S(f_x, f_y) \cdot H_1(f_x, f_y)) \otimes W_1(f_x, f_y) \end{aligned} \quad (24)$$

where the following are Fourier pairs:

$$i_1(x, y) \iff I_1(f_x, f_y), \quad (25)$$

$$h_1(x, y) \iff H_1(f_x, f_y), \quad (26)$$

$$w_1(x, y) \iff W_1(f_x, f_y), \quad (27)$$

$$s(x, y) \iff S(f_x, f_y). \quad (28)$$

Similarly, for $i_2(x, y)$

$$\begin{aligned} I_2(f_x, f_y) \otimes W_2(f_x, f_y) \\ = (S(f_x, f_y) \cdot H_2(f_x, f_y)) \otimes W_2(f_x, f_y). \end{aligned} \quad (29)$$

To perform the inverse filtering and isolate the convolution ratio from the scene, (29) is divided by (24)

$$\begin{aligned} \frac{I_2(f_x, f_y) \otimes W_2(f_x, f_y)}{I_1(f_x, f_y) \otimes W_1(f_x, f_y)} \\ = \frac{(S(f_x, f_y) \cdot H_2(f_x, f_y)) \otimes W_2(f_x, f_y)}{(S(f_x, f_y) \cdot H_1(f_x, f_y)) \otimes W_1(f_x, f_y)}. \end{aligned} \quad (30)$$

It is generally assumed that $W_1(f_x, f_y)$ and $W_2(f_x, f_y)$ have a small spread in the spatial frequency domain⁸

$$\frac{I_2(f_x, f_y)}{I_1(f_x, f_y)} \approx \frac{S(f_x, f_y) \cdot H_2(f_x, f_y)}{S(f_x, f_y) \cdot H_1(f_x, f_y)}, \quad (31)$$

$$\approx \frac{H_2(f_x, f_y)}{H_1(f_x, f_y)} = H_3(f_x, f_y). \quad (32)$$

⁸Although this causes problems that are mentioned later.

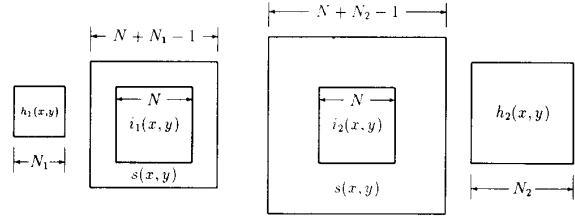


Fig. 3. Differences in scene area $s(x, y)$ when the same image areas for $i_1(x, y)$ and $i_2(x, y)$ are used. The dimensions of $i_1(x, y) = s(x, y) [\cdot] h_1(x, y)$ and $i_2(x, y) = s(x, y) [\cdot] h_2(x, y)$ are illustrated.

The function $H_3(f_x, f_y)$, which is the Fourier transform of the convolution ratio $h_3(x, y)$, has been isolated. From it, depth can be inferred.

1) *Problems with Inverse Filtering:* Inverse filtering, however, is fraught with several problems. If it is performed in the spatial frequency domain, the difficulties arise in accurately estimating the frequency spectra $I_1(f_x, f_y)$ and $I_2(f_x, f_y)$ of local regions $i_1(x, y)$ and $i_2(x, y)$. The major constraint is that since depth changes throughout the scene, it is desirable for these regions to be as small as possible, yet accurate spectral analysis benefits from larger local areas.

Another problem is caused by windowing. Looking at (31), $I_2(f_x, f_y)$ contains two types of zero crossings: those that arise from the scene $S(f_x, f_y)$ and those that arise from $H_2(f_x, f_y)$.

The result for $H_3(f_x, f_y)$ must be independent of the scene $S(f_x, f_y)$ since $H_1(f_x, f_y)$ and $H_2(f_x, f_y)$ are independent of the scene. Therefore, the zero crossings in $I_2(f_x, f_y)$ that are due to $S(f_x, f_y)$ must be matched by zero crossings in $I_1(f_x, f_y)$ so that they can cancel.⁹ Unfortunately, the result of convolving $I_2(f_x, f_y)$ with $W_2(f_x, f_y)$ and $I_1(f_x, f_y)$ with $W_1(f_x, f_y)$ is that the zero crossings shift slightly. This small movement can cause large variations in the quotient $H_3(f_x, f_y)$.

One way of attempting to avoid the problems associated with finding the spatial frequency domain representations is to perform the inverse filtering in the spatial domain, as done by Pentland [19]. The windowing problem, however, shows up in another way since the sharp bandpass filters required cannot be represented with a limited spatial kernel.

Another problem with inverse filtering by frequency domain division is selecting the most appropriate spatial area for analysis. According to conventional inverse filtering, the *same* sizes of $i_1(x, y)$ and $i_2(x, y)$ are used to calculate $I_1(f_x, f_y)$ and $I_2(f_x, f_y)$ so that in (31), frequency components of the same value can be divided. As shown by Fig. 3, using the same size $N \times N$ of $i_1(x, y)$ and $i_2(x, y)$ infers that *different* dimensions $(N + N_1 - 1)^2$ and $(N + N_2 - 1)^2$ of the scene $s(x, y)$ are used. Therefore, $i_2(x, y)$ may contain spurious data from bordering edges.

2) *Noise-Free Numerical Example:* Since the errors incurred in the spatial frequency domain are dependent on a number of factors, including the scene $s(x, y)$ and the windowing functions $w_1(x, y)$ and $w_2(x, y)$, it is difficult to provide a closed-form equation. By making some reasonable

⁹The zero crossings caused by $S(f_x, f_y)$ that may match the the zero crossings of $H_2(f_x, f_y)$ are, of course, excluded.

TABLE I
ERRORS RESULTING FROM WINDOWING A BLURRED
BLACK-TO-WHITE TRANSITION IN THE SPATIAL DOMAIN BEFORE
INVERSE FILTERING BY DIVISION IN THE FREQUENCY DOMAIN

N	% Error
4	65.8
8	23.7
16	6.4
32	1.1

assumptions, however, an example of some of the errors incurred can be shown.

It will be assumed that one defocus operator h_1 is an impulse function, and hence, from (17), $i_1 = s$. Dealing in 1-D for simplicity, let the scene be a simple black-to-white transition

$$s = i_1 = (0, \dots, 0, 0, 1, 1, 1, \dots, 1). \quad (33)$$

It will also be assumed that the other blurring operator is a gate function

$$h_2(x) = (1, 1, 1). \quad (34)$$

In addition, using (22)

$$h_3(x) = h_2(x) = (1, 1, 1). \quad (35)$$

The second blurred image can be calculated using (18)

$$i_2 = (0, \dots, 0, 1, 2, 3, 3, \dots, 3). \quad (36)$$

Using the Gaussian window¹⁰

$$w_1(x) = w_2(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \quad (37)$$

Working in the spatial frequency domain, if \mathcal{F}^{-1} is the inverse Fourier operator, it can be deduced from (30) and (31) that

$$\hat{h}_3(x) = \mathcal{F}^{-1} \left(\frac{W_2(f_x) \otimes I_2(f_x)}{W_1(f_x) \otimes I_1(f_x)} \right) \quad (38)$$

where $\hat{h}_3(x)$ is the estimated pattern for $h_3(x)$. The error ϵ_f , which is caused by calculating $\hat{h}_3(x)$ by inverse filtering, can be written as

$$\epsilon_f = \sum_{x=0}^{N-1} [h_3(x) - \hat{h}_3(x)]^2, \quad (39)$$

$$\epsilon_f = \sum_{x=0}^{N-1} \left[h_3(x) - \mathcal{F}^{-1} \left(\frac{W_2(f_x) \otimes I_2(f_x)}{W_1(f_x) \otimes I_1(f_x)} \right) \right]^2. \quad (40)$$

For various sizes N of data used, the rms errors calculated from (40) are shown in Table I.

The table shows that the error decreases as N increases; however, to drive the error down to 1%, it is necessary to observe a length that is about an *order of magnitude* larger

¹⁰The Gaussian is used because, as shown by Gabor [7], it is simultaneously optimally localized in both the spatial and spatial frequency domains.

than the blurring operator. Unfortunately, solving the depth-from-focus problem is a local operation; therefore, large image areas where features are at the same depth are generally not available.

3) *Constrained Inverse Filtering*: Using regularization, the inverse filtering solution in the frequency domain can also be constrained to make it more robust. For convenience, this solution will be derived using matrix notation;¹¹ however, it should not be confused with the matrix solution of the next section. This inverse filtering solution, although constrained, still has the problems detailed in Section II-A-1.

It has been shown [6] that the uniquely characteristic part of $H_3(f_x, f_y)$ is a downturned quadratic-type shape centered at $f_x = f_y = 0$. Therefore, $H_3(f_x, f_y)$ can be modeled by least squares fitting to a quadratic. The only problem with this approach is that the quadratic model is valid only within a limited circle around $f_x = f_y = 0$, and the radius of this circle is dependent on the width of the quadratic. Assuming that only points within the valid radius are used,¹² regularization [32], [2] can be used to find $H_3(f_x, f_y)$. The regularized form of (32) minimizes the functional

$$\| [I_{1DS}] \cdot \mathbf{H}_{3S} - \mathbf{I}_{2S} \|^2 + \lambda \| [\mathbf{C}] \cdot \mathbf{H}_{3S} \|^2 = \text{minimum} \quad (41)$$

where

- $[I_{1DS}]$ is a matrix with $I_1(f_x, f_y)$ stacked along its diagonal as shown by Gonzalez and Wintz [9].
- \mathbf{H}_{3S} is a stacked vector formed from $H_3(f_x, f_y)$.
- \mathbf{I}_{2S} is a stacked vector formed from $I_2(f_x, f_y)$.
- λ is a scalar parameter, which can be adjusted to be small and place more emphasis on fitting \mathbf{H}_{3S} to the data or large and place more emphasis on fitting \mathbf{H}_{3S} to a quadratic shape.
- $[\mathbf{C}]$ is a matrix that minimizes the magnitude of the second term if \mathbf{H}_{3S} has a quadratic shape.

The matrix $[\mathbf{C}]$ is relatively easy to find. Let $H_3(f_x, f_y)$ be represented by the quadratic

$$H_3(f_x, f_y) = a(f_x^2 + f_y^2) + b. \quad (42)$$

Then, in matrix form, (42) can be written as

$$\mathbf{H}_{3S} = [\mathbf{Q}] \cdot \mathbf{A} \quad (43)$$

¹¹The notation used is as follows:

- Bold letters represent vectors.
- Bold letters in square braces represent matrices.
- Subscripts, when present, indicate the nature of the matrix

- \mathbf{C} : circulant
- \mathbf{D} : diagonal
- \mathbf{T} : Toeplitz
- \mathbf{B} : block form
- \mathbf{S} : stacked vector.

¹²The experimental implementation in Section III-A-1 avoids the problem of iteratively finding the valid radius by calculating it based on the *known* depth value.

where

$$[Q] = \begin{bmatrix} f_0^2 + f_0^2 & 1 \\ f_1^2 + f_0^2 & 1 \\ \vdots & \vdots \\ f_{N-1}^2 + f_0^2 & 1 \\ f_0^2 + f_1^2 & 1 \\ f_1^2 + f_1^2 & 1 \\ \vdots & \vdots \\ f_{N-1}^2 + f_1^2 & 1 \\ \vdots & \vdots \\ f_{N-1}^2 + f_{N-1}^2 & 1 \end{bmatrix} \quad \text{and} \quad (44)$$

$$A = \begin{bmatrix} a \\ b \end{bmatrix}. \quad (45)$$

Using general least squares [23], (43) can be solved for A as

$$A = ([Q]^T [Q])^{-1} [Q]^T H_{3S}. \quad (46)$$

Now, A contains the parameters to the best fitted quadratic, and the best quadratic is

$$[Q]A. \quad (47)$$

Multiplying (46) by $[Q]$ on each side yields

$$[Q]A = [Q]([Q]^T [Q])^{-1} [Q]^T H_{3S}. \quad (48)$$

Then, the error between H_{3S} and the best quadratic is

$$\|H_{3S} - [Q]([Q]^T [Q])^{-1} [Q]^T H_{3S}\|^2 \quad \text{which} \\ = \|[I] - [Q]([Q]^T [Q])^{-1} [Q]^T\| H_{3S}^2 \quad (49)$$

where $[I]$ is the identity matrix. Comparing (49) with (41) reveals that

$$[C] = [I] - [Q]([Q]^T [Q])^{-1} [Q]^T. \quad (50)$$

Now that an expression for $[C]$ has been derived, the original regularization equation can be solved. The Euler equation [2] for (41) is solved for H_{3S} as follows:

$$[I_{1DS}]^T [I_{1DS}] H_{3S} - [I_{1DS}]^T I_{2S} + \lambda [C]^T [C] H_{3S} = 0, \\ ([I_{1DS}]^T [I_{1DS}] + \lambda [C]^T [C]) H_{3S} = [I_{1DS}]^T I_{2S}, \quad \text{and} \\ H_{3S} = ([I_{1DS}]^T [I_{1DS}] + \lambda [C]^T [C])^{-1} [I_{1DS}]^T I_{2S}. \quad (51)$$

Since the quadratic parameters A are desired from H_{3S} , however, using (46), (51) then becomes

$$A = ([Q]^T [Q])^{-1} [Q]^T ([I_{1DS}]^T [I_{1DS}] \\ + \lambda [C]^T [C])^{-1} [I_{1DS}]^T I_{2S}. \quad (52)$$

The last step in solving the defocus problem by constrained inverse filtering is to relate the quadratic parameters a and b in A to the distance to the scene D_o . This will be done in two steps: a and b will be related to the radius of the geometric

optics defocus operator R , and the relationship of R to D_o has already been established.

The most convenient way of relating a and b to R is to equate the zero crossing of the quadratic and the analogous pattern in $H_3(\rho_p)$. Using (32) and (13)

$$H_3(\rho) = \frac{R_1 \rho J_1(2\pi R_2 \rho)}{R_2 \rho J_1(2\pi R_1 \rho)}. \quad (53)$$

Then, substituting in (16)

$$H_3(\rho_p) = \frac{R_1 \rho_p J_1\left(\frac{2\pi R_2 \rho_p P}{N}\right)}{R_2 \rho_p J_1\left(\frac{2\pi R_1 \rho_p P}{N}\right)}. \quad (54)$$

The zero crossings of (54) are found when

$$J_1\left(\frac{2\pi R_2 \rho_p P}{N}\right) = 0. \quad (55)$$

The first zero crossing occurs when the argument of J_1 is equal to 3.83. Therefore

$$\frac{2\pi R_2 \rho_p P}{N} = 3.83, \quad \text{and} \quad (56)$$

$$R_2 = \frac{3.83N}{2\pi \rho_p P}. \quad (57)$$

The same zero crossing in the quadratic model can be found by equating (42) to zero

$$0 = a(f_x^2 + f_y^2) + b. \quad (58)$$

Then, from the definition of ρ in (10), if a and b are in pixel units

$$0 = a\rho_p^2 + b \quad \text{and} \quad (59)$$

$$\rho_p = \sqrt{-\frac{b}{a}}. \quad (60)$$

Substituting (60) into (57) yields

$$R_2 = \frac{3.83N}{2\pi P \sqrt{-\frac{b}{a}}}. \quad (61)$$

Finally, substituting (61) into (8) yields an equation for distance

$$D_o = \frac{F D_i}{2f_2 \frac{3.83N}{2\pi P \sqrt{-\frac{b}{a}}} + D_i - F}. \quad (62)$$

To summarize, constrained inverse filtering can be performed by taking two local regions $i_1(x, y)$ and $i_2(x, y)$ of the same size and windowing them with $w_1(x, y)$ and $w_2(x, y)$, respectively. After taking the discrete Fourier transform of both windowed regions, a regularized division and fit to a quadratic is done using (52). The quadratic parameters a and b in A are related to distance to the scene D_o using (62).

B. A Matrix-Based Solution for the Convolution Ratio

Solving the depth-from-focus problem using a matrix-based method avoids the difficult problems experienced with inverse filtering. The matrix-based technique will be presented as three methods of increasing complexity.

1) *Noise-Free Method:* The convolution equation (19) can be written as a matrix equation [1]

$$[\mathbf{i}_{1BT}] \cdot \mathbf{h}_{3S} = \mathbf{i}_{2S} \quad (63)$$

where $[\mathbf{i}_{1BT}]$ is an $N^2 \times N^2$ block Toeplitz matrix constructed from $i_1(x, y)$, and \mathbf{h}_{3S} , \mathbf{i}_{2S} are row-stacked vectors created from $h_3(x, y)$ and $i_2(x, y)$. Solving for \mathbf{h}_{3S}

$$\mathbf{h}_{3S} = [\mathbf{i}_{1BT}]^{-1} \cdot \mathbf{i}_{2S}. \quad (64)$$

Using a matrix-based method, the example given in Section II-A-2 can be solved *exactly* using dimensions as small as $N = 3$. A 1-D convolution matrix $[\mathbf{i}_{1T}]$ is constructed by inverting i_1 given by (33) and shifting it into each row

$$[\mathbf{i}_{1T}] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \quad (65)$$

To confirm that \mathbf{h}_3 can be calculated *exactly*, (65) and (36) can be substituted into the 1-D form of (63) and solved for \mathbf{h}_3

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \cdot \mathbf{h}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}. \quad (66)$$

$$\mathbf{h}_3 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad (67)$$

$$\mathbf{h}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \quad (68)$$

2) *A Simple Model for the Defocus Operator with Noise:* In the absence of noise and with one blurred image, (64) has shown that the deconvolution of segments from local images can be performed exactly using a matrix-based method. When noise is added to the model, solving (64) for \mathbf{h}_{3S} becomes very *ill posed*.

Hadamard [10] defined a mathematical problem to be *well posed* if

- an answer exists
- the answer is unique
- a small error induced into the data does not produce large fluctuations (instability) in the answer.

Because the depth-from-focus problem fails on the last point,¹³ it is advisable to apply *all* available information to the problem. An important, unused piece of information is that the defocus operators $h_1(x, y)$ and $h_2(x, y)$ have shapes that can be theoretically or experimentally derived. Referring to (22), it can be seen that although $h_3(x, y)$ is *unknown*, it must belong to a family of patterns that can be known *a priori*.

The regularized form of (63) minimizes the functional

$$\|[\mathbf{i}_{1BT}] \cdot \mathbf{h}_{3S} - \mathbf{i}_{2S}\|^2 + \lambda \|[\mathbf{C}] \cdot \mathbf{h}_{3S}\|^2 = \text{minimum} \quad (69)$$

where λ is a scalar parameter, and $[\mathbf{C}]$ is a matrix minimizing the magnitude of the second term if \mathbf{h}_{3S} belongs to the expected family of patterns.

¹³Small changes in the data $i_1(x, y)$ can produce large fluctuations in the matrix $[\mathbf{i}_{1BT}]^{-1}$ used in (64).

The parameter λ can be adjusted to be small and place more emphasis on fitting $h_3(x, y)$ to the data or large and place more emphasis on fitting $h_3(x, y)$ to the expected family of patterns.

The Euler equation [2] for (69) is solved for \mathbf{h}_{3S} as follows:

$$\begin{aligned} & [\mathbf{i}_{1BT}]^T [\mathbf{i}_{1BT}] \mathbf{h}_{3S} - [\mathbf{i}_{1BT}]^T \mathbf{i}_{2S} \\ & + \lambda [\mathbf{C}]^T [\mathbf{C}] \mathbf{h}_{3S} = 0. \\ \mathbf{h}_{3S} & = \left([\mathbf{i}_{1BT}]^T [\mathbf{i}_{1BT}] + \lambda [\mathbf{C}]^T [\mathbf{C}] \right)^{-1} [\mathbf{i}_{1BT}]^T \mathbf{i}_{2S}. \end{aligned} \quad (70)$$

In practice, solving (70) is computationally expensive. Additionally, finding $[\mathbf{C}]$ is difficult for anything other than simple parametric families (i.e., quadratic, cubic) of $h_3(x, y)$. Since $h_3(x, y)$ can be quite complex and not easily coerced into a parametric shape, it is generally not possible to find $[\mathbf{C}]$. Therefore, in a generalized implementation, the closed form (70) cannot be used. An alternative approach is to introduce an iterative technique.

3) *Any Model for the Defocus Operator with Noise:* Equations (7) and (2) have shown that assuming geometric optics, the defocus operator is a unique indicator of depth. Then, $h_1(x, y)$ and $h_2(x, y)$ correspond uniquely to scene depth. It can also be shown from (22), and assuming geometric optics, that $h_3(x, y)$ is also a unique indicator of depth.

Using (22), a table of $h_3(x, y)$ patterns can be calculated *a priori* to any desired detail. Each $h_3(x, y)$ pattern in the table corresponds to a known scene depth. Recall (19):

$$i_1(x, y) \otimes h_3(x, y) = i_2(x, y). \quad (71)$$

Therefore, for any given image regions $i_1(x, y)$ and $i_2(x, y)$, the table of $h_3(x, y)$ patterns can be searched iteratively for the best $\hat{h}_3(x, y)$ that minimizes

$$\sum_{x=0}^{N-k} \sum_{y=0}^{N-k} \left[i_1(x, y) \otimes \hat{h}_3(x, y) - i_2(x, y) \right]^2 = \text{minimum} \quad (72)$$

where the size of $i_1(x, y)$ is $N \times N$, the kernel $\hat{h}_3(x, y)$ is $k \times k$, and therefore, $i_2(x, y)$ is $(N - k + 1) \times (N - k + 1)$.

To implement the new method, a table of all possible $h_3(x, y)$ patterns may be calculated *a priori*. To calculate this table, a model of the defocus operators $h_1(x, y)$ and $h_2(x, y)$ must be assumed or experimentally derived. Since the scene features fall within a known range, then both $h_1(x, y)$ and $h_2(x, y)$ form a range of known patterns, from which the table of $h_3(x, y)$ patterns can be calculated. This calculation starts with (22):

$$h_1(x, y) \otimes h_3(x, y) = h_2(x, y). \quad (73)$$

It can be shown, by using the *projection slice theorem* [24], that (73) can be expressed in 1-D form

$$h_{1y}(y) \otimes h_{3y}(y) = h_{2y}(y) \quad (74)$$

where

$$h_{1y}(y) = \int_{-\infty}^{\infty} h_1(x, y) dx, \quad (75)$$

$$h_{2y}(y) = \int_{-\infty}^{\infty} h_2(x, y) dx, \quad (76)$$

$$h_{3y}(y) = \int_{-\infty}^{\infty} h_3(x, y) dx. \quad (77)$$

In matrix form, (22) becomes

$$[\mathbf{h}_{1C}] \cdot \mathbf{h}_3 = \mathbf{h}_2, \quad (78)$$

$$\mathbf{h}_3 = [\mathbf{h}_{1C}]^{-1} \cdot \mathbf{h}_2 \quad (79)$$

where \mathbf{h}_3 and \mathbf{h}_2 are vectors formed from $h_{3y}(y)$ and $h_{2y}(y)$, and $[\mathbf{h}_{1C}]$ is a circulant matrix formed from $h_{1y}(y)$, according to Gonzalez and Wintz [9].

Unfortunately, (79) will still not yield a usable result for \mathbf{h}_3 because the problem is ill posed, and \mathbf{h}_3 may have an unlimited spread. A better approach is the regularization equation

$$\|[\mathbf{h}_{1C}] \cdot \mathbf{h}_3 - \mathbf{h}_2\|^2 + \lambda_1 \|[\mathbf{C}_C] \cdot \mathbf{h}_3\|^2 + \lambda_2 \|[\mathbf{T}_D] \cdot \mathbf{h}_3\|^2 = \text{minimum} \quad (80)$$

where

- $[\mathbf{C}_C]$ can be any circulant matrix that detects the smoothness of a function
- λ_1 is a scalar parameter that controls the smoothness of \mathbf{h}_3
- $[\mathbf{T}_D]$ is a diagonal matrix formed from a function that penalizes \mathbf{h}_3 for having nonzero tails
- λ_2 is a scalar parameter that controls the emphasis on minimizing the tails of \mathbf{h}_3 .

The Euler equation [2] for (80) can be solved as follows:

$$\begin{aligned} & [\mathbf{h}_{1C}]^T [\mathbf{h}_{1C}] \mathbf{h}_3 - [\mathbf{h}_{1C}]^T \mathbf{h}_2 \\ & + \lambda_1 [\mathbf{C}_C]^T [\mathbf{C}_C] \mathbf{h}_3 + \lambda_2 [\mathbf{T}_D]^T [\mathbf{T}_D] \mathbf{h}_3 = 0, \\ & \mathbf{h}_3 = \left([\mathbf{h}_{1C}]^T [\mathbf{h}_{1C}] + \lambda_1 [\mathbf{C}_C]^T [\mathbf{C}_C] \right. \\ & \left. + \lambda_2 [\mathbf{T}_D]^T [\mathbf{T}_D] \right)^{-1} [\mathbf{h}_{1C}]^T \mathbf{h}_2. \end{aligned} \quad (81)$$

The design of \mathbf{h}_3 involves tradeoffs; however, the regularization technique controls how far \mathbf{h}_3 strays from the original data by the adjustment of λ_1 and λ_2 . The qualitative test of \mathbf{h}_3 is to substitute it back into (78) to get a new $\hat{\mathbf{h}}_2$:

$$\hat{\mathbf{h}}_2 = [\mathbf{h}_{1C}] \cdot \mathbf{h}_3 \quad (82)$$

which can be compared with the original \mathbf{h}_2 .

III. EXPERIMENTAL RESULTS

The experiments were performed with the point of sharpest focus set at 1 m from the camera and objects placed within the interval of $0.80 \leq D_o \leq 0.95$ m from the camera. The two aperture settings chosen were $f_2 = 1.3$ for the large defocus operator and $f_1 = 2.0$ for the small defocus operator. To help minimize the noise, 16 images were averaged to produce the image used for processing. The images were corrected radiometrically by calibration with a Kodak Gray Scale [17]. Cosine effects and vignetting [25] were also corrected by using experimental models.

The experimental equipment used for this research was the following:

- A JAVELIN Model JE2062 black-and-white digital CCD camera with 384 pixels horizontal and 485 pixels vertical resolution and RS-170 standard video output

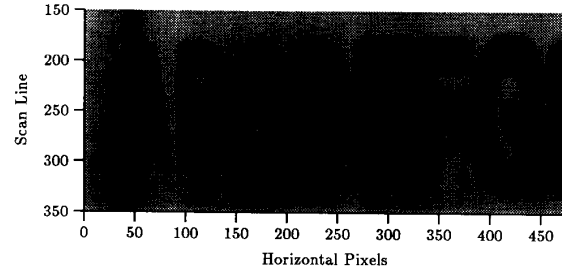


Fig. 4. Slanted image of alphabet at $f_1 = 2.0$.

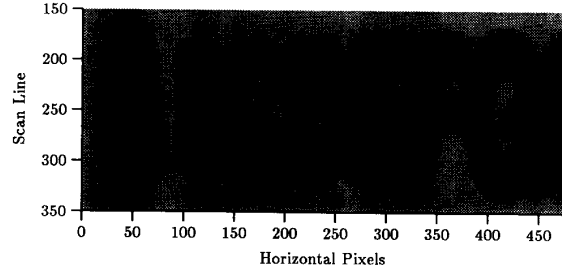


Fig. 5. Slanted image of alphabet at $f_2 = 1.3$.

- a 50-mm lens
- a PC VISION frame grabber with 512×512 resolution at 8 b
- an 8-MHz IBM PC XT for data acquisition control and processing.

A. Scene with 2-D Features

The first scene processed was a flat sheet of paper with letters of the alphabet. The paper was arranged so that the distance from the camera ranged linearly from 0.95 m at the left-hand side of the picture to 0.80 m at the right-hand side. The acquired images are shown in Figs. 4 and 5. Since the aperture settings were only one f number apart, the images show only a little difference in defocus between them.

1) *Constrained Inverse Filtering*: The first solution to the depth-from-focus problem was obtained using the constrained inverse filtering method of (52) and (62). A total of 242 local regions containing 'edges,' of size 64×64 pixels were processed, and one depth point was calculated for each local region. A mesh plot of the results is shown in Fig. 6. The dashed line on the side of the plot shows the expected depth values for the scene being imaged.

The most consistently accurate depth values in the plot are found on the left-hand side for both large and small scan lines. In the images of Figs. 4 and 5, this corresponds to the bottom and top of the letters "A" and "B." Here, the errors associated with windowing and border effects are minimal. Windowing errors are small because a majority of the outside perimeter of the local areas are about the same intensity. Border effects are also slight because the width of the defocus operator is relatively small.

The least accurate depth values are on the right-hand side of the plot, toward the center. Here, windowing effects are

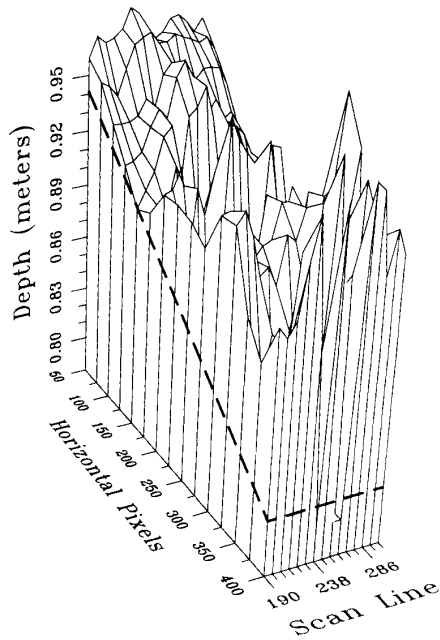


Fig. 6. Mesh plot of calculated depth values from using constrained inverse filtering on the "alphabet" images. The dashed line shows the expected depth values.

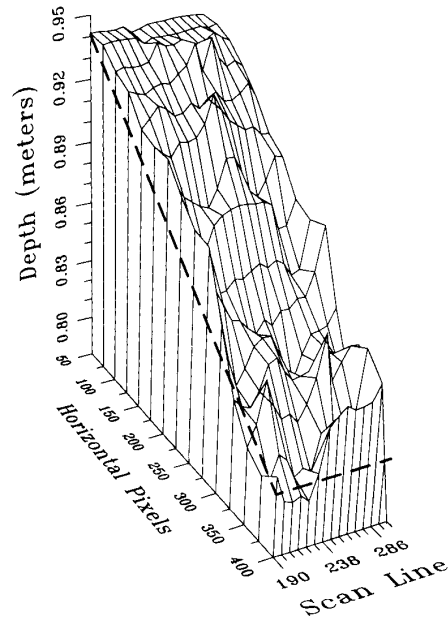


Fig. 7. Mesh plot of calculated depth values by using the iterative matrix solution on the "alphabet" images and assuming the geometric optics model of defocus. The dashed line shows the expected depth values.

large since a local area can likely have a dark intensity on one side and a light intensity on the other side. The border effects are also large since the defocus operators are wide, and edges from adjacent local areas "bleed" into each other. The average rms error for the scene was 42% over the *expected range* and 6.8% in terms of *distance from the camera*.

2) *Iterative Matrix Solution*: Using the geometric optics model of defocus, (7), (2), and (81), a table of $h_3(x, y)$ patterns was calculated over the expected range $0.80 \leq D_o \leq 0.95$ m. Depth values were calculated at the same local regions used in the previous section. A mesh plot of the derived depth values is shown in Fig. 7. No filtering of the depth values was required. The rms error over the *expected range* was 10%, and in terms of *distance from the camera* was 1.7%. These values represent a substantial improvement over the errors obtained using the constrained inverse filtering technique.

Using experimentally obtained patterns for $h_1(x, y)$ and $h_2(x, y)$ [18], another table of $h_3(x, y)$ patterns was calculated. The depth results, which are shown in Fig. 8, had an rms error of 7.4% in terms of *expected range* and 1.3% in terms of *distance from the camera*. This is a reduction from the error found while assuming the geometric optics model of defocus.

B. Scene with Various Edges

Although the images of the alphabet contained an assortment of 2-D features, they had only sharp black-to-white printed edges. The next scene contains three other types of edges:

1. A soft edge caused by a shadow (penumbra)
2. an occlusion edge
3. a printed, grey, ramp edge.

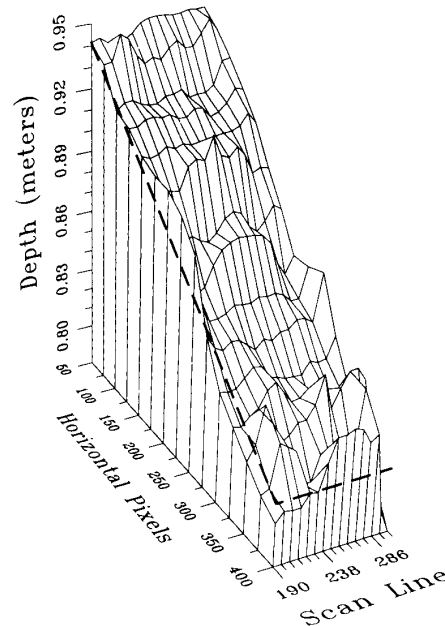


Fig. 8. Mesh plot of calculated depth values by using the iterative matrix solution on the "alphabet" images and an experimentally derived defocus operator. The dashed line shows the expected depth values.

To construct these three edges, a top view of the scene geometry is shown in Fig. 9.

The resulting images are shown in Figs. 10 and 11.

1) *Constrained Inverse Filtering*: Using constrained inverse filtering, 11 local regions on each of the three edges were processed. The errors in the results are shown in Table II.

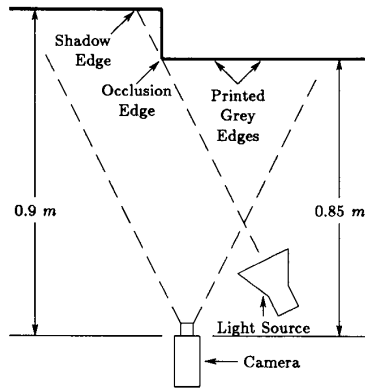


Fig. 9. Top view of scene geometry to obtain images with various edges.

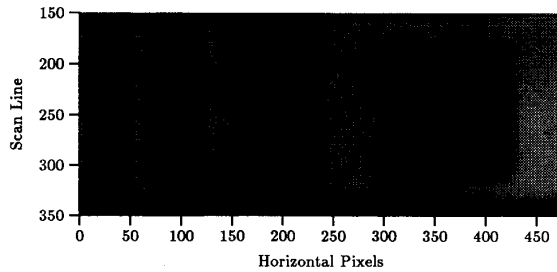
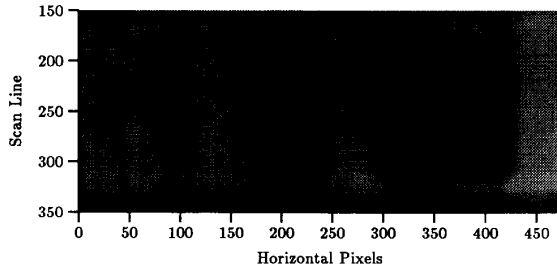
Fig. 10. Image with various edges at $f_1 = 2.0$.Fig. 11. Image with various edges at $f_2 = 1.3$.

TABLE II
RMS ERRORS IN DEPTH VALUES FROM USING CONSTRAINED
INVERSE FILTERING ON THE IMAGES OF VARIOUS EDGES

Scene edges	rms error over expected range (%)	rms error in distance to camera (%)	Number of points
Shadow edge	34.8	5.86	11
Occlusion edge	—	—	11
Printed grey edge	71.3	11.16	11
Total	49.7	8.1	33

These errors are somewhat greater than the errors found when processing the “alphabet” images. The occlusion edge yielded invalid results at all local regions tested.

2) *Iterative Matrix Solution:* The iterative matrix solution was also implemented on the images of various edges. For

TABLE III
RMS ERRORS IN DEPTH VALUES FROM USING ITERATIVE
MATRIX SOLUTION ON THE IMAGES OF VARIOUS EDGES

Scene feature	rms error over expected range (%)	rms error in distance to camera (%)	Number of points
Shadow edge	2.99	0.50	11
Occlusion	2.83	0.50	11
Printed grey edge	4.90	0.86	11
Total	3.70	0.64	33

this test, the experimentally derived table of $h_3(x, y)$ was used since it has been shown that it yielded superior results to the table formed assuming geometric optics. The results are shown in Table III.

The errors shown are even lower than those found while processing the “alphabet” images.

IV. CONCLUSIONS

It has been shown that the inverse filtering solution to the depth-from-focus problem contains fundamental limitations to accuracy. These limitations are as follows:

- Errors are made in estimating the frequency spectra.
- Windowing the data in the spatial domain leaves a residual convolution in the spatial frequency domain.
- The same size area of the two images is used, whereas the data is best contained in different sized areas.

These errors are exacerbated by the requirement that the local image regions be as small as possible since depth changes across the scene, and hence, the defocus operators vary throughout the images.

As an alternative, a matrix-based approach has been presented, which solves the problems above. This approach seeks to deconvolve the defocus operator from the images by characterizing the problem as a large system of linear equations. This approach is different from inverse filtering because it allows the depth-from-focus problem to be represented as a restricted convolution, where the defocus kernel is *not* convolved past the ends of a local region of the scene. The matrix-based approach remains accurate in small image areas.

Two images $i_1(x, y)$ and $i_2(x, y)$ of the same scene $s(x, y)$ are acquired with two different defocus operators $h_1(x, y)$ and $h_2(x, y)$. A convolution ratio $h_3(x, y)$ has been presented where

$$h_1(x, y) \otimes h_3(x, y) = h_2(x, y). \quad (83)$$

This convolution ratio is a unique indicator of depth.

Three methods of recovering $h_3(x, y)$ have been presented in order of increasing complexity. Assuming no noise, the first method solves for $h_3(x, y)$ by

$$\mathbf{h}_{3S} = [\mathbf{i}_{1BT}]^{-1} \cdot \mathbf{i}_{2S} \quad (84)$$

where $[\mathbf{i}_{1BT}]$ is an $N^2 \times N^2$ block Toeplitz matrix constructed from $i_1(x, y)$, and \mathbf{h}_{3S} , \mathbf{i}_{2S} are row-stacked vectors created

from $h_3(x, y)$ and $i_2(x, y)$. It was shown that this method produced accurate results with an order of magnitude less data than required by the inverse filtering method.

The second method uses regularization to deal with the noise, and the solution for $h_3(x, y)$ was given in (70). However, this solution requires that $h_1(x, y)$ and $h_2(x, y)$ be represented as simple parametric families.

The third and most general solution for $h_3(x, y)$ was given in (72). A regularized family of all solutions to $h_3(x, y)$ is modeled in a table of patterns. Using an iterative technique, this table is searched for the best pattern $\hat{h}_3(x, y)$, which, when convolved with $i_1(x, y)$ so that the borders of $\hat{h}_3(x, y)$ stay within $i_1(x, y)$, yields $i_2(x, y)$. The position of $\hat{h}_3(x, y)$ in the table gives a direct indication of distance.

This method is general for the following reasons:

1. **Minimal constraints on the scene:** The two images can be obtained under any differing conditions of defocus (change of aperture, position of image plane, or focal length), one image need not be obtained with a pinhole aperture, and no assumptions are made about the nature of the edges in the scene.
2. **Independent of the defocus operator model:** The defocus operator may take an assumed form for simplicity, or for more accuracy, it can be experimentally measured.
3. **Selective tradeoffs:** The tradeoffs in forming the convolution ratio $h_3(x, y)$ are resolved through regularization.

Experimental results of obtaining depth from focus using images obtained with two different apertures, where one is not a pinhole, have not previously been reported in the literature. However, the author's results of 1.3% rms still compare favorably with Pentland's measured error of 2.5% [19], where one image was acquired with a pinhole aperture.

REFERENCES

- [1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [2] M. Bertero and T. A. Poggio, "Ill-Posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869-889, Aug. 1988.
- [3] M. Born and E. Wolf, *Principles of Optics*. New York: Pergamon, 1980, 6th ed.
- [4] V. Michael Bove, Jr., "Discrete fourier transform based depth-from-focus," in *Image Understanding and Machine Vision, 1989 Tech. Dig. Series*, Opt. Soc. Amer. Air Force Office Sci. Res., June 1989, vol. 14, pp. 118-121.
- [5] R. N. Bracewell, *The Fourier Transform and Its Applications*. New York: McGraw-Hill, 1986, 2nd ed.
- [6] J. E. Ens, "An investigation of methods for determining depth from focus," Ph.D. thesis, Univ. of British Columbia, July 1990.
- [7] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, vol. 93, no. 429-457, 1946.
- [8] B. Girod and S. Scherrock, "Depth from defocus of structured light," *Optics Illumination Image Sensing Machine Vision IV*, vol. 1194, 1989.
- [9] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading MA: Addison-Wesley, 1987, 2nd ed.
- [10] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New York: Dover, 1952.
- [11] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51-83, Jan. 1978.
- [12] E. Hecht, *Optics*. Reading MA: Addison-Wesley, 1987, 2nd ed.
- [13] H. L. F. von Helmholtz, *Helmholtz's Treatise on Physiological Optics (translated from the 3rd German ed.)* (J. P. C. Southall, Ed.). Rochester, New York: Opt. Soc. Amer., 1924.
- [14] H. H. Hopkins, "The frequency response of a defocused optical system," *Proc. Royal Soc. London, A*, vol. 231, pp. 91-103, 1955.
- [15] B. K. P. Horn, "Focusing," A. I. Memo 160, Mass. Inst. Technol., May 1968.
- [16] ———, *Robot Vision*. Cambridge MA: MIT Press, 1986.
- [17] T. R. Hsing, "Correction of pixel nonuniformities for solid-state imagers," in *Processing Images Data Opt. Sensors*, vol. 292, pp. 218-224, 1981.
- [18] F. Lei and H. J. Tiziani, "A comparison of methods to measure the modulation transfer function of aerial survey lens systems from image structures," *Photogram. Eng. Remote Sensing*, vol. 54, no. 1, pp. 41-46, Jan. 1988.
- [19] A. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple, real-time range camera," in *IEEE Comput. Soc. Conf. Comput. Vision Patt. Recogn.*, 1989, pp. 256-261.
- [20] A. P. Pentland, "Depth of scene from depth of field," in *Proc. Image Understanding Workshop*, Apr. 1982, pp. 253-259.
- [21] ———, "A new sense of depth of field," in *Ninth Int. Joint Conf. AI*, Aug. 1985, pp. 988-994.
- [22] ———, "A new sense of depth of field," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-9, pp. 523-531, July 1987.
- [23] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*. Cambridge, UK: Cambridge University Press, 1987.
- [24] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982, vol. 1, 2nd ed.
- [25] W. F. Schreiber, *Fundamentals of Electronic Imaging Systems (Springer Series in Information Sciences)*. Berlin: Springer-Verlag, 1986, vol. 15.
- [26] M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Proc. Second Int. Conf. Comput. Vision*, Dec. 1988, pp. 149-155.
- [27] ———, "Direct recovery of depth-map," Tech. Rep. 87-02, Comput. Vision Graphics Lab., Dept. of Elect. Eng., State Univ. of New York, Stony Brook, Feb. 1987.
- [28] ———, "Direct recovery of depth-map I: Differential methods," in *Proc. IEEE Comput. Soc. Workshop Comput. Vision*, Dec. 1987, pp. 58-65.
- [29] ———, "Direct recovery of depth-map II: A new robust approach," Tech. Rep. 87-03, Comput. Vision Graphics Lab., Dept. of Elect. Eng., State Univ. of New York, Stony Brook, Apr. 1987.
- [30] ———, "Efficient depth recovery through inverse optics," in *Machine Vision for Inspection and Measurement* (H. Freeman, Ed.). New York: Academic, 1989, pp. 101-126.
- [31] M. Subbarao and N. Gurumoorthy, "Depth recovery from blurred edges," in *IEEE Comput. Soc. Conf. Comput. Vision Patt. Recogn.*, 1988, pp. 498-503.
- [32] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math.*, vol. 4, pp. 1035-1038, 1963 (translation of Doklady Akademii Nauk SSSR).



John Ens received the B.A.Sc. (honors) degree in electrical engineering from the University of British Columbia (UBC) in 1981. In 1990, he received the Ph.D. degree in electrical engineering from UBC. His thesis was on "Depth from Focus."

He spent four years developing medical devices at Vancouver General Hospital and holds two patents for his work. During the next two years, as a research and development engineer at MacDonald Dettwiler and Associates, he became involved with image processing and computer vision. The following two years, he was a Research Associate in the Computer Science Department of Simon Fraser University, Burnaby, Canada, researching parallel computer vision. He is presently a principal consultant with Intense Technologies, Richmond, Canada, where he specializes in image processing and computer vision.

Dr. Ens is a member of the Professional Engineers of British Columbia.



Peter Lawrence received the Bachelors degree in electrical engineering from the University of Toronto, Toronto, Canada, in 1965, the Masters degree in biomedical engineering from the University of Saskatchewan, Saskatoon, Canada, in 1967, and the Ph.D. degree in computing and information science from Case Western Reserve University, Cleveland, OH, in 1970.

Subsequently, he was a Guest Researcher with the Department of Applied Electronics, Chalmers University, Goteborg, Sweden, during 1970-1972 and a Lecturer and Research Staff Member with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, during 1972-1974. He is now a Professor with the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada. His research interests have included control of electromechanical assistive devices for the disabled and biomedical signal processing. His main research interests are currently in telerobotic control of heavy mobile machines such as excavator-based machines used in the construction, forestry, and mining industries. His current projects include human interface studies, binocular displays, image-based measurement systems, hydraulic control systems, computational kinematics and dynamics, and real-time on-board computing systems.