

Depth from Focusing and Defocusing

Yalin Xiong

Steven A. Shafer

CMU-RI-TR-93-07

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

March, 1993

©1993 Carnegie Mellon University

A short version of this technical report will appear in the *Proceedings of Computer Vision and Pattern Recognition 1993* and *Proceedings of Image Understanding Workshop 1993*.

This research was sponsored by the Avionics Lab, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U. S. Air Force, Wright-Patterson AFB, OH 45433-6543 under Contract F33615-90-C-1465, ARPA Order No. 7597. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DARPA or the U. S. Government.

Contents

1	Introduction	1
2	Depth From Focusing	1
2.1	Introduction	1
2.2	Fibonacci Search and Curve Fitting	2
2.3	Error Analysis	3
2.4	Implementation and Result	6
2.4.1	Hardware	6
2.4.2	Experiments and Results	7
3	Depth From Defocusing	8
3.1	Introduction	8
3.2	Basic Theory	9
3.3	Gabor Transform and Window Effect	10
3.4	Maximal Resemblance Estimation	11
3.5	Fitting Algorithm	12
3.6	Blurring Model	13
3.7	Implementation and Results	14
3.7.1	Simulation	14
3.7.2	Variation Measure and Thresholding	16
3.7.3	Calibration of Blurring Function	16
3.7.4	Blurring Model Calibration	18
3.7.5	σ^2 -Map and Shape Recovery	19
4	Summary	19

List of Figures

1	Step Edge Image as Target	3
2	Focus Measure Profile	4
3	Fibonacci Search	5
4	Curve Fitting	6
5	Minimal Differentiable Motor Displacement	7
6	Motor Position Distribution	8
7	Iterative Estimation of $\sigma_1^2 - \sigma_2^2$	15
8	Absolute Error of Estimation	15
9	Edge Bleeding	16
10	Intensity Image and Variation Map	17
11	Variation Measure and Threshold	17
12	Blurring Function Fitting	18
13	Blurring Model	19
14	Pictures of Different Objects	20
15	σ^2 -Map Recovery	21
16	σ^2 -Map Recovery Without Maximal Resemblance Estimation	21
17	Shape Recovery For the Convex Object	22

List of Tables

1	Different Definitions of Minimal Differentiable Motor Displacements	6
---	---	---

Abstract

This paper studies the problem of obtaining depth information from focusing and defocusing, which have long been noticed as important sources of depth information for human and machine vision. In depth from focusing, we try to eliminate the local maxima problem which is the main source of inaccuracy in focusing; in depth from defocusing, a new computational model is proposed to achieve higher accuracy.

The major contributions of this paper are: (1) In depth from focusing, instead of the popular Fibonacci search which is often trapped in local maxima, we propose the combination of Fibonacci search and curve fitting, which leads to an unprecedentedly accurate result; (2) New model of the blurring effect which takes the geometric blurring as well as the imaging blurring into consideration, and the calibration of the blurring model; (3) In spectrogram-based depth from defocusing, an iterative estimation method is proposed to decrease or eliminate the window effect.

This paper reports focus ranging with less than $1/1000$ error and the defocus ranging with about $1/200$ error. With this precision, depth from focus ranging is becoming competitive with stereo vision for reconstructing 3D depth information.

1 Introduction

Obtaining depth information by actively controlling camera parameters is becoming more and more important in machine vision, because it is passive and monocular. Compared with the popular stereo method for depth recovery, this focus method doesn't have the correspondence problem, therefore it is a valuable method as an alternative of the stereo method for depth recovery.

There are two distinct scenarios for using focus information for depth recovery:

- *Depth From Focus*: We try to determine distance to one point by taking many images in better and better focus. Also called "autofocus" or "software focus". Best reported result is 1/200 depth error at about 1 meter distance [14].
- *Depth From Defocus*: By taking small number of images under different lens parameters, we can determine depth at all points in the scene. This is a possible range image sensor, competing with laser range scanner or stereo vision. Best reported result is 1.3% RMS error in terms of distance from the camera when the target is about 0.9 m away [6].

Both methods have been limited in past by low precision hardware and imprecise mathematical models. In this paper, we will improve both:

- *Depth From Focus*: We propose a stronger search algorithm with its implementation on a high precision camera motor system.
- *Depth From Defocus*: We propose a new estimation method and a more realistic calibration model for the blurring effect.

With this new results, focus is becoming viable as technique for machine vision applications such as terrain mapping and object recognition.

2 Depth From Focusing

2.1 Introduction

Focusing has long been considered as one of major depth sources for human and machine vision. In this section, we will concentrate on the precision problem of focusing. we will approach high precision from both software and hardware directions, namely, stronger algorithms and more precise camera system.

Most previous research on depth from focusing concentrated on developments and evaluations of different focus measures, such as [17, 7, 9, 15]. Among them, [15] provided a theory for evaluating various focus measures based on OTF(Optical Transfer Function) instead of experimental evaluations which could be biased by specific scenes used for experiments. As described by all these researchers, an ideal focus measure should be unimodal, monotonic, and should reach the maximum only when the image is focused. But as pointed out in [18, 15], the focus measure profile has many

local maxima due to noises and/or the side-lobe effect ([15]) even after magnification compensation ([18]). This essentially requires a more complicated peak detection method compared with the Fibonacci search which is optimal under the unimodal assumption as in [1, 7]. In this paper, we use a recognized focus measure from the literature, which is the Tenengrad with zero threshold in [7] or M_2 method in [15]. Our major concern is to discover to what extent the precision of focus ranging can scale up with more precise camera systems and more sophisticated search algorithms. We knew the popular Fibonacci search method would not be enough when we learned that being trapped in local maxima is the major cause of focusing error. Instead, We propose the combination of Fibonacci search and curve fitting to detect the peak of focus measure profile precisely and quickly.

To evaluate the results from peak detections, an error analysis method is presented to analyze the uncertainty of the peak detection in the motor count space, and to convert the uncertainty in the motor count space into uncertainty in depth. We compute the variance of motor positions resulted from peak detections over equal depth targets. The Rayleigh criterion of resolution is applied to the distribution of motor positions to calculate the minimal differentiable motor displacement. With the assumption of local linearity of the mapping from the motor count space to focus depth, the minimal differentiable motor displacement is converted to the minimal differentiable depth.

The lack of high precision equipment has been a limiting factor to previous implementations of various focus ranging methods. Many implemented systems, such as SPARCS, have fairly low motor resolution, which actually prohibits more precise results. We will give brief description of the motor-driven camera system in Calibrated Imaging Lab later, and further details can be found in [19].

2.2 Fibonacci Search and Curve Fitting

When the focus motor resolution is high, we usually have a very large parameter space which prevents us from exhaustively searching all motor positions. Based on the unimodal assumption of focus measure profile, Fibonacci search was employed to narrow the parameter space down to the peak [1].

Assume the initial interval is $[x, y]$, and we know the focus measure profile is unimodal in this interval, if $x < x_1 < x_2 < y$ and $F(x_1) < F(x_2)$, where F is the focus measure function, then the peak can not be within interval $[x, x_1)$, otherwise the unimodal assumption will be violated. Therefore, if we can properly choose x_1 and x_2 , the peak can be found optimally. Fibonacci search is the optimal search under the unimodal assumption.

Figure 1 shows the target used for testing the focus measure, and Figure 2 is the focus measure profile of the target.

It is clear from Figure 2 that Fibonacci search will fail to detect the peak precisely because of the jagged profile. Fortunately, those local maxima are small in size, and therefore can be regarded as disturbances. From previous paragraphs, we know that the Fibonacci search only evaluates at two points within the interval, which gives

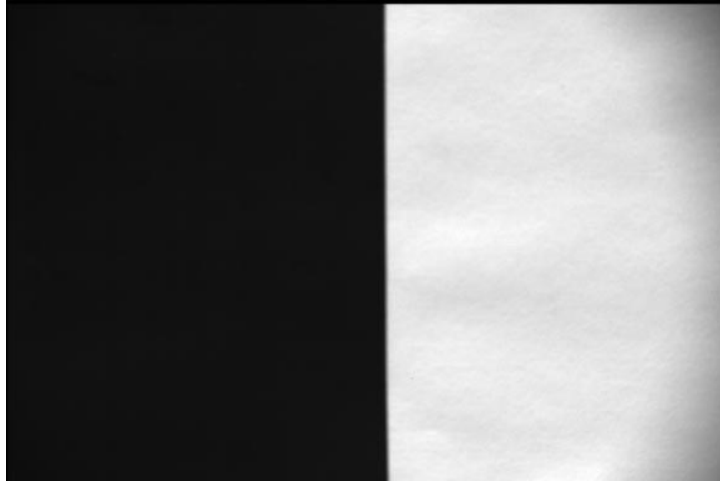


Figure 1: Step Edge Image as Target

rise to the hope that when the interval is large, Fibonacci search is still applicable because it will overlook those small ripples.

As the search goes on, the interval becomes smaller and smaller. Consequently, Fibonacci search must be aborted at some point when the search might be misleading. We can experimentally set up a threshold, when the length of the interval is less than the threshold, Fibonacci search is replaced by an exhaustive search. After the exhaustive search, a curve is fitted to the part of profile resulting from the exhaustive search.

In our experiments, we set the threshold to be 5 motor counts. So when the Fibonacci search narrows down the whole motor space to $[a, b]$, where $b - a \leq 5$, an exhaustive search is fired on the interval $[a - c, b + c]$, where c is a positive constant. A Gaussian function is fitted to the profile in the interval $[a - c, b + c]$ using the least square method described in [11].

Figure 3 shows the result when Fibonacci search alone is applied to the focus measure profile. Apparently, the search is trapped in a local maximum. Figure 4 shows the result from Gaussian function fitting. Both graphs show only a part of the whole motor space.

2.3 Error Analysis

The depth error from focus ranging generally results from two sources: error in peak detection and error in mapping from camera parameters to range. Here, we are primarily concerned with the error in peak detection, because in our experiments, the repeatability and accuracy of the camera motor system is so high that the error of the mapping calibration can be regarded as negligible compared with the error from the peak detection.

The major sources of error in peak detection are:

- *Image Noise*: Since the focus criterion is essentially a gradient operator, high

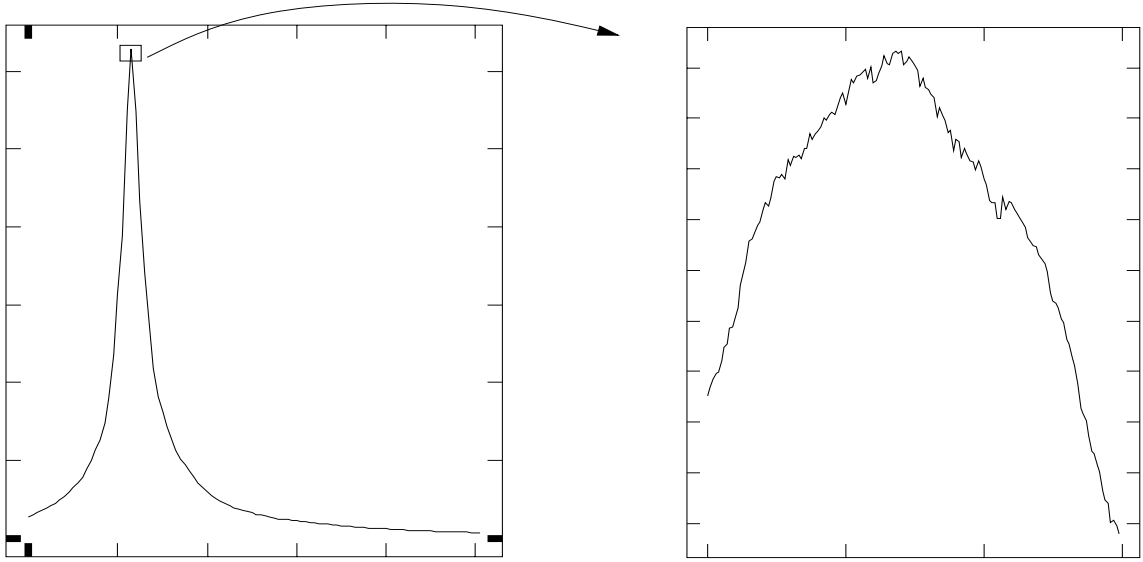


Figure 2: Focus Measure Profile

frequency noise may cause a problem. In our experiments, this is the major limiting factor.

- *Image Shift:* Focusing involves changes of camera parameters, which, in turn, may cause image misregistrations. There are two major kinds of image misregistrations: one is the magnification change due to different effective focal length, the other one is the image shift caused by the imperfection of the optical system, for example, if the axes of lenses are not exactly collinear, changing camera parameters may not only cause a magnification change of the image, but also will shift the whole image to the extent of several pixels [18].
- *Search Strategy:* As explained earlier, naive search might be trapped in local maxima.
- *Content of Image:* Since the focus measure is essentially a high-pass filter, it requires that the content within the evaluation window have enough high frequency components. In the extreme case, a uniform intensity pattern will provide no information about depth at all. Quantitatively, more high frequency components mean a sharper peak in Figure 2, and therefore, more precise peak detection.

As explained earlier, we are concerned about the error from searching, so other sources of error should be identified and minimized. Image magnification changes can be compensated by camera calibration as suggested in [4, 18]. To tolerate the image shift caused by optic system, we used a fairly large window 40x40. Because we use a target as in Figure 1, the image magnification change and the image shift can be ignored if the edge is in the middle of the window. Therefore, in our experiments,

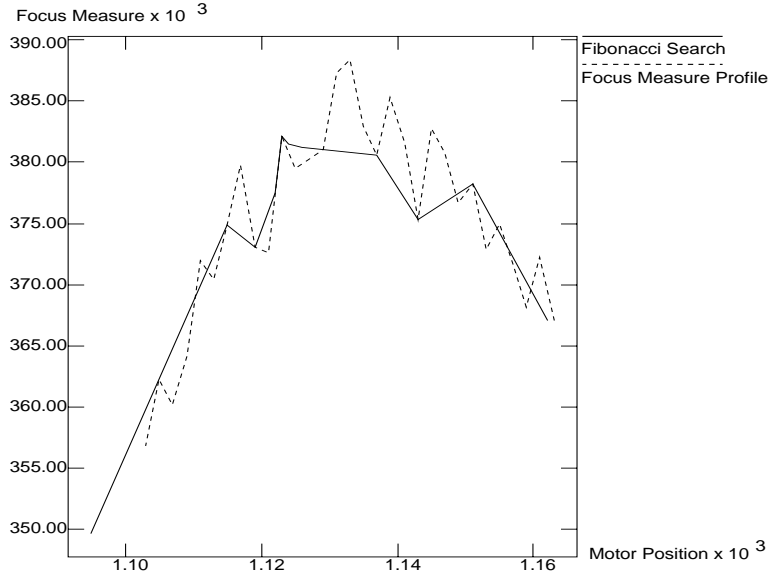


Figure 3: Fibonacci Search

we didn't use the magnification compensation. And the content of image certainly includes enough high frequency components.

Because of the depth accuracy we expected, a direct measurement of absolute depth is impossible. Instead, we prefer to use the minimal differentiable depth as an indication of depth accuracy. If we assume the peak motor positions resulting from the same repeated experiments have a Gaussian distribution, we can define the minimal differentiable motor displacement as the minimal difference of two motor counts which have pre-defined probability of representing different peaks. For example, in Figure 5, the Gaussian distribution is artificially cut at 90% line, so we can say, if we do one focus ranging experiment on a target at the depth corresponding to the peak a , there is a 90% of probability the motor count will be within the interval A .

There can be different pre-defined probability for the definition of minimal differentiable motor displacement. Table 1 is a list of probabilities and the corresponding motor displacements in terms of the Gaussian constant σ . We define the minimal differentiable motor displacement based on Rayleigh criterion for resolution [2] which specifies the saddle-to-peak ratio as $8/\pi^2$. In case of the Gaussian distribution, the cut-off line corresponding to the Rayleigh criterion is about 0.9σ .

There is a mapping from a motor count to an absolute depth value definitely. Assume $d = f(m)$ where d is the depth, m the motor count and f the mapping, we have

$$\frac{\Delta d}{\Delta m} = f'(m), \quad (1)$$

where $f'(m)$ is the first order derivative with respect to m . Because what we really want to know is the minimal differential depth or depth resolution Δd , and we already have the minimal differentiable motor displacement Δm , the only thing need to be calibrated is $f'(m)$. If we assume $f'(m)$ is a constant in the vicinity of $d = D$, and

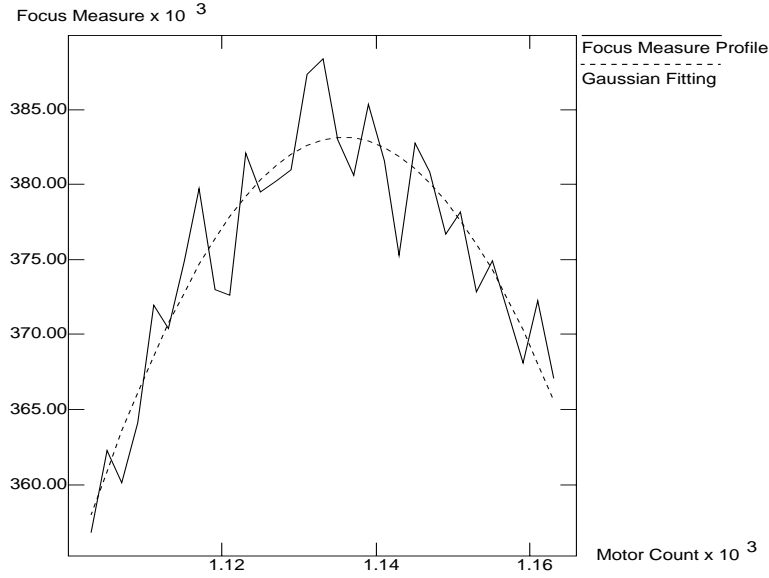


Figure 4: Curve Fitting

Table 1: Different Definitions of Minimal Differentiable Motor Displacements

Cut-off Line of Gaussian	Probability of correctness
0.68σ	50%
0.84σ	60%
1.03σ	70%
1.28σ	80%
1.64σ	90%

the motor count distribution has its center at $m = M$, then when the target is moved ΔD , the distribution center moves ΔM , and we will have the minimal differentiable depth

$$\Delta d = \frac{\Delta m}{\Delta M} \Delta D \quad (2)$$

where Δm is the minimal differentiable motor displacement.

2.4 Implementation and Result

2.4.1 Hardware

We implemented this focus ranging algorithm in the Calibrated Imaging Laboratory, using the Fujinon/Photometric camera system [19]. This system consists of a Fujinon ENG zoom lens mounted on a Photometrics Star I scientific camera. The camera can provide a 12-bit per pixel greyscale image. Color band can be selected by the filter wheel mounted in front of the lens.

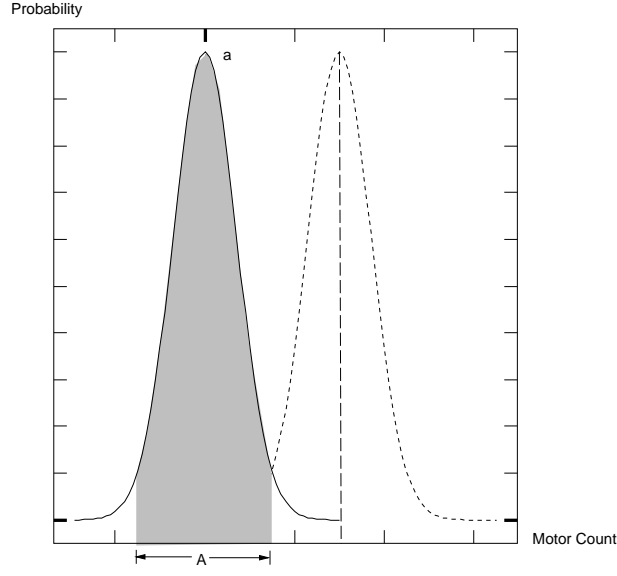


Figure 5: Minimal Differentiable Motor Displacement

The focal length can change between 10 mm to 130 mm with 11100 motor steps, and the focus distance can change from approximately 1 meter to infinity with 5100 motor steps, the aperture can change from F1.7 to completely closed with 2700 motor steps. The SNR of the camera can be as low as 400/1 because of the pixel by pixel digitization and the -40°C temperature of the sensor.

2.4.2 Experiments and Results

We put the target of Figure 1 at about 1.2 meters away from the front lens element of the camera. Maximal focal length and maximal aperture are employed to achieve the minimal depth of field. The evaluation window is 40x40, while the gradient operator is a 3x3 Sobel operator.

The distribution of motor positions are sketched in Figure 6 resulting from an experiment repeated 40 times. With the mean as the center of a Gaussian, and the standard deviation as σ of the Gaussian, we have the minimal differentiable motor displacement as $2 \times 0.9 \sigma = 4.5$ motor counts.

Then the target is moved toward the camera 1 centimeter, and we repeated the above experiments. The center of the motor count distribution moves 38 counts. Therefore, by Eq. 2, we have the minimal differentiable depth:

$$\Delta d = \frac{\Delta m}{\Delta M} \Delta D = \frac{4.5}{38} \times 1\text{cm} = 0.118\text{cm}. \quad (3)$$

And the relative depth error is about $0.118 / 120 = 0.098\%$.

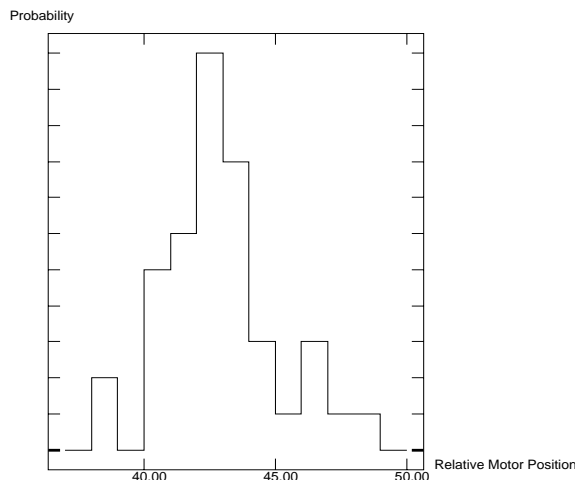


Figure 6: Motor Position Distribution

3 Depth From Defocusing

3.1 Introduction

The depth from defocusing method uses the direct relationships among the depth, camera parameters and the amount of blurring in images to derive the depth from parameters can be directly measured. In this part of the paper, we propose an iterative method to estimate the amount of defocusing accurately, and a calibration-based blurring model.

Because the blurring in an image can be caused by either the imaging process or the scene itself, it generally requires at least two images taken under different camera configurations to eliminate this ambiguity. Pentland solved this problem by taking one picture by a pin-hole camera, which can be regarded as the orthographic projection of the scene with zero imaging blurring, and another one by a wide aperture camera [10]. In this paper, we intend to employ two images which are defocused to different extents.

Window effects have largely been ignored in the literature of this field, except [6, 5], where the author derived a function of RMS depth error in terms of the size of window. For example, when the window is 4 pixels by 4 pixels, the RMS error from the window effect can be as large as 65.8%! The iterative method we propose is capable of eliminating the window effect. It is also noticed that the size of the window is the decisive factor that limits the resolution of depth maps if we try to obtain a dense depth map. Therefore if we can use smaller window without reducing the quality of the results, the resolution of dense depth maps can be much higher.

Previous work has employed oversimplified camera models to derive the relationship between blurring functions and camera configurations. In [10, 13, 3], the radius of blurring circles are derived from the ideal thin lens model. In this paper, we will propose a more sophisticated function which directly relates the blurring function

with camera motors. Experimental results are very consistent with this model as to be shown later.

3.2 Basic Theory

The depth from defocus method is based on the idea that, the amount of blurring change is directly related to the depth and camera parameters. Since the camera parameters can be calibrated, the depth can be expressed by the amount of blurring change correspondingly. To estimate the amount of blurring change, we need a model of the optical blurring. Traditionally, the blurring effect is modeled as the convolution of Gaussian in computer vision literature, partly due to its mathematical tractability. Here we will still assume a Gaussian model, i.e.:

$$I(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I_0(\xi, \eta) g_{\sigma(d, c)}(x - \xi, y - \eta) d\xi d\eta \quad (4)$$

$$g_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

where $I(x, y)$ is the intensity image, d is the depth of the object, c is the vector of camera parameters, $g_{\sigma(d, c)}$ is the blurring function.

For simplicity, we can instead consider the one dimensional case, in which Eq. 4 and Eq. 5 become

$$I(x) = \int_{-\infty}^{+\infty} I_0(\xi) g_{\sigma(d, c)}(x - \xi) d\xi \quad (6)$$

$$g_{\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (7)$$

The basic idea of the depth-from-defocus method is that, in Eq. 6, since the $I(x)$ which is the image and c which results from camera calibration are known, and d and $I_0(x)$ are unknown, we can take two images under different camera settings c_1 and c_2 , then at least theoretically d can be computed.

But since the Eq. 6 is not a linear equation with respect to unknown d , directly solving for d is either impossible or numerically unstable. Pentland proposed a method to solve d by Fourier transforms [10]:

If

$$I_1(x) = I_0(x) * g_{\sigma_1}(x) \quad (8)$$

$$I_2(x) = I_0(x) * g_{\sigma_2}(x) \quad (9)$$

then,

$$\ln \frac{\mathcal{F}[I_1(x)]}{\mathcal{F}[I_2(x)]} = \ln \frac{\mathcal{I}_1(f)}{\mathcal{I}_2(f)} = \ln \frac{\mathcal{I}_0(f) \mathcal{G}_{\sigma_1}(f)}{\mathcal{I}_0(f) \mathcal{G}_{\sigma_2}(f)} = \ln \frac{\mathcal{G}_{\sigma_1}(f)}{\mathcal{G}_{\sigma_2}(f)} = -\frac{1}{2} f^2 (\sigma_1^2 - \sigma_2^2) \quad (10)$$

$$\sigma_1 = \sigma(d, c_1) \quad (11)$$

$$\sigma_2 = \sigma(d, c_2) \quad (12)$$

Replacing Eq. 11 and Eq. 12 into Eq. 10, we can get:

$$\ln \frac{\mathcal{I}_1(f)}{\mathcal{I}_2(f)} = -\frac{1}{2} f^2 (\sigma^2(d, c_1) - \sigma^2(d, c_2)) \quad (13)$$

where the function σ can be calibrated.

Obviously, in Eq. 13, the only unknown is d , therefore, depth recovery from two images is straightforward.

3.3 Gabor Transform and Window Effect

The method explained above is based on $\mathcal{F}[I(x)]$, which is the Fourier transform of the entire image. Thus, only one d can be calculated from the entire image. If our goal is to obtain a dense depth map $d(x, y)$, we are forced to use the STFT (Short Time Fourier Transform) to preserve the depth locality. To eliminate the spurious high frequency components generated by the discontinuity at the window boundary, people usually multiply the window by a window function. Unfortunately, the elegant cancellation in Eq. 10 doesn't hold any more if we introduce the window function $W(x)$:

$$\ln \frac{\mathcal{F}[I_1(x)W(x)]}{\mathcal{F}[I_2(x)W(x)]} = \ln \frac{\mathcal{I}_1(f) * \mathcal{W}(f)}{\mathcal{I}_2(f) * \mathcal{W}(f)} = \ln \frac{(\mathcal{I}_0(f)\mathcal{G}_{\sigma_1}(f)) * \mathcal{W}(f)}{(\mathcal{I}_1(f)\mathcal{G}_{\sigma_2}(f)) * \mathcal{W}(f)} \quad (14)$$

The convolutions in Eq. 14 introduce blurring in both the time (space) and frequency domains. The Gabor transform [12], which uses a Gaussian as the window function, can minimize the production of spatial uncertainty and spectral uncertainty. Assuming a Gaussian function is used as $W(x) = g_\sigma(x)$, and its Fourier transform is $\mathcal{W}(f) = e^{-\sigma^2 f^2/2}$, we have:

$$\Delta f^2 = \frac{\int f^2 | \mathcal{W}(f) |^2 df}{\int | \mathcal{W}(f) |^2 df} = \frac{1}{2\sigma^2} \quad (15)$$

$$\Delta x^2 = \frac{\int x^2 | W(x) |^2 dx}{\int | W(x) |^2 dx} = \frac{\sigma^2}{2} \quad (16)$$

The above equations mean that, in the frequency domain, two frequency components Δf away can't be discriminated, and in the space domain, two impulses Δx away can't be discriminated either. Apparently, one interpretation of why Eq. 10 doesn't hold is that Δf is not zero. As σ approaches infinity, Δf approaches zero. In other words, as the window function approaches a constant function, the Gabor transform becomes more and more accurate in the frequency domain, but on the other hand, less and less accurate in the space domain as indicated in Eq. 16.

3.4 Maximal Resemblance Estimation

From observation, we know that when σ_1 is approaching σ_2 , Eq. 14 also approaches zero, in other words, when σ_1 almost equals σ_2 , the Eq. 10 can be a good approximation in terms of absolute error. This observation suggests an iterative method in which the blurring difference Δ is refined by blurring one image to resemble the other in the vicinity of one pixel. In symbols: (Assuming $\Delta_{(k)}$ is the the k th estimation of $\sigma_1^2 - \sigma_2^2$)

1. $I_1^{(0)} = I_1, I_2^{(0)} = I_2$ and $\Delta = 0.0, k = 0$;
2. $\mathcal{I}_1^{(k)} = \mathcal{F}[I_1^{(k)}W],$
 $\mathcal{I}_2^{(k)} = \mathcal{F}[I_2^{(k)}W].$
3. Fit a curve to $\ln \frac{\mathcal{I}_1^{(k)}}{\mathcal{I}_2^{(k)}} = -f^2 \Delta_{(k)}/2$. (Refer to Eq. 10)
4. $\Delta = \sum_{i=0}^k \Delta_{(i)}.$
5. If $\Delta > 0$, then
 $I_1^{(k+1)} = I_1;$
 $I_2^{(k+1)} = I_2 * G_{\sigma=\sqrt{\Delta}};$
 else,
 $I_1^{(k+1)} = I_1 * G_{\sigma=\sqrt{-\Delta}};$
 $I_2^{(k+1)} = I_2;$
- Note all these convolutions are done very locally because of the window function multiplication in step 2.
6. If the termination criteria are satisfied, exit.
7. $k = k+1$, go to step 2.

All above operations involve only local pixels, and don't require taking new pictures. Therefore, the computation can be done in parallel to all pixels to obtain a dense depth map.

Let's trace the above iterations, at the first cycle, we have (Assuming $\sigma_1 > \sigma_2$):

$$\Delta_{(0)} = (\sigma_1^2 - \sigma_2^2) + E_{(0)}$$

while $E_{(0)}$ is the error of estimating $\sigma_1^2 - \sigma_2^2$.

$$I_1^{(1)} = I_1 = I_0 * G_{\sigma_1}$$

$$I_2^{(1)} = I_2 * G_{\sqrt{\Delta}} = I_0 * G_{\sqrt{\sigma_1^2 + E_{(0)}}}$$

We can see, after the first iteration, we actually switched to estimate $E_{(0)}$. So after k iterations, we have

$$\Delta = \sum_{i=0}^k \Delta_{(i)} = \sigma_1^2 - \sigma_2^2 + E_{(k)} \quad (17)$$

Now the problem is whether the sequence $E_{(k)} (k = 0, 1, 2, \dots)$ converges to zero. Unfortunately, there is no way to prove this mathematically because it depends on the fitting method used in step 3. Notice that if in step 3, we get an estimate of $\sigma_1^2 - \sigma_2^2$ only based on one particular frequency component, $E_{(k)}$ may diverge. Previous depth from defocus methods usually counted on a pre-selected frequency band, such as in [10], sometimes this may cause a very large error if there is not enough energy of the image content within the frequency band.

3.5 Fitting Algorithm

Common to any frequency analysis, we need a robust algorithm to extract $\sigma_1^2 - \sigma_2^2$ in Eq. 10 in a noisy environment. Ignoring the phase information resulting from Gabor transform, Eq. 10 becomes:

$$\ln \frac{|\mathcal{I}_1(f)|^2}{|\mathcal{I}_2(f)|^2} = -f^2(\sigma_1^2 - \sigma_2^2) \quad (18)$$

Assuming an additive white noise model, we have:

$$\ln \frac{|\mathcal{I}_1(f)|^2 + n_1}{|\mathcal{I}_2(f)|^2 + n_2} = \ln(|\mathcal{I}_1(f)|^2 + n_1) - \ln(|\mathcal{I}_2(f)|^2 + n_2), \quad (19)$$

where n_1 and n_2 are energy of noises. Because $\ln(x + dx) \approx \ln x + \frac{1}{x}dx$, if we assume $|\mathcal{I}_1(f)|^2 \gg n_1$ and $|\mathcal{I}_2(f)|^2 \gg n_2$, Eq. 19 can be approximated as:

$$\ln(|\mathcal{I}_1(f)|^2 + n_1) - \ln(|\mathcal{I}_2(f)|^2 + n_2) \approx \ln \frac{|\mathcal{I}_1(f)|^2}{|\mathcal{I}_2(f)|^2} + \left(\frac{n_1}{|\mathcal{I}_1(f)|^2} - \frac{n_2}{|\mathcal{I}_2(f)|^2} \right) \quad (20)$$

Therefore, at each frequency, the left hand of Eq. 18 can be approximated by dividing corresponding spectral energy of two images at the specific frequency, provided that the energy values are much larger than the energy of noise. The deviation of this approximation can be expressed as:

$$\sigma_f = c_n \left(\frac{1}{|\mathcal{I}_1(f)|^2} + \frac{1}{|\mathcal{I}_2(f)|^2} \right) \quad (21)$$

where c_n is a constant related to the noise energy of the camera.

Certainly, Eq. 21 is an approximation to model the error distribution as an Gaussian. As an intuition, when $|\mathcal{I}_1(f)|^2$ or $|\mathcal{I}_2(f)|^2$ is large, i.e. the energy within the frequency is high, the deviation is small, and *vice versa*.

From Eq. 18, we know this estimation problem is a typical linear regression problem. With the uncertainty measurement approximated in Eq. 21, $\sigma_1^2 - \sigma_2^2$ can be estimated robustly. More details can be found in [11].

There is one more problem need to be addressed. Since we obtain images under different camera configurations, the total energy within the image is different. Usually, a brightness normalization is performed to every image before Gabor transforms, as in [16]. But since this normalization will have different effects over the noise in two images, which will complicate the uncertainty analysis, we prefer not to normalize the brightness in two images. Instead we assume:

$$I_1(x) = c_1 I_0(x) * g_{\sigma_1}(x) \quad (22)$$

$$I_2(x) = c_2 I_0(x) * g_{\sigma_2}(x), \quad (23)$$

where c_1 and C_2 are two unknown constants. Replacing the two equations into Eq. 18, we have:

$$\ln \frac{|\mathcal{I}_1(f)|^2}{|\mathcal{I}_2(f)|^2} = -f^2(\sigma_1^2 - \sigma_2^2) + 2 \ln \frac{c_2}{c_1}, \quad (24)$$

which is still a linear problem, while the uncertainty analysis still holds.

3.6 Blurring Model

Since the defocus ranging method derives the depth instead of searching for the depth, it requires a direct modeling of defocusing in terms of camera parameters and depth. Previous researchers usually derived the relation among lens parameters, the depth and the blurring radius, such as in [10, 13]. For example, in [10], by simple geometric optics, Pentland derived the formula:

$$D = \frac{Fv_0}{v_0 - F - \sigma kf} \quad (25)$$

where D is the depth, F the focal length, f the f -number of the lens, v_0 the distance between lens and image plane, σ the blurring circle radius, and k a constant.

The basic limitation of this approach is that those parameters are based on the ideal thin lens model and in fact, they can never be measured precisely on any camera. We desire a function which is in terms of motor counts, which are measurable and controllable. For instance, if we use m_z for zoom motor count, m_f for focus motor count, and m_a for aperture motor count, we wish to get a function in the form of $D = F(m_z, m_f, m_a, \sigma)$ or $\sigma = F(m_z, m_f, m_a, D)$. Since there is a depth ambiguity in the former form ([10], Appendix), we prefer to express the blurring radius σ as a function of motor counts and the depth.

From Eq. 25, we can express σ as:

$$\sigma = \frac{v_0 - F}{kf} - \frac{Fv_0/kf}{D} \quad (26)$$

Since all the lens parameters can be thought as derived from motor counts, we can rewrite Eq. 26 as:

$$\sigma = k_1(m_z, m_f, m_a) + \frac{k_2(m_z, m_f, m_a)}{D + k_3(m_z, m_f, m_a)} \quad (27)$$

Notice that there is another term k_3 added. Because D is the distance between lens and object, and the position of lens changes as camera parameters are changed, we intend to use a fixed plane perpendicular to the optical axis as the depth reference plane at $z = k_3$. From now on, we always refer to depth as the distance between the target and the depth reference plane.

Eq. 27 says, at some point, σ will drop to zero, which is the best focused point. But we knew that we can never get a *real* step edge, because there is always high frequency loss in the imaging process. It can be attributed to the pixel quantization, diffraction, etc. Similarly, we can model this as a convolution with a Gaussian independent of the geometric blurring which we already modeled. We use k_4 to model its width.

Since two consecutive convolutions with Gaussians are equivalent with one Gaussian convolution:

$$G_{\sigma_1} * G_{\sigma_2} = G_{\sqrt{\sigma_1^2 + \sigma_2^2}}, \quad (28)$$

we have our final blurring model expression as:

$$\sigma = \sqrt{\left(k_1(m_z, m_f, m_a) + \frac{k_2(m_z, m_f, m_a)}{D + k_3(m_z, m_f, m_a)}\right)^2 + k_4^2(m_z, m_f, m_a)} \quad (29)$$

3.7 Implementation and Results

3.7.1 Simulation

Our first simulation examines how precise the estimate of $\sigma_1^2 - \sigma_2^2$ can be. We use step function as I_0 , and convolve it with two different Gaussian G_{σ_1} and G_{σ_2} . The window function is also a Gaussian with σ equal to three pixel widths.¹ From Eq. 16, the locality of the window function is about 2 pixel widths.

The result of the iterative method is illustrated in Fig. 7. And we can see that, when the window function is narrow, how poor the first estimation can be. Experimentally, the final error is lower bounded by the discretization of the functions, that is, when the σ of the Gaussian function is too small with respect to the pixel width, the discrete Gaussian is no longer a good approximation of the real Gaussian function, and the results begin to degenerate.

Figure 8 shows the absolute error of the estimating $\sigma_1^2 - \sigma_2^2$ in the absence of noise. Generally, the errors are less than 1/1000 of true values. Remember that in those figures, the blurring radius σ due to defocus can be even much larger than the window size!

¹In this report, all σ values are in pixel width.

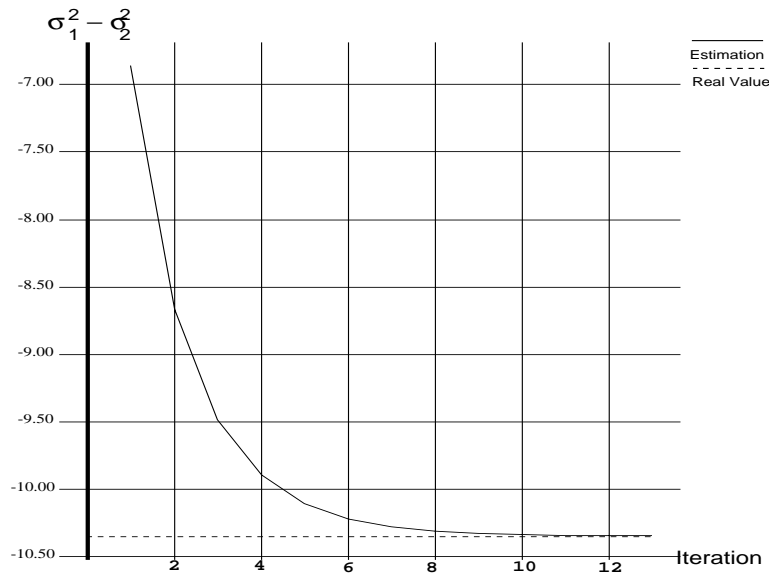


Figure 7: Iterative Estimation of $\sigma_1^2 - \sigma_2^2$

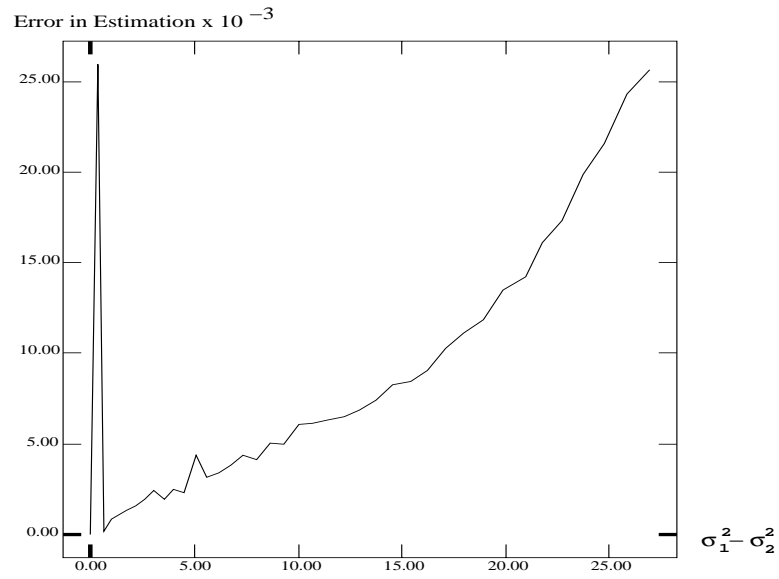


Figure 8: Absolute Error of Estimation

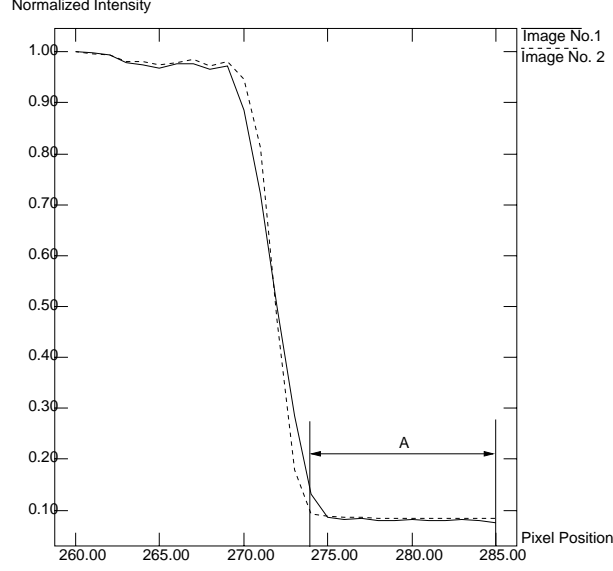


Figure 9: Edge Bleeding

3.7.2 Variation Measure and Thresholding

Certainly, if there is no texture or little texture within the image, we can not expect to obtain accurate estimation of depth. Therefore, we need a measure which can discriminate image patches with enough texture from those without enough texture. Another reason why we need a variation measure is because of the so-called edge bleeding [8]. For example, in Figure 9 case, where there is a step edge blurred by different amounts, apparently, image No. 1 contains less high frequency components than image No. 2 does, but if the window is at A , the image variation of A in image No. 1 is larger than that in image No. 2, and therefore wrong results will be deducted.

Assuming the image patch is $f(x, y)$, we have the variation measure expressed as in Eq. 30. Applying it to a real image, we have the variation map as in Figure 10, which properly quantifies the variation content within neighbors of pixel.

$$V = \iint \left(\left| \frac{\partial f(x, y)}{\partial x} \right|^2 + \left| \frac{\partial f(x, y)}{\partial y} \right|^2 \right) g_{\sigma_v}(x, y) dx dy \quad (30)$$

To better illustrate the relation between the variation measure and the result of estimate of $\sigma_1^2 - \sigma_2^2$, Figure 11 shows the selection of the threshold to exclude the effects of the low variation content and the edge bleeding.

3.7.3 Calibration of Blurring Function

First, we tried to confirm our assumption that the blurring function can be approximated by the Gaussian function. Ideally, if we have a point light source, the image of this light source should be the blurring function because

$$\delta(x) * F(x) = F(x). \quad (31)$$

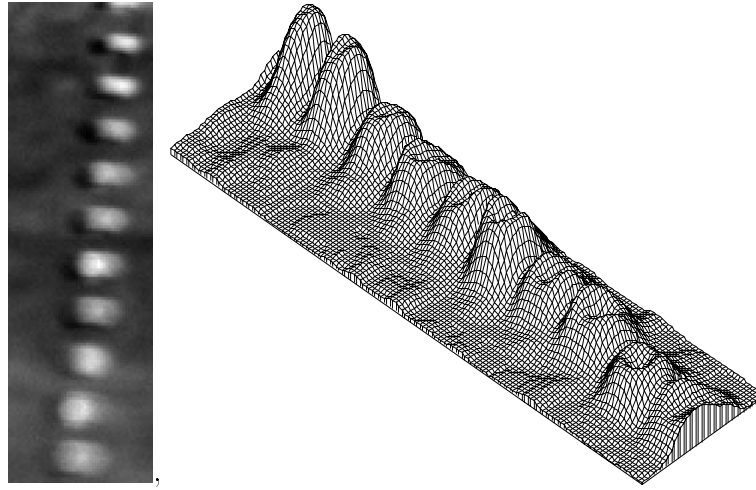


Figure 10: Intensity Image and Variation Map

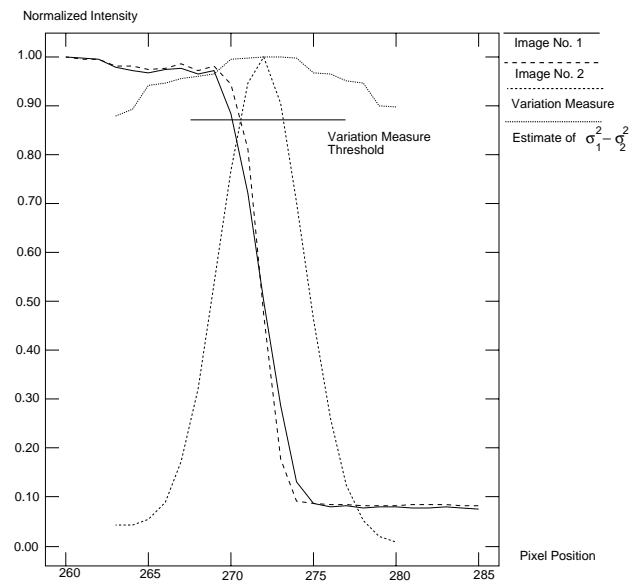


Figure 11: Variation Measure and Threshold

Due to the technical difficulty of an ideal point light source, a step edge image is used instead as shown in Figure 1. Assume the step function is $u(x)$, the image of the step edge should be

$$g_\sigma(x) * u(x) = \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-t)^2}{2\sigma^2}} dt = c_1 + c_2 \operatorname{Erf} \left(\frac{x}{\sqrt{2}\sigma} \right) \quad (32)$$

where Erf is the error function, c_1 and c_2 are two constants.

The Figure 12 illustrates the least square fitting results for a blurred step edge.

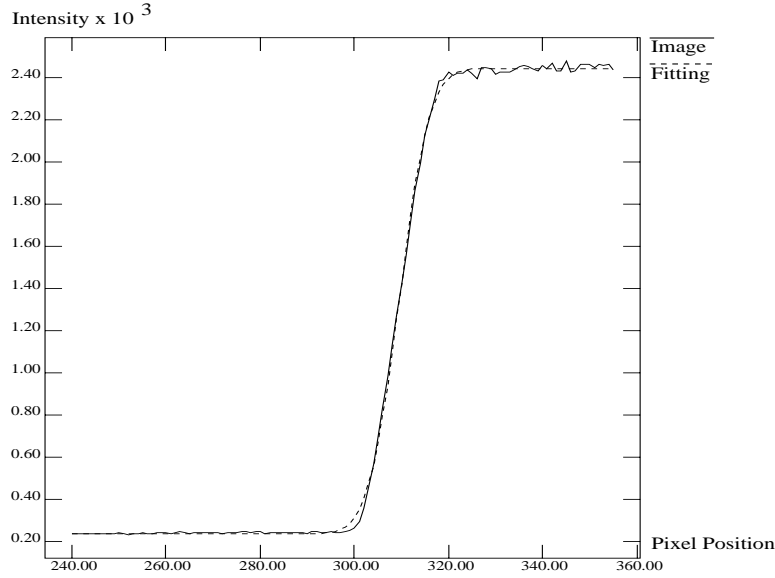


Figure 12: Blurring Function Fitting

3.7.4 Blurring Model Calibration

The coefficients k_1, k_2, k_3, k_4 are constants in Eq. 29 when motors are fixed. We can move the calibration target over four different places, and assume at the first place, the depth of the target is zero (Note the depth is w.r.t. the reference plane), we will have four non-linear equations with four unknowns. To suppress noise, we can measure at more than four places and fit the blurring model to the results.

Using the rail table in CIL ([19]), the whole process of calibration can be automated. The target moves from about 1.5 meter from the camera to about 3.5 meters, and the blurred edges are fed to the least square fitting described in the previous section, the resulting σ 's are, in turn, fitted against the model expressed in Eq. 29.

Experiments have shown very consistent results with the model as illustrated in Figure 13. The target is moved from far to near, at the furthest distance the rail motor position is zero. And when it moved through the whole range of the rail, the blurring circle first becomes smaller and smaller, then after a point, it becomes larger and larger. It is very clear that this effect can be well modeled by Eq. 29.

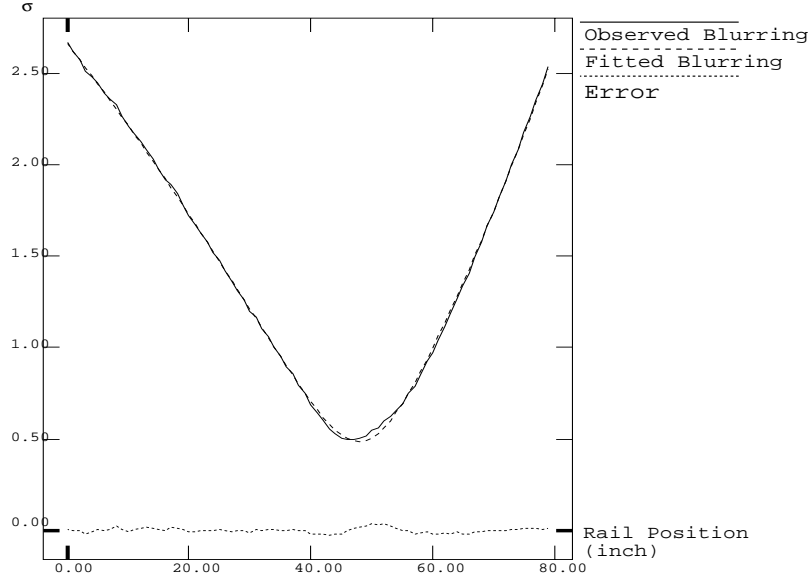


Figure 13: Blurring Model

3.7.5 σ^2 -Map and Shape Recovery

The first step toward a dense depth map is to compute $\sigma^2 = \sigma_1^2 - \sigma_2^2$, without loss of generality we assume $\sigma_1 \geq \sigma_2$, for every pixel, using the maximal resemblance estimation. In Figure 14, we bent a sheet of A4 paper in different directions about 1.0 inches and took images. The target is about 100 inches away from the camera. The focal length is 130mm, the f-number is f/4.7 for (a) and (c), f/8.1 for (b) and (d).

Then we recover σ^2 -map for those two objects. The rectangle in Figure 14 is the area for σ^2 -map. The σ_w for Gabor transform is 5.0 pixel size. Figure 15 shows the σ^2 -map recovery based on the images in Figure 14. The holes within the σ^2 maps are those patches without enough texture.

Compared with the σ^2 map recovery without iterative maximal resemblance estimation showed in Figure 16, we can see that results without iteration are much more noisy.

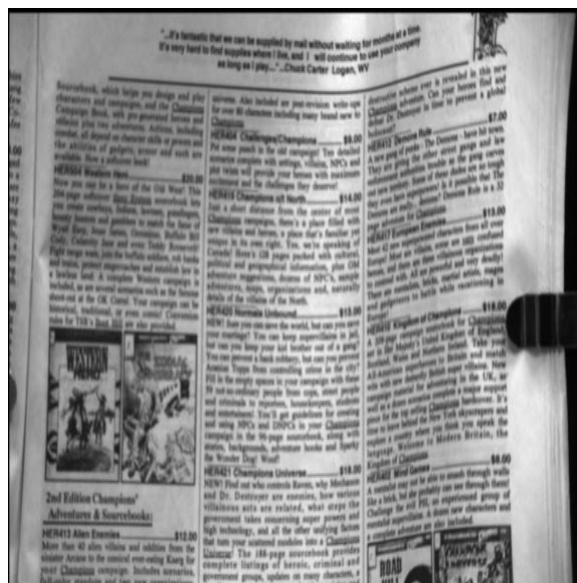
With σ^2 map recovered and the coefficients in Eq. 29 calibrated w.r.t. the two camera configurations, the depth map recovery is straightforward by using the Brent's method [11] to numerically solve the nonlinear equation. Figure 17 showed the depth map of the convex object in Figure 14 (c) and (d), with respect to the depth reference plane, which is behind the object.

4 Summary

In summary, we have described two sources of depth information—depth from focusing and depth from defocusing—separately. In depth from focusing, we pursued high accuracy from both the software and hardware directions, and experiments proved



(a) Concave Object Image No. 1



(b)Concave Object Image No.2

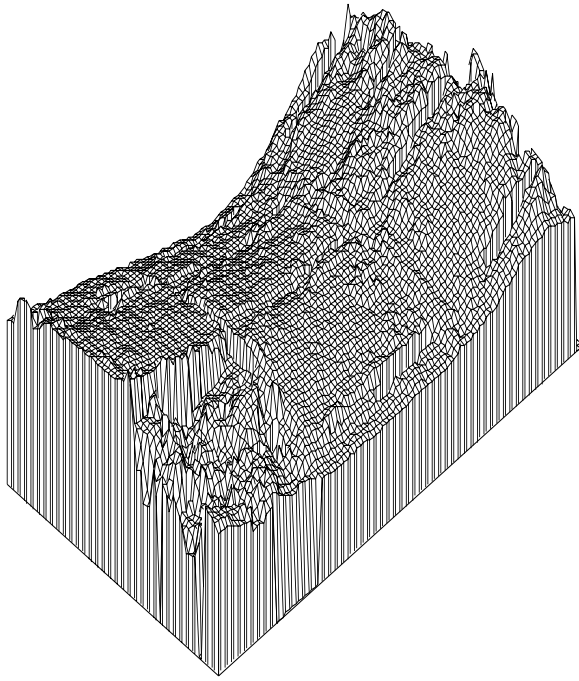


(c) Convex Object Image No. 1

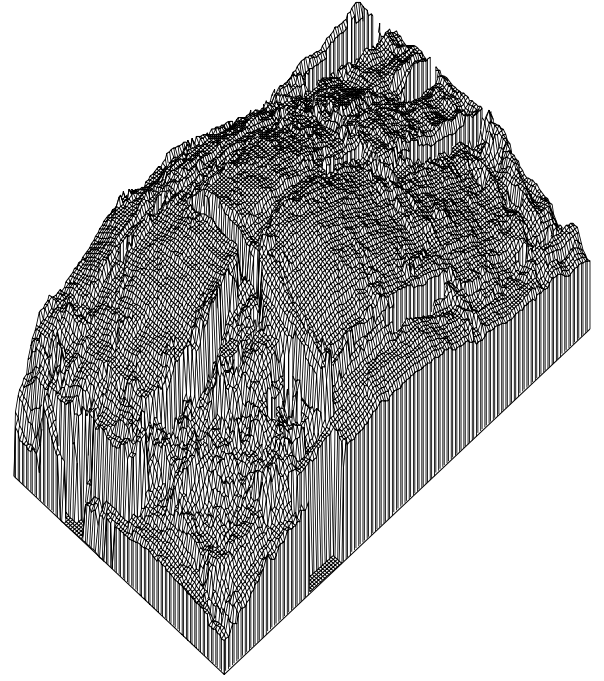


(d) Convex Object Image No. 2

Figure 14: Pictures of Different Objects



(a) Concave Object



(b) Convex Object

Figure 15: σ^2 -Map Recovery

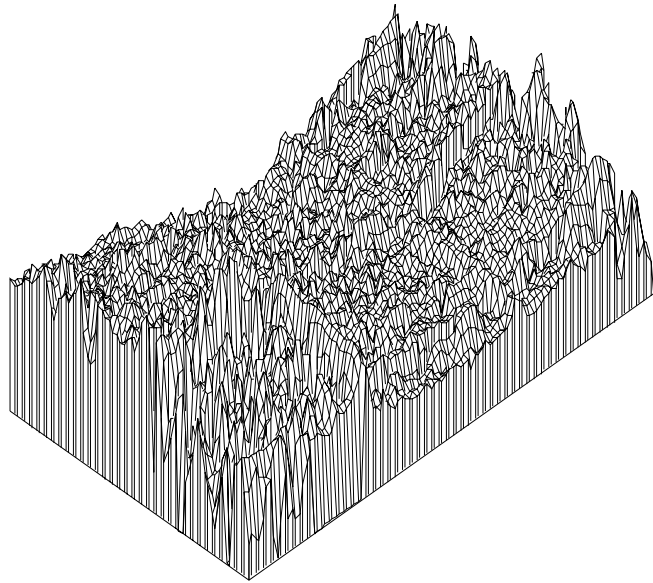


Figure 16: σ^2 -Map Recovery Without Maximal Resemblance Estimation

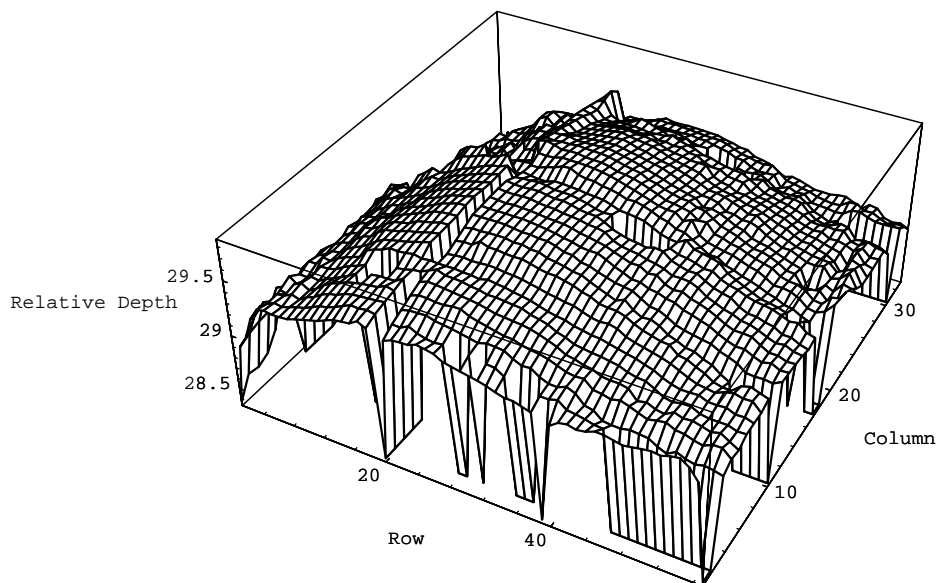


Figure 17: Shape Recovery For the Convex Object

that a great improvement was obtained. In depth from defocusing, we re-examined the whole underlying theory, from signal processing to camera calibration, and established a new computational model, which has been successfully demonstrated on both synthesized and real images.

The significance of these works is two-fold. First, there are few previous reports talking about the shape from defocusing or focusing, and the main reason of the inefficacy of shape from defocusing or focusing is its low precision. Demonstrated in this paper, the improvements on precisions of focus ranging and defocus ranging can lead to efficient shape recovery methods. Second, it has been shown that the iterative method proposed in this paper is capable of preserving the depth locality, which is also essential to obtain dense depth map.

References

- [1] G. S. Beveridge and R. S. Schechter. *Optimization: Theory and Practice*. McGraw-Hill, 1970.
- [2] Max Born and Emil Wolf. *Principles of Optics*. The MACMILLAN COMPANY, 1964.
- [3] V. Michael Bove, Jr. Discrete fourier transform based depth-from-focus. In *Proceedings OSA Topical Meeting on Image Understanding and Machine Vision*, 1989.
- [4] Trevor Darrell and Kwangyeon Wahn. Pyramid based depth from focus. In *Proceedings of CVPR*, pages 504–509, 1988.

- [5] John Ens. *An Investigation of Methods For Determining Depth From Focus*. PhD thesis, University of British Columbia, 1990.
- [6] John Ens and Peter Lawrence. A matrix based method for determining depth from focus. In *Proceedings of CVPR*, 1991.
- [7] Eric P. Krotkov. Focusing. *International Journal of Computer Vision*, pages 223–237, 1987.
- [8] Hari N. Nair and Charles V. Stewart. Robust focus ranging. In *Proceedings of CVPR*, pages 309–314, 1992.
- [9] Shree K. Nayar and Yasuo Nakagawa. Shape from focus: An effective approach for rough surfaces. In *International Conference on Robotics and Automation*, pages 218–225, 1990.
- [10] Alex P. Pentland. A new sense for depth of field. *IEEE Transactions on PAMI*, 9(4):523–531, 1987.
- [11] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [12] Olivier Rioul and Martin Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, pages 14–38, October 1991.
- [13] Murali Subbarao. Parallel depth recovery by changing camera parameters. In *2nd International Conference on Computer Vision*, pages 149–155, 1988.
- [14] Murali Subbarao. Presentation at the symposium on physics-based vision workshop. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1992.
- [15] Murali Subbarao, Tae Choi, and Arman Nikzad. Focusing techniques. Technical Report 92.09.04, Department of Electrical Engineering, State University of New York at Stony Brook, 1992.
- [16] Murali Subbarao and Tse-Chung Wei. Depth from defocus and rapid autofocus-ing: A practical approach. In *Proceedings of CVPR*, pages 773–776, 1992.
- [17] J. M. Tenenbaum. *Accommodation in Computer Vision*. PhD thesis, Stanford University, 1970.
- [18] Reg G. Willson and Steven A. Shafer. Dynamic lens compensation for active color imaging and constant magnification focusing. Technical Report CMU-RI-TR-91-26, The Robotics Institute, Carnegie Mellon University, 1991.
- [19] Reg G. Willson and Steven A. Shafer. Precision imaging and control for machine vision research at carnegie mellon university. Technical Report CMU-CS-92-118, School of Computer Science, Carnegie Mellon University, 1992.