Depth from Focusing and Defocusing

Yalin Xiong Steven A. Shafer

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

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. 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, a maximal resemblance estimation method is proposed to decrease or eliminate the window effect.

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 [8].
- *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 [3].

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

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 [4, 5, 9]. 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 the focus measure profile has many local maxima due to noises and/or the side-lobe effect ([9]) even after magnification compensation ([10]). This essentially requires a more complicated peak detection method compared with the Fibonacci search which is optimal under the unimodal assumption as in [4]. In this paper, we use a recognized focus measure from the literature, which is the Tenegrad with zero threshold in [4] or M_2 method in [9]. Our major concern is to discover to what extent

1063-6919/93 $03.00 © 1993 IEEE

the precision of focus ranging can scale up with more precise camera systems and more sophisticated search algorithms. 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 of depth.

The lack of high precision equipment has been a limiting factor to previous implementations of various focus ranging methods. We used the motor-driven camera system in CIL, and further details can be found in [11].

2.1 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 [4].

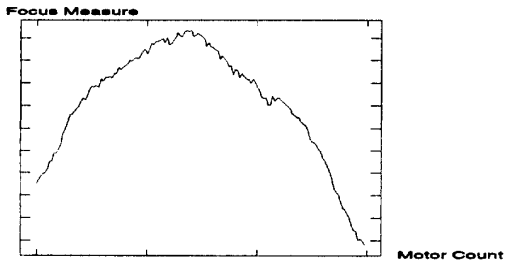


Figure 1: Focus Measure Profile

Figure 1 is the focus measure profile of the step edge target. It is clear from Figure 1 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 the process of Fibonacci search, we know that the Fibonacci search only evaluates at two points within the interval, which gives 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.

Figure 2 shows the result when Fibonacci search alone is applied to the focus measure profile. Apparently, the search is trapped in a local maximum. Figure 3 shows the result from Gaussian function fitting. Both graphs show only a part of the whole motor space.

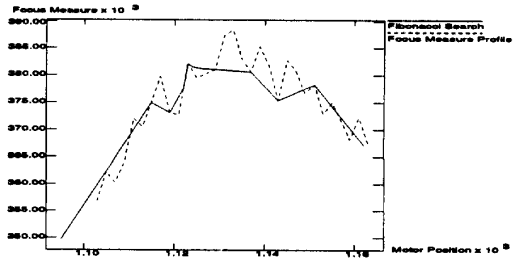


Figure 2: Fibonacci Search

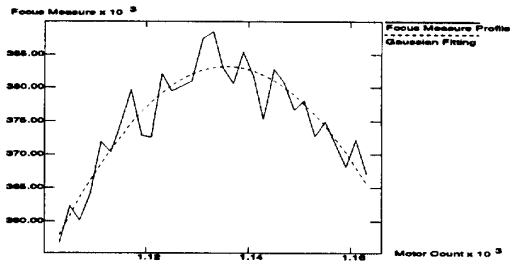


Figure 3: Curve Fitting

2.2 Error Analysis

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 the 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.

There can be different pre-defined probability for the definition of minimal differentiable motor displacement. We define the minimal differentiable motor dis-

placement based on Rayleigh criterion for resolution [1] which specifies the saddle-to-peak ratio as $8/\pi^2$.

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

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

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 $f'(m)$.

2.3 Implementation and Results

We put the step edge target 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 4 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 4.5 motor counts.

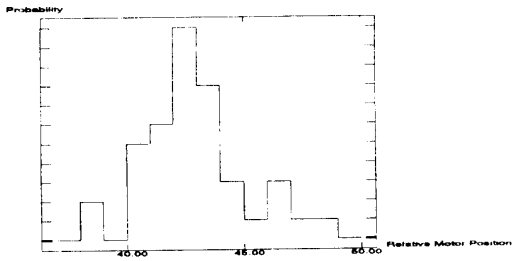


Figure 4: Motor Position Distribution

Then the target is moved toward the camera 1 centimeter, and we repeated the above experiments. The center of the motor count distribution moves 38.0 counts. Therefore, assuming the linearity of $f'(m)$ in the small interval, we have the minimal differentiable depth:

$$\Delta d = \frac{\Delta m}{\Delta M}\Delta D = \frac{4.5}{38} \times 1\text{cm} = 0.118\text{cm}. \tag{2}$$

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

3 Depth From Defocusing

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 which can be directly measured. In this part of the paper, we propose the maximal resemblance estimation method to estimate the amount of defocusing accurately, and a calibration-based blurring model.

Window effects have largely been ignored in the literature of this field, except [3], where the author derived a function of RMS depth error in terms of the size of window. The maximal resemblance estimation 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 [6, 7, 2], 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.

3.1 Maximal Resemblance Estimation

As explained in [6, 12], if we take two images $I_1(x)$ and $I_2(x)$ under different camera configurations, we can recover depth by the following equation:

$$\ln \frac{\mathcal{I}_1(f)}{\mathcal{I}_2(f)} = -\frac{1}{2}f^2(\sigma^2(d, c_1) - \sigma^2(d, c_2)) \tag{3}$$

where d is the depth value, $\mathcal{I}_1(f)$ and $\mathcal{I}_2(f)$ are the Fourier transforms of $I_1(x)$ and $I_2(x)$ respectively, c_1 and c_2 are two vectors of lens parameters, the function σ can be calibrated.

This method 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, we can no longer have the same elegant equation as Eq. 3. [12]

To deal with the window effect problem, we propose 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 kth 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. 3)
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.

Common to any frequency analysis, we need a robust algorithm to extract $\sigma_1^2 - \sigma_2^2$ in Eq. 3 in a noisy environment. For each frequency, the left hand of Eq. 3 can be approximated by dividing corresponding spectral energy of two images at the specific frequency, provided that the energy in that frequency is much larger than the energy of noise. The error of this energy division caused by noise can be expressed as: [12]

$$\sigma_f = c_n \left(\frac{1}{|\mathcal{I}_1(f)|^2} + \frac{1}{|\mathcal{I}_2(f)|^2} \right) \quad (4)$$

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

3.2 Blurring Model

Since the defocus ranging method derives the depth instead of searching for it, 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 [6, 7]. For example, in [6], by

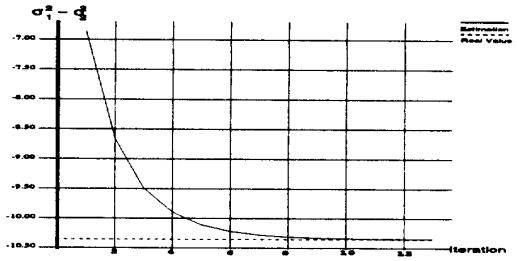


Figure 5: Iterative Estimation of $\sigma_1^2 - \sigma_2^2$

simple geometric optics, Pentland derived the formula:

$$D = \frac{F v_0}{v_0 - F - \sigma k f} \quad (5)$$

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. By taking pixel averaging, diffraction, and other implementational factors into consideration, we come up with a blurring model in motor space: [12]

$$\sigma^2 = \left(k_1(m_z, m_f, m_a) + \frac{k_2(m_z, m_f, m_a)}{D + k_3(m_z, m_f, m_a)} \right)^2 + k_4^2(m_z, m_f, m_a) \quad (6)$$

where we use m_z for zoom motor count, m_f for focus motor count, and m_a for aperture motor count.

3.3 Implementation and Results

3.3.1 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 equals to three pixel widths.[1]

The result of the iterative method is illustrated in Fig. 5. And we can see that, when the window function is narrow, how poor the first estimation can be. As the iteration goes on, the estimated value converges fast to the true value.

3.3.2 Calibration of The Blurring Model: The coefficients k_1, k_2, k_3, k_4 are constants in Eq. 6 when motors are fixed. We can therefore calibrate those

[1] In this paper, all σ values are in pixel width.

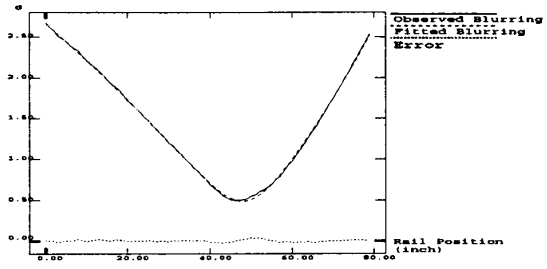


Figure 6: Blurring Model

constants by measuring the blurring amount of a step edge over several different depth.

Using the rail table in CIL ([11]), the whole process of calibrating blurring model 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, the resulting σ's are, in turn, fitted against the model expressed in Eq. 6.

3.3.3 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 assumed $\sigma_1 \geq \sigma_2$, for every pixel, using the maximal resemblance estimation. In Figure 7, we bent a sheet of A4 paper in different directions about 1.0 inchs and took images. The target is about 100 inchs 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 7 (a) is the area for σ^2-map. The σ_w for Gabor transform is 5.0 pixel size. Figure 8 shows the σ^2-map recovery based on the images in Figure 7. The holes within the σ^2-maps are those patchs without enough texture.

Compared with the σ^2-map recovery without iterative maximal resemblance estimation showed in Figure 9, we can see that results without iteration are much more noisy.

With σ^2-map recovered and the coefficients in Eq. 6 calibrated w.r.t. the two camera configurations, the depth map recovery is straightforward by using the Brent's method to numerically solve the nonlinear equation. Figure 10 showed the depth map (in inch) of the convex object in Figure 7 (c) and (d), with respect to the depth reference plane, which is behind the object. A conservative estimation of depth relative error is 1/200 when the target is 100 inchs away.

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 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 real images.

References

[1] Max Born and Emil Wolf. *Principles of Optics.* The MACMILLAN COMPANY, 1964.

[2] V. Michael Bove, Jr. Discrete fourier transform based depth-from-focus. In *Proceedings OSA Topical Meeting on Image Understanding and Machine Vision*, 1989.

[3] John Ens and Peter Lawrence. A matrix based method for determining depth from focus. In *Proceedings of CVPR*, 1991.

[4] Eric P. Krotkov. Focusing. *International Journal of Computer Vision*, pages 223–237, 1987.

[5] 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.

[6] Alex P. Pentland. A new sense for depth of field. *IEEE Transactions on PAMI*, 9(4):523–531, 1987.

[7] Murali Subbarao. Parallel depth recovery by changing camera parameters. In *2nd International Conference on Computer Vision*, pages 149–155, 1988.

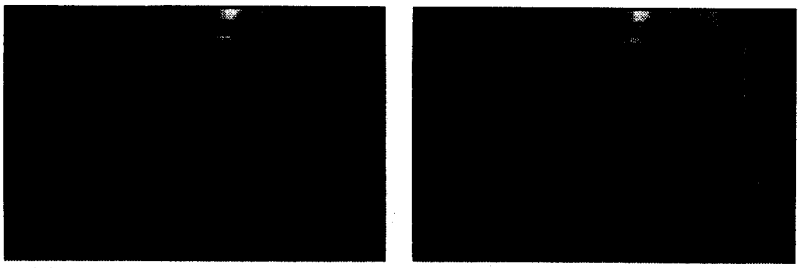
[8] Murali Subbarao. Presentation at the symposium on physics-based vision workshop. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1992.

[9] 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.

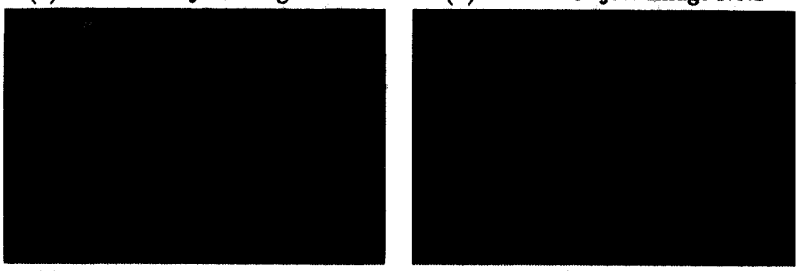
[10] Reg G. Willson and Steven A. Shafer. Dynamic lens compensation for active color imaging and constant magnification focusing. Technical Report CMU-RI-TR-91-26, The Robotics Institute, Carnegie Mellon University, 1991.

[11] 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.

[12] Yalin Xiong and Steven Shafer. Depth from focusing and defocusing. Technical Report CMU-RI-TR-93-07, The Robotics Institute, Carnegie Mellon University, 1993.

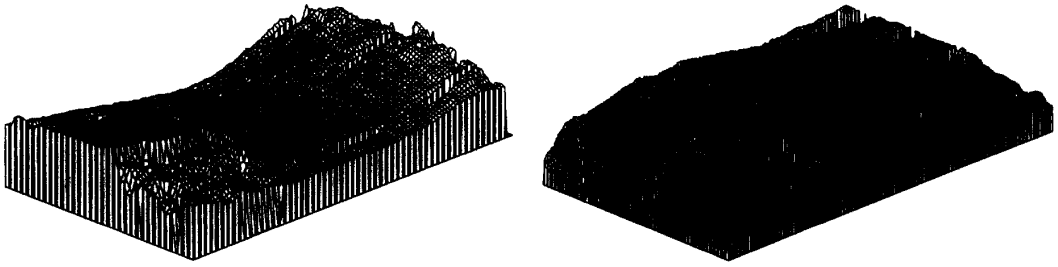


(a) Concave Object Image No. 1 (b)Concave Object Image No.2



(c) Convex Object Image No. 1 (d)Convex Object Image No. 2

Figure 7: Pictures of Different Objects



(a) Concave Object (b) Convex Object

Figure 8: σ^2-Map Recovery

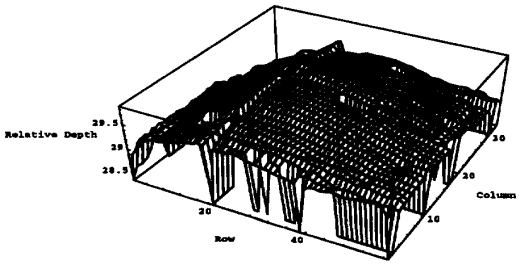


Figure 10: Shape Recovery For the Convex Object

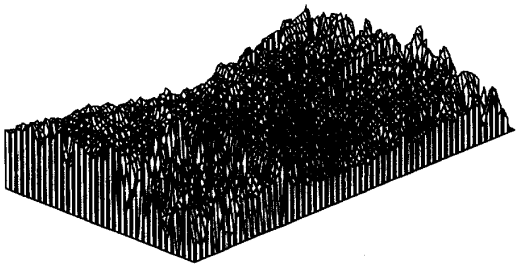


Figure 9: σ^2-Map Recovery Without Maximal Resemblance Estimation