# Depth from Defocus and Rapid Autofocusing: A Practical Approach

Murali Subbarao and Tse-Chung Wei
Dept. of Electrical Engg., State University of New York, Stony Brook, New York 11794-2350

## Abstract

*A new method named DFD1F of determining depth (range) from image defocus and rapid autofocusing of a camera is presented. It requires only two images in theory (but three images in our implementation). In contrast with a related prior method [?, ?], DFD1F is based on computing only one-dimensional Fourier coefficients as opposed two-dimensional Fourier coefficients, thus providing not only computational advantage but also robustness in practical applications. DFD1F is independent of the form of the Modulation Transfer Function of the camera.*

*DFD1F has been successfully implemented and tested on an actual camera system named SPARCS built in our laboratory. SPARCS can determine the distance of an object placed in front of it at any distance in the range 0.5 meter to infinity and successfully focus the object by moving the lens with a root-mean square error of less than 6% in terms of lens position. We believe this performance to be unmatched by any prior method (based on image defocus) we are aware of which uses only three or fewer images.*

## 1 Introduction

A well-known approach for determining the distance of an object in machine vision is what we call *Depth-from-Focus* or DFF. In this approach, the space of physical parameters of a camera is searched to find a set of camera parameter values which focuses the object. The distance is then determined based on the values of the camera parameters. The search usually requires a large number (infinite in theory but about 10 or more in practice) of images to be acquired at different camera parameter settings and processing. An example of recent work in DFF is [?].

Recently, some researchers [?, ?] have proposed methods for finding distance of an object which does not involve focusing the object. They take the level of defocus of the object into account in determining distance. Therefore, we call the approach taken by these methods to be *Depth-from-Defocus* or DFD. DFD methods do not involve any searching and they require only a few (two in theory but three or four in practice) images to be acquired and processed. The distance of all objects in a scene can be determined from only these few images, irrespective of whether the objects are focused or not in any of the images. Like depth-from-stereo, DFD approach is easily implemented in parallel, but unlike stereo, the problems of correspondence and occlusion can be avoided.

Several methods based on DFD approach have been proposed during the last 5 years (see CVPR/ICCV proceedings and IEEE-PAMI). These methods have one or more weaknesses such as restriction on the form of the point spread function of the camera systems, restriction on the camera parameters and appearance of objects, high noise sensitivity, limited range of effectiveness, etc. Demonstration of the practical utility of DFD approach has been far from satisfactory.

In this paper we outline (see [?] for details) a DFD method named DFD1F for arbitrary images which does not restrict the form of the point spread function of the camera systems. DFD1F has been correctly and successfully implemented on an actual camera system named SPARCS. Experimental results show that the method is efficient, robust, and useful in practical applications. The level of performance of the method is such that a major camera manufacturer is planning to apply the method for rapid autofocusing of consumer cameras.

DFD1F is derived from the DFD method proposed in [?] named DFD2F. While DFD2F is theoretically sound and complete, it is not efficient and not robust in the presence of noise. We did implement DFD2F on SPARCS successfully, but, in comparison with DFD1F, the performance was unsatisfactory.

DFD1F has been successfully applied for both ranging (i.e., determining distance) and rapid autofocusing. It is more efficient than DFD2F [?] because it involves computing only one-dimensional Fourier coefficients as opposed to two-dimensional Fourier coefficients (hence the suffixes 1F and 2F in their names). This facilitates easy hardware implementation of DFD1F. Also, DFD1F is robust in the presence of zero-mean noise (not necessarily random) because it involves summing grey levels of many pixels, and it uses a special technique for comparing computed values to pre-stored calibration data which makes the method practical.

## 2 DFD1F

Due to space limitation (imposed by the referees of this paper!), we refer the reader to [?, ?, ?] for details about DFD1F. However, we shall provide an outline of DFD1F with emphasis on practical implementation so that our experiments can be duplicated by others.

A schematic diagram of a camera system with variable camera parameters is shown in Fig. 1. DFD1F is implemented on a system named *Stonybrook Passive Autofocusing and Ranging Camera System* (SPARCS) built by us in our laboratory. Fig. 2 shows a schematic diagram of SPARCS.

SPARCS has a SONY XC-711 color CCD camera, an Olympus 35-70 mm motorized lens, a Contec mPIO24/24 digital I/O board for the control of lens movement, a Data-Translation QuickCapture DT2953 frame grabber, an IBM PS/2 computer, and a SONY PVM-1342Q color monitor for real-time image display. The lens motor is a stepper motor with 97 steps numbered from 0 to 96. The system is set up such that a C program running on the PS/2 computer can move the lens to any desired position specified by step number and take pictures and process them.

When the lens is at one extreme position corresponding to step number 0 of motor, an object at distance infinity will be in best focus. As the lens is moved gradually to the other extreme position, the step number increases from 0 to 96, and simultaneously the distances of objects in best focus decreases monotonically from infinity to about 50 cm. Based on this observation, we can associate with each step number a distance corresponding to best focus and vice versa. This relation between step number and object distance can be used to specify distances of objects in terms of step numbers. We shall do so since it is convenient in autofocusing. As an example, if the distance of an object is said to be step number 35, it means that the object's distance is such that the object would be in best focus if the lens is moved to step number 35. Incidentally, the relation between lens step number and the *reciprocal* of best focused object distance is almost linear [?]. It can be stored in a lookup table.

The overall operation of SPARCS for finding the distance of an object can be summarized as below. The lens is first moved to step 10 and a first image $g_{10}$ of the object of interest is recorded. The lens is then moved to step 40, and a second image $g_{40}$ of the object is recorded. (Lens position is only one of a set of camera parameters such as focal length and aperture diameter. Any one or more of these parameters may be changed for the second image.) Optionally, we can specify the number of image frames (typically 4) to be recorded which are then averaged (over time) to reduce electronic noise. Such frame averaging is particularly needed under low illuminations, and in the presence of flickering illumination such as fluorescent lamps. Bright incandescent lamps are highly recommended for this reason.

As in consumer cameras, the object to be focused is specified by specifying a region on the image. The default region is the center of the image but it can be changed. The size of the region is also an option and the default size is $128 \times 128$.

In order to reduce the effect of the *image overlap* problem [?] at the borders of an image, the image is weighted (i.e. multiplied) by a two-dimensional Gaussian centered at the center of the image and having a spread parameter $\sigma$ equal to about 1/3rd of the image size (i.e. about 40 for a $128 \times 128$ image).

The two images are then summed rowwise to obtain two one-dimensional sequences, say $g_{10}[i]$ and $g_{40}[i]$. This step is a major improvement over DFD2F. As a result of summing grey levels along rows, the effect of any zero-mean noise is greatly reduced. The noise need not even be random. In fact, our camera has a systematic periodic noise of vertical bar pattern of about 10 pixels period which had adverse effect on the performance of DFD2F.

The two 1D sequences are then normalized with respect to brightness. This is done by dividing each value of the sequence by the mean value of the entire sequence. At present, our implementation does not normalize the sequences with respect to other types of distortions such as vignetting and sensor response characteristics of the camera as these distortions were not significant. For the same reason, we ignored the magnification normalization. In SPARCS, the change in magnification due to change in lens position is only about 2%. If these distortions are not negligible, then they must be corrected for.

Next, the first 6 discrete Fourier coefficients corresponding to lowest 6 non-zero frequencies of $g_{10}$ and $g_{40}$ are computed. (Theoretically, a single Fourier coefficient suffices, but in practice, more are needed. The number 6 was chosen empirically based on noise and maximum allowable blur.) Let these be $G_{10}(\rho)$ and $G_{40}(\rho)$ for $\rho = 1, 2, \cdots, 6$. A computed table $T_c$ is obtained by calculating

$$T_c(\rho) \;=\; \frac{-2}{\rho^2} \ln \frac{G_{10}(\rho)}{G_{40}(\rho)}. \qquad (1)$$

The Modulation Transfer Function (MTF) of the lens system as a function of object distance $u$ and spatial frequency $\rho$ was provided to us by the lens manufacturer. This information was provided for both lens positions– step number 10 and 40. The MTF of our camera is circularly symmetric, and therefore the two MTFs can be denoted by $H_{10}(\rho, u)$ and $H_{40}(\rho, u)$ where $u$ is the object distance (expressed in lens step number corresponding to best focus). The manufacturer obtained this MTF data using a computer simulation of the lens system. The same data could be obtained through direct measurements on the lens system using special equipment. The manufacturer provided the data at intervals of 1 cycle/mm spatial frequencies starting from 1 cycle/mm to 15 cycles/mm. However, for our camera, each pixel corresponds to 0.601 cycles/mm (inter pixel distance: 0.013 mm). Therefore, the MTF data provided by the manufacturer is coarser (1 cycle/mm) than what we would like (0.601 cycle/mm).

A transform named log-by-rho-squared transform was applied to the the two MTFs to obtain two tables $T_{10}$ and $T_{40}$ defined by

$$T_{10}(\rho, u) \;=\; \frac{-2}{\rho^2} \ln H_{10}(\rho, u), \;\; T_{40}(\rho, u) \;=\; \frac{-2}{\rho^2} \ln H_{40}(\rho, u) \qquad (2)$$

The effect of the above transform is to make the new values nearly a constant with respect to $\rho$. The reason for this is that, for low frequencies, the MTF resembles a Gaussian. However, the fact that it is not a Gaussian exactly does not introduce any errors into DFD1F. This step is an important improvement over DFD2F. In order to obtain these table data at intervals of 0.601 cycles/mm from the data available at intervals of 1 cycle/mm, we used a linear interpolation scheme. Linear interpolation gives satisfactory results because the log-by-rho-squared transform makes the table values to be nearly constant.

For robustness against noise, we discarded data at points where the magnitude of the Fourier coefficients $G_{10}(\rho)$ or $G_{40}(\rho)$ was low. This threshold was arbitrarily chosen to be around 30% of the maximum magnitude. One effect of this restriction on limiting the data points used is that it restricts the maximum allowable blur in an image. This limitation is purely due to practical reasons and not theoretical. This problem is easily solved in practice by taking one or two additional images as described later.

Next we compute what we call a stored table $T_s$ defined as

$$T_s(\rho, u) \; = \; T_{10}(\rho, u) \; - \; T_{40}(\rho, u) \tag{3}$$

The most important relation which facilitates the determination of object distance regardless of the appearance of the object is

$$T_c(\rho) \; = \; T_s(\rho, u_0) \tag{4}$$

where $u_0$ is the actual distance of the object. (Direct use of an equivalent and simpler relation $G_{10}(\rho)/G_{40}(\rho) = H_{10}(\rho, u_0)/H_{40}(\rho, u_0)$ resulted in very poor performance.) Therefore, mean-square error (MSE) is computed between $T_c$ and $T_s$ for different values of $u$. The value of $u$ for which the MSE is a minimum is taken to be the estimated distance of the object. However, if the minimum error occurs for a distance corresponding to higher than step 60 then $g_{10}$ is considered to be too much blurred for reliable results. This is because an object at best focus distance step 60 or beyond would be highly blurred when the lens is moved to step 10. Therefore, in this case, a third image is taken at step position 70, and the images taken at steps 40 and 70 are used in estimating distance in a similar manner as before.

The distance of the object is printed on the computer terminal, and the lens is moved to the corresponding step number to focus the object, thus accomplishing autofocusing.

In another variation of the implementation, mean of $T_c(\rho)$ is computed over $\rho$ and it is compared with the mean values of $T_s(\rho, u)$ (again computed over $\rho$), and the value of $u$ for which the two means are closest is taken as the distance of the object. This method also performed almost as good as the MSE method.

## 3 Experiments

Experiments were performed under the following conditions: Camera setting: focal length = 35 mm, F-number = 4, camera gain control +6dB, White balance = off, Gamma compensation = off, illumination about 200 lux.

Three different objects, a human face (Fig. 3), text (Fig. 4), and a cartoon (not shown), were used. Each object was placed at 16 different distances, and for each distance, our program was run about 5 times. In each case, four image frames were averaged to reduce noise. The estimated distance of the object, expressed in terms of the corresponding best focused lens position (in step number) is shown in Fig. 5. This Figure represents the result of 255 experiments (many points in the plot overlap exactly and therefore are not distinguishable). A straight line was fitted to this data using the least-squares approach. The resulting straight line along with two parallel lines on either side of it at a distance equal to the RMS error is shown in the figure. The RMS error is 5.6 steps out of 97 steps which corresponds to about 6% error. The region enclosed by the two parallel lines gives an idea about the uncertainty in the measurement of distance using our method. We see that about 90% of the points are within the two parallel lines corresponding to the RMS error (5.6 steps). In these cases the quality of the focused images were very good as judged visually by humans.

Numerous informal experiments were carried out on a wide variety of objects at many different distances. The results were comparable to those above. Additional experiments were conducted under different illumination conditions and different objects. The results were again good except under very poor illumination (50 lux) [?]. (Ambient illumination in an office is about 200 lux).

## 4 Work in Progress

Further improvements to DFD1F have been done regarding computation, memory, image overlap problem, and hardware implementation. A user friendly computer simulation system called Image Defocus Simulator (IDS) has been developed [?] which can synthesize defocused images sensed by a CCD camera as a function of camera parameters and scene parameters. We have found this to be an extremely useful research tool for testing DFD methods.

An entirely new DFD method based on a new spatial-domain convolution/deconvolution transform (S transform) has been developed, implemented, and successfully demonstrated on SPARCS [?]. The performance of this method is approximately comparable to DFD1F.

**Acknowledgments:** The support of this research by the National Science Foundation and the Olympus Optical Corporation is gratefully acknowledged.

## References

[1] E. Krotkov, "Focusing", *Inter. Jour. of Computer Vision*, 1, 223-237, 1987.

[2] A. P. Pentland, "A new sense for depth of field", *IEEE Trans. on Patt. Anal. and Mach. Intel.*, Vol. PAMI-9, No. 4, pp. 523–531.

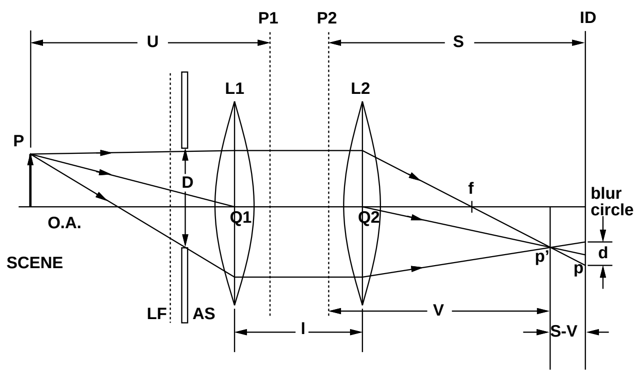
[3] M. Subbarao, "Parallel depth recovery by changing camera parameters", *Second International Conference on Computer Vision,* USA, pp. 149-155, Dec. 1988. (Patented.)

[4] M. Subbarao, U.S. patent application serial number 07/373,996, June 1989 (pending).

[5] M. Subbarao, N. Agarwal, and G. Surya, "Application of Spatial-Domain Convolution/Deconvolution Transform for Determining Distance from Image Defocus", Tech. Report No. 92.01.18, Computer Vision Laboratory, Dept. of Electrical Engg., State University of New York, Stony Brook, NY 11794-2350, 1992. (US Patent pending.)

[6] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing: A Practical Approach", Tech. Report No. 92.01.17, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992. (US Patent pending.)

[7] M. Subbarao, and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus", Tech. Report No. 92.01.16, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992.




P: Object Point
LF: Light Filter
AS: Aperture Stop
L1: First Lens
L2: Second Lens
O.A.: Optical Axis
P1: First Principal Plane
P2: Second Principal Plane
Q1: First Principal Point
Q2: Second Principal Point
ID: Image Detector
S,f,D: Camera Parameters
V: Dist of Image Focus
p': Focused Image
p: Blurred Image

Fig. 1 Camera Model and Camera Parameters

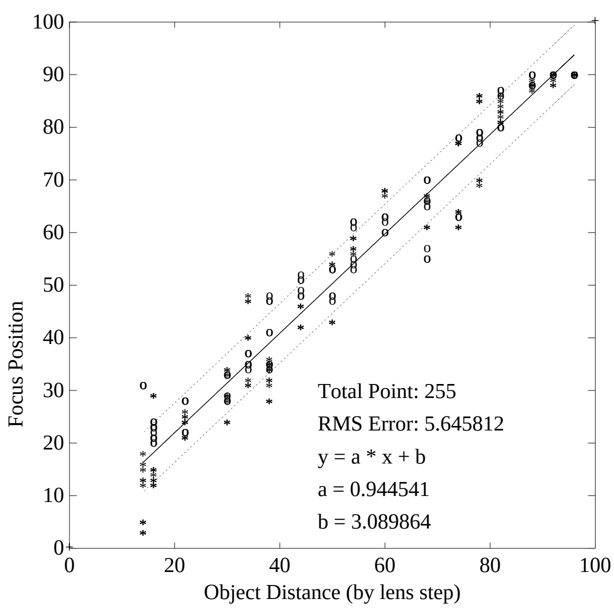
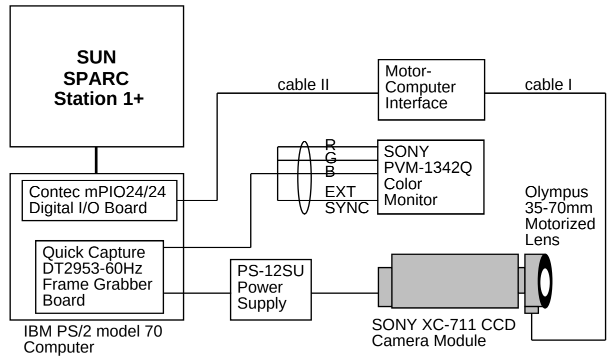



Fig. 5 Experimental Results




Stonybrook Passive Autofocusing and Ranging Camera System-SPARCS - is a prototype camera system developed at the Computer Vision Labatory for experimental research in robotic vision, State University of New York at Stony Brook

Fig. 2 SPARCS

Fig. 3 Test Image: Human Face


Fig. 4 Test Image: Text