

RANGE FROM FOCUS

Eric Krotkov and Jean-Paul Martin

GRASP Laboratory
Department of Computer and Information Science
University of Pennsylvania
Philadelphia PA 19104

ABSTRACT

One method for improving the vision system of a robot is to allow the robot to automatically focus its cameras and to use focus as a depth cue. In this paper, the problem of automatic focusing is reduced, by a theoretical analysis of defocus as a wave aberration which attenuates high spatial-frequencies, to the problem of measuring the high spatial-frequency content of an image. The problem of recovering range from focus is treated using a focus criterion function along with knowledge of the lens parameters to solve the lens law for range. The accuracy of the range measurements is approximately 10% of object distances between 1 and 2 m.

1. Introduction

The vision system of a robot is limited by the quality of the images it obtains. Defocused images inherently have less information in them than sharply focused images, so most vertebrate animals having eyes with a refractive lens use focus to improve the quality of their vision. The goal of our research is to give a robot this ability. To this end, we have designed and built an active camera system, a stereo pair of moveable CCD cameras, whose focal lengths and apertures are under computer control.

In addition to improving image quality, focus can be a depth cue, providing information about the absolute distances of objects. Again, many animals use this cue. Using focus to compute range is very attractive because there is no need to solve the correspondence problem, which is generally a slow, error-prone process.

This paper will first describe the problems in determining range from focus, then present practical methods for automatically focusing and computing range, and finally show some range computation results.

2. Problem statement and related research

This paper will discuss two specific problems. First, it will address the problem of automatic focusing for a standard CCD camera with a servo-controlled focus ring: Given the projection $P' = (u, v)$ onto the focal plane of an object point $P = (x, y, z)$ (z unknown), what focal length f produces the sharpest definition of P' ? This problem has received the attention of numerous camera manufacturers [2], Horn [3], Schlag et. al. [9], Jarvis [4], and Tenenbaum [10]. Several workable methods have been reported, but there is no consensus on them.

Given the focal length f bringing P' into sharpest focus, the second problem is to recover the z component of P , which is its range or absolute distance from the center of the lens. Some related work on *focal gradients* has been reported by Pentland [8]. His approach requires measuring the amount of blur at a particular image point. This is difficult because blur is a function both of the characteristics of the scene and those of the lens. A related approach is mentioned by Jarvis [5] although apparently never developed.

3. Automatic focusing

To present the method in detail this section discusses the theory of defocus, practical methods for computing the

quality of focus, and control strategies for automatic focusing.

3.1. Theory of defocus

Simply put, defocusing an incoherent optical system causes blurring of image points. However, this is too simplistic, because blurring is also caused by geometric and diffraction effects. To properly account for defocus the theory of wavefront aberrations must be employed.

An aberration is a departure of the exit pupil wavefront from its ideal spherical form (Figure 1). It is possible to treat defocus as a wave aberration, and to show exactly how defocusing attenuates and distorts different spatial-frequencies. The attenuation is described by the optical transfer function (OTF). The OTF is a complex function that describes how each spatial frequency component in an object is attenuated by the lens as it forms an image, by specifying the complex weighting factor applied by the system to the frequency component (f_x, f_y) relative to the weighting factor applied to the zero-frequency component. The OTF can be written (apart from a constant factor) as the auto-correlation function of the pupil function G :

$$OTF(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\xi + \xi', \eta + \eta') G^*(\xi', \eta') d\xi' d\eta' \quad (1)$$

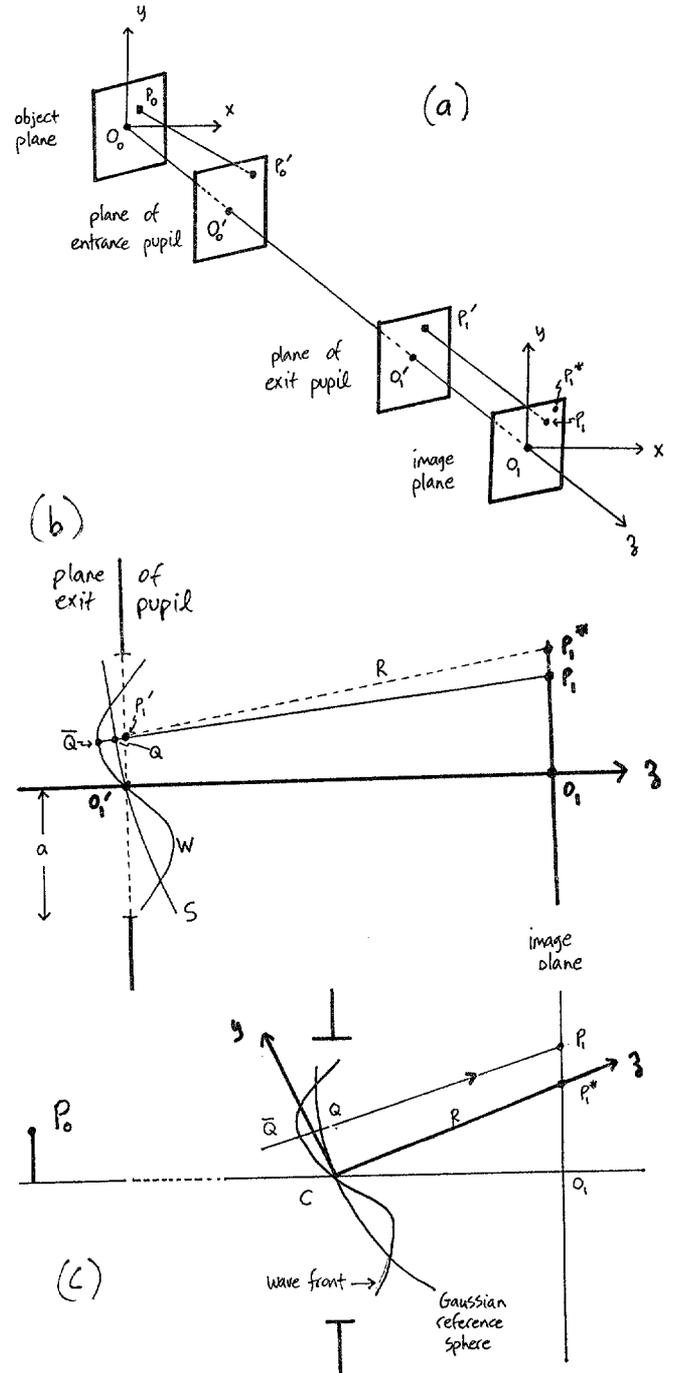
Aberrations will in general lower the contrast of each spatial-frequency component of image intensity; severe aberrations will significantly reduce the high-frequency portions of the OTF. Let us call the aberration function Φ , and assume that there are no aberrations of the lens. The effect of the wave aberration Φ (Figure 1b) on the light passing through a lens (even a thin lens) is to change the path length of each ray, changing the optical path length by

$$\Phi = P_0P - P_0P_1^* = P_0Q - P_0\bar{Q} = \bar{Q}Q \quad (2)$$

Referring to Figure 1c, the displacement of the detector plane by a small amount z in the positive z -direction is formally equivalent [1, p. 489] to the introduction of a wave aberration of amount

Figure 1. Wave aberration.

- The effect of the wave aberration $\Phi = \bar{Q}Q$ is to change the optical path length from object point P_0 to image point P_1^* .
- (a) The object plane, the image plane, and the pupil planes.
- (b) The wave aberration $\Phi = \bar{Q}Q$.
- (c) A different choice of reference system for (b).



$$\Phi(\xi, \eta) = \frac{z}{2R^2} (\xi^2 + \eta^2) \quad \left[\xi^2 + \eta^2 \leq a^2 \right] \quad (3)$$

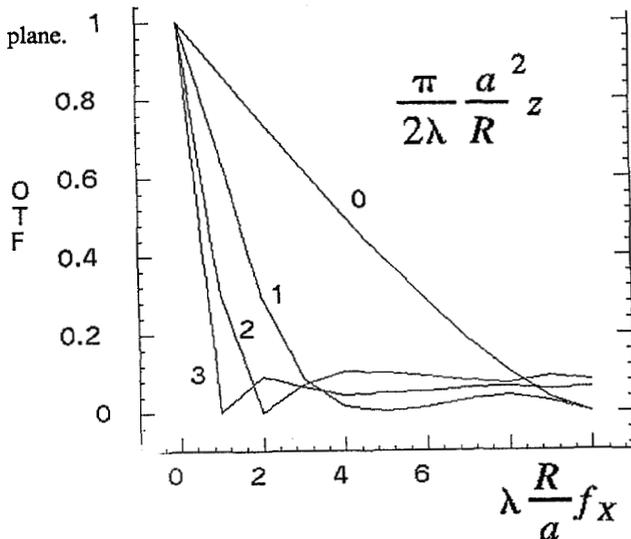
This produces a phase shift $k\Phi(\xi, \eta)$, which causes the pupil function G , which is zero outside the pupil and unity inside the pupil for an ideal lens, to become

$$G(\xi, \eta) = e^{ik\Phi(\xi, \eta)} \quad (4)$$

If this pupil function is substituted into Equation (1) for the OTF, the spatial-frequency response is given directly in terms of the amount of defocus (phase shift introduced). The resulting function can be expressed as a power series in Bessel and Struve functions and evaluated analytically, but it is simpler to evaluate numerically. It can be observed that a gradual attenuation of contrast and a number of contrast reversals are obtained for increasing spatial-frequency, as illustrated in Figure 2, which plots the frequency response of a defocused optical system free of other aberrations as a function of the defect z of focus. From the figure it is clear that $z > \frac{2\lambda}{\pi} \left[\frac{R}{a} \right]^2$ causes a rapid deterioration of the response of the system for higher frequencies.

Figure 2. Cross-section of the OTF of a defocused incoherent system.

A one-dimensional cross-section of the OTF. The number on each curve is the value of $\frac{\pi}{2\lambda} \left(\frac{a}{R} \right)^2 z$, where z is the distance between the detector plane and the Gaussian focal plane.



3.2. Computing Focus Quality

In the previous section we saw that defocus results in attenuation of high spatial-frequencies. It follows that the quality of focus (conversely, the degree of defocus) is proportional to the amount of high-frequency energy present in the spatial power spectrum. We have experimented with a wide variety of focus criterion functions which either estimate or measure directly this high spatial-frequency energy.

Let I represent the intensity function, $I(x, y)$ represent a grey-level, f_x denote a spatial-frequency, F denote a Fourier transform, and $P(I(x, y))$ represent the probability of the occurrence of a grey-level. Some of the focus criterion functions we have examined [7] include:

- the sum of first differences on each scan line

$$\sum_x \sum_y |I(x, y) - I(x, y-1)| \quad (5)$$

- the entropy E of the grey-level histogram

$$E = - \sum_{I(x, y)} P(I(x, y)) \ln(P(I(x, y))) \quad (6)$$

- the sum of the gradient magnitudes above a threshold

$$\sum_x \sum_y \nabla I(x, y) \quad \text{for } \nabla I > T \quad (7)$$

- the grey-level variance

$$\sum_x \sum_y (I(x, y) - \mu)^2 \quad (8)$$

- the high-frequency energy in the power spectrum

$$F(f_x, f_y) \quad \text{for } \sqrt{f_x^2 + f_y^2} > T \quad (9)$$

- the slope of a least-squares fit to a histogram of local

$$H(|I(x, y) - I(x', y')|) \quad (10)$$

Some important criterion function implementation details: each method is defined on a given search window (usually 10x10); since changing the focal length also changes the lens magnification and consequently the image coordinates of a given feature, "window tracking" must be employed to guarantee that the point to be focused upon stays within the search window; finally, temporal averaging is used to filter the noisy intensity signals.

To compare the methods, a sequence of images of a

static scene is taken with different focus parameters, varying from too close to just right to too far, and the behavior of the focus criterion functions recorded. From a number of experiments with different window sizes and a wide variety of scenes, the function which performs best empirically-- in speed, robustness in the presence of temporal noise corrupting the intensity signals, repeatability, and accuracy-- is the first (Equation 5), which in practice is unimodal, varies monotonically with focal length on either side of the mode, and is speedily computed.

3.3. Search strategy

This section describes a control strategy for automatically focusing on a point in a static scene. A hill-climbing (gradient ascent) technique (described independently by Sanderson [9] and Jarvis [4]) is to compute the criterion function, move the lens, recompute the criterion function, and look at the sign of the difference of the criterion. If the sign is negative, the lens was moved in the wrong direction. This technique is in principle adequate, but because each evaluation of the criterion function requires taking a picture, an expensive operation, it is important to minimize the number of evaluations required.

Since the focus criterion function is unimodal, we may employ a numerical method for finding the extrema in a function of one variable (focal length). The *Fibonacci search technique* is the most efficient of any restricted search [6]. The search method (described formally in Figure 3) is to successively narrow the search interval until its size is a given fraction of the initial search region. For each interval, the criterion function is evaluated at two points which are determined by the Fibonacci sequence. The next interval is determined by which point has a higher criterion value. Since the focal length is represented as 10 bits, the number of different focal lengths is 2^{10} . For an accuracy of one unit in 2^{10} , 16 iterations of the algorithms are required to select the extrema; this is considerably less than the number of iterations required by gradient ascent.

FIGURE 3. Fibonacci search technique.

The Fibonacci sequence may be defined by the boundary conditions

$$F_0 = F_1 = 1$$

and the recurrence relation

$$F_N = F_{N-1} + F_{N-2}$$

The size of the initial search interval L_0 must be known in advance; in our case $L_0 = 2^{16}$. The number of iterations required to find the mode of the unimodal and monotonic function (here called *criterion*) is the least integer N such that $F_N \geq L_0$. In our case $N = 23$. The search algorithm is illustrated below:

```

 $b_0 - a_0 = L_0 \quad k = 1$ 
while ( $k \leq N$ ) do
 $L_k = L_{k-1} \frac{F_{N-(k-1)}}{F_{N-(k-2)}}$ 
 $l_k = L_k \frac{F_{N-(k+1)}}{F_{N-(k-1)}}$ 
 $x_1 = a_k + l_k$ 
 $y_1 = \text{criterion}(x_1)$ 
 $x_2 = b_k - l_k$ 
 $y_2 = \text{criterion}(x_2)$ 
if ( $y_1 > y_2$ ) then  $a_{k+1} = a_k \quad b_{k+1} = x_1$ 
else  $a_{k+1} = x_2 \quad b_{k+1} = b_k$ 
 $k = k + 1$ 
endwhile

```

4. Computing Range

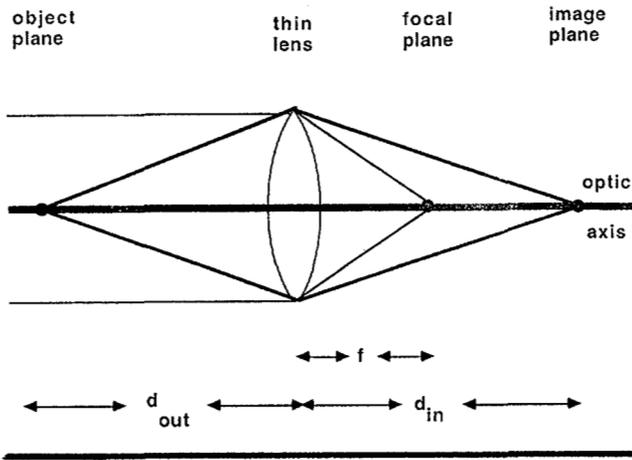
This section describes how the distance to a visible object point can be computed from the lens parameters and a sequence of images of a static scene. It also presents some obvious limitations of the method.

4.1. Method

For the single thin lens illustrated in Figure 4, the lens law holds:

$$\frac{1}{d_{out}} + \frac{1}{d_{in}} - \frac{1}{f} = \epsilon \quad (11)$$

FIGURE 4. Thin lens geometry.



The lens law also holds for thick lenses, provided that the distances are measured not from the center of the lens but from its principal points. When the lens is properly focused, $\epsilon = 0$, and the lens law may be solved as

$$d_{out} = \frac{d_{in}f}{d_{in}-f} \quad (12)$$

giving a closed-form expression for d_{out} , which depends only on f and d_{in} . To find the distance d_{out} to the object point (x,y,z) projecting to image point (u,v) :

1. maximize the sharpness of focus at (u,v) by changing f so that the detector plane coincides with the image plane, i.e., $d_{in} = d_{det}$
2. read f
3. solve Equation (12)

4.2. Limitations

There are a number of limitations of this method. Perhaps the most significant is that it applies to only one point at a time. Thus with a single camera, range from focus can not be computed in parallel. Other limitations include the accuracy with which d_{in} and f can be physically measured, the performance of the criterion function with noisy signals, the validity of the thin lens model describing a compound lens, and the nonlinearity of Equation (12).

Spatial quantization is another limitation, one which can not be circumvented by more precise measurements or

better equipment or algorithms. Because the photoreceptors have finite area, an object point may lie at a number of different distances and still be imaged sharply on the same receptor. The distance in object space between the nearest plane and the farthest plane at which satisfactory definition is obtained (the *depth of field*) is given [7] by

$$DOF = \frac{2Dafc(D-f)}{a^2f^2 - c^2(D-f)^2} \quad (13)$$

where f is focal length, D is object distance, a is aperture diameter, and c is the smallest dimension of the photoreceptor. The fundamental impact of spatial quantization, then, is that the range d_{out} computed is accurate to $\pm DOF$ mm.

5. Results

First the focus motors are calibrated with a point source target so that the relationship between motor position and focal length is known. Next an object point (on an ophthalmologist eye chart) at a known distance is indicated, and Fibonacci search for the optimal focal length is undertaken. Then the optimal focal length is used to compute range. The results are presented in the Tables.

The results were gathered from only a few experiments and should be interpreted as illustrative; their repeatability is probably quite poor. Nevertheless, the results do suggest that error does not vary significantly with object distance. For the objects used the uncertainty caused by the depth of field of the cameras is very small. The range computation errors are primarily due to the performance of the focus criterion function. The magnitude of the errors in range computed from focus is on the order of 10% of the object distance for distances between 1 and 2 m.

6. Discussion

This paper has described a method for improving the vision system of a robot, namely giving the robot the ability to actively focus its cameras and to use focus as a depth cue. The problem of automatically measuring the quality of focus is reduced to the problem of measuring the high spatial-

frequency content of an image, by an analysis of defocus as a wave aberration which attenuates high spatial-frequencies. The problem of recovering range from focus is treated by using a focus criterion function along with knowledge of the lens parameters to solve the lens law for range. The results are unexceptional, but clearly demonstrate that computing range from focus is possible and deserves more attention in order to be done more accurately.

Research continues in calibrating the focus motors to more precise focal lengths (using point source targets), and making the focus criterion function more robust in the presence of noise on the intensity signals (by modeling the noise and filtering it intelligently). In the future, the effects of image magnification caused by changing the focal length and zooming must be incorporated into the range computation.

Table 1. Focus data.

Point	$f_{ideal}(mm)$	$f_{computed}(mm)$
A	84.8	102.2
B	79.8	82.2
C	64.2	64.8
D	57.4	56.3
E	48.6	48.9
F	33.1	30.1

f_{ideal} is the calibrated focal length. $f_{computed}$ is the focal length determined by the Fibonacci search technique.

Table 2. Range data.

Point	$D(mm)$	$f_{ideal}(mm)$	$f_{computed}(mm)$
A	762	571	2.5
B	1066	1103	5.4
C	1371	1449	11.5
D	1676	1891	19.5
E	1981	2112	32.4
F	2286	2510	64.0

D is the measured distance in millimeters. D' is the computed distance. The depth of field DOF is computed using Equation 13 with $c = 12 \mu m$ and $a = 58.33 mm$.

References

1. Born, M. and E. Wolf, *Principles of Optics, Fifth Edition*, Pergamon Press, New York, 1975.
2. Goldberg, N., "Inside Autofocus: How the Magic Works," *Popular Photography*, pp. 77-83, February, 1982.
3. Horn, B. K. P., "Focusing," *MIT A.I. Memo No. 160*, May, 1968.
4. Jarvis, R. A., "Focus optimisation criteria for computer image processing," *Microscope*, vol. 24, no. 2, pp. 163-180, 1976.
5. Jarvis, R. A., "A Perspective on Range Finding Techniques for Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 2, pp. 122-139, March, 1983.
6. Kiefer, J., "Sequential Minimax Search for a Maximum," *Proc. Am. Math. Soc.*, no. 4, pp. 502-506, 1953.
7. Krotkov, E. P., "Focusing," *University of Pennsylvania T.R.*, 1986.
8. Pentland, A., "A New Sense for Depth of Field," *Proceedings IJCAI 85*, pp. 988-994, Los Angeles, August, 1985.
9. Schlag, J. F., A. C. Sanderson, C. P. Neumann, and F. C. Wimberly, "Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control," *CMU-RI-TR-83-14*, August 1983.
10. Tenenbaum, J. M., "Accommodation in Computer Vision," *Stanford University Ph.D. Thesis*, November, 1970.

Table 3. Range errors.

Point	f_{error}	D_{error}	DOF_{error}
A	20.5	25.1	0.33
B	3.0	-3.5	0.51
C	0.9	-5.6	0.84
D	1.9	-12.8	1.16
E	0.6	-11.7	1.64
F	9.1	-9.8	2.80

f_{error} is the quantity $100 \frac{f_{ideal} - f_{computed}}{f_{ideal}}$. D_{error} is the quantity $100 \frac{D - D'}{D}$. DOF_{error} is the quantity $100 \frac{DOF}{D}$.