

TR-88-026

Range Based on Focusing

中研院資訊所圖書室



3 0330 03 000099 1

0099

# Range Based on Focusing

Jun S. Huang

Computer Vision Laboratory

Institute of Information Science

Academia Sinica

Taipei, Taiwan, Rep. of China

October, 1988

Abstract:

In this paper we address the problem of calculating the range of an object point based on its focused distance by a fixed lens. The details of the lens properties are discussed and a thin spherical lens with long focal lens and large aperture is proposed for the use of range calculation. The error analysis of the Gaussian lens formular  $1/p + 1/q = 1/f$  is discussed and analyzed, where  $p$  is the object distance, and  $q$  is the image distance, and  $f$  is the focal length. It is found that the magnitude of the error of inferred range based on a fixed focal lens is proportionally increasing with the object distance subjected to a multiplied random error, which is the precision of focusing. The depth of focus (DOF) is also discussed for a fixed resolution of  $512 * 512$  CID chip, and it is found that for a moderate range the DOF is tolerable for proper aperture and focal length.

## 1. INTRODUCTION

Machine stereo vision is critical to the success of 3D vision and it has wide applications including autonomous vehicles, 3D object robot grasping and machine control, etc. However, the most difficult unsolved problem in stereo vision is the correspondence problem, where a point in the left image must be matched to a right image's point corresponding to the same object point in 3D space. In [1] Huang has proposed a dynamic stereo camera model to solve this problem, where the two stereo cameras are controlled to focus at the same object point and a complicated formular for computing the searching path on the right image is developed. In that formular the classical imaging equation (Gaussian lens law)  $1/p + 1/q = 1/f$  is needed where  $p$  is the object distance,  $q$  is the image distance and  $f$  is the focal length. The validity of this equation can be checked on the general optics textbook [2]. However the applicability of this equation in real cases is questionable, and here I will give a detailed discussion of it.

The main usage of this imaging equation here is to infer the object distance  $p$  once the image distance  $q$  is computed subjected to some error. That is, once the object point is properly focused the image distance  $q$  can be measured by some specific methods and then the object distance  $p$  is computed by the imaging equation, but  $p$  may suffer large errors depends on how accurate of the fabrication of the lens, how long of the focal length  $f$ , how large of the true  $p$  and finally how accurate of measuring  $q$ .

## 2. FOCUSING AND THE VALIDITY OF THE IMAGING EQUATION

Auto focusing has been studied by Krotkov [3] in detail, and he concludes that the gradient method is the best method among eight known methods for general imaging systems. However he did not try the image with smooth surfaces. Here we try this experiment in the hope that the lens can produce the microtexture image of a surface. The painted wall looks smooth from the human eyes because human eyes have the capability of filtering the high frequency. We take thirteen images of the same point on the smooth wall where the positions of these images are very near the exact focusing point, which is judged by human eyes. The sum of gradients of each point within a  $10 * 10$  window is computed for each image and the result is plotted on Fig. 1. The camera used here is a COHU-C5000 T.V. camera. The whole experiment is carried three times to see whether there is any temporal variation. Each image is the average of sixteen images taken continuously at the same time. This is designed to reduce the lighting variation and the sensor noises. Fixing the focusing position we repeat to take eight images of the same wall and compute the sum of gradients within  $10 * 10$  window of each image. We find the range of sensor variation is  $34944 - 34786 = 158$  and the estimated standard deviation is 44.5. Because of this high variation of the image sensor, after three experiments of computing the sum of gradients we conclude that there is temporal variation and there is no evidence of indicating the focusing of the microtexture of the wall.

Sum Of Gradients

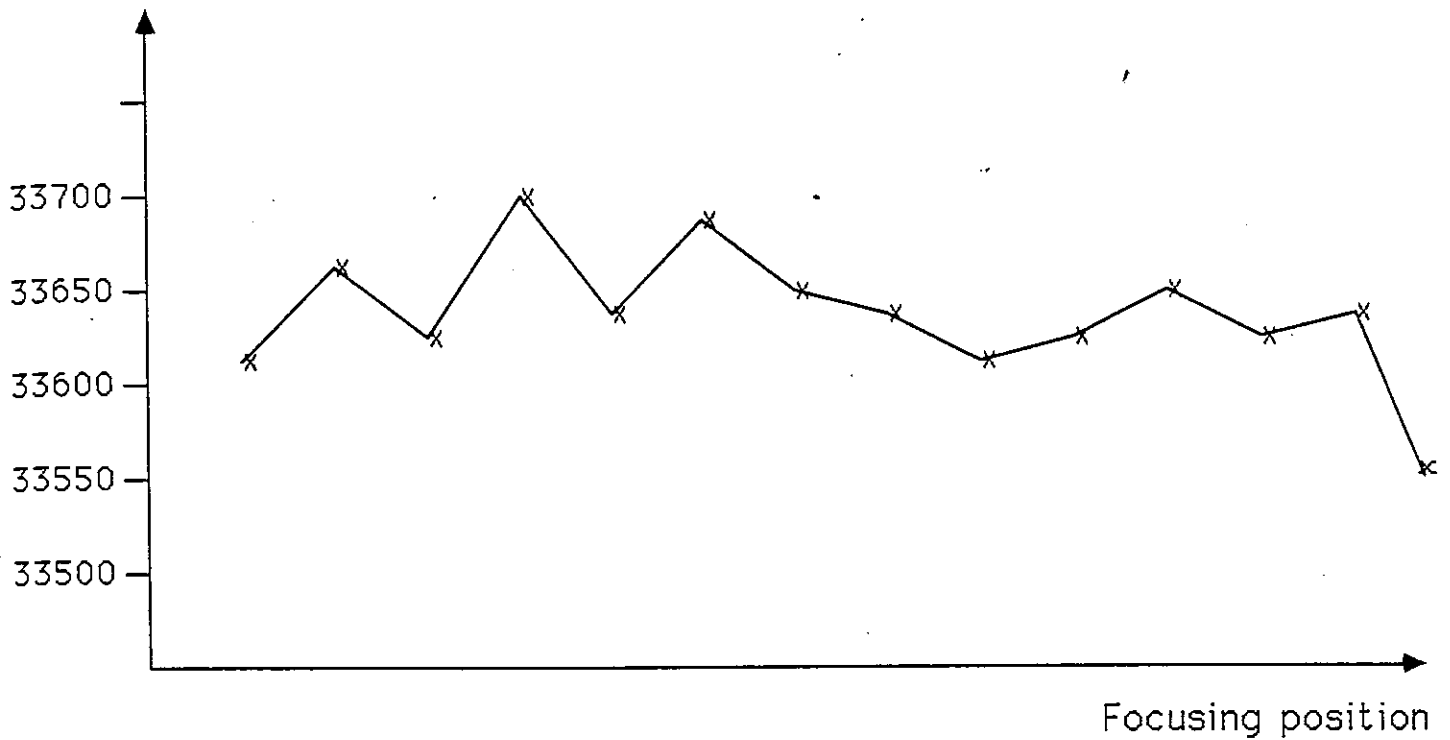


Fig. 1 Change of the sum of gradients within a 10 \* 10 window

From Huygen's principle (or eikonal equation, or Fermat's principle) we can derive the three laws of geometrical optics, and from these laws we can derive lens maker's equation for the thin lens (or the imaging equation) :

$$\frac{1}{p} + \frac{1}{q} = \left( \frac{n'}{n} - 1 \right) \left( \frac{1}{R_1} + \frac{1}{R_2} \right) = \frac{1}{f}$$

where  $p$  is the object distance (The object point must lie on the optical axis);  $q$  is the image distance;  $n'$  is the refractive index of the lens material;  $n$  is the refractive index of the surrounding medium;  $R_1$  and  $R_2$  are the radiuses of the refracting surfaces.

The validity of this equation clearly depends on the following factors :

1. The lens material has uniform density so that  $n'$  is constant every where in the lens.

2. The refracting surfaces are perfect spherical so that  $R_1$  and  $R_2$  have real geometrical meanings.
3. Different wavelengths have different refractive indexes and thus produce different  $f$ 's.

In practical situations factor (1) can be approximated in a precise way. However factor (2) is pretty hard to be approximated in precise way. Modern optical technology uses the techniques of holography to fabricate spherical lenses with very high accuracy and very large aperture. Such lenses are used in IC industry and are extremely expensive. Factor (3) can be tackled by using RGB three image sensors to separate the effects of different wavelengths.

The commercial camera is composed of multiple lenses, which are conceptually equivalent to a single thick lens. The idea of using multiple special designed lenses is to minimize the lens aberrations, which includes spherical aberration, chromatic aberration, astigmatism, and coma, etc. The performance of an ideal thick lens is similar to the thin lens. The imaging equation is still valid except that  $p$  is the distance from the object point lying on the optical axis to  $P_1$ , the primary principle points inside the lens, and  $q$  is the distance from the image point to  $P_2$ , the secondary principle points inside the lens (see Fig. 2).

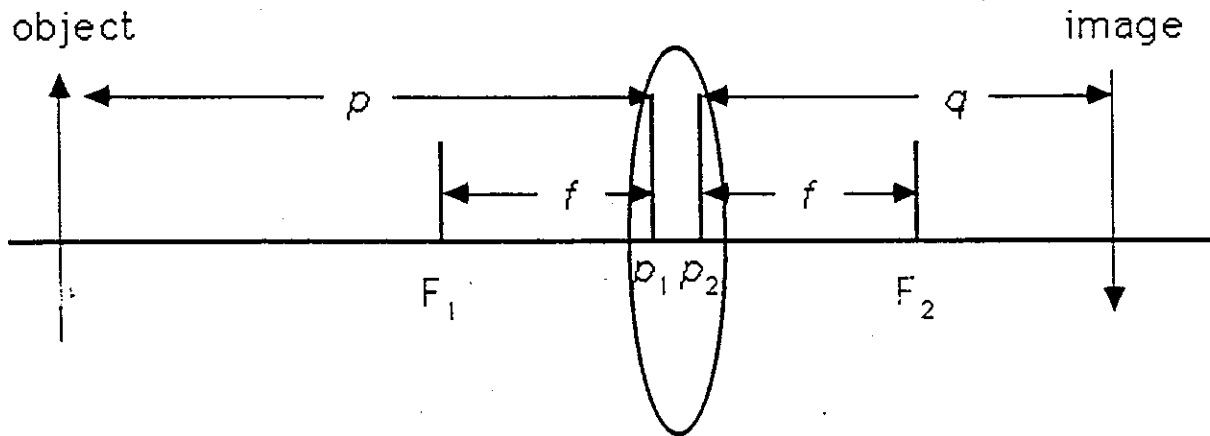


Fig. 2. A thick lens, where  $P_1$  is primary principle point and  $P_2$  is secondary principle point.

The spatial resolution of the image sensor array is a very important factor for the image quality. The higher resolution gets the better or finer quality. For a fixed resolution the accuracy of calculating the range is also limited. Because the sensor elements have finite area, an object point may lie at a number of different distances and still be imaged onto the same sensor element. The distance between the nearest and the farthest object points at which satisfactory definition is obtained is called the depth of field (DOF). The DOF has been discussed by Krotkov [2] but the formular in [2] is not correct. Here we give a correct formular of DOF.

From Fig. 3 we can see that rays from  $x$  converge to the detector of size  $c^2$  at  $y$ ,



while rays from a nearer point  $x_2$  would converge to a point  $y_2$  behind the detector, and rays from a farther point  $x_1$  would converge to  $y_1$  in front of the detector. For a circular lens, the distribution of light on the detector is approximately circular, the circle formed by the intersection of the rays from an object point with the detector is the circle of confusion (or blur circle). In Fig. 3  $x_1$  and  $x_2$  are at the limits of the depth of field, i.e., they produce circles of confusion with diameters no greater than  $c$ . From Fig. 3, it is easily seen that

$$\frac{c}{a} = \frac{y - y_1}{y} = \frac{y_2 - y}{y}$$

Using the imaging equation to transform each  $y$  in image space into a corresponding  $x$  in object space and solving for object distances yields

$$x_1 = \frac{x f (a - c)}{a f - c x}, \quad x_2 = \frac{x f (a + c)}{a f + c x}$$

The far distance is  $D_1 = x_1 - x$  and the near distance is  $D_2 = x - x_2$  and the depth of field is

$$\text{DOF} = D_1 + D_2 = \frac{2 x f a c (x - f)}{(a f)^2 - (c x)^2}$$

Thus the range computation based on a fixed sensor is related to DOF, which is proportional to  $c$ ,  $x^2$  and inversely proportional to  $\frac{(a f)^2 - (c x)^2}{a f}$ , or approximately  $a f$  if  $c x$  is small. That is, if the aperture  $a$  is twiced or the focal lens  $f$  is doubled then the DOF for a fixed  $x$  is halfed. If  $c$  is reduced to half, then the DOF is also reduced to half.

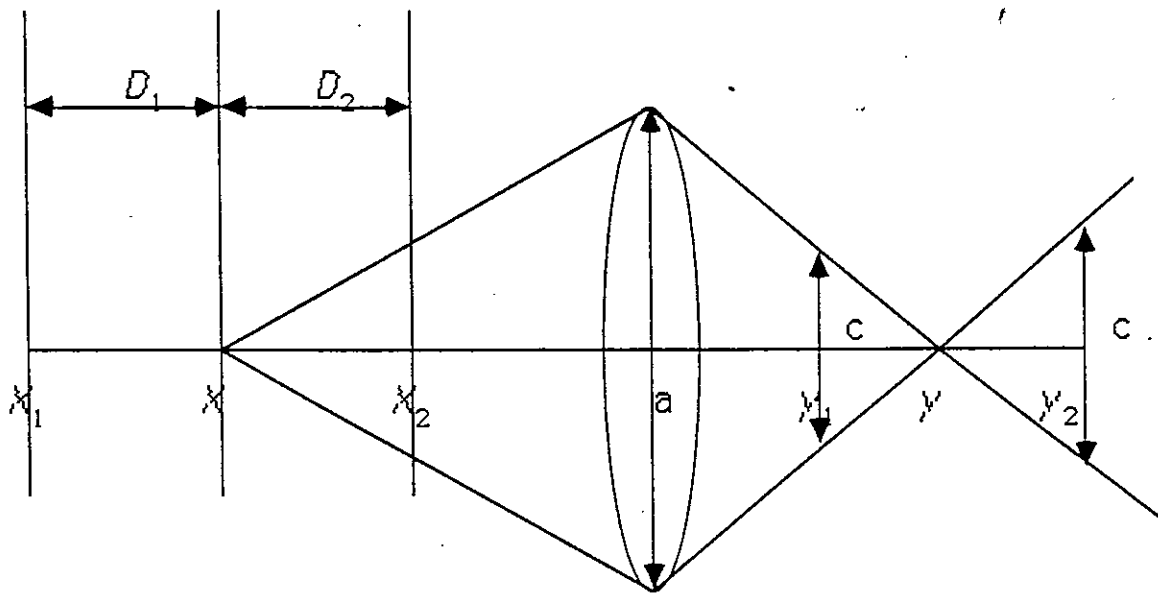


Fig. 3. The depth of field is  $D_1 + D_2$

For an excellent image sensor CID TN2250 available in industry, which has 512 \* 512 square sensing elements and is about 10 mm wide. In this case  $c = 0.0195$  mm. The following table shows a number of DOF's for a number of different  $x$ 's and  $f$ 's.

	$f = 500$	$f = 200$
$X$	$a = 80$	$a = 40$
50000	2414.56	12321.77
10000	92.60	478.03
5000	21.94	117.02
4000	13.65	74.11
3000	9.11	41.00
2000	2.95	16.75

In general, for an autonomous vehicle the general viewing distance is around 50000 mm to 2000 mm. To get high accuracy of range calculation the DOF's must be very small. For  $f = 1500$  mm; this is long enough, and the aperture is increased to  $a = 480$  mm, a large aperture, then the DOF is 99.7 mm. But in order to achieve the DOF's values less than 10 mm  $c$  must be reduced to 10 times smaller, which means we have to fabricate an IC sensor such that there are 512 elements within 1 mm. We know that current technology can fabricate an IC to a degree of submicron ( $< 10^{-3}$  mm) and so this is possible to do. That means, we can fabricate an IC sensor such that the central part has very high density ( $c = .00195$ ) and the outer part has low density ( $c = .0195$ ), and this sensor is suitable for our range computation. The machine designed in this fashion must be extremely expensive.

### 3. ERROR ANALYSIS OF IMAGING EQUATION

Rewriting the imaging equation  $1/p + 1/q = 1/f$  into the forms :

$$p = \frac{f q}{q - f} \quad \text{and} \quad q = \frac{p f}{p - f},$$

it is easy to see that for a given fixed  $p$ , the larger  $f$  (hence the larger  $q$ ) will give larger  $q - f$  and hence better avoidance of the dividing error. In practical applications we would like to have  $p$  with range  $1000 \text{ mm} \leq p \leq 10000 \text{ mm}$ . Let the measured image distance  $\hat{q}$  by autofocusing methods (e.g. gradient method) be  $q + \xi$ , where  $\xi$  is the measuring random error. Now let  $f = \alpha p$ ,  $0 < \alpha < 0.1$  we want to see how  $\alpha$  influences on the error of the

estimated  $p$ . From the above equations :  $q = p \frac{\alpha}{1 - \alpha}$ ,

and hence :

$$\hat{p} = \frac{\alpha p \left[ \left( \frac{\alpha}{1 - \alpha} \right) p + \xi \right]}{\frac{\alpha}{1 - \alpha} p - \alpha p + \xi},$$

$$= p \left[ \frac{\alpha^2 p + \alpha(1 - \alpha)\xi}{\alpha^2 p + (1 - \alpha)\xi} \right],$$

$$= p \left[ 1 + \frac{-(1 - \alpha)^2 \xi}{\alpha^2 p + (1 - \alpha)\xi} \right].$$

Hence the error of estimated  $p$  is  $\Delta p$  :

$$\Delta p = \hat{p} - p = -p\xi \frac{(1 - \alpha)^2}{\alpha^2 p + (1 - \alpha)\xi}.$$

From this equation, if  $\alpha \rightarrow 0$  then  $\Delta p \rightarrow -p$ . And  $|\Delta p|$  decreases as  $\alpha$  increases. To see this we differentiate  $|\Delta p / p\xi|$  with respect to  $\alpha$  and we see that after differentiation the numerator becomes :

$$\begin{aligned} \text{numerator} &= -[(2 - 2\alpha)(\alpha^2 p + (1 - \alpha)\xi) - (2\alpha - \alpha^2 - 1)(2\alpha p - \xi)] \\ &= -[(-2\alpha^2 + 2\alpha)p + (1 - \alpha)^2 \xi] \\ &= -[2\alpha p + (1 - \alpha)\xi](1 - \alpha) < 0. \end{aligned}$$

This means the larger  $\alpha$  value (or larger  $f$ ) the smaller error  $|\Delta p|$ . However in practices the larger value of  $f$  the larger magnitude of the measuring error  $\xi$  since it is very difficult

to locate the precise focusing position. Hence a compensation should be made on the choice of  $\alpha$ .

The following table gives the error  $\Delta p$  in terms of different values of  $p$ ,  $\xi$  with fixed focal length  $f = 200$  mm.

$p$	$\xi$		$\pm 1$	$\pm .5$	$\pm .1$
	$\alpha$				
10000	.02		1928.5	1069.5	234.4
			-3180.1	-2736.2	-246.1
5000	.04		482.1	254.7	53.4
			-613.6	-287.2	-54.7
4000	.05		329.7	172.2	35.8
			-398.9	-189.6	-36.4
3000	.067		166.2	85.6	17.6
			-188.5	-91.1	-17.8
2000	.1		77.51	39.6	8.06
			-84.81	-41.4	-8.14

From this table we see that with  $p = 10000$  even  $\xi$  is very small ( $\pm .1$  mm) the magnitude of error of  $p$  is still large. This means  $f = 200$  is impractical for  $p = 10000$ . However for  $p \leq 5000$  and  $\xi = \pm .1$  the magnitude of error of  $p$  is somehow not too large and tolerable in some engineering applications. Krotkov [3, p.57] has achieved the absolute error of around  $60 \pm 20$  with  $2000 \leq p \leq 3000$  by using gradient method on focusing. Comparing with this table his system should be able to measure the focusing error within  $\pm .5$  mm. However in the Computer Vision Laboratory of Academia Sinica, the T.V. camera is not good enough

and noise sensitive; the focusing is manual so the experiment is not tried. For the purpose of stereo correspondence, the error of  $p$  must be very small, say a few millimeters. The only available device that can measure the object distance within such accuracy so far on the market is designed by laser light phase shift calculation, which needs to place a reflection mirror on the object point, and this is not suitable for our application. But how can we increase the precision of the estimate of  $p$  to within, say, 4 mm? There are several ways to go and one way is to use a large number of cameras well arranged in order. For example, if  $p = 2000$ ,  $\xi = \pm .1$ , then the range of error of  $p$  is  $(-8.14, 8.06)$  which has roughly a standard deviation of 4. If we require that the estimate  $\hat{p}$  have a standard deviation 1 mm, then the number of cameras required to set up is  $n = 4^2 = 16$ . Another way is to use very ideal lens and very high resolution sensor with very low noises so that the microtexture of the object point can be imaged and thus  $\xi$  can be measured with .0001 mm. From this error analysis we can conclude that the work of Subbarao [4] is almost impossible to be realized, where he is trying to use the change of power spectrum to get the range data of each imaging point. Thus the use of laser range finder (by triangulation or time of flight) is still necessary in many practical applications although the laser range finder is not flexible as the stereo imaging system.

#### 4. CONCLUSION

From our discussions above we can make a conclusion that the imaging equation is generally subjected to some error and the error of range calculation by autofocusing is tolerable when the object distance is not too large compared with the focal length. Also to get small value of depth of focusing a large aperture and a long focal length are needed. This result indicates the use of laser range finder is necessary in many practical

applications.

#### REFERENCES

1. Jun S. Huang, *Machine stereo vision*, TR-86-008 Technical Report of IIS, Academia Sinica, December 1986.
2. M. Klein, *Optics*, John Wiley & Sons, Inc., New York, 1970.
3. E. P. Krotkov, *Exploratory visual sensing for determining spatial layout with an agile stereo camera system*, MS-CIS-87-29, University of Pennsylvania, 1987.
4. M. Subbarao, *Direct recovery of depth-map I : differential methods*, Proceedings of the IEEE Workshop on Computer Vision, 1987, p.58-65.