Range from focus-error

M. W. Siegel and M. L. Leary

Measurement and Control Laboratory
The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh PA 15213

ABSTRACT

We derive theoretically and demonstrate experimentally an approach to range-from-focus with an important improvement over all previous methods. Previous methods rely on subjective measures of sharpness to focus a selected locale of the image. Our method uses measured physical features of the optical signal to generate an objective focus-error distance map. To compute range-from-focus-error distance it is not necessary to focus any part of the image: range is calculated directly from the lens formula by substituting the difference between the lens-to-sensor distance and the focus-error distance for the usual lens-to-image distance. Our method senses focus-error distance in parallel for all locales of the image, thus providing a complete range image. The method is based on our recognition that when an image sensor is driven in longitudinal oscillation ("dithered") the Fourier amplitude of the first harmonic component of the signal is proportional to the first power of the ratio of dither amplitude to focus-error distance, whereas the Fourier amplitude of the second harmonic component is proportional to the square of this ratio. The ratio of the first harmonic $\sin \omega t$ amplitude A_ω to the second harmonic $\cos 2\omega t$ amplitude $B_{2\omega}$ is thus a constant (−4) multiple of the ratio of the focus-error distance to the dither amplitude. The focus-error distance measurement via the ratio of the first-to-second harmonic amplitudes is extremely robust in the sense that the scene's gray level structure, the spatial and temporal structure of the illumination, and technical noise sources (most of which affect the Fourier amplitudes multiplicatively) all appear identically in both amplitudes, thus cancelling in the ratio. Extracting the two Fourier amplitudes and taking their ratio could be accomplished, pixel-by-pixel, by some ambitious but not outrageous analog computing circuitry that we describe. We derive the method for a point scene model, and we demonstrate the method with apparatus that instantiates this model.

2. BACKGROUND

Conventional imaging captures a two dimensional projection of the scene; the third dimension, range, must be reconstructed either by a knowledge based process involving understanding of scene content, or by a physical process commonly involving additional cameras and triangulation,[1] a separate ranging device,[2] or active focus by optimizing some sharpness criteria and calculating range-from-focus.[3] Despite impressive recent progress in brute-force correlation to measure parallax disparity between binocular stereoscopic image pairs, these techniques suffer from a variety of faults: deficiencies in range accuracy, density of range data, aliasing difficulties with the correlation methods, etc, are common.[4] While the problems differ for each method they do have one common source: the range is found off-line, not captured in intimate connection with recording the image. We describe a new on-line approach to range-from-focus that overcomes this and several related difficulties with current range-from-focus techniques.

Image focusing[3] is conventionally regarded as a spatial-domain activity: the focus-controlling parameter (lens-to-sensor distance in a camera, focal length in the eye) is presumed to be adjusted with the goal of maximizing the amplitudes of the high spatial frequency image components. The amplitudes of the appropriate spatial frequencies are derived from pixel-to-pixel signal differences. The focusing information available from these differences is in reality weak and noise prone. Thus most practical focusing, *e.g.*, in film and video photography, is done indirectly, without reference to the image, by an open-loop method using a rangefinder (*e.g.*, a parallax based split-image method) coupled to the lens-to-sensor distance by a calibration scale. In humans, depth perception is known to be derived from the fusion of focus and binocular parallax cues. However the focusing cue is easily discounted: most people have no trouble understanding 3D-stereoscopic photos even though focus is confined to the screen-plane, while the conflicting convergence cues are controlled by the offset between corresponding points in the left and right eyes's images.[5]

In principle the ideal way to capture the focus information and thus the range directly would be to use a three dimensional imaging sensor. Imagine this hypothetical three dimensional sensor as a stack of planar imaging sensors with each layer thin and transparent. Local lens-to-image distance, and thus local range-from-focus, could be found by an algorithm that located the plane in which the value of some transverse focusing criteria is optimized. Although attempts at building processing layers behind sensing layers are underway,[6] we are unaware of any efforts to build three dimensional sensing lattices.

But it is possible to simulate a three dimensional sensor's key features by moving a single conventional sensor back and forth along the lens axis. This motion eliminates the transparency and thinness requirements, but the problem of finding where each scene locale is in best focus remains. Explicitly optimizing a focusing criteria would entail laboriously focusing each locale in the image. Our contribution to the range-from-focus domain is a hardware signal processing method that measures the focus-error distance directly. From focus-error distance and lens-to-sensor distance we calculate lens-to-image-distance, then apply the lens equation

$$\frac{1}{z}+\frac{1}{z'}=\frac{1}{f} \tag{1}$$

(where z and z' are object and image distances and f is focal length) to find range. We can thus measure range to all scene locales simultaneously, pixel-by-pixel, in parallel.

Our method of measuring the focus-error distance without explicitly focusing is the main innovation of this work.[7] We oscillate ("dither") the sensor array longitudinally and measure two key Fourier component amplitudes of the resulting oscillating signals.[8] We show that the focus-error distance is captured in the ratio of the two key Fourier amplitudes. The individual Fourier amplitudes depend on the scene, illumination, and noise, but the ratio is insensitive to both gray level and to many kinds of noise, making the method extremely robust.

3. EXPERIMENT

We will later present the theory in a simplified intuitive form; a detailed but necessarily abstract mathematical model was presented in an earlier technical report.[9] In the simplified case we will model the optical signal received by a small but finite size circular pixel transversely centered (but obviously not longitudinally centered) in the light cone converging toward and then diverging from the point image of a point scene. Our experiments instantiate the assumptions of the simplified point scene point image model.

The scene is a bright red LED behind a pinhole. It is viewed at a range of about 350 mm by a 50 mm f/1.2 lens approximately but not exactly focusing the scene on another pinhole in front of a phototransistor optical sensor. The receiving pinhole and phototransistor assembly are mounted on an audio speaker whose cone can be displaced relative to the scene and lens by application of appropriate voltage to the driving coils. The voltage that we apply to the speaker consists of a slowly varying component that scans the lens-to-sensor distance over about one millimeter in about one minute, and a rapidly varying component that dithers the lens-to-sensor distance through about one-tenth of a millimeter three hundred times per second.

Four lock-in amplifiers[10] monitor the signal from the phototransistor. Two monitor the optical signal's Fourier amplitudes in phase and in quadrature phase with the dither. The other two monitor the optical signal amplitude and phase at the second harmonic of the dither frequency. The theory says that if we define the initial phase such that the dither is represented by $\sin \omega t$ then we expect to see substantial Fourier amplitude in phase but none in quadrature phase $\cos \omega t$, and we expect to see substantial second harmonic Fourier amplitude at $\cos 2\omega t$ but none at its quadrature phase $\sin 2\omega t$. The two "extra" lock-in amplifiers are used only to confirm that there is no significant Fourier amplitude where the theory says there should be none.

The slowly varying outputs of the lock-in amplifiers are digitized by a general purpose PC computer analog/digital I/O card, and the ratio of the two interesting Fourier amplitudes is taken in software.[11] The ratio is plotted vs lens-to-sensor distance. As expected, it is found to be linear (within understandable limits), and it has the expected slope. The zero crossing of the

ratio gives the range to the scene with impressive accuracy. The zero crossing can be located physically (which corresponds to finding the location of the sharply focused image), but since it is linear and its slope is known from first principles it is straightforward and adequate to find it by extrapolation.

4. THEORY

Figure 1 shows a simple lens imaging an on-axis point scene to an on-axis point image in the geometrical optics approximation. Imagine the output of an on-axis optical sensor, a single circular pixel of small but non-zero diameter, as a function of the sensor-to-image (or focus-error) distance ζ_o. For positive values of ζ_o the image falls in front of the sensor (i.e., the image is between the lens and the sensor), and for negative values of ζ_o the image falls in behind of the sensor (i.e, the sensor is between the lens and the image).

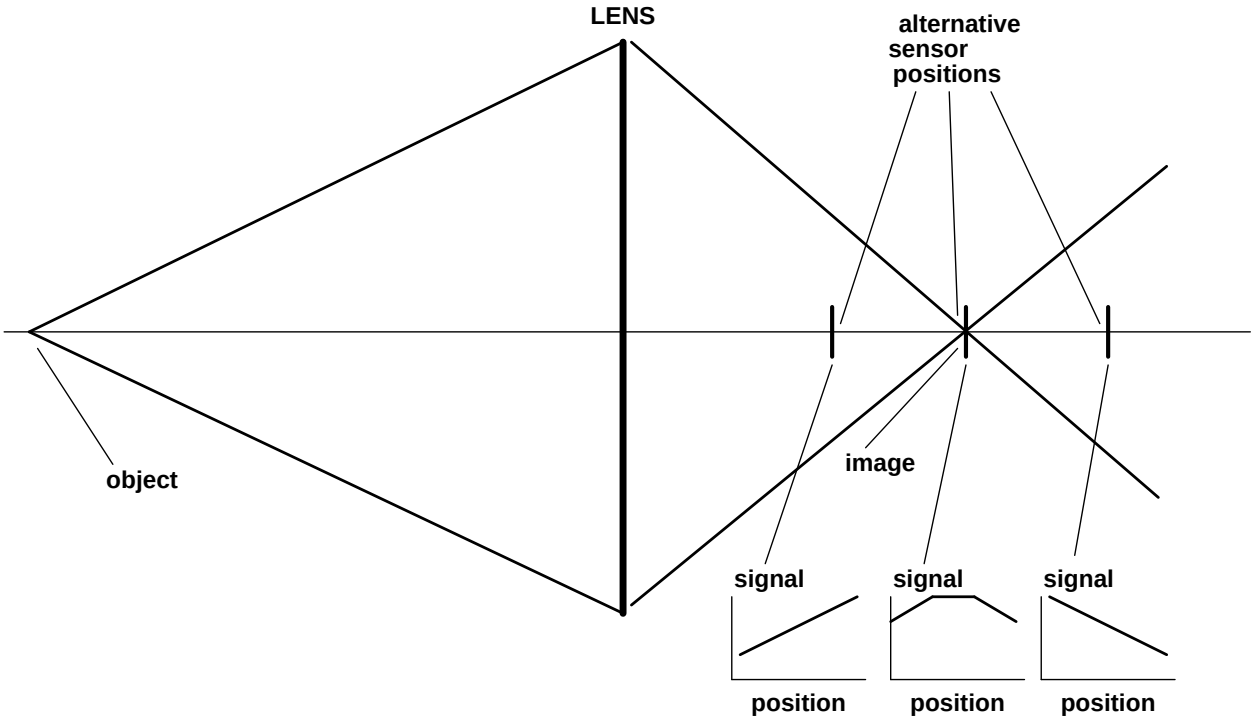

Figure 1: Geometry of the intuitive model

Now imagine a small oscillatory motion $\alpha \sin \omega t$ superimposed on ζ_o so the total image-to-sensor distance (or focus-error distance) is

$$\zeta = \zeta_o + \alpha \sin \omega t \tag{2}$$

We will now describe an analog signal processing approach to measuring $\frac{\zeta_o}{\alpha}$. Since α is known, ζ_o is therefore known. Since z'' is known, $z' = \zeta_o + z''$ is therefore. Then from the lens-equation (equation 1), the range z is known.

There are three limiting cases where how the signal varies during each oscillation period is particularly straightforward, and can be understood just by inspection of Figure 1.

When ζ_o is negative and α is small enough that ζ is strictly negative, so the sensor is always between the lens and the image, the sensor intercepts a larger solid angle of the light cone as it moves toward the image plane. The result is a signal *in phase* with the sensor motion:

$$S(t) = B_0 + A_\omega \sin \omega t \tag{3}$$

where B_0 and A_ω are positive constants. This behavior is illustrated by the left inset in Figure 1.

In contrast, when ζ_o is positive and α is small enough that ζ is strictly positive, so the sensor is always behind the image, the sensor intercepts a smaller solid angle of the light cone as it moves away from the image plane. The result is a signal *180^o out of phase* with the sensor motion:

$$S(t) = B_0 - A_\omega \sin \omega t \tag{4}$$

where B_0 and A_ω are positive constants. This behavior is illustrated by the right inset in Figure 1.

Alternatively both these cases can be summarized by:

$$S(t) = B_0 - SIGN(\zeta_o) A_\omega \sin \omega t \tag{5}$$

where B_0 is a positive constant and A_ω is a constant whose sign is opposite to the sign of ζ_o. It thus follows that the Fourier amplitude of the in-phase signal is zero when the focus-error distance ζ_o is zero.

This leads us to ask what happens when the dither drives the sensor *through* the image plane.

Inspection of the middle inset in Fig. 1 suggests that when this is the case we should expect to see a signal component that rises as the sensor moves toward the image plane and falls as the sensor moves away from the image plane *in either direction*, i.e., a signal component *at twice the frequency of the dither*, and by symmetry, having a *cosine* shape about ζ_o:

$$S(t) = B_0 + B_{2\omega} \cos 2\omega t \tag{6}$$

where $B_{2\omega}$ is a positive constant.

Combining all three special cases leads to:

$$S(t) = B_0 - SIGN\{\zeta_o\} A_\omega \sin \omega t + B_{2\,\omega} \cos 2\,\omega t \tag{7}$$

Finally, we will now show that, in agreement with these intuitive expectations:

$$A_\omega = -\frac{2\,\alpha}{\zeta_o} \quad \textit{and} \quad B_{2\omega} = \frac{\alpha^2}{2\,\zeta_o{}^2} \tag{8}$$

Let W_o *watts* be the optical power incident on the lens, let R be the lens radius, let r be the circular pixel radius, and let z be the lens-to-image distance. Then, by examination of the geometry shown in Figure 1, the optical power collected by the pixel is:

$$W = W_o \frac{r^2 z^2}{R^2 \zeta^2} \tag{9}$$

Substituting $\zeta_o + \alpha \sin \omega t$ for ζ, making the assumption[12] $\alpha << \zeta_o$, and invoking the approximation $(1+x)^{-1} \approx (1-x)$ for small x is follows that:

$$W = W_o \frac{r^2 z^2}{R^2 \zeta_o{}^2} \left(1 - \frac{2\alpha}{\zeta_o} \sin \omega t + \frac{\alpha^2}{\zeta_o{}^2} \sin^2 \omega t\right) \tag{10}$$

Invoking the trigometric identity $\sin^2 \omega t = \frac{1 - \cos 2\omega t}{2}$ then gives:

$$W = W_o \frac{r^2 z^2}{R^2 \zeta_o{}^2} \left(\left(1 - \frac{\alpha^2}{2\zeta_o}\right) - \frac{2\alpha}{\zeta_o} \sin \omega t + \frac{\alpha^2}{2\zeta_o{}^2} \cos 2\omega t\right) \tag{11}$$

The amplitudes of the in phase first harmonic and the dominant second harmonic signal amplitudes are thus, by inspection, exactly those disclosed in equation 8.

It follows, as anticipated, that if the dither amplitude is known and the amplitudes A_ω and $B_{2\omega}$ are measured then the focus-error is known:

$$\zeta_o = -\frac{\alpha A_\omega}{4 B_{2\omega}} \tag{12}$$

To emphasize that conceptually the ratio of the Fourier amplitudes is measured directly, e.g., using an analog ratio circuit rather than measuring the amplitudes individually, digitizing, and dividing, we define

$$q = \frac{A_\omega}{B_{2\omega}} \tag{13}$$

and write

$$\zeta_o = -\frac{\alpha q}{4} \tag{14}$$

We think of the dither amplitude α as a setable parameter, q as the measured variable, and the focus-error distance ζ_o as the calculated result.

The sensitivity of the method $\frac{\delta \zeta_o}{\delta q} = -\frac{\alpha}{4}$, increases linearly with dither amplitude. The experimental data confirm this predicted increase in sensitivity with increasing dither amplitude, as well as "the other side of the coin," a corresponding decrease of dynamic range, which is limited by the dynamic range of the amplifiers, the limits of validity of the approximations used in the derivation, and the additive noise that dominates when ζ_o is large, causing q to appear to saturate at a constant value.

The scale factor $W_o \frac{r^2 z^2}{R^2 \zeta_o{}^2}$ that multiplies all the Fourier coefficients can be understood in terms of four physically meaningful components:

- W_o: the signal scales with the optical power received from the scene, almost a tautology, since we have tacitly (and correctly) assumed that signal voltage is proportional to incident optical power;
- $\frac{z^2}{R^2}$: the signal scales with the square of the reciprical of the *effective f–number* of the lens;
- r^2: the signal scales with the active area of the pixel;
- $\zeta_o{}^{-2}$: the signal scales as the inverse square of the image-to-sensor distance.

5. RESULTS

The results of our experiments are illustrated in Fig. 2, 3, and 4. These figures each show data obtained during a single run. They are typical of results obtained over a series of repeated runs.

Signal vs. Focus-Error: Near focus the dither amplitude, the offset from the image plane to the sensor plane, and the ratio of the first to second harmonic amplitudes in the sensor output obey the expected relationship; far from focus the model breaks down. This is illustrated by Fig. 2. From the slope near $q = 0$ in Fig. 2 we estimate

$$\alpha = -4\left(\frac{dq}{d\zeta_o}\right)^{-1} \approx 229\,\mu m \tag{15}$$

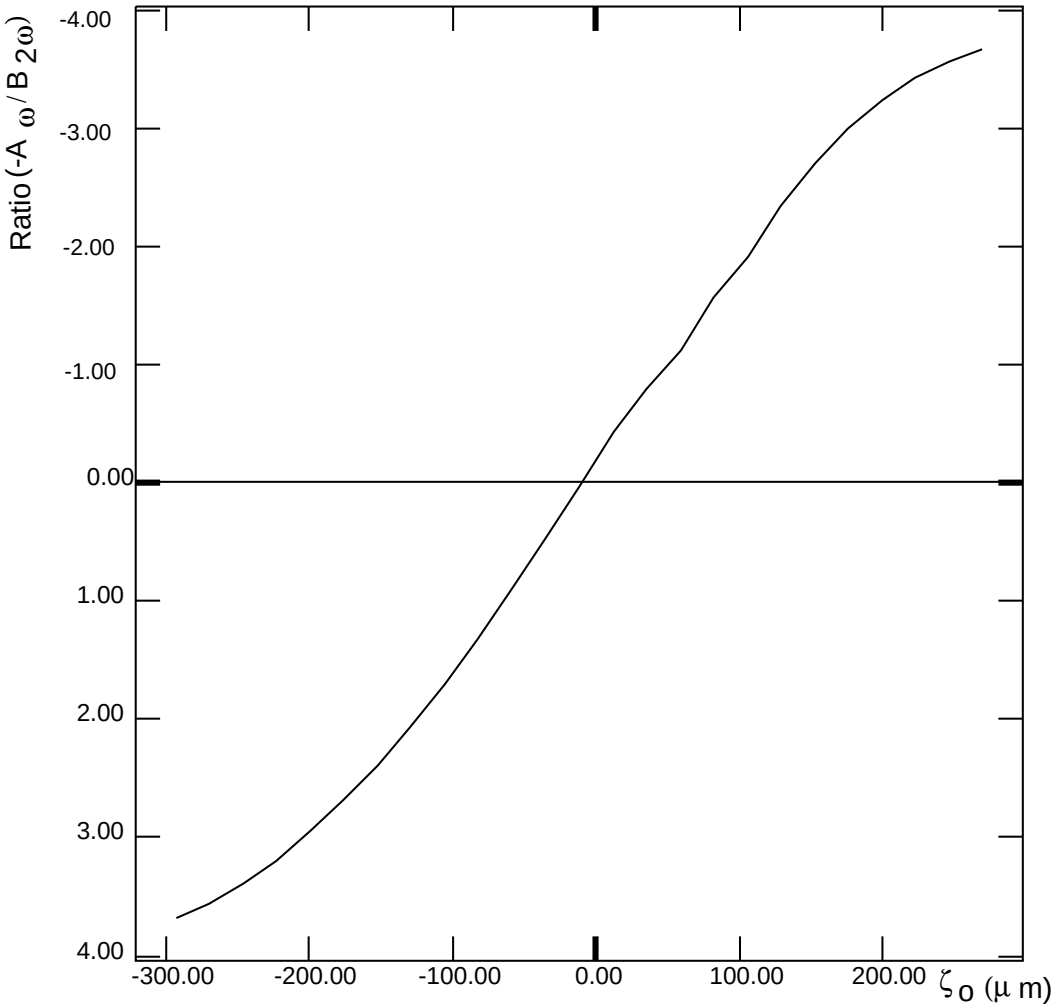

Figure 2: Ratio of first harmonic to second harmonic *vs* focus error

This result is in reasonable agreement with an independent estimate of 266 μm obtained by calculating the standard deviation of the position-sensitive detector output.

Signal vs. Range: Changing the range to the object produces a corresponding change in the image position. The four curves in Fig. 3 were generated using identical experimental conditions except for range to the object, which is noted beside each curve. The result is summarized in Table 1.

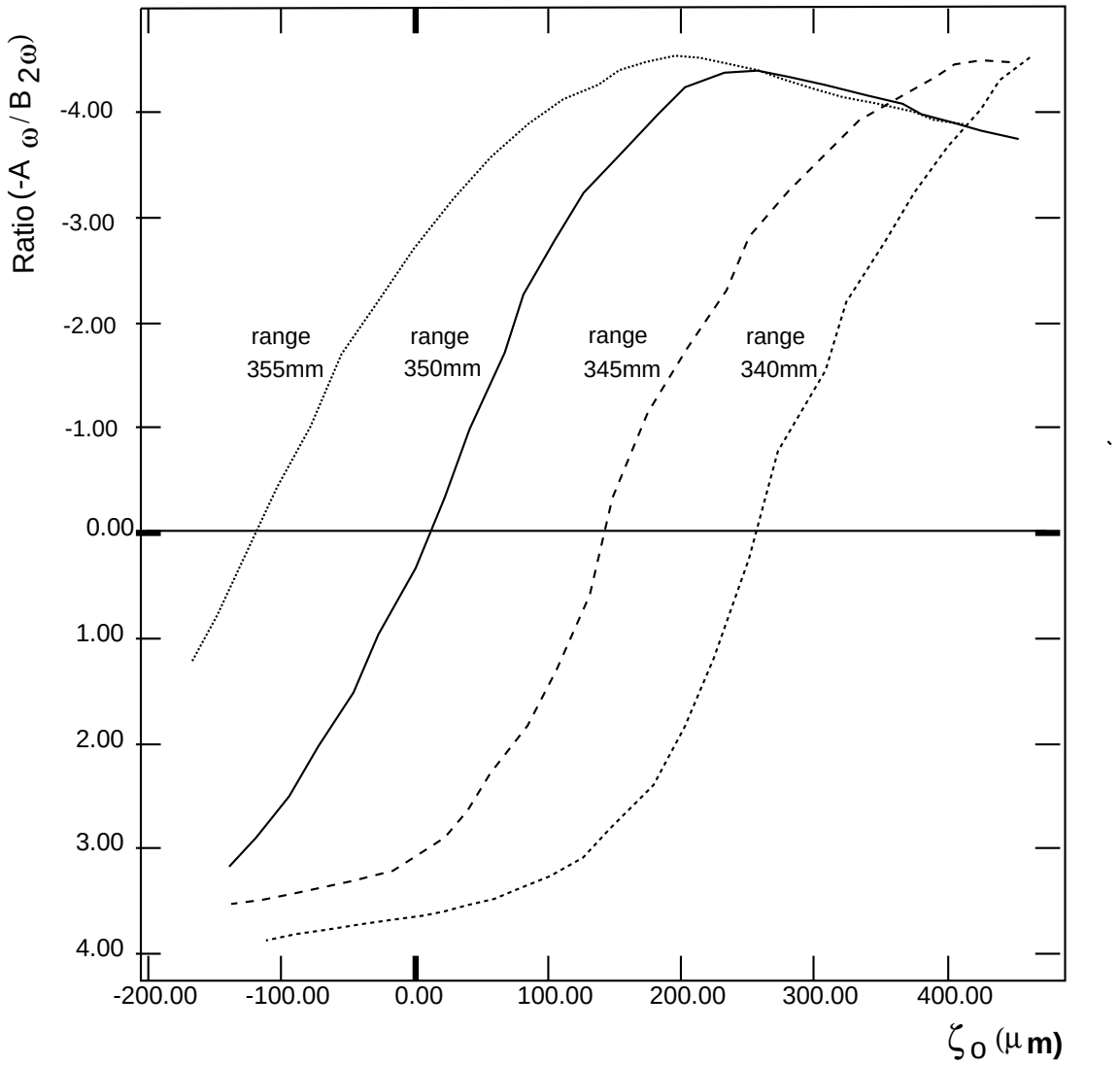

Figure 3: Range to the object *vs* object image location

Measurement Range (mm)	50 mm Lens calculated image distance (mm)	50 mm Lens calculated image distance difference (μm)	48 mm Lens calculated image distance (mm)	48 mm Lens calculated image distance difference (μm)	measured image distance difference (μm)
355	58.197		55.505		
350	58.333	136	55.629	124	122
345	58.475	142	55.758	128	128
340	58.621	146	55.890	133	133

Table 1

The 3.5 mm span of the position-sensitive detector is too small to measure the image distances, approximately 55 mm. Thus the calculated and measured image distance differences are of greatest utility. Our measurements are respectively 10.3%, 9.9% and 8.9% too low, indicating a systematic rather than a statistical error. This is probably due either to an error in the effective focal length in the lens or to an error in the calibration factor of the position-sensitive detector. The right side of Table I shows the effect of repeating the calculation with an assumed focal length of 48 mm instead of the nominal 50 mm: the discrepancy between measurement and calculation disappears. It would similarly disappear if we reduced the calibration factor of the position-sensitive detector by 10%.

Effect of dither amplitude change: Changing the dither amplitude affects both the dynamic range and the precision of the measurements. This is shown in Fig. 4, where the two curves were taken under identical experimental conditions with the exception that the curve marked "large dither" is taken at twice the dither voltage of the one marked "small dither". From the curves we calculate $\alpha = 307$ mm in the "large dither" case and $\alpha = 154$ mm in the "small dither" case, in exactly the predicted 2:1 ratio.

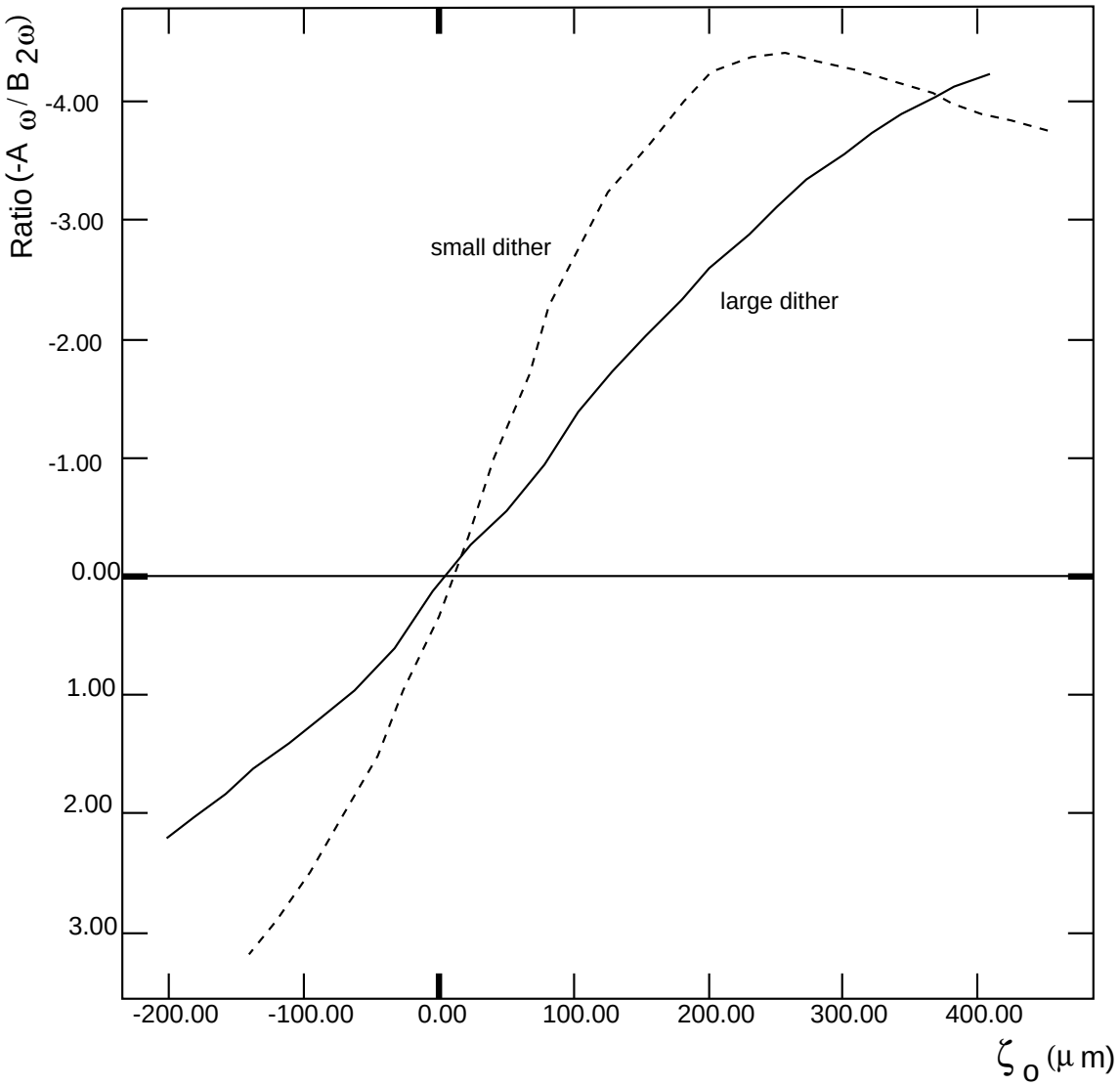

Figure 4: Dither amplitude *vs* dynamic range

The difference in zero crossing between the two curves is about 6 mm, which translates into about 0.25 mm range uncertainty at a range of 350 mm. We imagine that by utilizing this trade-off between dynamic range and precision we could optimize practical range measurements in response to the scene or specific scene features.

6. CONCLUSIONS

These experiments indicate that our first order model is valid: within a useful range of the distance between the image plane and the sensor plane the harmonic content of the signal from the longitudinally dithered sensor predicts the distance. Analyzing the data in the context of our models' predictions allow us to make the following conclusions:

The method works as predicted. The ratio of the first to second harmonic amplitudes of the sensor output exhibits a linear relationship to the longitudinal focus-error when the sensor is near the image plane.This linear relationship allows us to measure range to an object without the need to actually move the sensor to the point of exact focus, thereby suggesting the possibility of measuring range to all scene regions in parallel with the normal imaging process. This is in contrast to conventional range-from-focus methods which measure the range to each scene region separately by moving the sensor plane to optimize some focusing criteria. In future experiments we plan to use an array sensor to explore the hypothesized ability to measure the range to all scene regions in parallel.

The method is very accurate. The accuracy of the range measurements using our method approaches that of known passive single camera methods even though our initial experiment utilizes relative imprecise devices to create and measure movement of the sensor.

The precision/dynamic range space is easily navigated. By changing the amplitude of the dither we can control the trade-off between the precision and the dynamic range of the range measurements: signal = $A_\omega/B_{2\omega} = -4\zeta o/\alpha$.

Thus increasing dither amplitude α will decrease the signal corresponding to a given focus-error, thereby increasing dynamic range at the expense of precision. Decreasing α will increase precision at the expense of dynamic range, i.e., the method will work only for small values of focus-error. It would be easy, and it might be useful in some applications, to make this adjustment dynamically in response to the scene or specific scene features.

7. FUTURE WORK

A small array of sensors with some on-chip processing has been designed and fabricated using VLSI technology. It is now being tested in the apparatus described here.

8. ACKNOWLEDGEMENTS

Thanks to Nicole Desvignes for work on the early prototype and to Alan Guisewite for extensive laboratory and editorial assistance.

9. REFERENCES

1. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

2. R. A. Jarvis, ''A Perspective on Range Finding Techniques for Computer Vision'', *IEEE Trans. PAMI-5(2)*, Mar 1983, pp. 122-139.

3. E. Krotkov, ''Focusing'', *International Journal of Computer Vision*, Vol. 1, No. 3, October 1987, pp. 223-37.

4. P. R. Cohen and E. A. Feigenbaum, *Handbook of Artificial Intelligence, Chap. XIII*, William Kaufman, Los Altos, CA, Vol. IV, 1982.

5. Reuel A. Sherman, ''Benefits to Vision Through Stereoscopic Films'', *Journal of the SMPTE*, Vol. 61, September 1953, pp. 294-308.

6. Shoei Kataoka, ''An Attempt Towards an Artificial Retina: 3-D IC Technology for an Intelligent Image Sensor'', *Transducers '85*, IEEE, IEEE, Piscataway, NJ 08854, June 1985, pp. 440-5.

7. B. H. Wilcox, et al., ''A Vision System for the Mars Rover'', *Mobile Robots II*, SPIE, Vol. 852, 1987.

8. Conceptually the signals are available to us from all pixels continuously in parallel; the method would be applicable to a scanned CCD, vidicon, etc., only if the dither frequency were many times the frame frequency, e.g., 1 Hz. We envisage fabricating sensor arrays with the necessary signal processing replicated behind each pixel, so scanning pixels would deliver combined gray level and range images at the normal frame rate.

9. M. W. Siegel, ‘‘Image Focusing in Space and Time’’, Technical Report CMU-RI-TR-88-2, The Robotics Institute, Carnegie Mellon University, February 1988.

10. Lock-in amplifiers are also variously known as synchronous amplifiers, synchronous rectifiers, and phase sensitive amplifiers. This analog signal processing technique is extremely powerful for extracting small signals from overpowering backgrounds. A reference oscillator modulates the signal source (in this case by dithering the lens-to-sensor distance), causing an amplitude modulation in the detector signal that is synchonous with the reference oscillator. The analog-computed product of the (perhaps phase shifted) reference signal and the detector signal is analog-integrated over many modulation cycles. The integrator output is a DC voltage proportional to the Fourier amplitude of the reference frequency in the detector signal. Unmodulated components of the detector signal have random phases with respect to the reference signal, and thus their net contribution to the integral is zero. Lock-in detection methods powerfully squeeze out continuum noise by extreme narrow-banding, but unlike open-loop narrow-band methods they are insensitive to small drifts in the reference frequency.

11. Taking the ratio in software is appropriate in the present context of demonstrating the concept while retaining diagnostic access to the lowest level measurements. We would want a fieldable sensor to take the ratio using an analog dividing circuit whose output we could digitize.

12. In fact this assumption is violated by the third exemplary case, wherein ζ_o is nearly or exactly zero, but this treatment is intended to be intuition building, not mathematically rigorous.