Image Focusing in Space and Time

M. W. Siegel

February 1988

CMU-RI-TR-88-XXX

Intelligent Sensors Laboratory

The Robotics Institute
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Abstract

The integral form of the instrument transmission function for a one-dimensional pixel in a two-dimensional optical system is presented. The integral is solved explicitly in the paraxial ray approximation for a single spatial Fourier component of a Lambertian object. The difference between signals from adjacent pixels is derived. It is shown to have zero derivative with respect to focusing error when the focusing error is zero, *i.e.*, it is a weak source of range-from-focus information. Describing the instantaneous focusing error as the sum of a fixed offset and a time-domain sinusoidal dither, the power spectrum of the signal from each individual pixel is shown to contain large first and second harmonic terms for physically reasonable values of the parameters. The first harmonic signal is proportional to the product of the dither amplitude and the offset. The second harmonic signal is proportional to the square of the dither amplitude and is independent of offset. The two coefficients are identical except for an integral numerical factor. It is suggested that the ratio of second harmonic to first harmonic signals is thus potentially a powerful measure of offset, *i.e.*, of focusing error in the limit of zero dither, and thus of range-from-focus *pixel-by-pixel*. Extending the model to three dimensions, removing the approximations, extending the model to natural scenes, and verifying and implementing the results experimentally are outlined briefly.

2 Introduction

Image focusing [7] is conventionally regarded as a spatial-domain activity: the focus-controlling parameter (lens-to-sensor plane 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 focusing signal, *i.e.*, these amplitudes, is derived from pixel-to-pixel signal differences. The focusing information available from these differences is in reality weak. 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) arbitrarily coupled to the image distance. 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 stereo 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' images [10].

In this report I partially model the image, *i.e.*, the signal associated with each pixel in the sensor plane, simply and approximately described by:

- the object modeled as a Fourier amplitude for an arbitrary spatial frequency and phase;
- the object distance z, the image distance z', and their relationship via the lens equation;
- the sensor plane distance z'' and the pixel diameter 2ρ

in two dimensions, *i.e.,* for cylindrical optics.

The result shows explicitly why pixel-to-pixel signal differences are a weak source of focusing information: the derivative of the pixel-to-pixel signal difference with respect to sensor plane distance z'' is zero precisely at perfect focus $z'' = z'$, which makes it operationally difficult to find the exact focus using only the spatial domain information.

I then examine the predictions that the model makes in the longitudinal direction. This is conveniently imagined as an experiment in the time domain: the signal from *each pixel* is modulated by dithering the sensor plane distance as $\sin \omega t$. The dominant AC signal appears at the fundamental dither frequency, and precisely in phase with it, and the next largest harmonic is the second, corresponding to a $\cos 2\omega t$ term. The fundamental signal is proportional to the product of the dither amplitude and the offset distance between the image plane and the sensor plane, whereas the second harmonic signal is proportional to the square of the dither amplitude, and is independent of the offset distance. The proportionality constant for the second harmonic is exactly one-fourth the proportionality constant for the fundamental. I then use this model to show how a pair of synchronous amplifiers [8] tuned to $\sin \omega t$ and $\cos 2\omega t$ could be used in a ratio mode to detect focus precisely, and thus robustly to deduce range-from-focus *pixel-by-pixel*.

3 Model

For geometrical simplicity, and for the accompanying simplicity in the degree and limits of the integrals representing the instrument transmission function[1], in this introductory report I will model a

[1] Idealized instruments use point detectors to look at point sources through infinitesimal apertures; real instruments report the integral over their own detector area of signal received from a finite source area through finite sized apertures. The instrument transmission function is a description, in terms of integrals over aperture and detector dimensions, of the signal that will be seen for any specified source description.

cylindrical rather than a spherical optical system. This will affect the power law behavior of some variables, *e.g.*, the "exposure time" will be linear rather than quadratic in the f-number, and some numerical coefficients may differ in the two cases by factors the order of unity, but the essential conclusions should be the same in cylindrical and spherical models. To keep notation simple and physical concreteness in the forefront, throughout the report I will illustrate with geometrically special cases that involve no loss of physical generality.

The model optical system is depicted in Figure 1. It consists of a simple, thin, aberration free lens of aperture $2R$, an object plane at distance z measured to the left of lens center, the corresponding image plane at distance z' measured to the right of lens center, and a sensor plane at distance z'' also measured to the right of lens center. Locations in the object, corresponding image, and sensor planes are measured by x, x', and x'' respectively, with the positive direction of x physically opposite to the positive directions of x' and x''. The optical model is geometrical, ignoring diffraction entirely.

The object plane is characterized by a source function $W(x, \theta)$ that in this introduction I will take as an angularly Lambertian, spatially sinusoidal grating[2] representing one Fourier component of the optical power emitted or reflected by the object:

$$W(x, \theta) \;=\; W_o \cos\,(kx + \phi) \cos\,\theta \quad \textit{watts–meter}^{-1}\textit{–radian}^{-1} \tag{1}$$

The constant $k = \frac{2\pi}{\lambda}$, where λ is the spatial wavelength of the sinusoidal object feature. The constant ϕ is a phase factor that describes the symmetry (or lack of symmetry) of the sinusoid about the optical axis. $\phi = 0$ is the special case of a sine function (antisymmetrical), and $\phi = \frac{\pi}{2}$ is the special case of a cosine function (symmetrical). Direction angle θ is with respect to the object plane normal.

Figure 1 also shows a typical ray connecting x_o'', the center of a pixel that extends from $x_o'' - \rho$ to $x_o'' + \rho$, to the object plane, which it intersects at location x_o and angle θ. The power collected by this pixel[3] is

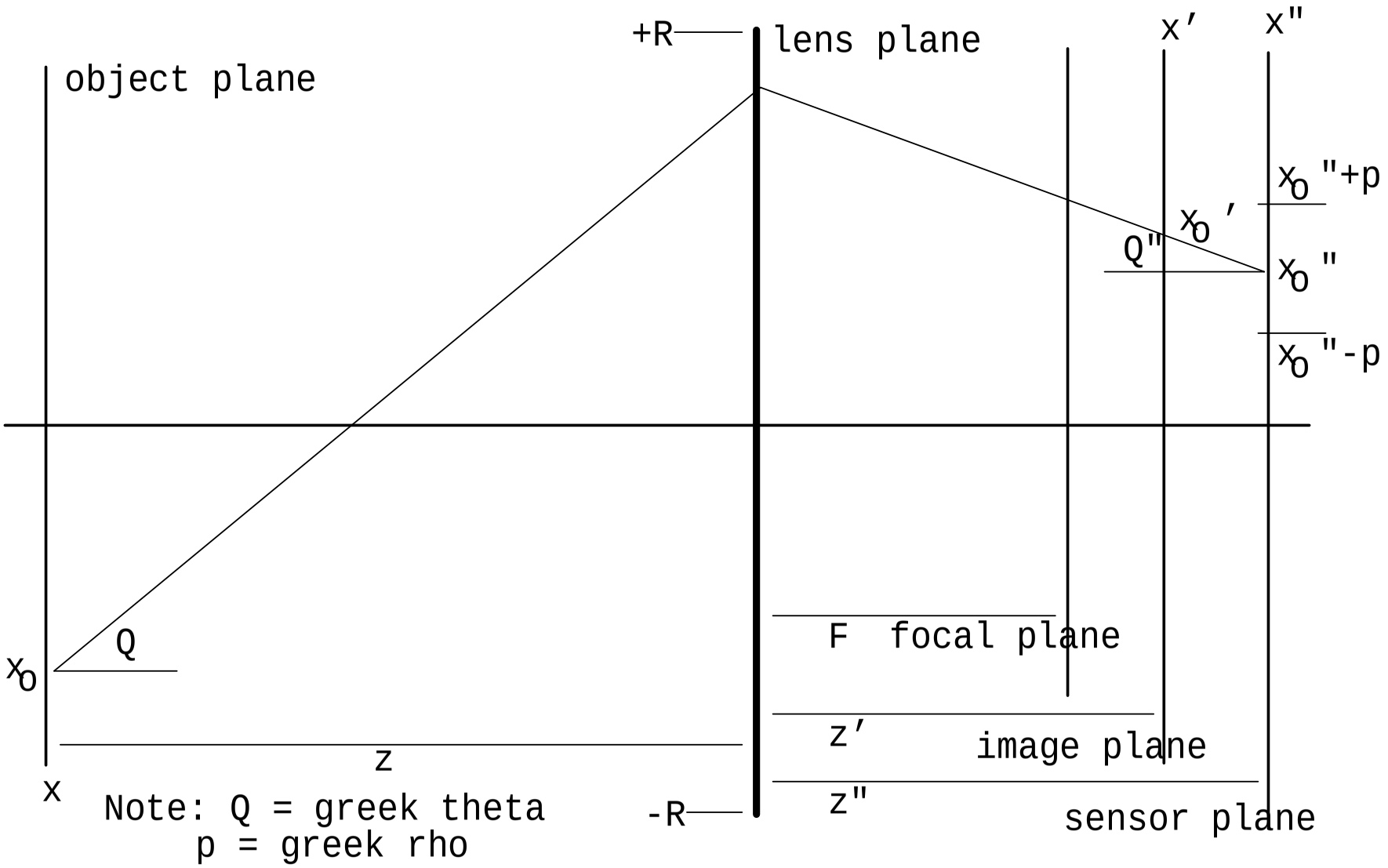
$$S \;=\; \int_{x_o''-\rho}^{x_o''+\rho} \int_{-\text{atan}\frac{R+x_o}{z''}}^{+\text{atan}\frac{R-x_o}{z''}} W(x(x'', \theta''), \theta(x'', \theta''))\;\; d\theta''\;\; dx''\;\; \textit{watts} \tag{2}$$

As is often the case in modeling instrument transmission functions, the key features of the problem reside in the limits of the integrals that describe the physical averaging performed by the various apertures in the system.

---

[2] A uniform background, structureless in space and time, can be superimposed in the reader's mind if the negative values sometimes assumed by this function are disconcerting. The background makes no net contribution to the spatial difference and temporal derivative signals of interest in this report.

[3] Depending on the physical mechanisms underlying transduction, sensors may or may not generate output signals more-or-less linear in the incident optical power. In practice optical detectors, both electronic and photochemical, when used under the conditions recommended by their manufacturers, deliver electrical voltage or developed optical density signals whose amplitudes are approximately linear in the product of incident optical powers and integrating time, *i.e.,* these detectors are *incident energy* sensitive. This functional relationship is not required in any fundamental sense: a detector could in principle respond to the electric field strength (wave amplitude) rather than to power (wave intensity, essentially amplitude squared). The distinction fortuitously evaporates in the usual case of incoherent illumination: averaging over random phases leaves rms power proportional to electric field amplitude. However for coherent (laser) illumination, where this averaging does not occur, the distinction is important. It is also important in sonar ranging: typical modern acoustic transducers, *e.g.,* the ubiquitous Polaroid [9] product, are amplitude (diaphragm displacement) sensitive, whereas typical older transducers, *e.g.,* the carbon granule microphones in telephone mouthpieces, are (I suppose) power sensitive. Sonar ranging modules that compensate for attenuation with range by using an amplifier whose gain is ramped linearly with time are relying on the amplitude sensitivity of the detector. Because photodetectors are in practice intensity sensitive, time linear compensation does not work with optical imaging, *e.g.,* for a given flashbulb or strobe lamp energy pulse the product of f-number (reciprocal square root of exposure) and object distance is a constant, the "guide number".



Note: Q = greek theta
p = greek rho

Figure 1

From the geometry

$$x \;=\; x'' + (z'' - z')\tan\theta'' \quad meters \tag{3}$$

and

$$\theta \;=\; \text{atan}\,\frac{x + x'' + z''\tan\theta''}{z} \quad radians \tag{4}$$

From the simple lens equation [5]

$$x \;=\; \frac{z}{z'}x' \quad meters \tag{5}$$

Substituting these relationships, and making small angle, near axis approximations

$$S \;=\; \int_{x_o''-\rho}^{x_o''+\rho} \int_{\frac{-R}{z''}}^{\frac{+R}{z''}} W_o \cos\left(\frac{k\,z(x'' + (z'' - z')\,\theta'')}{z'} + \phi\right) d\theta''\; dx'' \quad watts \tag{6}$$

which integrates to

$$S \;=\; \frac{4\,W_o\,\rho R}{z''}\;\frac{\sin\frac{k\,z\,\rho}{z'}}{\frac{k\,z\,\rho}{z'}}\;\frac{\sin\frac{k\,z\,R\,(z'' - z')}{z'\,z''}}{\frac{k\,z\,R\,(z'' - z')}{z'\,z''}}\;\cos\left(\frac{k\,z\,x_o''}{z'} + \phi\right) \quad watts \tag{7}$$

Substituting some convenient definitions:

- $k\frac{z}{z'} \;\equiv\; k'$, the object feature spatial frequency in the image plane;
- $\frac{R}{z''} \;\equiv\; A$, the half-aperture, approximately $\frac{1}{2f}$, and exactly $\frac{F}{2\,z''\,f}$ where F is the focal length, and f is the conventional f-number;
- $k'A \;\equiv\; k''$, the object feature spatial frequency in the image plane times the half-aperture, and thus a measure of the depth of field;
- $z'' - z' \;\equiv\; \zeta$, the offset between the image and sensor planes;

the result is simply expressed as

$$S \;=\; 4\,W_o\,\rho\,A\;\frac{\sin k'\rho}{k'\rho}\;\frac{\sin k''\zeta}{k''\zeta}\;\cos\left(k'\,x_o'' + \phi\right) \quad watts \tag{8}$$

This equation for the instrument transmission function of a pixel is the model in the approximation stated. It describes the optical power received by a pixel as the product of several physically sensible terms:

- the image function $W_o \cos\left(k'\,x_o'' + \phi\right)$ corresponding to the source function *equation 1*, where in the small angle approximation $\cos\theta = 1$;
- the full pixel height $2\,\rho$;
- the full lens aperture $2A$;
- a transverse spatial filter $\frac{\sin k'\rho}{k'\rho}$;
- and a longitudinal spatial filter $\frac{\sin k''\zeta}{k''\zeta}$.

The features of this result that I want to investigate in this report are its transverse pixel-to-pixel differences and its longitudinal derivatives (conveniently modeled as the temporal frequency spectrum when ζ undergoes forced oscillation). In a future report I will discuss the corresponding three dimensional model, integration over multiple spatial frequencies (thus admitting realistic object descriptions), and the effects of removing the paraxial and other smallness approximations.

4 Example

For concreteness I will assign "typical" values to the parameters, and use these values for illustration and comparison throughout the rest of this report:

- source function $W_o = 1000\,watts\text{–}meter^{-1}\text{–}radian^{-1}$;
- pixel size $2\rho = 13\,\mu m$;
- lens focal length $F = 20\,mm$, lens aperture $2R = 10\,mm$, thus $f\text{–}number = 2$ and $A = 0.25$;
- object distance $z = 2\,meters$, *i.e.,* magnification 0.01.

An interesting choice for the object feature size is the one for which $k'\rho = \frac{\pi}{2}$, so that for the $13\,\mu m$ pixel size $k' = 2.4166 \text{ x } 10^5 \ meters^{-1}$ and $k = 2.4166 \text{ x } 10^3 \ meters^{-1}$. These correspond to a spatial wavelength in the object plane of $2.6\,mm$ or $26\,\mu m$ in the image plane, *i.e.,* exactly two pixel widths: a light band falling on one pixel and a dark band falling on an adjacent pixel maximumize contrast. The pixelation is then an optimally matched filter for the spatial wavelength.

Finally, $k'' = k'A = 6.042 \text{ x } 10^4\,meters^{-1}$, corresponding to a longitudinal wavelength (for the specified f-number) of $104\,\mu m$, twice the f-number times the transverse wavelength. "Small" in the longitudinal direction means small with respect to this distance.

5 Differences Between Adjacent Pixels

Consider a pixel centered on axis at $x_o'' = 0$ and an adjacent pixel centered at $x_o'' = 2\rho$. By symmetry, when the sensor plane has a pixel centered on-axis (in contrast to having two pixels straddle the axis), the most visible object features will be those for which $\phi = 0$. The difference in signal between the on-axis pixel and an adjacent pixel is then

$$\Delta S \ = \ 4\,W_o\,\rho\,A\,\frac{\sin k'\rho}{k'\rho}\ \frac{\sin k''\zeta}{k''\zeta}\ (\cos 0 \ - \ \cos 2k'\rho)\ \ watts \tag{9}$$

which expands exactly to

$$\Delta S \ = \ \frac{8\,W_o\,A}{k'}\ \sin^3 k'\rho\ \ \frac{\sin k''\zeta}{k''\zeta}\ \ watts \tag{10}$$

and for small focusing errors $k''\zeta << 1$

$$\Delta S \ \approx \ \frac{8\,W_o\,A}{k'}\ \sin^3 k'\rho\ \ [\,1 - \frac{(k''\zeta)^2}{3!}\,]\ \ watts \tag{11}$$

The absolute value of the difference signal clearly has a local transverse extremum for any integer n satisfying $k'\rho = \frac{n\pi}{2}$ and a local longitudinal extremum for any integer m satisfying $k''\zeta = \frac{m\pi}{2}$. Because of the k' in the denominator of the numerical coefficient the difference signal has a global maximum when $n = 1$. *The best contrast between adjacent pixels is obtained when the image plane feature size has a spatial half-wavelength equal to the pixel diameter.*

For this best-contrast condition, with the feature size optimally matched to the pixel size $k'\rho = \frac{\pi}{2}$ but perhaps away from precise focus:

$$\Delta S_{matched} \ = \ \frac{16\,W_o\,A\,\rho}{\pi}\ \ \frac{\sin k''\zeta}{k''\zeta}\ \ watts \tag{12}$$

When the sensor plane also coincides with the image plane $\zeta = 0$ the value of $\frac{\sin k''\zeta}{k''\zeta}$ is unity so

$$\Delta S_{matched}^{focused} = \frac{16\, W_o A \rho}{\pi} \quad watts \tag{13}$$

This is the largest difference signal that can ever be obtained between adjacent pixels (for an object with a single sinusoidal feature). It is thus convenient to use $S_o \equiv \frac{\Delta S_{matched}^{focused}}{2}$ as the unit relative to which to measure other signal powers.

With this notation the instrument transmission function is

$$S = \frac{\pi S_o}{2} \frac{\sin k' \rho}{k' \rho} \frac{\sin k'' \zeta}{k'' \zeta} \cos (k' x_o'' + \phi) \quad watts \tag{14}$$

and the difference between signals from adjacent pixels for object features optimally matched to pixelation is exactly

$$\Delta S_{matched} = 2 S_o \frac{\sin k'' \zeta}{k'' \zeta} \quad watts \tag{15}$$

and in the limit of small $k'' \zeta$

$$\Delta S_{matched} \approx 2 S_o \left[1 - \frac{(k'' \zeta)^2}{3!}\right] \quad watts \tag{16}$$

The physical interpretation is that at focus, with the feature's spatial wavelength and phase matched to the pixelation, the on-axis pixel sees S_o, an adjacent pixel sees $-S_o$, and when the sensor plane fails to coincide with the image plane the difference signal is attenuated as $\frac{\sin k'' \zeta}{k'' \zeta}$, or approximately quadratically in the focusing error.

The matched condition is optimal for focusing on contrast. Its sensitivity to ζ is given by the derivative

$$\frac{d\Delta S_{matched}}{d\zeta} = 2 S_o k'' \left(\frac{\cos k'' \zeta}{k'' \zeta} - \frac{\sin k' \zeta}{(k'' \zeta)^2}\right) \quad watts\text{–}meter^{-1} \tag{17}$$

which to first-order in ζ, and recalling $k'' = k' A$, is

$$\frac{d\Delta S_{matched}}{d\zeta} = \frac{-2 S_o A^2 k'^2 \zeta}{3} \quad watts\text{–}meter^{-1} \tag{18}$$

which, of course, could have alternatively been obtained by directly differentiating *equation 16*.

Two points are worth noting:

- the sensitivity improves rapidly with increasing aperture (and might be predicted to do so even more rapidly with spherical optics);
- *nevertheless*, the situation is hopeless at $\zeta = 0$: the effect we would use to detect a discrepancy between the image plane and the sensor plane has zero slope when the discrepancy is zero.

The last point is not so serious if the goal of focusing is just to obtain a sharp image: that the derivative of the difference signal is small simply says that the endpoint is not critical. *But if part of the goal of focusing is to obtain range-from-focus, this result makes the prospects seem grim indeed.*

Returning to the ongoing numerical example $S_o = 4.138\, mW$, so $\Delta S_{matched} = 8.276 \frac{\sin k'' \zeta}{k'' \zeta} mW$. Then how big does the focusing error ζ have to be to make a one-bit difference in ΔS? Suppose (to be generous) that we can digitize *the difference* to 8-bits when ΔS is a half scale signal. We want to know the value of $k'' \zeta$ that makes ΔS_o differ from unity by $1/128$. The answer (obtained graphically) is $k'' \zeta \approx 0.216$, which for $k'' = 6.042 \times 10^4 \ meters^{-1}$ corresponds to a focusing error $\zeta = 3.6\, \mu m$. The

corresponding range error is found by resubstituting the lens equation into its own derivative with respect to z':

$$\frac{\Delta z}{z} = \frac{\Delta z'}{z'} \left[1 - \frac{z}{F}\right] \tag{19}$$

To a good approximation, the fractional range error is the fractional focusing error times the reciprocal of the magnification, 100 in this example. Thus a one bit signal change that corresponds to a $3.6\mu m$ focusing error in a $20mm$ focal length corresponds to 1.8% range error, or $36\,mm$ range error in $2\,meters$. In any but the lowest precision real world application this range error would be unacceptable.

The rest of this report suggests a class of data collection and processing technologies that show on-paper promise of being able to use the longitudinal structure of the image to obtain range-from-focus with high accuracy.

6 The Temporal Dimension

I now investigate the signal observed in the time domain from a single pixel when the sensor plane is both offset from the image plane and is driven in a small amplitude oscillation:

$$\zeta = \zeta_o + \alpha \sin \omega t \quad meters \tag{20}$$

In the absence of practical three dimensional image sensors[4], imagining a pixel plane in longitudinal oscillation, especially with the recognition that synchronous detection can then be employed, is a useful expository tool as well as a proposal for a practical implementation.

For a pixel on-axis $x_o'' = 0$, an object symmetrical about the axis $\phi = 0$, and object feature and pixel sizes satisfying $k'\rho = \frac{\pi}{2}$

$$S(t) = S_o \frac{\sin k''(\zeta_o + \alpha \sin \omega t)}{k''(\zeta_o + \alpha \sin \omega t)} \quad watts \tag{21}$$

For small ζ

$$S(t) = S_o\left[1 - \frac{k''^2(\zeta_o + \alpha \sin \omega t)^2}{3!}\right] \quad watts \tag{22}$$

which expands to

$$S(t) = S_o\left[1 - \frac{k''^2}{3!}(\zeta_o^2 + 2\zeta_o \alpha \sin \omega t + \alpha^2 \sin^2 \omega t)\right] \quad watts \tag{23}$$

In units of S_o we then have for the power spectrum:

- a *DC* term $1 - \frac{(k''\zeta_o)^2}{3!}$, which is just half the adjacent pixel difference *equation 16*;
- an *AC* term synchronous with the driving term $\frac{-2k''^2\zeta_o\alpha}{3!}$ (which can alternatively be interpreted as a positive amplitude and a phase shift of π with respect to the driving term);

[4]Some attempts at building processing layers behind sensing layers are underway [6], but I am unaware of any efforts to build three dimensional sensing lattices.

- a term whose time dependence corresponds to $\sin^2 \omega t$ and *whose amplitude* $\frac{-(k'' \alpha)^2}{3!2}$; is independent of ζ_o, *i.e., independent of focus, and is thus a measure of the product of all the uncertain intensity and geometry related weights.*

Noting that $\sin^2 \omega t = \frac{1 - \cos 2\omega t}{2}$, the $\sin^2 \omega t$ amplitude is further interpreted as

- another *DC* term of power $\frac{-(k'' \alpha)^2}{3!2}$;
- an *AC* term of power $\frac{(k'' \alpha)^2}{3!\,2}$ with frequency and phase corresponding to $\cos 2\omega t$.

The net power accounting is then:

- for the *DC* term: $P_o = 1 - \frac{(k'' \zeta_o)^2}{3!} - \frac{(k'' \alpha)^2}{3!\,2}$
- for the $\sin \omega t$ term: $P_\omega = \frac{-2 k''^2 \alpha \zeta_o}{3!}$
- for the $\cos 2\omega t$ term: $P_{2\omega} = \frac{(k'' \alpha)^2}{3!\,2}$

The ratio of the signals at the second harmonic and the fundamental driving frequencies is $\frac{-\alpha}{4 \zeta_o}$, which becomes arbitrarily large as ζ_o approaches zero, *i.e.*, as focus is achieved. This ratio might thus provide a high sensitivity, high accuracy focusing criterion.

How much useful *AC* signal there is depends on k'', the depth of field in the image distance in relation to the object feature size. If the transverse matching condition $k' \rho = \frac{\pi}{2}$ is satisfied then the longitudinal condition is $k'' = \frac{\pi}{4 \rho f}$. Recall that in the ongoing numerical example, corresponding to these conditions and some typical parameters, $k'' \approx 6.042 x 10^4 \ meters^{-1}$, or $\frac{1}{k''} \approx 16.55 \ \mu m$. For offset and dither amplitudes of this order-of-magnitude the longitudinal smallness approximation is valid to about 1%. As a practical matter, displacements of this size could be easily obtained piezoelectrically. Then continuing the example, if we take $\alpha = \zeta_o = \frac{1}{k''} = \frac{4 \rho f}{\pi}$ the relative signal powers are $P_o \bullet P_\omega \bullet P_{2\omega} = 0.75 \bullet -0.3333 \bullet 0.0833$. Since $P_{2\omega}$ grows quadratically as α, even higher modulation fractions are obtainable within the realm of plausible electronically driven sensor plane displacements.

7 Temporal-Longitudinal *vs* Spatial-Transverse Domains

Combining the results of the two previous sections, the ratio of the maximum difference between signals from adjacent pixels to the time domain single pixel signals at DC, first harmonic, and second harmonic is

$$\Delta S_{matched} \bullet P_o \bullet P_\omega \bullet P_{2\omega} = 2 [1 - \frac{(k'' \zeta)^2}{3!}] \bullet [1 - \frac{(k'' \zeta_o)^2}{3!} - \frac{(k'' \alpha)^2}{3!\,2}] \bullet \frac{-2 k''^2 \alpha \zeta_o}{3!} \bullet \frac{(k'' \alpha)^2}{3!\,2} \quad (24)$$

Substituting $\zeta = \zeta_o + \alpha \sin \omega t$ into the expression for $\Delta S_{matched}$ and averaging over time, *i.e.*, making a DC measurement, shows that $2 P_o$ is effectively the same as $\Delta S_{matched}$.

The DC components $\Delta S_{matched}$ and P_o are unable to distinguish between signal due to focus error and extraneous signals (noise) induced by motion and vibration of the object or the camera, changes in

illumination, changes in thermal dark current in the sensor, electronic noise in the detection system, etc. In contrast the AC signals from individual pixels have several properties that make them potentially immune to fluctuations and noise, and thus sensitive to focus with a high signal-to-noise ratio:

- *ratiometric measurement*: the ratio of the first harmonic signal to the second harmonic signal is $\frac{-\zeta_o}{4\alpha}$, independent of illumination, optical system uncertainties, small motions, vibrations, fluctuations, etc;
- *zero crossing*: the first harmonic signal and the ratio signal have zero crossings at offset $\zeta_o = 0$, *i.e.,* at exact focus, and this is a desirable condition for detectability;
- *synchronous detection*: the first and second harmonic signals are phase-locked to the dither; synchronous detection methods[5] can thus cleanly extract these signals from noisy environments;
- *insensitivity to flicker noise*: modulation (dither) and AC detection move the measurement from near DC to a higher frequency regime in which flicker (or $\frac{1}{f}$) noise may be dramatically lower.

Hands-on experience in different but analogous problem domains [11] leads me to expect that these methods could yield an advantage of several orders-of-magnitude with even cursory attention to good engineering practice.

8 Extensions and Pitfalls

Extending the model to spherical optics looks straightforward, although the additional integration over an azimuthal coordinate may involve some messy intermediate algebra. I expect the result will look very similar to the one presented here, with the complication of a transverse filter term for the y-direction, signal proportional to the square of the lens aperture, and slightly different numerical coefficients.

Removing the paraxial ray approximations should be straightforward, although the more general results are usually regrettably less revealing of the intuitive physics and geometry.

Removing the smallness approximation on the arguments of the transverse and longitudinal spatial filter functions should similarly be straightforward. The result could bring some surprises, especially longitudinally, since the longitudinal scale distance $\frac{2\pi}{k''}$ is typically rather small.

By far the most important restriction to remove is the description of the object as a simple sinusoidal grating. When the object space is described as a Fourier integral over a spatial frequency continuum instead of as a single spatial frequency, will the effect be to wash out the structures I am counting on detecting, *i.e.*, as many optical interference effects are washed out when a monochromatic light source is replaced by a polychromatic light source, or will the pixelation act as *a matched spatial filter that selects exactly what is needed to focus on surface texture?* If the answer is washing out rather than selecting out, then my method will be useful only for artificially simple scenes.

---

[5]Also known in different implementations and contexts as lock-in amplification, phase sensitive amplification, and synchronous rectification, the methods are powerful for extracting small signals with *line* spectra from overpowering backgrounds with *continuum* spectra. The signal from a reference oscillator both modulates the source (in this case: dithers the sensor plane) and in effect dials in the center frequency of the detector input filter. Synchronous detection methods are robust in that they squeeze out continuum noise by extreme narrow-banding, but because of their inherent tracking ability they exhibit no line spectrum signal loss with reference oscillator frequency drift.

Experimental verification can be envisioned as real-time and direct, by electromechanically driving the sensor plane (or, more-or-less equivalently, the lens or even the focal length) at an audio frequency and analog parallel processing of the signals from a few pixels. High temporal bandwidth detectors, *e.g.,* photodiodes, would be desirable. Alternatively, an indirect, non-real-time equivalent would involve stepping the sensor plane a fraction of the longitudinal scale distance between successive frames from a conventional video camera, with after-the-fact digital analysis. The direct real-time approach is preferred: it could take full advantage of synchronous detection, whereas the indirect simulation, with little practical prospect for averaging over many cycles, could easily be disabled by fluctuation noise.

A final potential pitfall is that real photosensors do not necessarily stop all the light incident on them in effectively zero thickness. The sensor thickness is manifested as an averaging operation in the longitudinal direction that attenuates the signals developed by the method proposed. With some sensor types this attenuation-by-thickness might be a fatal flaw.

9 Acknowledgements

I find it hard to believe that this simple model and its straightforward exploitation could be news to the machine vision community. My colleagues Robert H. Thibadeau, Steven A. Shafer, Takeo Kanade and Raj Reddy, assure me that it is news to them, and so probably news to the community. I thank them for encouraging me to flesh out and submit for publication the early drafts I shared with them. However if further scrutiny shows all this to be old hat, I accept all responsibility for failing to discover it.

10 References

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

2. Paul J. Besl. Range Imaging Sensors. Research Publication GMR-6090, General Motors Research Laboratories, March, 1988. Submitted for publication in the book *Advances in Machine Vision: Applications and Architectures*, J. Sanz, Editor, Springer-Verlag, New York.

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

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

5. Francis A. Jenkins and Harvey E. White. *Fundamentals of Optics.* McGraw-Hill Book Company, New York, 1957.

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

7. E. Krotkov. "Focusing". *International Journal of Computer Vision 1*, 3 (October 1987), 223-37.

8. John C. Fisher. Lock in the Devil, Educe Him, or Take Him for the Last Ride in a Boxcar? Princeton Applied Research Corporation, Tek Talk, Volume 6, Number 1. Princeton, NJ 08540, around 1964.

9. Anon. Sonar Ranging Module. Polaroid Corporation, Cambridge MA, instruction manual for a ranging module.

10. Reuel A. Sherman. "Benefits to Vision Through Stereoscopic Films". *Journal of the SMPTE 61* (September 1953), 294-308.

11. M. W. Siegel. "Cross Sections for Production of O_3^+, O_2^+ and O^+ by Electron Impact Ionization of Ozone Between Threshold and 100 eV". *International Journal of Mass Spectrometry and Ion Physics 44* (1982), 19-36.

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

13. B. H. Wilcox, et al. A Vision System for the Mars Rover. Mobile Robots II, 1987.

Table of Contents