

A NEW SENSE FOR DEPTH OF FIELD

Alex P. Pentland

Artificial Intelligence Center, SRI International
333 Ravenswood Ave, Menlo Park, CA 94025
and
Center for the Study of Language and Information
Stanford University, Stanford CA 94038

ABSTRACT

One of the major unsolved problems in designing an autonomous agent [robot] that must function in a complex, moving environment is obtaining reliable, real-time depth information, preferably without the limitations of active scanners. Stereo remains computationally intensive and prone to severe errors, the use of motion information is still quite experimental, and autofocus schemes can measure depth at only one point at a time. We examine a novel source of depth information: focal gradients resulting from the limited depth of field inherent in most optical systems. We prove that this source of information can be used to make reliable depth maps of useful accuracy with relatively minimal computation. Experiments with realistic imagery show that measurement of these optical gradients can potentially provide depth information roughly comparable to stereo disparity or motion parallax, while avoiding image-to-image matching problems. A potentially real-time version of this algorithm is described.

I. INTRODUCTION

Our subjective impression is that we view our surroundings in sharp, clear focus. This impression is reinforced by the virtually universal photographic tradition** to make images that are everywhere in focus, i.e., that have infinite depth of field. Unfortunately, both this photographic tradition and our feeling of a sharply focused world seems to have lead vision researchers — in both human and machine vision — to largely ignore the fact that in biological systems the images that fall on the retina are typically quite badly focused everywhere except within the central fovea [1,2]. There is a *gradient* of focus, ranging from nearly perfect focus at the point of regard to almost complete blur at points on distant objects.

It is puzzling that biological visual systems first employ an optical system that produces a degraded image, and then go to great lengths to undo this blurring and present us with a subjective impression of sharp focus. This is especially peculiar because it is just as easy to start out with everything in perfect focus. Why, then, does Nature prefer to employ a lens system in which most of the image is blurred?

In this paper we report the finding that this gradient of focus inherent in biological and most other optical systems is a useful source of depth information, prove that these focal gradients may be used to recover a depth map (i.e., distances between viewer and points in the scene) by means of a few, simple transformations of the image, and that with additional computation the reliability of this depth information may be internally checked. This source of depth information (which differs markedly from that used in automatic focusing methods) has not previously been described in the human vision literature, and we have been unable to find any investigation of it in the somewhat more scattered machine vision literature. The performance of a practical technique has been demonstrated on realistic imagery, and an inexpensive, real-time version of the algorithm is described. Finally, we report experiments showing that people make significant use of this depth information.

This novel method of obtaining a depth map is important because there is currently no passive sensing method for obtaining depth information that is simultaneously fast enough, reliable enough, and produces a sufficiently dense depth map to support the requirements of a robot moving in a complex environment. Stereopsis, despite huge investment, remains computationally intensive and prone to severe errors, the use of motion information is still in an experimental stage, and autofocus schemes can measure depth at only one point at a time. We believe that this research, therefore, will prove a significant advance in solving the problem of real-time acquisition of reliable depth maps without the limitations inherent in active scanners (e.g., laser rangefinders).

II. THE FOCAL GRADIENT

Most biological lens systems are exactly focused* at only one distance along each radius from the lens into the scene. The locus of exactly focused points forms a doubly curved, approximately spherical surface in three-dimensional space. Only when objects in the scene intersect this surface is their image exactly in focus; objects distant from this surface of exact focus are blurred, an effect familiar to photographers as depth of field.

The amount of defocus or blurring depends solely on the distance to the surface of exact focus and the characteristics of the lens system; as the distance between the imaged point and the surface of exact focus increases, the imaged objects become progressively more defocused. If we could measure the amount of blurring at a given point in the image, therefore, it seems possible that we could use our knowledge of the parameters of the lens system to compute the distance to the corresponding point in the scene.

* This research was made possible in part by a grant from the Systems Development Foundation, and by a grant from the National Science Foundation, Grant No. DCR-83-12766, and by Defense Advanced Research Projects Agency contract no. MDA 903-83-C-0027

** A practice established in large part by Ansel Adams and others in the famous "f/64 Club"

* "Exact focus" is taken here to mean "has the minimum variance point spread function," the phrase "measurement of focus" is taken to mean "characterize the point spread function."

The distance D to an imaged point is related to the parameters of the lens system and the amount of defocus by the following equation, which is developed in the appendix.

$$D = \frac{Fv_0}{v_0 - F - \sigma f} \quad (1)$$

where v_0 is the distance between the lens and the image plane (e.g., the film location in a camera), f the f-number of the lens system, F the focal length of the lens system, and σ the spatial constant of the point spread function (i.e., the radius of the imaged point's "blur circle") which describes how an image point is blurred by the imaging optics. The point spread function may be usefully approximated by a two-dimensional Gaussian $G(r, \sigma)$ with a spatial constant σ and radial distance r . The validity of using a Gaussian to describe the point spread function is discussed in the appendix.

In most situations, the only unknown on the right-hand side of Equation (1) is σ , the point spread function's spatial parameter. Thus, we can use Equation (1) to solve for absolute distance given only that we can measure σ , i.e., the amount of blur at a particular image point.

Measurement of σ presents a problem, however, for the image data is the result of both the characteristics of the scene and those of the lens system. To disentangle these factors, we can either look for places in the image with known characteristics (e.g., sharp edges), or we can observe what happens when we change some aspect of the lens system. In the following discussion both of these two general strategies for measurement of σ are described: the use of sharp edges, and comparison across different aperture settings. Both approaches require only one view of the scene.

A. Using Sharp Discontinuities

Image data are determined both by scene characteristics and the properties of the lens system, e.g., how fast image intensity changes depends upon both how scene radiance changes and the diameter of the blur circle. If we are to measure blur circle, therefore, we must already know the scene's contribution to the image. At edges — sharp discontinuities in the image formation process — the rate of change we observe in the image is due primarily to the point spread function; because we can often recognize sharp discontinuities with some degree of confidence [3,4] we can use image data surrounding them to determine the focus. These observations lead to the following scheme for recovering the viewer-to-scene* distance at points of discontinuity.

Mathematical Details. To calculate the spatial constant of the point spread function we require a measure of the rate at which image intensity is changing; the wide-spread use of zero-crossings of the Laplacian to find edges [5] suggests using slope of the Laplacian across the zero-crossing as a measure of rate of change.

Consider a vertical step edge in the image of magnitude δ at position x_0 . In this case the values $C(x, y)$ resulting from the convolution of image intensities $I(x, y)$ with the Laplacian of a Gaussian $\nabla^2 G(r, \sigma)$ (as in [5]) have the form

$$\begin{aligned} C(x, y) &= \nabla^2 G(r, \sigma) \otimes I(x, y) \\ &= \int \int \nabla^2 G(\sqrt{(x-u)^2 + (y-v)^2}, \sigma) I(u, v) du dv \\ &= \delta(dG(x - x_0, \sigma)/dx) \end{aligned} \quad (2)$$

where $G(x - x_0, \sigma)$ is a one-dimensional Gaussian centered at point x_0 , and σ is the spatial constant of the point spread function at that point in the image. For such an edge the slope of the function $C(x, y)$ at the

*When the discontinuity is in depth, as at an occluding contour, the distance measured is to the nearer side of the discontinuity.

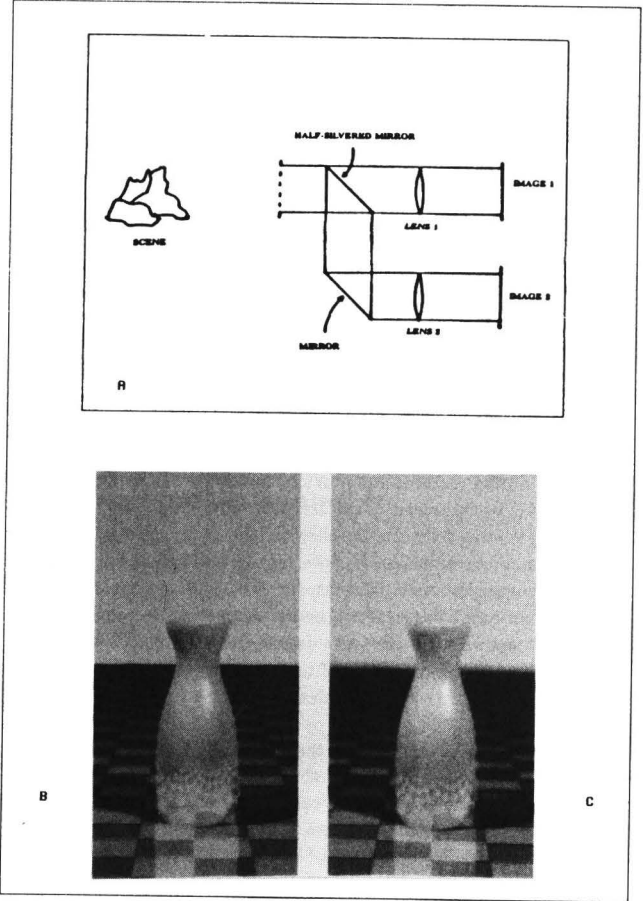


Figure 1. Images Identical Except for Depth of Field. (a) Production: The light from a single view is split into two identical images and directed through two lens systems with different aperture size. Alternatively, one can vary the aperture between alternate frames from a standard video or CCD camera. In either case the two resulting images are identical except for depth of field, as shown in Figure 1 (b) and (c). These images are of a mirrored bottle on a checkered plain.

point of the zero-crossing is equal to the maximum rate of change in image intensity, and so we can use it to estimate σ .

An estimate of σ can be formed as follows:

$$C(x, y) = \delta \frac{dG(x, \sigma)}{dx} = \frac{-\delta x}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3)$$

where x , y and δ are as before, and for convenience x_0 is taken to be zero. Taking the absolute value and then the natural log, we find

$$\ln \frac{\delta}{\sqrt{2\pi}\sigma^3} - \frac{x^2}{2\sigma^2} = \ln |C(x, y)| \quad (4)$$

We can formulate Equation (4) as

$$Ax^2 + B = C \quad (5)$$

where

$$A = -\frac{1}{2\sigma^2} \quad B = \ln \frac{\delta}{\sqrt{2\pi}\sigma^3} \quad C = \ln |C(x, y)|$$

If we interpret Equation (5) as a linear regression in x^2 we can then obtain a maximum-likelihood estimate of the constants A and B ,

and thus obtain σ . The solution of this linear regression is

$$A = \frac{\sum_i (x_i^2 - \bar{x}^2) C_i}{\sum_i (x_i^2 - \bar{x}^2)^2} \quad B = \bar{C} - \bar{x}^2 A \quad (6)$$

where $\bar{x}$ is the mean of the x_i , and $\bar{C}$ is the mean of the C_i . From A in Equation (6) we can obtain the following estimate of the value of the spatial constant σ :

$$\sigma = (-2A)^{-2}$$

Having estimated σ we can now use Equation (1) to find the distance to the imaged point; note that there are two solutions, one corresponding to a point in front of the locus of exact focus, the other corresponding to a point behind it. This ambiguity is generally unimportant because we can usually arrange things so that the surface of exact focus is nearer to the sensor than any of the objects in the field of view.

B. Comparison Across Differing Apertures

The limiting factor in the previous method is the requirement that we must know the scene characteristics before we can measure the focus; this restricts the applicability of the method to special points such as step discontinuities. If, however, we had two images of exactly the same scene, but with different depth of field, we could factor out the contribution of the scene to the two images (as the contribution is the same), and measure the focus directly.

Figure 1 shows one method of taking a single view of the scene and producing two images that are identical except for aperture size and therefore depth of field. This lens system uses a half-silvered mirror (or comparable contrivance) to split the original image into two identical images, which are then directed through lens systems with different aperture size. Because change in aperture does not affect the position of image features, the result is two images that are identical except* for their focal gradient (amount of depth of field), and so there is no difficulty in matching points in one image to points in the other. Figures 1 (b) and (c) show a pair of such images. Alternatively, one could rig a video or CCD camera so that alternate frames employ a different aperture; as long as no significant motion occurs between frames the result will again be two images identical except for depth of field.

Because differing aperture size causes differing focal gradients, the same point will be focused differently in the two images; for our purposes the critical fact is that the magnitude of this difference is a simple function of the distance between the viewer and the imaged point. To obtain an estimate of depth, therefore, we need only compare corresponding points in the two images and measure this change in focus. Because the two images are identical except for aperture size they may be compared directly; i.e., there is no matching problem as there is with stereo or motion algorithms. Thus we can then recover the absolute distance D by simple point-by-point comparison of the two images, as described below.

Mathematical Details. We start by taking a patch $f_1(r, \theta)$ centered at (x_0, y_0) within the first image $I_1(x, y)$:

$$f_1(r, \theta) = I_1(x_0 + r \cos \theta, y_0 + r \sin \theta)$$

and calculate its two-dimensional Fourier transform $\mathcal{F}_1(t, \theta)$. The same is done for a patch $f_2(r, \theta)$ at the corresponding point in the second image, giving us $\mathcal{F}_2(t, \theta)$. Again, note that there is no matching problem, as the images are identical except for depth of field.

Now consider the relation of f_1 to f_2 . Both cover the same region in the image, so that if there were no blurring both would be equal to the same intensity function $f_0(r, \theta)$. However, because there is blurring

(with spatial constants σ_1 and σ_2), we have

$$\frac{f_1(r, \theta)}{f_2(r, \theta)} = \frac{f_0(r, \theta) \otimes G(r, \sigma_1)}{f_0(r, \theta) \otimes G(r, \sigma_2)} \quad (7)$$

[One point of caution is that Equation (7) may be substantially in error in cases with a large amount of defocus, as points neighboring the patches f_1, f_2 will be "spread out" into the patches by differing amounts. This problem can be minimized by using patches whose edges trail off smoothly, e.g., $f_1(r, \theta) = I(x_0 + r \cos \theta, y_0 + r \sin \theta) G(r, \omega)$ for appropriate spatial parameter ω .]

Noting that

$$f(r, \theta) = e^{-\pi r^2} \quad \mathcal{F}(\lambda, \theta) = e^{-\pi \lambda^2}$$

are a Fourier pair and that if $f(r, \theta)$ and $\mathcal{F}(\lambda, \theta)$ are a Fourier pair then so are

$$f(\alpha r, \theta) \quad \frac{1}{|\alpha|} \mathcal{F}\left(\frac{\lambda}{\alpha}, \theta\right)$$

we see that we may use Equation (7) to derive the following relationship between $\mathcal{F}_1$ and $\mathcal{F}_2$ (the Fourier transforms of image patches f_1 and f_2) and $\mathcal{F}_0$ (the transform of the [hypothetical] unblurred image patch f_0):

$$\mathcal{F}_1(\lambda, \theta) = \frac{\mathcal{F}_0(\lambda, \theta) G(\lambda, \frac{1}{\sqrt{2\pi}\sigma_1})}{\sqrt{2\pi}\sigma_1} \quad \mathcal{F}_2(\lambda, \theta) = \frac{\mathcal{F}_0(\lambda, \theta) G(\lambda, \frac{1}{\sqrt{2\pi}\sigma_2})}{\sqrt{2\pi}\sigma_2} \quad (8)$$

Thus*,

$$\frac{\mathcal{F}_1(\lambda)}{\mathcal{F}_2(\lambda)} = \frac{G(\lambda, \sigma_1)\sigma_2}{G(\lambda, \sigma_2)\sigma_1} = \frac{\sigma_2^2}{\sigma_1^2} \exp(\lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2)) \quad (9)$$

where

$$\mathcal{F}(\lambda) = \int_{-\pi}^{\pi} \mathcal{F}(\lambda, \theta) d\theta$$

Thus, given $\mathcal{F}_1$ and $\mathcal{F}_2$ we can find σ_1 and σ_2 , as follows. Taking the natural log of Equation (9) we obtain

$$\ln \frac{\sigma_2^2}{\sigma_1^2} + \lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2) = \ln \mathcal{F}_1(\lambda) - \ln \mathcal{F}_2(\lambda)$$

We may formulate this as $A\lambda^2 + B = C$ where

$$A = 2\pi^2(\sigma_2^2 - \sigma_1^2) \quad B = \ln \frac{\sigma_2^2}{\sigma_1^2} \quad C = \ln \mathcal{F}_1(\lambda) - \ln \mathcal{F}_2(\lambda)$$

i.e., as a linear regression equation in λ^2 . The solution to this regression equation is the same as shown in the last example, and gives us maximum-likelihood estimates of A and B . Solving A and B for σ_1 and σ_2 yields

$$\sigma_1 = \sqrt{\frac{A}{2\pi^2(e^B - 1)}} \quad \sigma_2 = \sqrt{\frac{Ae^B}{2\pi^2(e^B - 1)}} \quad (10)$$

We may now use these estimates of σ_1 and σ_2 to calculate absolute distance to the imaged surface patch. Using Equation (1) for each of the two images, we see that we now have

$$D = \frac{Fv_0}{v_0 - F - \sigma_1 f_1} \quad D = \frac{Fv_0}{v_0 - F - \sigma_2 f_2} \quad (11)$$

where f_1 and f_2 are the f -numbers for the two halves of the imaging system.

*Note that we need only consider the amplitude of the transforms in these calculations.

*Their overall brightness might also differ.

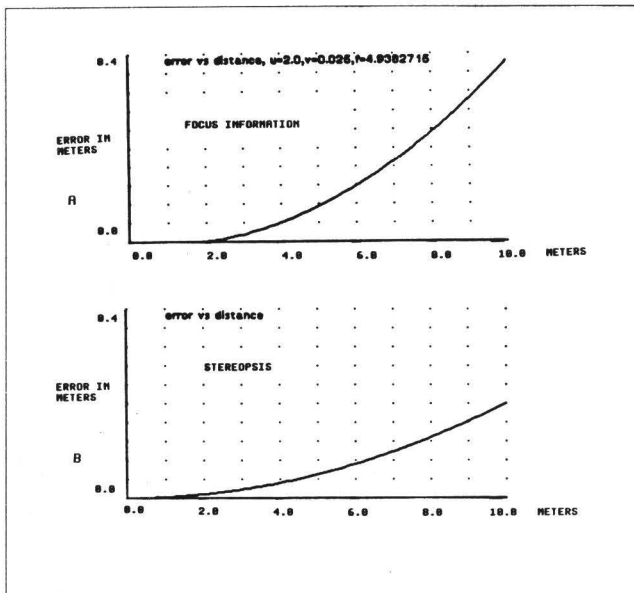


Figure 2. Accuracy at estimating distance, assuming human visual system parameters, using (a) focal gradient information, and (b) stereopsis.

C. Checking the answer: overconstraint

We may solve either of the two equations in (11) for D , the distance to the imaged surface patch. Thus the solution is overconstrained; both solutions must produce the same estimate of distance—otherwise the estimates of σ_1 and σ_2 must be in error. This can occur, for instance, when there is insufficient high-frequency information in the image patch to enable the change in focus to be calculated. The important point is that this overconstraint allows us to check our answer: if the equations disagree, then we know not to trust our answer. If, on the other hand, both equations agree then we can know (to within measurement error) that our answer *must* be correct.

D. Accuracy

Possibly the major question concerning the usefulness of focal gradient information is whether such information can be sufficiently accurate. There are two major issues to be addressed: first, can we estimate the variance σ of the point spread function with sufficient accuracy, and second, does this translate into a reasonable degree of accuracy in the estimation of depth.

Recent research aimed at estimating the point spread function has shown that it may be accurately recovered from unfamiliar images despite the presence of normal image noise [6,7]. Further, it appears that humans can estimate the width of the point spread function to within a few percent [8,9]. These findings, together with the results of estimating σ reported in the next section, show that accurate estimation of σ is practical given sufficient image resolution.

The second issue is whether the available accuracy at estimating σ translates into a reasonable accuracy in estimating depth. Figure 2 (a) show the theoretical error curve for the human eye, assuming the accuracy at estimating σ reported in [4]. It can be seen that reasonable accuracy is available out to several meters. This curve should be compared to the accuracy curve for stereopsis, shown in Figure 2 (b), again assuming human parameters. It can be seen that the accuracies are comparable.

E. Human Perception

We have recently reported evidence demonstrating that people make use of the depth information contained in focal gradients [9]; interestingly, the ecological salience of this optical gradient does not appear to have been previously reported in the scientific literature. The hypothesis that the human visual system makes significant use of this cue to depth has been investigated in two experiments.

In the first experiment, pictures of naturalistic scenes were presented with various magnitude of focal gradient information. It was found that increasing the magnitude of the focal gradient results in increasing subjective depth. In the second experiment, subjects were shown a rightward rotating wireframe (Nekker) cube displayed in perspective on a CRT. Such a display may be perceived as either as a rigid object rotating to the right, or (surprisingly) as wobbling, non-rigid object rotating to the left. Normally subjects see the rigid interpretations most of the time, but when we introduced a focal gradient that favored the non-rigid interpretations, the non-rigid interpretations was seen almost as often as the rigid one.

An experiment demonstrating the importance of depth of field in human perception can be easily performed by the reader. First make a pinhole camera by poking a small, clean hole through a piece of stiff paper or metal. Imposition of a pinhole in the line of sight causes the depth of field to be very large, thus effectively removing this depth cue from the image. Close one eye and view the world through the pinhole, holding it as close as possible to the surface of your eye, and note your impression of depth (for those of you with glasses, things will look sharper if you are doing it correctly). Now quickly remove the pinhole and view the world normally (still using only one eye). The change in the sense of depth is remarkable; many observers report that the change is nearly comparable to the difference between monocular and binocular viewing, or the change which occurs when a stationary object begins to move.

III. IMPLEMENTATION AND EVALUATION

A. Using sharp edges

The first method of deriving depth from the focal gradient, by measuring apparent blur near sharp discontinuities, was implemented in a straightforward manner (convolution values near zero-crossings were employed in Equations (4) - (6)) and evaluated on the image shown in Figure 3. In this image the optical system had a smaller depth of field than is currently typical in vision research; this was done because the algorithm requires that the digitization adequately resolve the point spread function.

Figure 3 also shows the depth estimates which were obtained when the algorithm was applied to this image. Part (a) of this Figure 3 shows all the sharp discontinuities identified [2]. It was found that there was considerable variability in the depth estimates obtained along these contours, perhaps resulting from the substantial noise (3 of 8 bits) which was present in the digitized image values. To minimize this variability the zero-crossing contours were segmented at points of high curvature, and the depth values were averaged within the zero-crossing segments. Figures 3 (b), (c), and (d) show the zero-crossing segments that have large, medium, and small depth values, respectively. It can be seen that the image is properly segmented with respect to depth, with the exception of one small segment near the top of (c). This example demonstrates that this depth estimation technique — which requires little computation beyond the calculation of zero-crossings — can be employed to order sharp edges by their depth values.

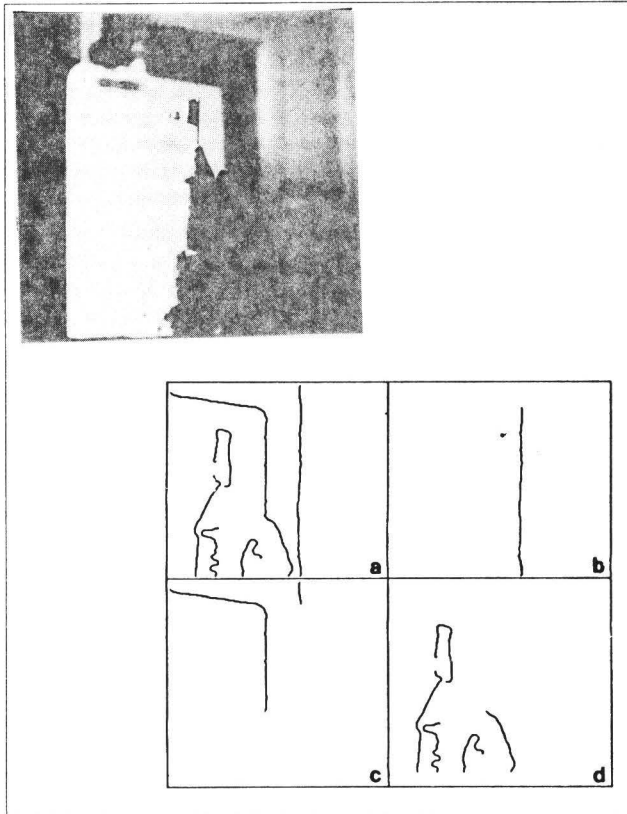


Figure 3. An Indoor Image of a Sand Castle, Refrigerator, and Door, Together with Depth Estimates for its Zero-Crossing Segments. Part (a) of this figure shows all the sharp discontinuities found. Parts (b), (c), and (d) show the zero-crossing segments that have large, medium, and small depth values, respectively. It can be seen that the image is properly segmented with respect to depth, with the exception of one small segment near the top of (c).

B. Comparison of different apertures

The second technique, comparing two images identical except for aperture, can be implemented in many different ways. We will report a very simple version of the algorithm that is amenable to an inexpensive real-time implementation.

In this algorithm two images are acquired as shown in Figure 1 (a); they are identical except for their depth of field and thus the amount of focal gradient present, as shown in Figures 1 (b) and (c). These images are then convolved with a small Laplacian filter, providing an estimate of their local high-frequency content. The output of the Laplacian filters are then summed over a small area and normalized by dividing them by the mean local image brightness, obtained by convolving the original images with a Gaussian filter. It appears that a region as small as 4×4 pixels is sufficient to obtain stable estimates of high-frequency content. Figures 4 (a) and (b) show the normalized high-frequency content of Figures 1 (b) and (c).

Finally, the estimated high-frequency content of the blurry, large-aperture image is divided by that of the sharp, small-aperture image, i.e., each point of Figure 4 (a) is divided by the corresponding point in Figure 4(b). This produces a "focal disparity" map, analogous to a stereo disparity map, that measures the change in focus between the two images and whose values are monotonically related to depth by Equation (1). Figure 4 (c) shows the disparity map produced from Figures 2 (b) and 2 (c); intensity in this figure is proportional to depth.

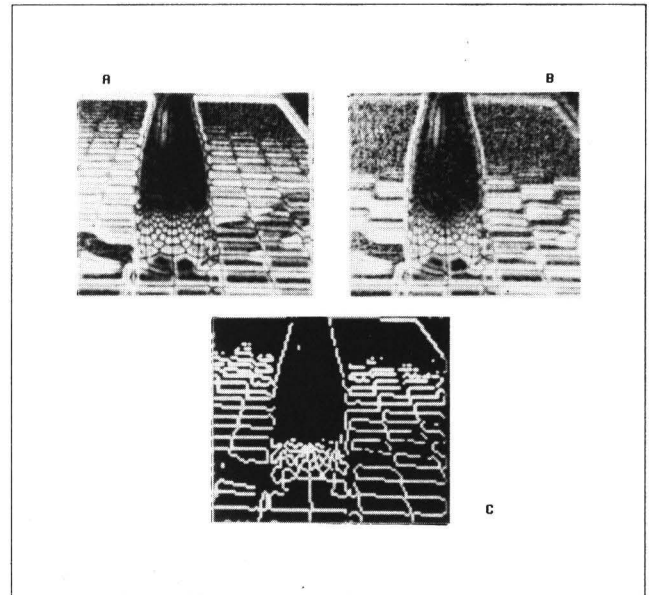


Figure 4. (a) and (b) show the normalized high-frequency content of Figures 2 (b) and (c), respectively. (c) shows the focal disparity map (analogous to a stereo disparity map) obtained by comparing (a) and (b); brightness is proportional to depth.

Areas of 4 (c) that are black have insufficient high-frequency energy in the sharp-focus image to make an estimate of depth.

It can be seen that this disparity map is fairly accurate. Note that points reflected in the bottle are estimated as further than points along the edge of the bottle; this is not a mistake, for these points the distance traveled by the light is further than for those along the edge of the bottle. This algorithm, in common with stereo and motion algorithms, does not "know" about mirrored surfaces.

C. Design for a real-time implementation

A minimum of one convolution per image is required for this technique, together with a left shift and four subtractions for the Laplacian, and three divides for the normalization and comparison. If special convolution hardware is available, one can use two convolutions — one Laplacian and one Gaussian — per image, leaving only three divides* for the normalization and comparison. Frame buffers that can convolve image data in parallel with image acquisition are now available at a reasonable price, leaving as few as 3 operations per pixel to calculate the disparity map. For a 256×256 image, this can be accomplished in as little as 0.35 seconds with currently available microcomputers.

IV. DISCUSSION

The most striking aspect of this source of depth information is that absolute range can be estimated from a single view with no image-to-image matching problem, perhaps the major source of error in stereo and motion algorithms. Furthermore, no special scene characteristics need be assumed, so that the techniques utilizing this cue to depth can be generally applicable. The second most striking fact is the simplicity of these algorithms: it appears that a real-time implementation can be accomplished relatively cheaply.

Measurement of the focal gradients associated with limited depth of field appears to be capable of producing depth estimates that are at least roughly comparable to edge- or feature-based stereo and motion

* which can be done by table lookup.

algorithms. The mathematics of the aperture-comparison technique shows it to be potentially more reliable than stereo or motion — i.e., there is no correspondence problem, and one can obtain an internal check on the answer — although (as discussed above) it has somewhat less accuracy.

The sharp-edge algorithm appears to have potential for useful depth-plane segmentation, although it is probably not accurate enough to produce a depth map. I believe that this algorithm will be of some interest because most of the work — finding and measuring the slope of zero-crossings — is often already being done for other purposes. Thus this type of depth-plane segmentation can be done almost as a side effect of edge finding or other operations.

The aperture-comparison algorithm provides considerably stronger information about the scene because it overconstrains scene depth, allowing an internal check on the algorithm's answer. Thus it provides depth information with a reliability comparable to the best that is theoretically available from three-or-more image stereo and motion algorithms, although it has somewhat less depth resolution. The major limitation in measuring focal gradient depth information in this manner appears to be insuring sufficient high-frequency information to measure the change between images; this requires having both adequate image resolution and high-frequency scene content.

Summary. In summary, we have described a new source of depth information — the focal gradient — that can provide depth information at least roughly comparable to stereo disparity or motion parallax, while avoiding the image-to-image matching problems that have made stereo and motion algorithms unreliable. We have shown that the limited depth of field inherent in most optical systems can be used to make depth maps of useful accuracy with relatively minimal computation, and have successfully demonstrated a potentially real-time technique for recovering depth maps from realistic imagery. It is our hope, therefore, that this research will prove to be a substantial advance towards building a robot that can function in complex, moving natural environments.

REFERENCES

- [1] H. Crane, *A Theoretical Analysis of the Visual Accommodation System in Humans*, Final Report NAS 2-2760, NASA Ames Research Center (1966).
- [2] M. Born and E. Wolf, *Principles of Optics*, Pergamon, London (1965).
- [3] A. Pentland, *The Visual Inference of Shape: Computation from Local Features*, Ph.D. thesis, Massachusetts Institute of Technology (1982).
- [4] A. Witkin, *Intensity-Based Edge Classification*, Proceedings of the American Association for Artificial Intelligence, August 1982, Pittsburgh, Penn.
- [5] E. Hildreth, *Implementation of a Theory of Edge Detection*, M.I.T. AI Laboratory Technical Report 579 (April 1980).
- [6] K.T. Knox and B.J. Thomson, *Recovery of Images from Atmospherically Degraded Short-Exposure Photographs*, *Astrophys. J.* 193, L45-L48 (1974).
- [7] J.B. Morton and H.C. Andrews, *A Posteriori Method of Image Restoration*, *Opt. Soc. Am.* 69, 2 (1979) 280-290.
- [8] A. Pentland, *Uniform Extrafoveal Sensitivity To Pattern Differences*, *Journal of the Optical Society of America*, November 1978.
- [9] A. Pentland, *The Focal Gradient: Optics Ecologically Salient*, Supplement to Investigative Ophthalmology and Visual Science, April 1985.

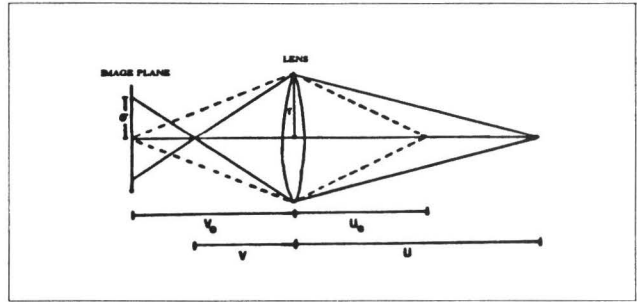


Figure 5. Geometry of Imaging. v_0 is the distance between the image plane and the lens, u_0 is the distance between the lens and the locus of perfect focus, and r is the radius of the lens. When a point at distance $u > u_0$ is projected through the lens, it focuses at a distance $v < v_0$, so that a blur circle is formed.

APPENDIX

For a thin lens,

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{F} \quad (12)$$

where u is the distance between a point in the scene and the lens, v the distance between the lens and the plane on which the image is in perfect focus, and F the focal length of the lens. Thus,

$$u = \frac{Fv}{v - F} \quad (13)$$

For a particular lens, F is a constant. If we then fix the distance v between the lens and the image plane to the value $v = v_0$, we have also determined a locus of points at distance $u = u_0$ that will be in perfect focus, i.e.,

$$u_0 = \frac{Fv_0}{v_0 - F} \quad (14)$$

We may now explore what happens when a point at a distance $u > u_0$ is imaged. Figure 5 shows the situation in which a lens of radius r is used to project a point at distance u onto an image plane at distance v_0 behind the lens. Given this configuration, the point would be focused at distance v behind the lens—but in front of the image plane. Thus, a blur circle is formed on the image plane. Note that a point at distance $u < u_0$ also forms a blur circle; throughout this paper we assume that the lens system is focused on the nearest point so that u is always greater than u_0 . This restriction is not necessary in the second algorithm, as overconstraint on the distance solution allows determination of whether $D = u > u_0$ or $D = u < u_0$.

From the geometry of Figure 5 we see that

$$\tan \theta = \frac{r}{v} = \frac{\sigma}{v_0 - v} \quad (15)$$

Combining Equations (13) and (15) and substituting the distance D for the variable u we obtain

$$D = \frac{Fr v_0}{r v_0 - F(r + \sigma)}$$

or

$$D = \frac{F v_0}{v_0 - F - \sigma f}$$

where f is the f-number of the lens.

The blurring of the image is better described by the point spread function than by a blur circle, although the blurring is bounded by the blur circle radius in the sense that the point spread function is less than some threshold outside of the blur circle. The point spread function is due primarily to diffraction effects, which for any particular

wavelength produce wave cancellation and reinforcement resulting in intensity patterns qualitatively similar to the sinc function, $\frac{\sin x}{x}$, but with different amplitudes and periods for the "rings" around the central peak [2].

The "rings" produced by this function vary in amplitude, width and position with different states of focus and with different wavelengths. As wavelength varies these rings change position by as much as 90 degrees, so that the blue light troughs become positioned over the red light peaks, etc. Further, change in wavelength results in substantial changes in the amplitude of the various rings. Although this point spread function is quite complex, and the sum over different wavelengths even more so, our analysis shows that for white light the sum of the various functions obtained at different wavelengths has the general shape of a two-dimensional Gaussian.

Sampling effects caused by digitization are typically next in importance after the diffraction effects. The effect of sampling may be accounted for in the point spread function by convolving the above diffraction-produced point spread function with functions of the form $\frac{\sin x}{x}$. Other factors such as chromatic aberration, movement, and diffusion of photographic emulsion may also be accounted for in the final point spread function by additional convolutions.

The net effect, in light of the central limit theorem and our analysis of the sum of single-wavelength focus patterns, is almost certainly best described by a two-dimensional Gaussian $G(r, \sigma)$ with spatial constant σ . The spatial constant σ of the point spread function will be proportional to the radius of the blur circle; however, the constant of proportionality will depend on the particulars of the optics, sampling, etc. In this paper the radius of the blur circle and the spatial constant of the point spread function have been treated as identical; in practical application where recovery of absolute distance is desired the constant of proportionality k must be determined for the system and included in Equation (1) as follows:

$$D = \frac{F v_0}{v_0 - F - \sigma k f}$$