Paper submitted to
IEEE Computer Society Workshop on
COMPUTER VISION
Miami Beach, Florida
November 30 - December 2, 1987.

Type of paper: Regular

Direct Recovery of Depth-map I: Differential Methods

Muralidhara Subbarao
Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, NY 11794

Index terms: 3-D from 2-D, depth-map recovery, point spread function of a lens.

Direct Recovery of Depth-map I: Differential Methods

Abstract

Three new methods are described for recovering the *depth-map* (i.e. distance of visible surfaces from a camera) of a scene from images formed by a convex lense. The recovery is based on observing the change in the scene's image due to a *small* known change in one of the three intrinsic camera parameters: (i) distance between the lens and the image detector plane, (ii) focal length of the lens, and (iii) diameter of the lens aperture. No assumptions are made about the scene being analyzed. The recovery process is *parallel* involving simple *local computations.* In comparision with some shape recovery processes such as stereo vision and motion analysis, the methods are *direct* in the sense that three-dimensional scene geometry is recovered directly from intensity images of the scene; spatial properties of the intensity distribution (e.g. the *raw primal sketch* described by Marr, 1982) are not computed as an intermediate step, and further the *correspondence* problem does not arise. These methods are relevant to both machine vision and human vision.

1. Introduction

1.1 Lens based inverse-optics is a well-posed problem

One of the early goals of a visual system is to recover the three-dimensional geometry of scenes. In the area of computer vision most of the research for recovering the scene geometry is based on a a pin-hole camera model (e.g.: Ballard and Brown, 1982; Rosenfeld and Kak, 1982; Horn, 1986). The image of a pin-hole camera completely lacks information about the distance of visible surfaces in the scene along the viewing direction. Therefore any analysis based on a pin-hole camera model has to use heuristic assumptions about the scenes to be analyzed. For example, in the shape-from-

shading process, assumptions are made about the reflectance and geometric properties of visible surfaces (e.g.: the Lambertian reflectance model and ‘‘smoothness’’ of surface structure). Practical camera systems, including the human eye, are not pin-hole cameras but consist of convex lenses. The image of a convex lens, in contrast to the image formed by a pin-hole, has complete information about scene geometry. The position of a point in the scene and the position of its *focused image* are related by the *lens formula*

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{1}$$

where f is the focal length, u is the distance of the object from the lens plane, and v is the distance of the focused image from the lens plane (see Figure 1). (Informally, an image is in focus if it is ‘‘sharp’’.) Given the position of the focused image of a point, its position in the scene is uniquely determined. In fact the position a point-object and its image are *interchangeable*, i.e. the image of the image is the object itself. Now, if we think of the visible surfaces in a scene to be comprised of a set of points, then the focused images of these points define another surface behind the lens. We can think of this surface and the intensity distribution on it as the *focused image* of the scene. The geometry of visible surfaces in the scene and the geometry of the surface defined by the focused image have a *one to one correspondence* defined by the lens formula (1). Therefore, for a convex-lens camera, the stage of *early vision* which is often defined to be *inverse optics* (Poggio, Torre, and Koch, 1985) is a *well-posed problem,* though, perhaps, *ill-conditioned.* We find it surprising that sufficient attention has not been paid to this source of depth information in computer vision research.

1.2 Three new methods for depth-map recovery

Recovering the depth-map of a scene is an important task in robot vision. Here we describe three new methods for depth-map recovery using a convex-lens-camera. The methods are based on mesuring the change in an image due to a *small* known change in one of the three intrinsic camera parameters: (i) distance between the image detector plane and the lens, (ii) focal length of the lens, and (iii) diameter of the lens aperture. In

Subbarao (1987a), a fourth method is described for recovering depth-map by moving the entire camera (lens and the image detector plane) along the optical axis. This method requires solving the same *correspondence* problem encountered in stereo vision and motion analysis. Since there exist only heuristic solutions to the correspondence problem, we choose not to describe this method here.

In the past, the lens formula (1) has been used for finding the distance of objects whose images are in focus (Horn, 1968; Jarvis, 1983). Many approaches exist for focusing a given part of an image. These approaches are primarily found in the autofocusing literature for cameras and microscopes (e.g.: Ligthart and Groen, 1982; Schlag *et al,* 1983; Krotkov,1986). In this method, the camera parameter setting for focusing an object is different for objects at different distances in the scene. Therefore this method is *sequential.* The depth along a given direction can also be obtained by some *active ranging* device such as a sonar or laser range finder. In this method the scene is scanned *sequentially* along different viewing directions to obtain a complete depth-map. In comparison with these two methods, the methods described here can obtain the depth map of the *entire scene at once, irrespective of whether any part of the image is in focus or not.* The depth-map recovery process is *parallel* and involves only simple *local computations.* In comparision with some shape recovery processes such as stereo vision and motion analysis, the methods are *direct* in the sense that three-dimensional scene geometry is recovered directly from intensity images of the scene; spatial properties of the intensity distribution (e.g. the *raw primal sketch* described by Marr, 1982) are not computed as an intermediate step, and further the *correspondence* problem does not arise.

In the approach described here, no assumptions are made about the scene being analyzed. The only requirement is that we ought to know the camera parameters and camera characteristics beforehand. This information can be acquired once and for all intially by a suitable camera calibration procedure. For the purpose of mathematical analysis, we have taken the point-spread-function of the camera to be a Gaussian function. Some justification for this choice is provided later. This choice has also been advocated by others (e.g.: Horn, 1986; Pentland, 1987). However, the methods and ideas of

the analysis presented here can be extended to other point spread functions. Therefore the significance of this work is perhaps not in the actual equations derived, but in the demonstration that given the point spread of the camera, a method can be devised to obtain depth-map.

A limitation of the methods described here is that the depth-maps obtained are not exact, but have some (usually negligible) uncertainty associated with them. This limitation arises because of image-overlap that occurs in blurred images (a sort of ‘‘aliasing’’). This problem is discussed later.

2. Previous work

Pentland (1982,’85,’87) was perhaps the first person to investigate depth-map recovery in parallel from images formed by a lens. Apart from his work, there is very little previous literature on this problem. Pentland (1987) says:

> ‘‘Surprisingly, the idea of using focal gradients to infer depth appears to have never been investigated (several authors have, however, mentioned the theoretical possibility of such information): we have been unable to discover any investigation of it in either the human vision literature or in the somewhat more scattered machine vision literature.’’

Pentland proposed two methods for finding the depth-map of a scene. The first method was based on measuring the ‘‘blur’’ (or slope) of edges which are step discontinuities in the focused image. Recently Grossman (1987) has reported the results of some experiments based on this same principle. Pentland tested his method on a natural scene and showed that edges could be classified qualitatively as having small, medium, or large depth values. This method requires the knowledge of the location and magnitude of step edges in the focused image. This information is rarely available in practical situations and therefore this method is not our main concern here.

Pentland’s second method which is of primary concern here is based on comparing two images formed with different aperture diameter settings. Pentland (1985)

demonstrated that two views could be used to obtain a depth-map of a scene. However, an algebraic error in his derivation lead Pentland to the incorrect conclusion that his method could apply to any two aperture settings. He later corrected the algebraic error (Pentland, 1987) and found that solution could be obtained only if one of the two aperture setting had near-zero diameter.

In deriving his two methods Pentland (1987) has used a point spread function for the lens whose ''volume'' is not unity but is dependent on the spread parameter of the point spread function (the volume is $\sqrt{2\pi}\sigma$ where σ is the spread parameter). For actual lenses the volume is unity (Horn, 1986). In particular, the volume sholud be independent of the spread parameter because the spread parameter itself depends on distances of objects in the scene. Therefore Pentland's both methods need to be rederived using the correct point spread function. However the final equations derived by Pentland (1987) in both his methods are correct; an error in the derivation resulted in correct equations although the point spread function was incorrect.

The main ideas in this paper were developed independently by the author and reported in Subbarao (1987a). These ideas have been further developed (Subbarao, 1987b,c) to obtain robust methods for recovering both *shape and motion* of objects in a scene. Our work shows that Pentland's second method is only one of a class of possible methods to obtain depth-map by comparing images formed by different camera parameter settings. This paper describes only three methods where the change in the camera parameters is restricted to be *small.* These methods have been extended for the case of *large* change in camera parameters in Subbarao (1987b). One of our method based on changing the diameter of the lens aperture is a more general version of Pentland's second method. In this method an additional constraint is derived for the unknowns so that Pentland's (1987) requirement of at least one image formed by a pin-hole camera is removed.

The experimental results reported by Pentland (1987) and Grossman (1987), (and our own crude preliminary experiments) indicate that our approach can provide very useful information in practical applications. Rigorous experimental evaluation of our

approach has been delayed due to the unavailability of a specialised camera system whose parameter setting can be controlled precisely.

3. Point spread function of a convex lens

In this section, first we derive an expression for the point-spread function of a lens based on geometric considerations, and then we modify it to take into account other factors such as *diffraction* and lens aberrations. The references Goodman (1968), Horn (1968), Pentland (1987), and Horn (1986) together contain most of the discussion in this section.

Let P be a point on a visible surface in the scene and p be its focused image (see Figure 1). If P is not in focus then it gives rise to a circular image on the image detector plane. In this case we call the circular patch the *blurred image* of p. From simple plane geometry (see Figure 1) and equation (1) we can show that the diameter d of the circular image is given by

$$d = D\left[s\left(\frac{1}{f}-\frac{1}{u}\right)-1\right] \tag{2}$$

where s is the distance of the image detector plane from the lens and D is the diameter of the lens. Note that d can be either positive or negative depending on whether $s \geq v$ or $s < v$. In the former case the image detector plane is *behind* the focused image of P and in the latter case it is *in front* of the focused image of P. The intensity within the circular patch is approximately constant and is proportional to the intensity of the focused image at p. Therefore the blurred image of P can be thought of as the result of convolving its focused image with a *point spread function* $h_1(x,y)$ where

$$h_1(x,y) = \begin{cases} \dfrac{4}{\pi d^2} & \text{if } x^2+y^2 \leq \dfrac{d^2}{4} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

This function has the form of a ‘‘pill-box’’ shown in Figure 2. Therefore, assuming the camera to be a *linear shift-invariant system,* an image acquired by the camera system can

be thought of as the result of convolving a focused image with a point spread function $h_1(x,y)$. The *focused image* of a scene for a given position of the image detector plane can be defined as follows. For any point (x,y) on the image consider a line through that point and the optical center. Let Q be the point where this line intersects a visible surface in the scene and let q be its focused image. Then the intensity at (x,y) is the intensity of the focused image at q.

Note that $h_1(x,y)$ is defined in terms of d and therefore has different spread parameter for points at different distances from the lens plane.

The form of the point spread function derived above is based purely on geometric considerations. For practical camera systems various other effects come into picture. Ignoring lens aberrations, the primary source of distortion is due to *diffraction* caused by the wave nature of light. For *coherent* monochromatic illumination, the effect of diffraction is to produce a ripple-like intensity pattern qualitatively similar to the square of the sinc function: $\sin^2 x/x^2$. (The light from objects which subtend a very small angle at the optical center of the lens is mostly coherent; otherwise it is usually incoherent.) The actual expression for the intensity pattern is $\left[\frac{2J_1(R\rho)}{R\rho}\right]^2$ where J_1 is Bessel-function of order 1, R is the radius of the blur circle, and ρ is the frequency in radians per unit distance. A cross section of this intensity pattern is shown in Figure 3 (the pattern is circularly symmetric). The amplitude, frequency, and the position of ripples in the intensity pattern are dependent on the wave length of light. The corresponding optical transfer function (i.e. the Fourier transform of the point-spread function) has the form of a pill-box with a diameter of $D/\lambda v$ in the focal plane of the image. For *incoherent* monochromatic illumination, the optical transfer function is given in Goodman (1968) to be

$$H(\rho) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}\left(\frac{\rho}{2\rho_0}\right) - \left(\frac{\rho}{2\rho_0}\right)\sqrt{1-\left(\frac{\rho}{2\rho_0}\right)^2}\right] & \rho \le 2\rho_0 \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The form of this function is shown in Figure 4. In the above equation, ρ is the frequency variable and ρ_0 is the cutoff frequency of the coherent system,

$$\rho_0 = \frac{D}{2\lambda v}. \tag{5}$$

(The optical transfer function for an incoherent system extends to a frequency that is twice the coherent cutoff frequency). For white light the overall intensity pattern is due to the cumulative effect of intensity patterns produced by lights of many different wave lengths. Distortion is also caused by many other factors. Therefore, intuitively, the net effect is probably best described by a Gaussian point spread function whose spread parameter is proportional to the diameter of the blur circle. Therefore, we shall consider the point spread function of the camera to be of the form

$$h_2(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}} \tag{6}$$

where σ is the spread parameter such that

$$\sigma = k\, d \quad \text{for } 0<k\leq 0.5. \tag{7a}$$

The actual value of *k* is characteristic of a given camera which is determined by an appropriate calibration procedure. From equation (2), the above equation can be written as

$$\sigma = kD\left[s\left(\frac{1}{f}-\frac{1}{u}\right)-1\right]. \tag{7b}$$

Note that the ‘‘volume’’ of the function defined in equation (3) is unity; we can also show that the function defined by equation (6) has unit volume. The Fourier transform corresponding to equation (6) is

$$H(\omega,\nu) = e^{-\frac{1}{2}(\omega^2+\nu^2)\sigma^2}. \tag{8}$$

A cross section of the above function is shown in Figure 5 (the function is circularly symmetric). The form of this function appears to agree with a function obtained by summing and normalizing many functions of the form (4) (shown in Figure 4) for different

wavelengths of light. (The summing should be weighted according to the spectral density of different wave length components in the reflected light.)

4. Power spectral density and blur parameter

Let $g(x,y)$ be an image of a scene on the image detector plane, $f(x,y)$ be the corresponding focused image, and $h(x,y)$ be the point spread function of the camera. The camera is assumed to be a linear shift-invariant system (see Rosenfeld and Kak, 1982). Let $G(\omega,\nu)$, $F(\omega,\nu)$ and $H(\omega,\nu)$ be the corresponding Fourier transforms. We have

$$g = h*f \tag{9}$$

where * denotes the convolution operation. Recalling that convolution in the spatial domain is equivalent to multiplication in the Fourier domain, equation (9) can be expressed in the Fourier domain as

$$G = H\ F. \tag{10}$$

Therefore, the power spectral density $P(\omega,\nu)$ of G is

$$P(\omega,\nu) = G\ G^{*}. \tag{11}$$

Noting that $G^{*} = H^{*}\ F^{*}$, the above expression can be written as

$$P(\omega,\nu) = H\ H^{*}\ \ F\ F^{*}. \tag{12}$$

Assuming that H is as in equation (8), the power spectrum of a blurred image region is

$$P(\omega,\nu) = e^{-(\omega^2+\nu^2)\sigma^2}\ F\ F^{*}. \tag{13}$$

In the above equation, the blur parameter σ is different for objects in the scene at different distances from the camera. Therefore, in the following discussion we restrict our analysis to small image regions in which the blur parameter σ is approximately constant. This limits the resolution of the depth-map that can be obtained by this method. Further, an image region cannot be analyzed in isolation because, due to blurring (caused by the finite spread of the point-spread-function), the intensity at the border of the region is affected by the intensity immediately outside the region. We call this the *image overlap*

problem because the intensity distribution produced by adjecent patches of visible surfaces in the scene overlap on the image detector plane. See Subbarao (1987c) for more discussion of this problem. In order to reduce the image overlap problem, the image intensity is first weighted (or multiplied) by an appropriate two dimensional Gaussian function centered at the region of interest. The resulting weighted image is used for depth-map recovery. Because the weights are higher at the center than at the periphery, this scheme gives a depth estimate which is approximately the depth along the center of the field of view. Alternative methods of dealing with the image overlap problem are being considered.

Weighting an image by a Gaussian is equivalent to convolving the Fourier spectrum of the image by another Gaussian (with a very small spread parameter). The error introduced in the depth measurement by such a weighting scheme is under investigation.

5. Direct depth-map recovery

Theorem : Let the point spread function of a convex lens camera be given by equation (6). Then the spread parameter σ is related to the power spectral density $P(\omega,\nu)$ of the image by the following relation.

$$\frac{1}{P}\,\frac{dP}{d\sigma} = -2(\omega^2+\nu^2)\sigma\,. \tag{14}$$

Proof : From equation (13) we have

$$\frac{dP}{d\sigma} = -2(\omega^2+\nu^2)\sigma e^{-(\omega^2+\nu^2)\sigma^2}\; F\,F^*\,. \tag{15}$$

From equations (13) and (15) we get equation (14). ●

The blur parameter σ of an image region can be changed by changing one of the camera parameters: *s*, *f*, and *D* (see equation 7b). Corresponding to each of these paratemers we shall describe one method of obtaining the depth-map.

5.1 Depth map by changing the position of the image detector plane

Lemma 1 :

$$\sigma(\sigma+kD) = -\frac{1}{2}\frac{s}{\omega^2+\nu^2}\ \frac{1}{P}\frac{dP}{ds}\ . \tag{16}$$

Proof : We have

$$\frac{1}{P}\frac{dP}{ds} = \frac{1}{P}\frac{dP}{d\sigma}\ \frac{d\sigma}{ds}\ . \tag{17}$$

From equation (7b) we get

$$\frac{d\sigma}{ds} = \frac{\sigma+kD}{s}\ . \tag{18}$$

From equations (14,17,18) we can derive equation (16). •

The above lemma says that, if we know the power spectral density of a frequency component (ω,ν) and the change in it's power spectral density for a small displacement of the image detector plane, then σ can be determined by solving a quadratic equation. Let c_1 denote the right hand side quantity of equation (16). c is a constant for all frequency components (ω,ν) because the left hand side of equation (16) does not depend on (ω,ν). c_1 can be computed as the mean value over some region as below.

$$c_1 = -\frac{1}{2}\frac{s}{(\omega_2-\omega_1)(\nu_2-\nu_1)}\int_{\omega_1}^{\omega_2}\int_{\nu_1}^{\nu_2}\frac{1}{\omega^2+\nu^2}\ \frac{1}{P}\frac{dP}{ds}\ d\omega\, d\nu. \tag{19}$$

Equation (16) is quadratic in σ and therefore gives two solutions for σ. However we shall see that the two solutions have opposite signs and that the correct solution is determined by the sign of the quantity $\frac{dP}{ds}$.

Lemma 2 If σ_0 is a solution of equation (16) which corresponds to the correct physical interpretation, then the two solutions of equation (16) are

$$\sigma_0,\ -(\sigma_0+kD). \tag{20}$$

Proof : Equation (16) can be written as

$$\sigma(\sigma+kD) = c_1. \tag{21}$$

Since σ_0 is a solution of the above equation, we have

$$\sigma_0(\sigma_0+kD) = c_1. \tag{22}$$

Therefore, from the above two relations we have

$$\sigma(\sigma+kD) = \sigma_0(\sigma_0+kD). \tag{23}$$

The roots of the above quadratic equation are as given in (20).

Lemma 3: If $\frac{dP}{ds} \leq 0$ then $s \geq v$ and $\sigma \geq 0$. If $\frac{dP}{ds} > 0$ then $s < v$ and $\sigma < 0$.

Proof : For a small increase in s the image blur increases (i.e. the size of the blur circle of a point increases) only if initially $s \geq v$ (see Figure 1). This implies that if $\frac{dP}{ds} \leq 0$ then $s \geq v$, i.e. the image detector plane is behind the focused image. In this case $\sigma \geq 0$. Similarly we can argue that when $\frac{dP}{ds} > 0$ then $s < v$ and $\sigma < 0$. This case corresponds to the the situation where the image detector plane is in front of the focused image. •

From the above two lemmas we see that (i) the sum of the roots of equation (16) is always $-kD$, (ii) if $dP/ds \leq 0$ then the unique positive root of equation (16) gives the correct σ, and (iii) if $dP/ds > 0$ then there can be up to two negative roots for equation (16) both of which are acceptable solutions. In the latter case it can be shown that both roots always satisfy the condition $|\sigma| \leq kD$. The degenerate case of $\sigma = -kD$ occurs when either the image detector plane coincides with the lens plane or when the object in the scene is at a distance equal to the focal length of the lens. To obtain a unique interpretation from the two solutions, we will have to use some additional information. For example, if we assume that the image is not blurred too much, say $|\sigma| \leq 0.5kD$, then a unique solution can be obtained.

Having determined σ, we can obtain the location u of the visible surface from relation (7b). In summary, the above lemmas state that if the intensity distribution in a small field of view is given for two image detector plane positions which are a small distance apart then the position of the visible surface in that field of view can be determined.

The position of the image of a point changes when the image detector plane is moved (see Figure 6). Therefore, to find the change in the power spectral density after the image detector plane is displaced, it will be necessary to know correspondence between image regions. This problem is discussed in the Appendix.

We could have based our analysis of a blurred image on its intensity function (or its Fourier transform) rather than its power spectral density. It is also possible to base the analysis on some other function of the image. We have chosen power spectral density because they have physical interpretations in signal processing.

5.1.1 Error sensitivity

Here we consider the error of the above method due to uncertainty in the measurement of the distance of image detector plane from the lens. From equation (1) we can get

$$\frac{|du|}{u} \leq \frac{u}{v}\frac{|dv|}{v} \tag{24}$$

We see that the error is a maximum when v is a minimum. The minimum value of v is f. For visible surfaces which are more than $5f$ distance away from the lens v is approximately equal to f. In this case the above formula can be approximated as

$$\frac{|du|}{u} \leq \frac{u}{f}\frac{|dv|}{f}. \tag{25}$$

Above we see that the percentage error in the estimated distance is proportional to the actual distance. If the distance between the lens and the image detector plane can be measured to an accuracy of f/n units then

$$\frac{|du|}{u} \leq \frac{1}{n}\frac{u}{f}. \tag{26}$$

Therefore, if f/n=0.001 units, we get a maximum of ten percent error for a surface which is at a distance of one hundred times the focal length.

Analysis of error sensitivity due to quantization of gray values and discrete sampling rate needs to be done in the future.

5.2 Depth map by changing focal length

In the human visual system accommodation is by changing the focal length of the eye's lens. Below we state a lemma which suggests a method of estimating the spread parameter of the Gaussian point spread function by observing the change in an image due to a small change in the focal length of the lens. From a knowledge of the spread parameter, depth-map can be obtained using equations (7b).

Lemma 4 :

$$\sigma = \frac{1}{2kDs} \frac{f^2}{\omega^2+\nu^2} \frac{1}{P} \frac{dP}{df} . \tag{27}$$

Proof : We have

$$\frac{1}{P} \frac{dP}{df} = \frac{1}{P} \frac{dP}{d\sigma} \frac{d\sigma}{df} . \tag{28}$$

From equation (7b) we have

$$\frac{d\sigma}{df} = \frac{-kDs}{f^2} . \tag{29}$$

From equations (14,28,29) we can derive equation (27). ●

In this case the spread parameter σ (and hence the depth) is uniquely determined. Previously we have seen that the right hand side of equation (16) can be computed as the mean over some region in the frequency domain given by equation (19). Similarly, in this case the right hand side of equation (27) can be estimated as

$$c_2 = \frac{f^2}{2kDs} \frac{1}{(\omega_2-\omega_1)(\nu_2-\nu_1)} \int_{\omega_1}^{\omega_2}\int_{\nu_1}^{\nu_2} \frac{1}{\omega^2+\nu^2} \frac{1}{P} \frac{dP}{df} \, d\omega \, d\nu. \tag{30}$$

5.3 Depth map by changing aperture diameter

Lemma 5 :

$$\sigma^2 = -\frac{1}{2}\frac{D}{\omega^2+\nu^2}\ \frac{1}{P}\frac{dP}{dD}\ . \qquad \textbf{(31)}$$

Proof : We have

$$\frac{1}{P}\frac{dP}{dD} = \frac{1}{P}\frac{dP}{d\sigma}\ \frac{d\sigma}{dD}\ . \qquad (32)$$

From equation (7b) we have

$$\frac{d\sigma}{dD} = \frac{\sigma}{D} \qquad (33)$$

From equations (14,32,33) we can derive equation (31). •

In this case, except when the right hand side of equation (31) equals zero (which is the case when the image is in focus (i.e. $s=v$), or $D=0$), there are two solutions for σ, one positive, and another negative. However, if the image detector plane is fixed at $s=f$ then σ is always negative and a unique interpretation is obtained. As in the case of the previous two lemmas, in this case too the right hand side of equation (31) can be estimated as

$$c_3 = -\frac{D}{2}\frac{1}{(\omega_2-\omega_1)(\nu_2-\nu_1)}\int_{\omega_1}^{\omega_2}\int_{\nu_1}^{\nu_2}\frac{1}{\omega^2+\nu^2}\ \frac{1}{P}\frac{dP}{dD}\ d\omega\, d\nu. \qquad (34)$$

Note: changing aperture diameter changes the overall brighteness of the image on the image detector plane. The gray values of the pixels should be normalized by the overall image brighteness to compensate for this effect.

6. Relevance to machine vision and human vision

Of the three methods described for monocular depth-map recovery, the first method based on changing the distance between the lens and the image detector plane has immediate application for machine vision systems. In a camera system where the focusing (in a small field of view) occurs by a negative feed-back servo mechanism, the distance between the lens and the image detector plane naturally oscillates around a mean value with a small amplitude. For example, Horn (1968) reports that, for the camera used in his experiments, this distance oscillated with an amplitude of 0.02 cms and a

frequency of about 0.4 cycles per second. These oscillations can be taken advantage of to observe how an image changes due to a small change in the position of the image detector plane. Thus a complete depth map can be recovered at no deliberate physical effort of moving the image detector plane. The second method based on changing the focal length is relevant to biological vision systems. It suggests that an organism can, in principle, perceive depth everywhere in the field of view even though only a small field of view is in focus. There is evidence in support of the fact that the human eye deliberately exhibits small fluctuations in the focal length of the lens. The following paragraph is quoted from Weale (1982) (page 18):

> ‘‘... the state of accommodation of the un stimulated eye is not stationary, but exhibits micro fluctuations with an amplitude of approximately 0.1 D (diopter: a unit of lens power given by the reciprocal of focal length expressed in meters) and a temporal frequency of 0.5 cycles/second. He (Cambell, 1960) demonstrated convincingly that these were not a manifestation of instrumental noise, since they occurred synchronously in both eyes. It follows that their origin is central.’’

Our work shows that such fluctuations could be used to percieve depth in the entire scene simultaneously.

7. Conclusion

We have shown that a monocular camera can recover the depth-map of a scene in parallel without any assumptions about the scene. One major question concerning the approach described here could be its accuracy. Pentland (1987) has made some important observations about his approach which are directly relevant to our methods. He has argued that depth-map recovery based on approaches similar to ours could be comparable in accuracy to that based on stereopsis or motion parallax (for example, in the case of the human visual system). In addition, unlike stereopsis and motion analysis, our approach does not require any heuristic assumptions about the scene. Another effective way of

obtaining accurate depth-maps is to have several cameras with different camera parameters such that each camera is ‘‘tuned’’ to recover depth more accurately in a particular range than out side of this range. For example, for a robot vision system based on obtaining depth-map by changing the distance between the lens and the image detector plane, several cameras with lenses of different focal lengths can be used. Cameras with smaller focal length lenses help to recover accurately the depth variations at shorter distances and those with larger focal length lenses help to recover accurately the depth variations at longer distances.

Recent progress in this area is reported in Subbarao (1987b,c). The depth-map recovery methods described here have been extended to the case of *large* changes in camera parameters. These methods are expected to be more robust than the ones described here. Further, it has been found that, in addition to depth-map a monocular convex-lens-camera can recover directly the *motion* of objects parallel to the image detector plane. Also, it has been shown that, by appropriately configuring and controlling a binocular camera system, both the *depth-map* of a scene and the *motion* of objects in the scene can be recovered much more accurately than a comparable monocular system. These results suggest a new machine architecture for a robot vision system which is similar in many respects to the human visual system. We are planning to build an actual system to verify our approach. At present we are in the process of acquiring a specialized camera system necessary to conduct experimental studies of our approach.

Acknowledgements: This research was supported from the summer research fund provided by Department of Electrical Engineering, State University of New York at Stony Brook. I thank my friends H. Dhadwal, H.S. Don, G. Natarajan, and Dr. Alex Pentland for useful discussions in the final stages of this research.

Appendix: Region correspondence problem

In order to obtain the distance of a surface patch from the camera we measure the change in the two images of the surface patch formed by different camara parameter

settings. For this measurement we need to know the corresponding regions on the two images where the image of the surface patch is formed. Here we briefly outline how region correspondence can be established.

First we observe that region correspondence can be solved if point correspondence can be solved (e.g. by finding corresponding points for points on the boundary of the region). Next we note that, for a thin convex lens, the image of a scene point always lies at the intersection of the image detector plane and the line passing through the scene point and the optical center. These observations imply that the image position of a scene point is not changed if either the focal length of the lens is changed or if the aperture diameter of the lens is changed. Therefore the correspondence is trivial in these two cases. Now consider the case where the lens to image detector plane distance is changed. This situation is shown in Figure 6. The perspective transformation (see Figure 7) is given by $x = -s\frac{X}{Z}$ and $y = -s\frac{Y}{Z}$. Suppose that the lens to image detector plane distance is changed to $s' = s+\Delta s$ then we have

$$x' = -\left(1+\frac{\Delta s}{s}\right)s\frac{X}{Z} \quad \text{and} \quad y' = -\left(1+\frac{\Delta s}{s}\right)s\frac{Y}{Z} \tag{28}$$

or,

$$x' = \left(1+\frac{\Delta s}{s}\right)x \quad \text{and} \quad y' = \left(1+\frac{\Delta s}{s}\right)y. \tag{29}$$

Using the above relations correspondence is established.

References

Ballard, D. H., and C. M. Brown. 1982. *Computer Vision.* Prentice-Hall, Inc. Englewood Cliffs, New Jersey. Section 2.2.2.

Cambell, F. W. 1960. Correlation of accommodation between the two eyes. *Journal of Optical Society of America,* 50, 738.

Goodman, J.W. 1968. *Introduction to Fourier Optics.* McGraw-Hill, Inc.

Grossman, P. 1987 (Jan.). Depth from focus. *Pattern Recognition Letters 5*. pp. 63-69.

Horn, B.K.P. 1968. Focusing. Artificial Intelligence Memo No. 160, MIT.

Horn, B.K.P. 1986. *Robot Vision.* McGraw-Hill Book Company.

Jarvis, R. A. March 1983. A perspective on range finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-5, No. 2, pp. 122-139.

Krotkov, E. 1986. Focusing. MS-CIS-86-22. Grasp lab 63. Dept. of Computer and Information Science. University of Pennsylvania.

Ligthart, G. and F. C. A. Groen. 1982. A comparison of different autofocus algorithms. *Proceedings of the International Conference on Pattern Recognition.*

Marr, D. 1982. *Vision.* San Francisco: Freeman.

Pentland, A.P. 1982. Depth of scene from depth of field. *Procedings of DARPA Image Understanding Workshop.* Palo Alto.

Pentland, A. P. 1985. A new sense for depth of field. *Proceedings of International Joint Conference on Artificial Intelligence,* pp. 988-994.

Pentland, A.P. 1987. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. PAMI-9, No. 4, pp. 523-531.

Poggio, T., V. Torre, and C. Koch. 1985 (September). Computational Vision and Regularization Theory. *Nature,* Vol. 317, No. 6035, pp. 314-319.

Rosenfeld, A., and A.C. Kak. 1982. *Digital Picture Processing,* Academic Press, Inc. Section 1.2.1.

Schlag, J.F., A.C. Sanderson, C.P. Neuman, and F.C. Wimberly. 1983. Implementation of automatic focusing algorithms for a computer vision system with camera control. CMU-RI-TR-83-14. Robotics Institute. Carnegie-Mellon University.

Subbarao, M. 1987a (February). Direct recovery of depth-map. Tech. Report 87-02, Computer Vision and Graphics Laboratory, Dept. of Electrical Engineering, SUNY at Stony Brook.

Subbarao, M. 1987b (April). Direct Recovery of Depth-map II: A New Robust Approach. Technical report 87-03. Computer Vision and Graphics Laboratory, Department of Electrical Engineering, State University of New York at Stony Brook.

Subbarao, M. 1987c (May). Progress in research on direct recovery of depth and motion. Technical report 87-03. Computer Vision and Graphics Laboratory, Department of Electrical Engineering, State University of New York at Stony Brook.

Tenenbaum, J. M. November 1970. Accommodation in Computer Vision, Ph.D. Dissertation, Stanford University.

Weale, R.A. 1982. *Focus on Vision.* Harvard University Press, Cambridge, Massachusetts.