Active Photometric Stereo

James J. Clark
Division of Applied Sciences
Harvard University
Cambridge, MA

Abstract

This paper describes an active vision technique for determining the absolute depth of surfaces. The algorithm assumes a very general model for the reflectance properties of the surface, and is valid for most of the shading models commonly used in computer vision work. The algorithm relies on the controlled motion of a point light source. In this approach the light source is not at infinity, but is relatively close to the surface and to the camera. The sensitivity of the computed depth values to errors in the measured quantities is derived allowing a confidence measure for the depth to be determined. These confidence measures can be used to aid in the estimation of accurate depth values from multiple image measurements taken over time. We present a method based on robust estimation that permits an unbiased estimate of the depth values to be obtained. Finally the results of experiments on synthetic and real-world imagery are presented, illustrating the efficacy of the active photometric stereo algorithm.

1 Introduction

It has long been a goal of computer vision researchers to develop methods of obtaining 3D shape information from photometric information only. Horn [3] introduced the now classic shape from shading technique which provided surface gradient information from image irradiance. Subsequently, Woodham [7] developed a method, which he termed Photometric Stereo, for obtaining surface gradient information from a series of images acquired with the light source in different positions. Both of these methods required point light sources located at infinity, and do not give any absolute depth information.

In [4] Iwahori *et al* presented a method for obtaining absolute depth information from a series of images of an object illuminated by a point source near the object. Their method requires that the object have a Lambertian reflectance and required solving a system of nonlinear equations in four unknowns - depth, albedo and surface slant and tilt. There was no analysis of the uncertainty in the depth values provided by their algorithm, although an experimental result was given.

In [6] Wolff described an approach for extracting the slant and tilt of a Lambertian surface given the *photometric flow fields* induced by angular changes in the position of an illuminant. His approach assumes the light source to be at infinity, and requires the solution of a nonlinear set of equations (to find the intersection of conic sections). The absolute depth to the surface is not computed with this approach.

In this paper we present a new method for obtaining absolute depth values from a series of images of a surface illuminated by a point source near the object. Unlike the techniques described above,[1] our method involves solving a linear equation and we present a detailed analysis of the uncertainty in the resulting depth estimate. In addition, our approach is applicable to a wide range of reflectance models and, in particular, is *not* restricted to surfaces with Lambertian reflectance.

2 Depth From Light Source Motion

Let us begin the description of our algorithm by considering the geometry of the imaging process, depicted in figure 1. The direction of the ray of light emanating from a point source, located at a position $\vec{t} = (t_x, t_y, t_z)^T$ with respect to a coordinate system centered on the focal point of the camera, that intersects a surface at the point $(X, Y, Z)^T = (-Zx/f, -Zy/f, Z)^T = -Z\vec{\chi}$ is given by

$$\hat{s} = \frac{Z\vec{\chi} + \vec{t}}{|Z\vec{\chi} + \vec{t}|} \tag{1}$$

The vector $\vec{\chi} = (x/f, y/f, -1)^T$ indicates the position of the image plane point $\vec{x} = (x, y)^T$ with respect to the world coordinate system centered on the camera focal point.

We take as a general model for relation between the measured image irradiance and light source position:

$$E(\vec{x}) = K\frac{R(\hat{s})}{|Z\vec{\chi} + \vec{t}|^2 |Z\vec{\chi}|^2} \tag{2}$$

The $|Z\vec{\chi}+\vec{t}|^2$ term represents the reduction in light per unit area over the distance from the light source to the body, the $|Z\vec{\chi}|^2$ represents the dimunition of light from the body to the camera, and K is a proportionality constant factoring in, among other things, the intensity of the light source. The reflectance will in general depend on the viewing direction and the surface normal, but as we are concerned with only a particular point on the surface, the only *variable* quantity that

[1] Personal communication with Dr. Iwahori has indicated that his group has developed a method that is similar to the one described in this paper.

0-8186-2855-3/92 $3.00 ©1992 IEEE

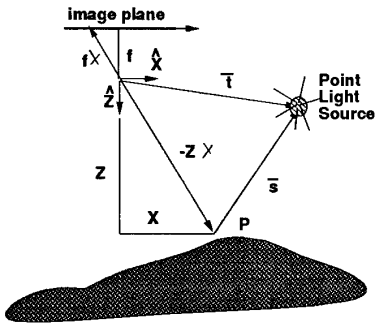

Figure 1: Geometry of the active photometric stereo scheme.

the reflectance depends on is the light source direction, $\hat{s}$. Note that we are assuming that the reflectance does not depend on the distance of the light source to the surface point, and that the dependence of the image irradiance on the distance of the light source to the surface is handled by the $|Z\vec{\chi}+\vec{t}|^2$ term in the equation above. This assumption is not very restrictive and is valid for most of the shading models currently used in the Computer Vision and Graphics communities.

Let us examine how the image irradiance varies with an infinitesimal change in the position of the light source. We can create a vector $\nabla_t E$ whose components are the rates of change of the image irradiance with respect to change in the components of the light source position vector. That is:

$$\nabla_t E = \left(\frac{\partial E}{\partial t_x}, \frac{\partial E}{\partial t_y}, \frac{\partial E}{\partial t_z}\right)^T \tag{3}$$

Differentiation of equation (2) yields:

$$\nabla_t E = \frac{\nabla_t R(\hat{s})}{|Z\vec{\chi}|^2 |Z\vec{\chi}+\vec{t}|^2} - \frac{2R(\hat{s})\nabla_t |Z\vec{\chi}+\vec{t}|}{|Z\vec{\chi}|^2 |Z\vec{\chi}+\vec{t}|^3} \tag{4}$$

where

$$\nabla_t |Z\vec{\chi}+\vec{t}| = \frac{Z\vec{\chi}+\vec{t}}{|Z\vec{\chi}+\vec{t}|} = \hat{s} \tag{5}$$

If we take the dot product of both sides of the above equation with the illumination direction vector, $\hat{s}$, scaled by $|Z\vec{\chi}+\vec{t}|$, and then use equation (2) we get the following relation:

$$\begin{aligned}(\nabla_t E)^T \hat{s} |Z\vec{\chi}+\vec{t}| &= \frac{(\nabla_t R(\hat{s}))^T \hat{s}}{|Z\vec{\chi}|^2 |Z\vec{\chi}+\vec{t}|} - \frac{2R(\hat{s})}{|Z\vec{\chi}|^2 |Z\vec{\chi}+\vec{t}|^2} \\ &= \frac{(\nabla_t R(\hat{s}))^T \hat{s}}{|Z\vec{\chi}|^2 |Z\vec{\chi}+\vec{t}|} - 2E\end{aligned} \tag{6}$$

Let us compute the quantity $(\nabla_t R(\hat{s}))^T \hat{s}$.

$$(\nabla_t R(\hat{s}))^T \hat{s} = (\nabla_{\hat{s}} R)^T \left(\frac{\partial \hat{s}}{\partial \vec{t}}\right)^T \hat{s} \tag{7}$$

The transpose of the Jacobian matrix, $\left(\frac{\partial \hat{s}}{\partial \vec{t}}\right)^T$ is given by:

$$\left(\frac{\partial \hat{s}}{\partial \vec{t}}\right)^T = \frac{(I - \hat{s}\hat{s}^T)}{|Z\vec{\chi}+\vec{t}|} \tag{8}$$

From the vector identity $(I - \hat{a}\hat{a}^T)\hat{a} = 0$ (where $\hat{a}$ is an arbitrary unit vector) we can see that:

$$(\nabla_t R(\hat{s}))^T \hat{s} = (\nabla_s R)^T \frac{(I - \hat{s}\hat{s}^T)}{|Z\vec{\chi}+\vec{t}|} \hat{s} = 0 \tag{9}$$

Thus, from equations (6) and (9), we have the following, rather surprising, result:

$$(\nabla_t E)^T (Z\vec{\chi}+\vec{t}) = -2E \tag{10}$$

(where we have used $\hat{s}|Z\vec{\chi}+\vec{t}| = Z\vec{\chi}+\vec{t}$). Solving this linear equation for Z gives us:

$$Z = -\frac{((\nabla_t E)^T \vec{t} + 2E)}{(\nabla_t E)^T \vec{\chi}} \tag{11}$$

An intuitive explanation of the above result can be had by considering that the gradient of R with respect to the light source motion can be broken into two orthogonal components, the first along the illuminant direction $\hat{s}$ and the second perpendicular to this direction. It is easily seen that, because of the shading law that we have assumed, the component of the gradient of R in the direction of $\hat{s}$ will be zero, as $\hat{s}$ is not changed by motion of the light source in this direction. Thus $(\nabla R)^T \hat{s}$ is zero.

Note that, with the above equation, the absolute depth can be obtained without knowledge of the surface albedo or even the particular shading model obeyed by the surface (as long as equation (2) holds).

3 Least Squares Depth Estimation

The depth measurement depends on a number of image measurements (E and $\nabla_t E$) and system parameters ($f, \vec{\chi}, \vec{t}$) all of which are known only to within some error level. If we assume that the errors in the parameters and measurements are small, we can do the following linearization to get the error in the depth value:

$$\delta Z = \sum_i \frac{\partial Z}{\partial \phi_i} \delta \phi_i \tag{12}$$

where the ϕ_i are the parameters and measurements listed above. If we assume that the parameter and measurement errors are Gaussian distributed with zero mean and variance σ_ϕ^2 (not a good assumption for some of the parameters, but not bad for modelling the noise

in the image and its time derivative) then the depth error is also Gaussian and we can write its variance as:

$$\sigma_Z^2 = \sum_i \left(\frac{\partial Z}{\partial \phi_i}\right)^2 \sigma_{\phi_i}^2 \tag{13}$$

In general, the sensitivities of the computed depth values to uncertainties in the camera parameters and measured image quantities are such as to make the error in the depth values unacceptably high. In order for the active photometric stereo process to be useful one must provide some way to reduce the effect of measurement noise and parameter uncertainty on the result.

One possible approach is to get multiple images of the same scene and average them to reduce the image noise, then compute the surface estimate based on the average images. As an alternative, one can compute the depth using each of the individual image sets and use some sort of estimation scheme to determine a more accurate depth map. The most straightforward way to do this is use a weighted least squares estimate. In this approach we take as our estimate, $\hat{Z}$, of the depth to be that which minimizes the following cost function (which essentially finds the estimate of the mean which minimizes the variance):

$$C = \sum_{k=1}^{N} \frac{(\hat{Z} - Z(k))^2}{\sigma_Z^2(k)} \tag{14}$$

where $Z(k) = -\frac{2E(k)+(\nabla_t E(k))^T \vec{t}}{(\nabla_t E(k))^T \vec{\chi}}$ is the k^{th} depth measurement.

In the above, the $\vec{\chi}$ and $\vec{t}$ vectors are fixed and assumed to be known (this means that we are looking at a single point on the surface and estimating its depth, and that the nominal light source position is fixed). The $\hat{Z}$ which minimizes C can be determined by differentiating C with respect to $\hat{Z}$ and setting to zero. This yields:

$$\hat{Z} = \frac{\sum_{k=1}^{N} Z(k)/\sigma_Z^2(k)}{\sum_{k=1}^{N} 1/\sigma_Z^2(k)} \tag{15}$$

This formula requires that all data points be available before an estimate can be computed. One can also implement the estimator in a sequential or recursive fashion, where an estimate is available at each time step. The variance, $\sigma_Z^2(k)$ of the depth measurement can be obtained from the sensitivities described in the previous section, if we know the variances of the image noise and camera parameter values.

4 Biases in the Depth Estimate

The weighted least squares estimate given above is the optimal estimate when the error in the depth measurements are Gaussian. The estimate is also *unbiased* in this case (i.e. the mean of the estimation error is zero). However, the error in the depth estimate is NOT Gaussian, except in the special case when there is no image noise and the vector χ is known exactly. In the previous section we assumed that the parameter and measurement errors were small, allowing a linearization of the error, resulting in a Gaussian error. In general, our errors will not be small and this approximation is not valid. We must therefore look at the nonlinear relationship between the parameter and measurement errors and the depth error. If all of the image measurement errors and imaging parameter uncertainties are Gaussian, then the depth measurement is the ratio of two Gaussian distributed random variables, each having non-zero mean. It is well known that such a quantity is non-Gaussian [2, 3]. One of the implications of this is that the estimate of the depth obtained using the least squares approach will be *biased*. This means that the mean of the estimation error is non-zero. In the case where the variances of each of the depth measurements are equal (i.e. $\sigma_Z^2(i) = \sigma_Z^2(j) = \sigma_Z^2$) the bias in the estimator is equal to the mean of the distribution of the depth measurements. Thus, if we could compute the mean of this distribution we could still use the least squares estimator and subtract off the bias.

In general, computation of the moments of the distribution of quotients of correlated Gaussian random variables with non-zero means is difficult and only approximate expressions can be obtained [1, 2, 3]. Another possible approach is to do an empirical study of the distribution of the depth measurements. This is impractical in our application, as the form of this distribution depends on the parameters $\vec{t}$ and $\vec{\chi}$. We would then have to do this empirical study for the product of every point in the image and every different light source position used!

To handle the non-Gaussianity of the depth measurement noise, robust estimation techniques can be used [5]. In particular the median is useful in our application since the median of a random variable formed by the ratio of two Gaussian random variables is a good approximation to the ratio of the means of these two random variables. Thus the median of the sequence of depth measurements will provide a nearly unbiased estimate of the true depth. The use of the median estimator, although it handles the bias problem of the mean estimator, does not overcome the effects of the non-stationarity of the noise.

5 Simulation of the Algorithm

We performed a series of simulations of this algorithm on a synthetic surface. Figure 2c) shows a depth map of the surface that we used. It is an exponentially damped product of sine waves. It has a nominal depth of 150 cm and an amplitude of 10 cm. The camera parameters used to generate the images were a focal length of 5.0 cm and a pixel size of 17.2 microns. The image shown in figure 2a) was mapped onto the surface to provide the albedo variation. Figure 2b) shows the image, obtained using a Lambertian shading model, of the surface with this albedo variation, for the case of $\vec{t} = (0, 0, 0)^T$. Figure 2d) shows the result of integrating the results of 20 estimated depth maps, using the least squares estimator. It should be noted that figure 2c) and 2d) show only the variation of the surface about the nominal depth of 150 cm. For the purposes

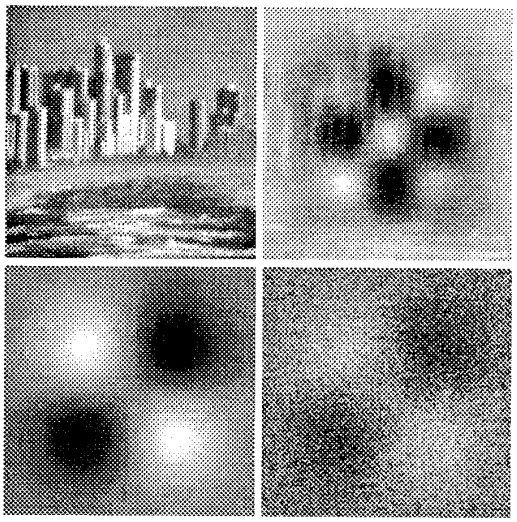

Figure 2: Clockwise from upper left: a) Albedo map; b) image of surface; c) true depth map; d) estimated depth map.

of gauging the amount of noise in the result one should really display the depth maps over a range scale of 0-160 cm or so, rather than the scale of 130-170 cm that was actually used.

The following set of graphs depict the comparative performance of the weighted least squares and median estimators. Figure 3 shows the variation in the mean of the estimation error for the weighted least squares estimator, for six different levels of image noise variance (indicated in the box below the graph; these levels were 1,2,5,10,20, and 50. To give some idea of the scale of the additive image noise component, the image values ranged from 3050 to 4550, with a variance of 297. Hence for the case of the noise variance being 50, the signal to noise ratio (ratio of standard deviations) was about 40). The experiments therefore show the extreme sensitivity of the depth estimate to image noise, even with the multi-image estimation processes.

Figure 4 shows the reduction in the (natural log of the) variance of the least squares estimate. Figure 5 shows the mean of the depth estimation error when the median estimator is used. Note the absence of a bias in this case, as well as the rapid reduction of the mean. The final graph shows the effect of the non-stationarity on the estimation process. In the experiments that provided the previous graphs the nominal position of the light source was held fixed. In the experiments that produced figure 6, the light source is moved in a logarithmic spiral in the plane perpendicular to the optic axis and including the focal point (or pinhole position). Figure 6 shows the estimation error means. Three different image noise variances were used. Note that, in the case of the least squares estimator, the

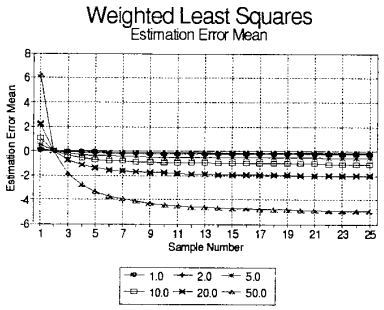

Figure 3:

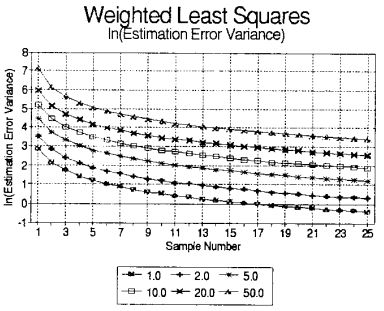

Figure 4:

bias in the estimate does not settle down to a fixed value, but continues to grow. The bias of the median estimate seems to remain near zero.

6 Experiments on Real Imagery

In order to test our algorithm on real imagery the following experiment was performed. A scene containing 3D objects of different reflectance and varying albedo were arranged in front of a standard video camera. The camera had a lense with a focal length of 50mm and a pixel size of 17.2 microns. A high stability bright light source was connected to a fiber optic light pipe with a diffusing end cap to provide the point light source. The extent of the light emitting region was 3mm. The light source was affixed to a Klinger high precision position stage (with a precision of 0.1mm) to allow accurate movements of the light source. The nominal light source position vector was $\vec{t} = (13, 75, -124)^T$ cm. The image of the scene with the light source in the nominal position is shown in figure 7a). The distance from the camera focal point to the front of the basketball was 124 cm, while the distance to the wall behind the ball was 172 cm.

The experiments were performed in a darkened area, however, there was still a small ambient (diffuse) illumination component. As this component was presumably fixed, the image obtained from this component alone was measured (by turning the point light

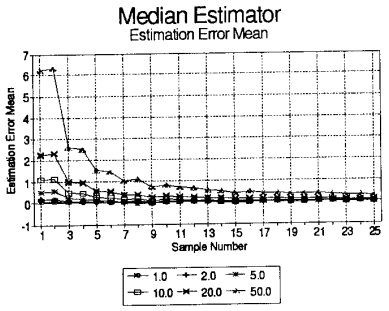

Figure 5:

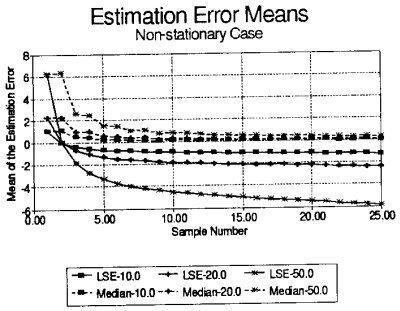

Figure 6:

source off) and was subtracted from all subsequent image measurements. This technique could, in principle, be applied for arbitrary ambient light conditions. If the image sensor becomes saturated or otherwise loses dynamic range due to the common mode ambient illumination then one may need to reduce the ambient light level.

To obtain the derivatives with respect to the light source position the light source was moved by an increment of 1 cm in the positive and negative directions of the X, Y, and Z axes. A first order central difference approximation to the derivatives was used. Thus, for each depth measurement, seven images were needed. Because of time and memory limitations only two sets of images were acquired. The images were digitized to a resolution of 485x512. These images were reduced down to 256x256 by averaging 2x2 blocks of pixels. Thus, adding the two subsampled sets of images is equivalent to reducing the image noise in a 256x256 image by a factor of $1/\sqrt{8}$. The depth equation (12) was then used to provide a 256x256 raw depth map. This depth map was very noisy. As we did not have a large sequence of depth maps to use in obtaining a depth estimate, we applied a 5x5 median filter to the depth map. The estimate that results would be similar to that obtained by applying the median estimator to a sequence of 25 depth maps provided the depth map does not significantly vary spatially.

The result of the estimation process is shown in figure 7d). Dark regions indicate areas close to the camera while lighter regions indicate parts of the scene that are further away. The logarithm of the estimated sensitivity of the depth estimate to image noise is depicted in figure 7c). Dark regions indicate areas in which the variance of the depth estimate is expected to be small while lighter areas represent regions in which the depth estimate is highly uncertain. The result in the shadow region below the ball cannot be trusted to give meaningful depth or sensitivity values as the image in those regions was mainly due to sensor noise. The algorithm had difficulties near specularities presumably because saturation of the camera resulted in small derivatives while the image was very bright. Thus the depth values were overestimated. Fortunately these effects show up as an increase in the sensitivity. To see the level of uncertainty in various areas of the depth map examine the histograms of the depth values in certain regions of the image. Some of these are shown in figure 8 (they correspond to the regions outlined in figure 7b). Again one should keep in mind that one is really measuring depths over a range of 0 to 200 cm and not merely the 140 to 200 cm range depicted in the histograms.

Histogram 1 is of a relatively low sensitivity area on the ball, and has a mean depth of around 133 cm. There is some spread (from a Gaussian shape) due to the variation of depth in this region. Histogram 2 is of a relatively low sensitivity area on the wall, and has a mean depth of around 170 cm. The standard deviation in this region is about 10 cm. Histogram 3 is of a smaller low sensitivity region on the wall. Its mean depth is 170 cm with a standard deviation of 12 cm. Histogram 4 is of a region containing the depth discontinuity between the ball and the wall. It contains two peaks, one near 128 cm and the other near 200 cm. The variance in the higher peak is larger, as is expected, since the sensitivity in the wall area in this region is relatively high. Histogram 5 is of a region of low to intermediate sensitivity on the wall. It has a mean of 174 cm and a standard deviation of around 13 cm. Finally histogram 6 is of a region of the image that has a high sensitivity to image noise, and is located on the ball. The mean of the depth values in this region is 132 cm with a standard deviation around 28 cm.

The results of this, admittedly primitive, experiment are encouraging in that the uncertainty in the depth estimate in the low sensitivity regions of the image are adequate for many purposes (such as segmentation or possibly even coarse positioning of robot manipulators), but are sobering in view of the large portions of the depth map that are too uncertain to be useful. Certainly integration of longer sequences of images will improve this performance, as will the ability to make large movements of the light source.

7 Summary

We have presented a photometric method for determining the absolute depth to points on the surface of an object. Unlike previous methods (e.g. that of [4]) this method only requires the solution of a linear equation in the depth value, and holds for a very general class of reflectance models. A detailed analysis of

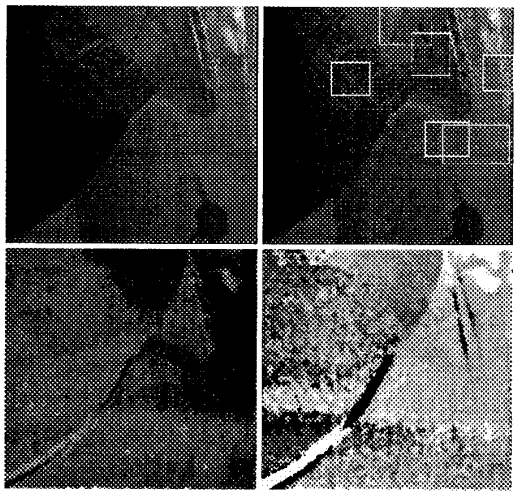

Figure 7: Clockwise from upper left: Original image; Overlay of the regions whose histograms are shown in figure 8; Sensitivity map; Median filtered depth map.

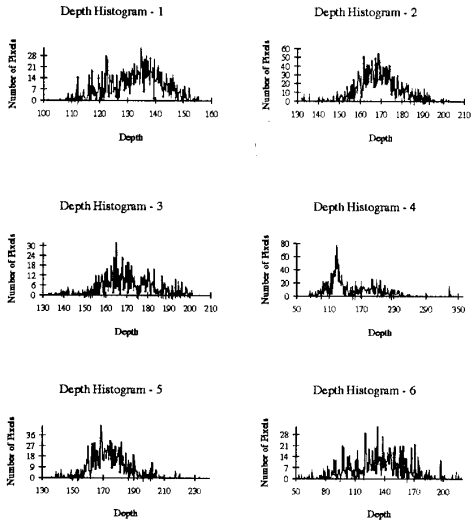

Figure 8: Histograms of the median filtered depth map in the regions shown in figure 7b).

the sensitivity of the computed depth value to errors in image measurements and camera parameters was made, and a robust estimation process was proposed. The robust estimation process had the advantage that the estimate was unbiased. Results of simulations on synthetic data and on real imagery were shown that demonstrate the efficacy of the estimation process.

Acknowledgments

This research was supported in part by the National Science Foundation through grants IRI-9003306 and CDR-85-00108. The Digital Equipment Corporation Parallel Research Initiative provided support in the form of computing equipment.

References

[1] Ferrier, N.J.,**Trajectory Control of Active Vision Systems,** Ph.D. Thesis, Division of Applied Sciences, Harvard University, 1992.

[2] Geary, R.C., "The frequency distribution of the quotient of two normal variables", Royal Statistical Society Series A, Vol. 93, pp 442-446, 1930

[3] Hinkley, D.V., "On the ratio of two correlated normal random variables", Biometrika, Vol. 24, pp 428-440, 1932

[3] Horn, B.K.P, "Obtaining shape from shading information", Chapter 4 in **The Psychology of Computer Vision,** P.H. Winston, ed., McGraw-Hill, New York, 1975

[4] Iwahori, Y., Sugie, H., and Ishii, N., "Reconstructing shape from shading images under point light source illumination", Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, pp 83-87, June 1990

[5] Kassam, S.A., and Poor, H.V., "Robust techniques for signal processing: A survey", Proceedings of the IEEE, vol. 73, pp 433-481, March 1985

[6] Wolff, L.B., "Shape understanding from Lambertian photometric flow fields", Proceedings of the 1989 Computer Vision and Pattern Recognition Conference, pp 46-52

[7] Woodham, R.J., "Photometric method for determining surface orientation from multiple images", Optical Engineering, Vol. 19, No. 1, pp 139-144, 1980