COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 32, 337–358 (1985)

Experiments with the Intensity Ratio Depth Sensor

BRIAN CARRIHILL AND ROBERT HUMMEL*

Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, New York 10012

Received February 26, 1985; revised June 12, 1985

This paper discusses a modification of the "plane-of-light" approach to range data acquisition which utilizes the ratio of two intensity images. The modification allows depth determination at each image pixel, avoiding the need to scan the plane of light, requiring only the acquisition and processing of two or three intensity images. The depth equation together with three experimental methods for its calculation are presented. Results of the three sensor implementations are given for test scenes. © 1985 Academic Press, Inc.

1. THE INTENSITY RATIO DEPTH SENSOR

Many computer vision applications require information about the three dimensional shape of objects in a scene. A central research interest in many vision studies concerns the extraction of three dimensional information given one or several two dimensional grayscale images. We mention "shape-from-shading," "shape-from-texture," stereo matching, "shape-from-motion," and occlusion cues as different approaches to this problem [1–6]. In robotics applications, where the environment and lighting can be controlled, it is possible to directly sample range data. Typically, range data is given by an image whose values $R(i, j)$ represent the distance from the sensor to the point in the scene imaged at the pixel (i, j). Methods for obtaining range data include laser range finders, ultrasonic and radar reflection schemes, interferometry and fringe counting, Moiré techniques, spot scanning schemes, and plane-of-light sensors. A survey of depth sensors is given in [7].

Let us review the plane-of-light concept, described by Shirai [1,8]. Light is projected through a slit, thus illuminating a known plane in three-space. This plane intersects surfaces in the scene, and yields an image consisting of curve segments (see Fig. 1). Each point on the image plane comes from a point in the scene which lies on a ray determined by the pixel location and the focal point of the camera system. Thus every point on an observed curve segment in the image can be associated with a point in space lying on a fixed plane and the line corresponding to that point. A simple triangulation leads to the (x, y, z) coordinate location of each point on the observed curve. By scanning the slit projection system so that the plane of light sweeps out the space of the scene, (x, y, z)-coordinate locations can be computed for all surface elements visible from the camera system illuminated during the sweep.

An elegant idea, due to Jacob Schwartz, shows that the plane-of-light scheme can be modified to obtain the 3D data using only two views of the scene, without sweeping a slit across the scene [9]. The idea is best understood in terms of the "gradient light" version of the sensor. In this version, the first view is taken with the scene illuminated by the projector system using a constant field filter—i.e., constant illumination. The second view uses a linear wedge filter in the same projector, so that

*This research was supported in part by ONR Grant N00014-82-K-0381, support from IBM Corporation, NSF Grant DCR-83-20085, and NSF Grant DCR-84-03300.

0734-189X/85 $3.00

the scene is illuminated brightly on one side, but only half brightly on the other side. We suppose that the illumination emerging from the projector in both views is constant along vertical lines. A ratio of the two views is formed. The observed ratio value at a given location determines the plane in which the corresponding point in the scene lies. The ratios of the illumination intensities code planes in three-space that fill the region of the scene. All reflectivity, foreshortening, and other intensity variations are cancelled by the ratio. There are *no* lambertian reflection assumptions. A change in the intensity between the two views must be due to the change in illumination intensity. Accordingly, a simple computation at each pixel in the scene, depending only on the pixel's coordinates and the observed intensity ratio at the location, can calculate the (x, y, z) coordinate of the corresponding point in the scene.

We have assumed that the reflected intensity is proportional to the illumination intensity (a most valid assumption), and that self-illumination effects are not too important (an assumption whose validity depends on the scene). The accuracy of the intensity ratio depth sensor is limited by the spatial resolution and fidelity of the camera system, the quantization of the intensity values, and the extent to which self-illumination and projector optics blur the constant ratio planes emitted from the light source system.

Variations of the gradient light scheme are possible, using different illumination filters for the pair of views of the scene; additional views can also be used for increased accuracy. For example, it is not necessary to assume that constant ratio surfaces are vertical planes.

The intensity ratio depth sensor enjoys several clear advantages over the plane-of-light range acquisition method. There are essentially no moving parts required with the intensity ratio method. By using electro-optical filters, it should be possible to acquire the two views rapidly. The triangulation computation required at each pixel is extremely simple, and could be carried out in parallel. Further, a depth value is obtained for every point which is in both the camera's and projector's field of view. Thus it should be possible to build a hardware intensity ratio depth sensor which supplies depth images updated at video rates or faster.

In the next section, we sketch the equations governing the calculation of depth using the intensity ratio depth sensor, and explain the parameters used in experiments. Section 3 presents some results, and Section 4 some comments.

2. FORMULATION

The formulas for computing depth from the ratio value and the pixel coordinates are straightforward. We will develop these formulas under some simplifying assumptions, in order to make explicit the parameters of the system. Formulations and treatments with fewer assumptions will be given in [10].

We will fix the coordinate system with the origin at the focal point of the camera, with the z axis in the direction in which the camera is pointed, which we assume to be horizontal. We take the x and y axes to be horizontal and vertical respectively (see Fig. 1). We will assume that the projection system lies in the same x–z horizontal plane as the camera, and that the position of the projector's focal point is accurately known at $(x_0, 0, z_0)$.

We further assume that the surfaces of constant illumination ratio formed by the projector and filter system form vertical planes passing through the point $(x_0, 0, z_0)$.

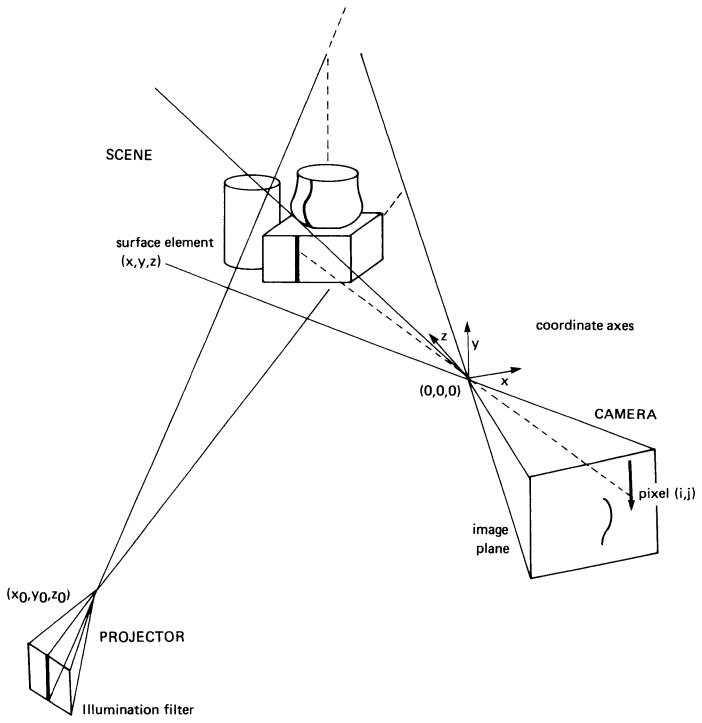


FIG. 1. Plane-of-light Depth Sensor.

We represent by $d(\rho)$ the z intercept, i.e., the depth of the intersection of the plane of constant ratio ρ and the z axis. We represent by $h(\rho, \xi)$ the x-coordinate intercept on the plane $z \equiv \xi$ (see Fig. 2). To avoid ambiguity in the depth calculation, it is necessary that each intensity ratio ρ be associated with a unique vertical plane. This implies that the functions $d(\rho)$ and $h(\rho, \xi)$ should be strictly monotonic functions in ρ. It can be shown that

$$h(\rho, \xi) = \frac{d(\rho) - \xi}{d(\rho) - z_0} x_0$$

for all ρ, and $\xi \geq z_0$.

In particular, $h(\rho, z_0) \equiv x_0$. The equation of the vertical plane of constant ratio ρ can be written in several ways. For example,

$$z = \frac{z_0 - d(\rho)}{x_0} \cdot x + d(\rho). \tag{1a}$$

In terms of $h(\rho, \xi)$, for $\xi = z_1 > z_0$,

$$z = \frac{z_1 - z_0}{h(\rho, z_1) - x_0}(x - h(\rho, z_1)) + z_1. \tag{1b}$$

If we consider two planes $z \equiv z_1$ and $z \equiv z_2$, it is possible to eliminate x_0 and z_0

from the equation of the plane, to yield

$$z = \frac{z_1 - z_2}{h(\rho, z_1) - h(\rho, z_2)} \cdot (x - h(\rho, z_1)) + z_1. \tag{1c}$$

This is useful for calibration purposes when the position (x_0, z_0) is not known with high accuracy.

Next, we assume that each point in the image plane is imaged by the camera system and is associated with a unique ray in space passing through the origin $(0, 0, 0)$. We define a coordinate system (u, v) on the image plane by requiring that the ray associated with the point (u, v) lie on the line $(x(t), y(t), z(t)) = (ut, vt, t)$. If the optics are perfect, u lies in the x direction, v in the y direction, and u and v measure distance on the image in units such that the focal length is 1. Since the image will be sampled at a grid of points, $u = u(i, j)$ and $v = v(i, j)$ will be functions of pixel coordinates (i, j), and will generally be close to linear functions. The location of a point imaged at (u, v) and having observed ratio ρ lies on the intersection of the line corresponding to (u, v) and the plane of constant illumination ratio ρ. The z coordinate of this point is obtained by substituting $x = uz$ into the equation for the plane of constant ratio ρ. This gives for the alternate forms (1a), (1b), (1c):

$$z = z(u, v, \rho) = \frac{d(\rho)}{1 - \left(\frac{z_0 - d(\rho)}{x_0}\right)u}, \tag{2a}$$

$$z = z(u, v, \rho) = \frac{z_1 h(\rho, z_1) + z_0 x_0}{h(\rho, z_1) - x_0 - u(z_1 - z_0)}, \tag{2b}$$

$$z = z(u, v, \rho) = \frac{z_1 h(\rho, z_1) + z_2 h(\rho, z_2)}{h(\rho, z_1) - h(\rho, z_2) - u(z_1 - z_2)}, \tag{2c}$$

respectively. In all cases,

$$x = x(u, v, \rho) = u \cdot z(u, v, \rho)$$

and

$$y = y(u, v, \rho) = v \cdot z(u, v, \rho).$$

The value $z(u, v, \rho)$ is the *depth* of the point imaged at (u, v), in distinction to the *range*, which is $R = \sqrt{x^2 + y^2 + z^2}$. Obviously, once the depth is known, the other coordinates and the range are easy to calculate.

Note that $z(u, v, \rho)$ is a function of u and ρ alone, and independent of v. This is due to our assumption that constant ratio planes lie vertically. Any implementation of the intensity ratio depth sensor requires the representation of the function $z(u, v, \rho)$. Although $z(u, v, \rho)$ represents an infinite number of parameters, we are forced to model the function by a finite set of values. Below we discuss three implementation methods, in order of increasing numbers of parameters in the representation of $z(u, v, \rho)$. The additional parameters to represent the functions

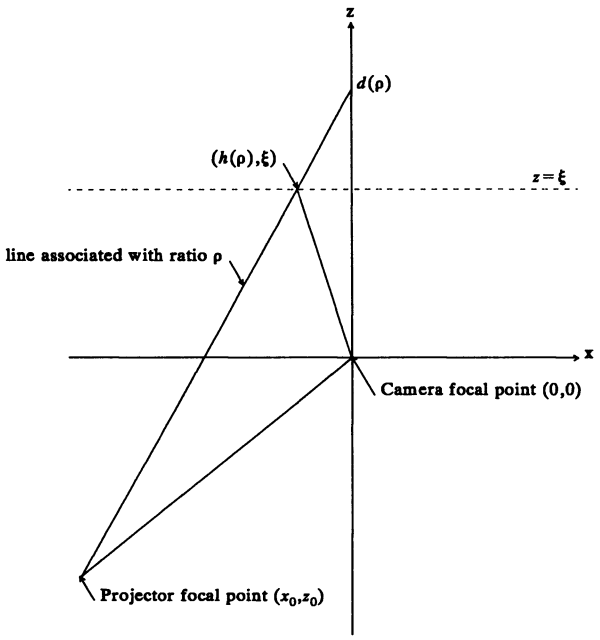


FIG. 2. Triangulation geometry after projection onto the x–z plane.

$u(i, j)$ and $v(i, j)$, used to determine the (u, v) coordinates at each pixel (i, j), are discussed separately in terms of the camera calibration in Section 3.

The first method, Method 1, makes use of Eq. (2a). The parameters of this method are x_0, z_0, and the function $d(\rho)$. Note that if the projector illumination system is at or near infinity ($|(x_0, z_0)| \to \infty$) and the transmissivity varies linearly on the filter used for the gradient view, then $d(\rho)$ will be a linear function, say $d(\rho) \approx A\rho + B$. This gives us four parameters: x_0, z_0, A, and B. Calibration of the projector system determines (x_0, z_0) as the location of the projector's focal point. The calibration of A and B is achieved by computing a linear regression of observed $d(\rho)$ data. The data may be collected directly by placing a flat calibration screen at a known sequence of depths and observing the intensity ratios along the vertical column $u = 0$. Averaging the ratios in the column leads to an assigned ratio value at the given depth. Inverting this data leads to a set of depth-ratio pairs, which are used to obtain A and B.

The second method, Method 2, utilizes both Eqs. (2b) and (2c). The parameters of this method are once again x_0, z_0 together with the functions $h(\rho, z_1)$ and $h(\rho, z_2)$ for a fixed pair z_1 and z_2. A linear fit of $d(\rho)$ or equivalently the function $h(\rho, \xi)$ in ρ, may not be appropriate when the projector is at a finite distance from the workspace ($|(x_0, z_0)| < \infty$) or the variation of the transmissivity of the filter used for the gradient illumination is far from linear. We therefore use a few hundred parameters to represent both $h(\rho, z_1)$ and $h(\rho, z_2)$ as piecewise linear functions of ρ. The calibration of $h(\rho, z_1)$ is performed by placing the flat calibration screen at the fixed depth $z = z_1$ and observing the ratio image $\rho(u, v)$. Now the ratio image, $\rho(u, v)$, is independent of v since the planes of constant illumination ratio are

vertical. Also, recalling that the x coordinate of the point at depth $z = z_1$ with image coordinates (u, v) is given by $x = uz_1$, we have the relation

$$h(\rho(u, v), z_1) = uz_1. \tag{3}$$

Therefore for each image column we have a ratio ρ and x intercept $h(\rho, z_1)$ data pair. In order to model $h(\rho, z_1)$ as a continuous function in ρ we linearly interpolate between the data pairs. A similar calibration of $h(\rho, z_2)$ is performed by placing the flat calibration screen at the fixed depth $z = z_2$.

The domains of $h(\rho, z_1)$ and $h(\rho, z_2)$ are thus determined by the range of observed ratios in the calibration ratio images. The evaluation of the depth function $z(u, v, \rho)$, for those ratio values ρ which lie in the intersection of the domains, makes use of Eq. (2c). However, for ratio values outside of this interval, extrapolation of either $h(\rho, z_1)$ or $h(\rho, z_2)$ is required. Within the interval for which the depth error due to extrapolation exceeds the depth error due to the projector calibration inaccuracies, Eq. (2b) is used.

The third method, Method 3, eliminates the need of the observed ratios $\rho(u, v)$ of the flat calibration screen being constant in columns. That is, the surfaces of constant ratio illumination are no longer required to form vertical planes. In this case the depth function $z(u, v, \rho)$ is modeled as a large spatially dependent lookup table in (u, v, ρ) involving hundreds of thousands of parameters. A practical implementation of the depth sensor would probably regard the depth computation as a local operation, to be computed at all sites in parallel. If enough local memory is available, $z(u, v, \rho) = z_{u,v}(\rho)$ can be implemented as a lookup function. Separate tables are needed at each pixel site (u, v), and these can be calibrated by exhaustively determining ratios as a function of depth at each pixel, and inverting the relation. However, the amount of memory needed for such a scheme is staggering. It is more reasonable, instead, to represent the local lookup table as a parameterized function of the single ratio variable. In our Method 3 experiments, we assume that each local function is well approximated by a quadratic function of ρ:

$$z(u, v, \rho) \approx A_{u,v}\rho^2 + B_{u,v}\rho + C_{u,v}.$$

The parameters of this scheme consist of three (floating point) images to store the $A_{u,v}$'s, $B_{u,v}$'s, and $C_{u,v}$'s.

It is important to realize that our Method 3 essentially abandons Eq. (1). The assumption that surfaces of constant ratio lie in vertical planes is replaced by the requirement that ratio values as a function of depth be independently invertible for each pixel ray. If surfaces of constant ratio happen to lie in vertical planes, then Method 3 uses quadratic equations to approximate the functions at each pixel (u, v) of the form.

$$z_{u,v}(\rho) = d(\rho) \Big/ \left(1 - \left(\frac{x_0 - d(\rho)}{z_0}\right) u\right).$$

An exact fit will in general be impossible.[1] Further, if the planes of constant ratio are

[1]An exception occurs if $d(\rho)$ is linear and x_0 and z_0 are infinite, as we assumed in Method 1. In that case, local linear equations would suffice (i.e., $A_{u,v} \equiv 0$).

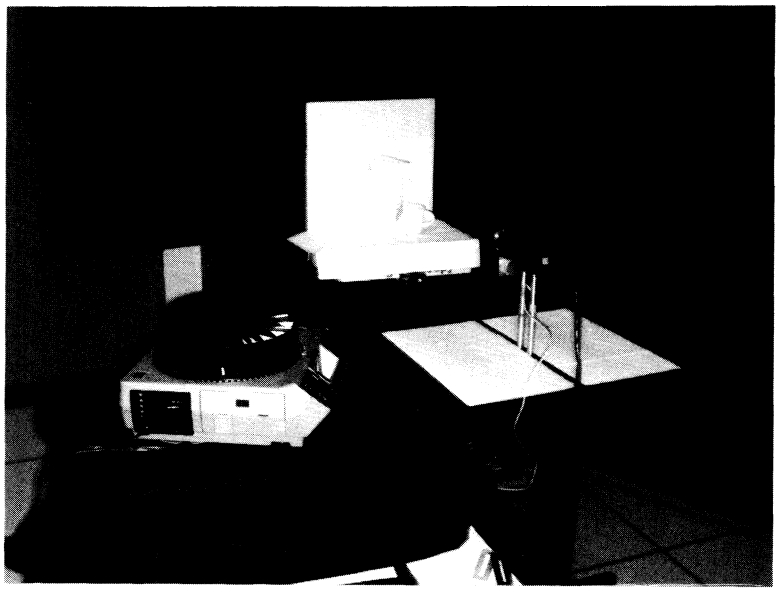


FIG. 3. Experimental setup of the intensity ratio depth sensor.

not vertical, and if x_0 and z_0 are not known accurately, then the pixel dependent quadratics may model more general functions.

The parameters are the coefficients of each pixel's quadratic function, and can be obtained from a quadratic regression at each pixel using several calibration planes placed at fixed, known depths. Thus, suppose at a given pixel (u, v), we observe ratios $\rho_1, \ldots, \rho_n$ at depths $z_1, \ldots, z_n$, respectively. Let $\overline{\rho^k}$ denote $(1/n)\Sigma\rho_i^k$, and $\overline{z\rho^k}$ denote $(1/n)\Sigma z_i\rho_i^k$. It is easy to show that the coefficients $A = A_{u,v}$, $B = B_{u,v}$, $C = C_{u,v}$ of the quadratic that best fits the (ρ_i, z_i) data (in the mean square sense), satisfy the 3×3 linear system

$$\begin{pmatrix} \overline{\rho^4} & \overline{\rho^3} & \overline{\rho^2} \\ \overline{\rho^3} & \overline{\rho^2} & \overline{\rho} \\ \overline{\rho^2} & \overline{\rho} & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \overline{z\rho^2} \\ \overline{z\rho} \\ \overline{z} \end{pmatrix}.$$

This system must be solved separately for each pixel, using the different observed ratio values at each site.

3. EXPERIMENTS

Figure 3 shows a photograph of the workbench and experimental setup for the intensity ratio depth sensor. The slide projector provides the illumination for the two views. The scene is viewed by a Fairchild 3000 CCD camera, which contains 380 by 480 elements, and the resulting video signal is digitized to a 479 by 512 pixel array, with each pixel having eight bits of quantization significance. In the rear of the workbench area, the flat white calibration plane can be seen.

In order to measure the coordinate value u in terms of pixel row and column number, we image a flat grid with score lines 2.7 centimeters apart, at a known depth $z = 57.72$ cm. It turns out that u is quite accurately determined as a linear function of the pixel column number, with $\Delta u = 7.765\text{E} - 4$ cm between successive columns. We define the $u = 0$ line to be the center vertical column (column 256). Likewise, for future reference, we determine the row number to v transformation, and discovered that $\Delta v = 6.826\text{E} - 4$ between rows.

In order to display the transmissivity function of the wedge filter, we place the

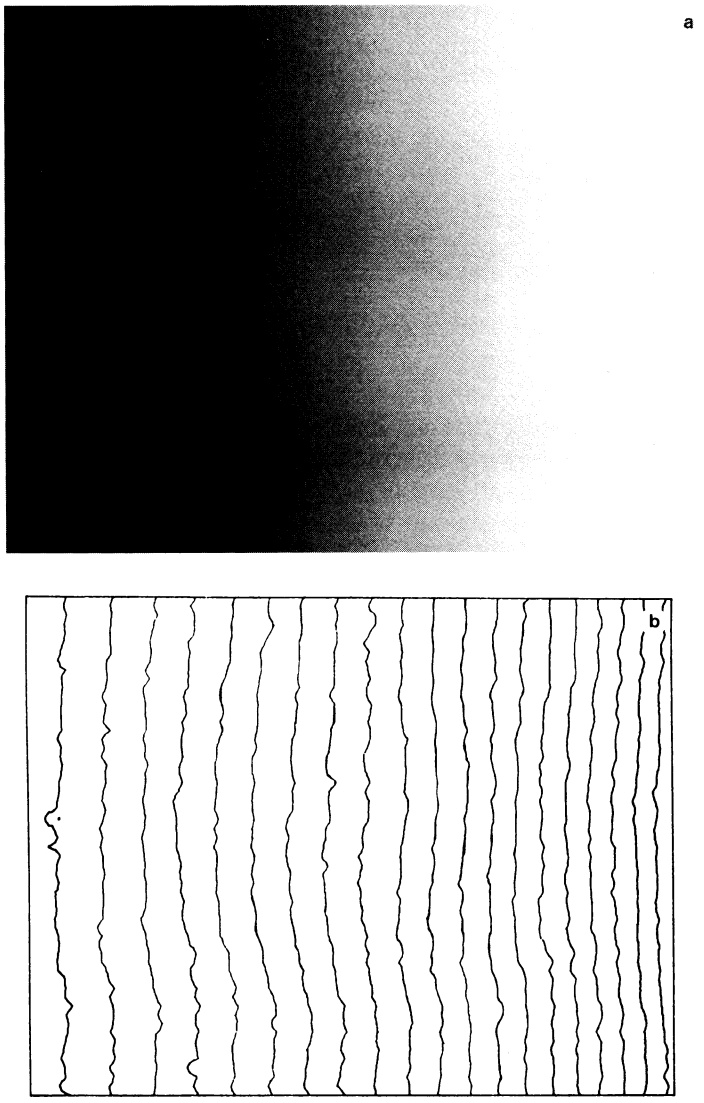


FIG. 4. (a) The ratio of the image formed by the wedge filter to the image formed by the constant filter. (b) The contour plot of the ratio displayed in Fig. 4a. (c) Ratio along a scan line (row 220). (d) Depth at observed ratio values together with the linear regression line (slope $A = 61.965$, intercept $B = 12.201$) .

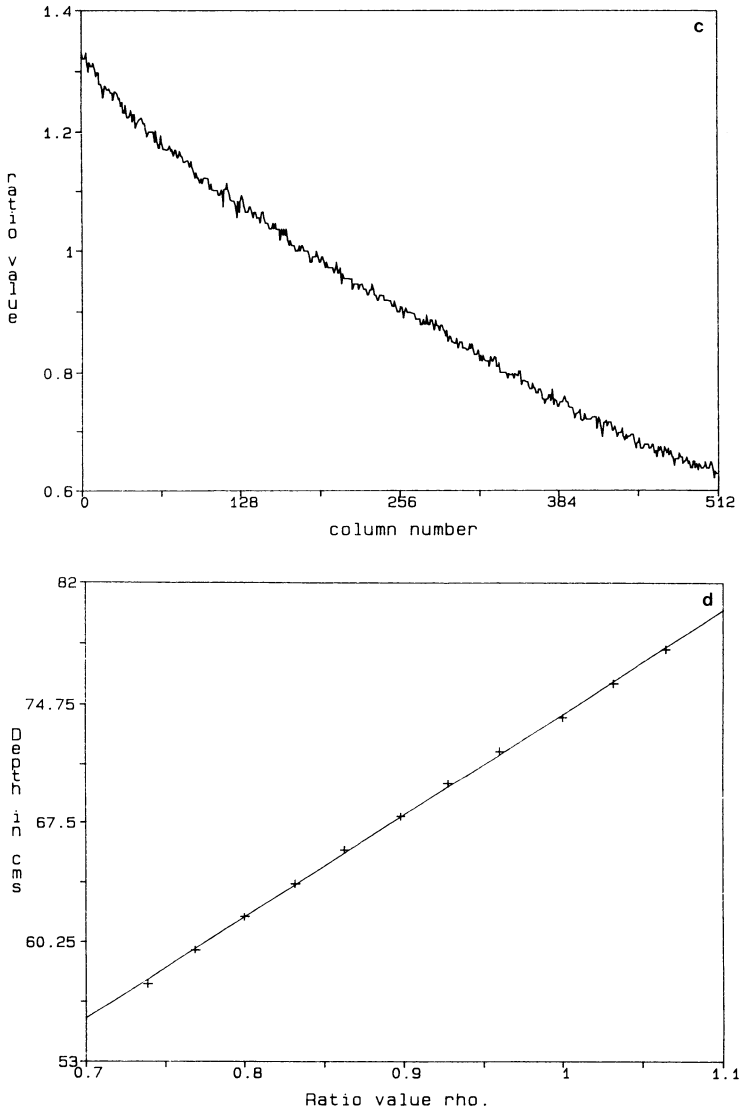


FIG. 4—*Continued*

projector very close to the camera ($x_0 = 0, z_0 = 0$), and image the calibration plane. Figure 4a shows the ratio of the image formed by the wedge filter to the image formed by the constant filter. This can be thought of as the ratio values projected by the illumination system as produced by the two views. Figure 4b is a contour plot of Figure 4a, giving the iso-ratio contours of the projection system. If the ratio of the transmissivities of the two filters is constant in vertical lines, then the constant ratio planes and contours will be vertical (providing the projector is properly aligned). Figure 4c plots the ratio along one of the scan lines. If the transmissivity is linear across the gradient filter and $|(x_0, z_0)| = \infty$, then the intensity ratio plot will be linear, and $d(\rho)$ will then be linear. Even though the function plotted is not perfectly

linear, our Method 1 implementation models $d(\rho)$ as a linear function.

In Figure 4d, we have plotted a number of depth values against ratio values. The ratio values were obtained by averaging the observed ratio values (excepting outliers) along the line $u = 0$ with the calibration plane placed at the corresponding depth. The linear regression line yielding the best mean square fit to this data is also plotted. The slope and intercept of this line form the constants A and B used to model the $d(\rho)$ function, used for the Method 1 experiments.

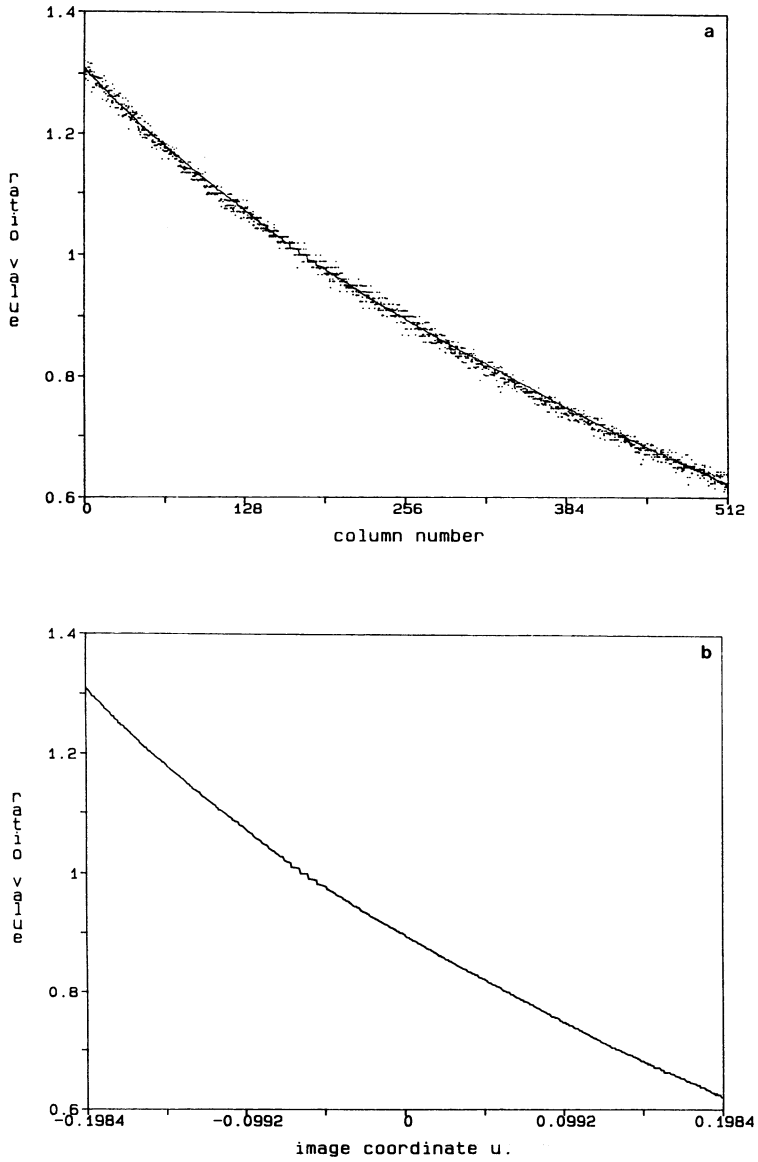


FIG. 5. (a) Ratio data from several scan lines together with the column averaged curve. (b) Ratio ρ as a function of the horizontal coordinate u with column averaging. (c) Piecewise linear monotonic function $\bar{\rho}(u)$.

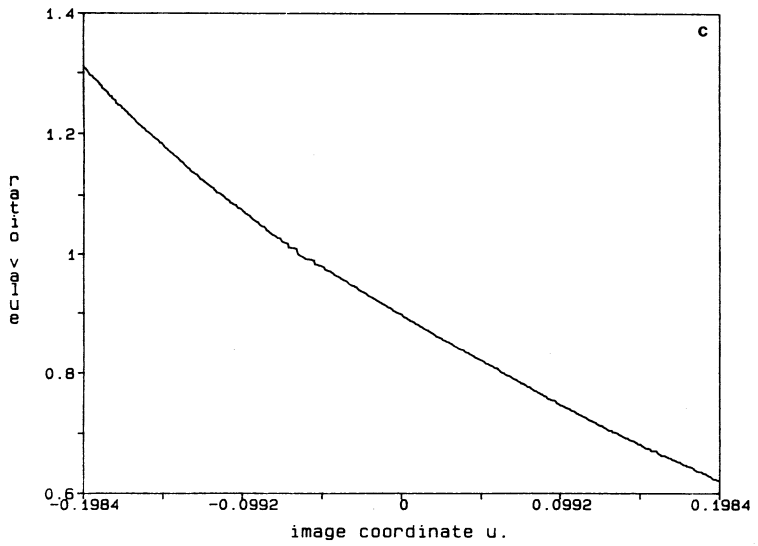


FIG. 5—*Continued*

For the Method 2 calibration, we placed the flat calibration screen at $z \equiv z_1 =$ 57.721 cm and $z \equiv z_2 = 78.041$ cm, to observe ratio images. The ratio images should be constant in columns and monotonic in the image coordinate u, in which case it is easy to assign the $h(\rho, z_1)$ and $h(\rho, z_2)$ functions according to Eq. (3). Suppose that $\rho(u, v)$ is the observed ratio image for $z = z_1$. Figure 5a shows several plots of scan lines of the observed ratio image $\rho(u, v)$. To factor out slight vertical variations, we average non-outliers in columns, and plot the resulting average ratio value $\bar{\rho}(u)$ as a function of the horizontal coordinate in Fig. 5b. Although this function is nearly monotonic, slight local variations must be smoothed to achieve strict monotonicity, which is necessary to assign the $h(\rho, z_1)$ data. We used the simple operation of eliminating $\bar{\rho}(u)$ data whenever there exists an observation $u_1 > u$ such that $\bar{\rho}(u_1) \leq \rho(u)$, resulting in

$$\bar{\bar{\rho}}(u) = \sup_{s \leq u} \bar{\rho}(s)$$

which guarantees a monotonic $\bar{\bar{\rho}}(u)$. Then using Eq. (3),

$$h\left(\bar{\bar{\rho}}(u), z_1\right) = u \cdot z_1$$

we can define $h(\rho, z_1)$ for certain values of ρ (the range of $\bar{\bar{\rho}}$ over the columns), and interpolate linearly to complete the definition of $h(\rho, z_1)$, as shown in Fig. 5c. The process is then repeated for the ratio image for $z = z_2$ in order to define the function $h(\rho, z_2)$.

The x coordinate x_0 and z coordinate z_0 of the projector focal point are required by both Method 1 and Method 2. The point (x_0, y_0, z_0) is calibrated with the use of a patterned filter in place of the filters used for the depth calculation. The flat screen,

placed at a sequence of known depths, is then illuminated by the projector with the patterned filter and an image of the screen is recorded by the camera at each fixed depth. The patterned illumination results in regions of high contrast in the images which can be easily identified and matched between images. Associated with each region there is a point determined by the image coordinates of the centroid of the region and the known depth at which the screen was positioned for the particular image. Every pair of points, corresponding to a single region matched between two images, define a line which passes through the projector focal point. The point

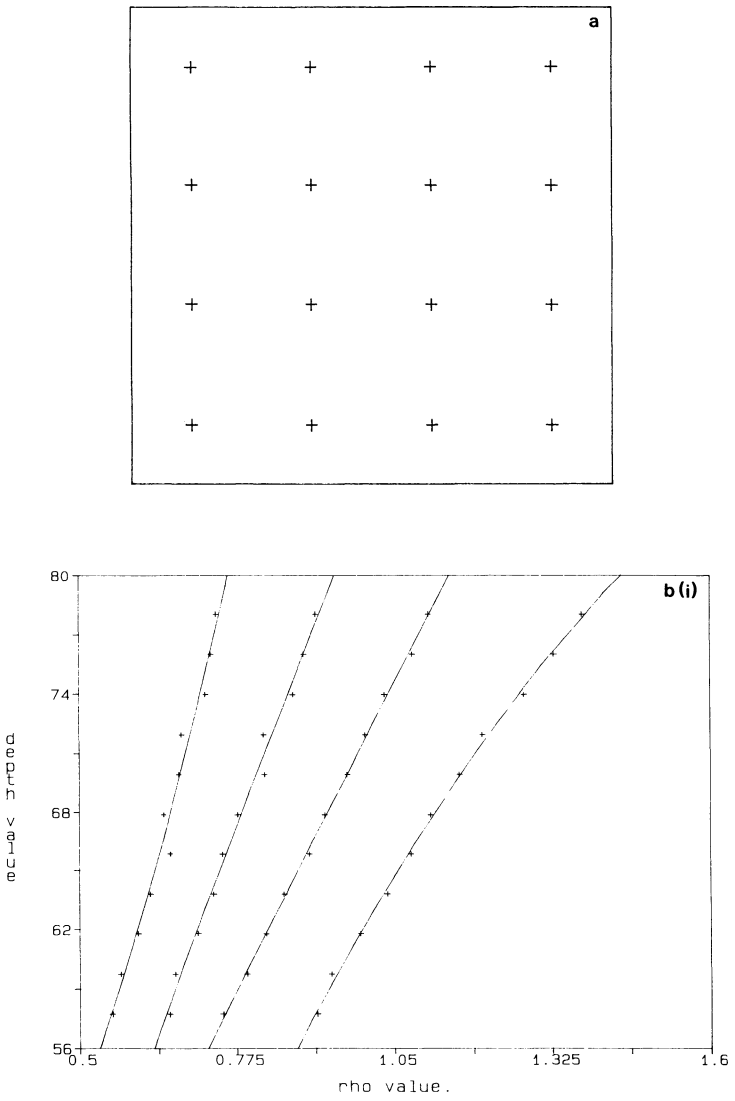


FIG. 6. (a) The pixel locations for which the fitted quadratic functions are plotted in Fig. 6b. (b) The (depth, ratio) data points with the fitted quadratic curves for the sampled pixels. The graphs (i)–(iv) each contain four plots, corresponding (in right to left order) to the four points on a single row in Fig. 6a (from left to right).

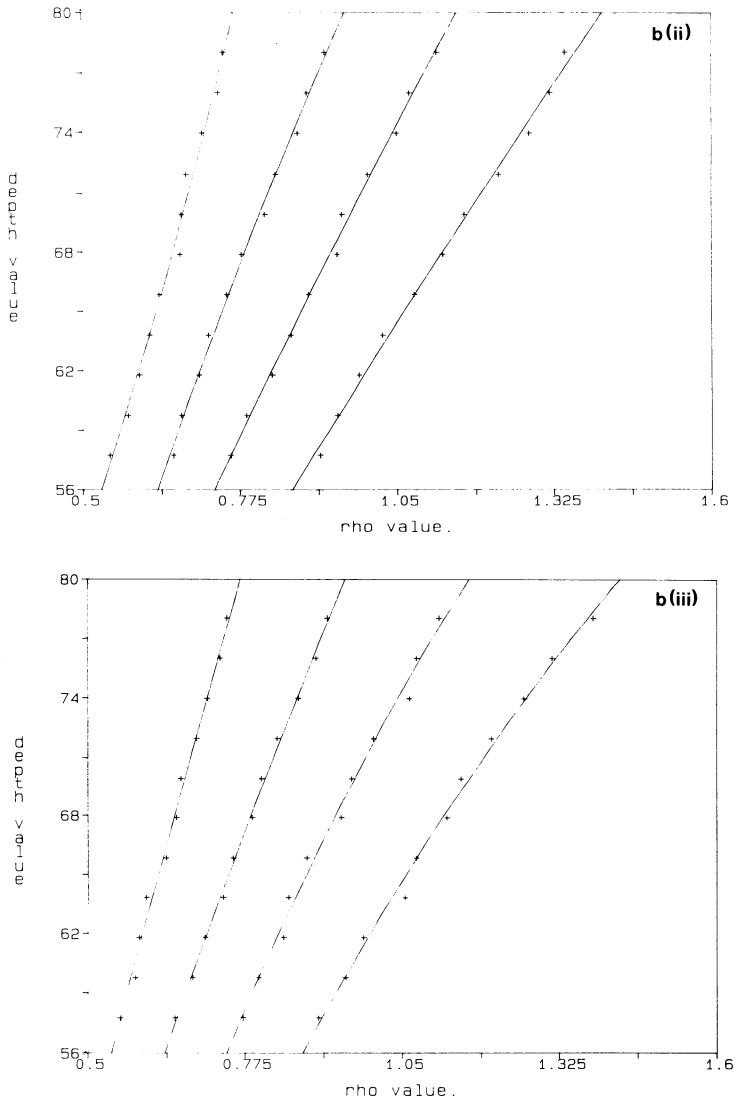


FIG. 6—*Continued*

(x_0, y_0, z_0) is then determined by the intersection of these lines. In our experiments this calibration gave $x_0 = -103.866$ cm and $z_0 = -91.977$ cm.

The Method 3 calibration involves fitting quadratic polynomials to (depth, ratio) data at each pixel. In Fig. 6b, we plot (depth, ratio) pairs at several pixels whose locations are shown in Fig. 6a. The quadratic polynomials fitting this data are superimposed on the plots. We used 11 depth values to obtain corresponding ratio values at each pixel. The coefficients of the quadratic polynomials are stored as three images A, B, and C, and depth values are computed from the ratio image ρ by the full frame image multiply and add operations, associated in the form $(A\rho + B)\rho + C$.

To test the accuracy of the depth calculations using Methods 1, 2, and 3, we manufactured an object containing both flat and curved surfaces, concave and

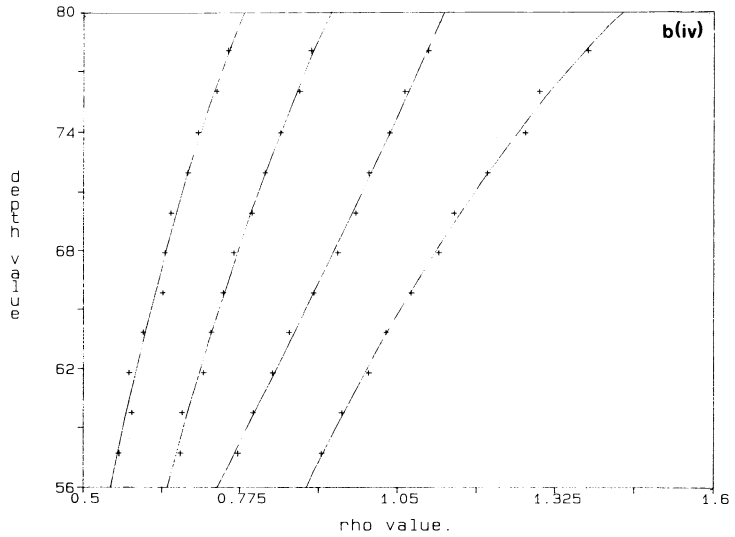


FIG. 6—*Continued*

convex surface edges, and occlusion boundaries, with precisely known shape. The object, shown in a photograph in Fig. 7a, can be positioned accurately on the workbench, so that the true depth values are known. Figures 7b(i, ii, iii) show resulting depth calculations using Methods 1, 2, and 3, respectively, displayed as 3D perspective plots. Calculations were done using floating point arithmetic, using ratio values and parameters represented with full available significance. Subtracting known depth values from the calculated depth values yields absolute errors, displayed in Fig. 7c(i, ii, iii) using a scale which shows larger errors as brighter points. The average absolute errors using Methods 1, 2, and 3 are 2.041, 1.321, 1.246 cm, with no more than 5% of the pixels showing absolute errors greater than 3.658, 2.658, 2.497 cm, respectively.

Relative errors can be defined in various ways. We used the interior of the large flat facing region of the object to define a uniform corrective depth translation which causes the observed depth values to match the known depth (in a minimum mean square sense). The resulting depth values fit the shape of the object somewhat better, with average absolute errors of 0.7168, 0.6972, and 0.6622 cm, with no more than 5% of the pixels having absolute error greater than 1.798, 1.703, 1.665 cm for Methods 1, 2, and 3, respectively. These error values may also be expressed relative to the depth to which objects were measured. In the experiments objects were measured with the depths in a range of 57 to 78 cm. The average relative errors are then computed to be 0.9185%, 0.8934%, and 0.8485%, with no more than 5% of the pixels having errors greater than 2.304%, 2.182%, and 2.134% for Methods 1, 2, and 3, respectively.

Note that the largest errors occur on the borders of the objects, where blurring and misalignment cause large inaccuracies. There are also errors in regions where light reflects off one object surface to illuminate another object surface, such as the far left portion of the sphere. Self illumination causes depth inaccuracies, but is not a problem for single convex bodies, nor for the large flat facing region used to define

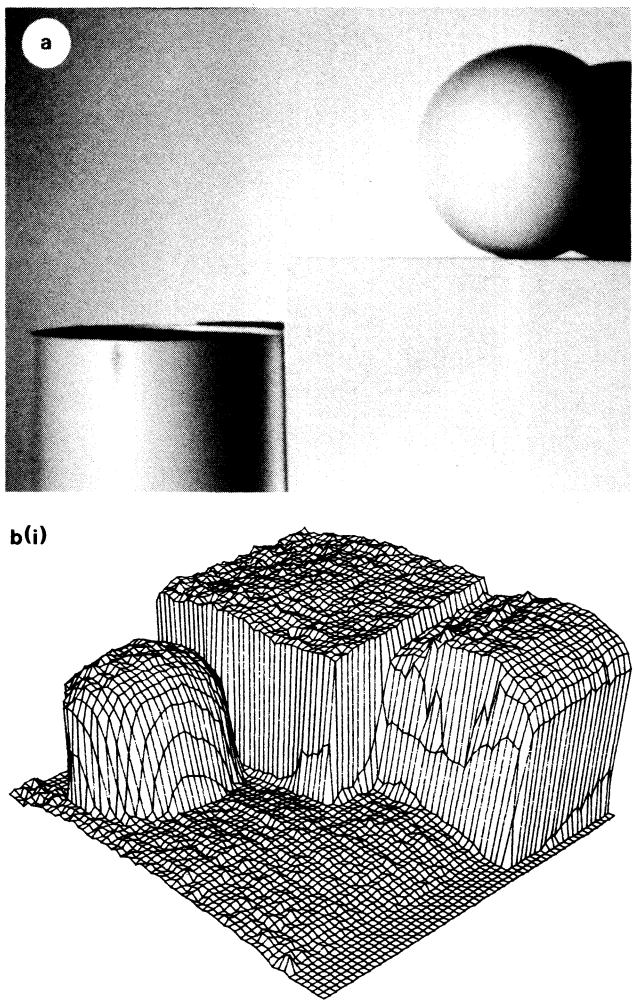
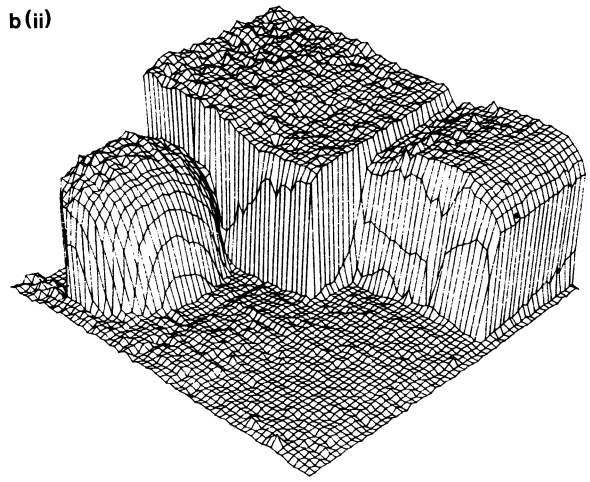
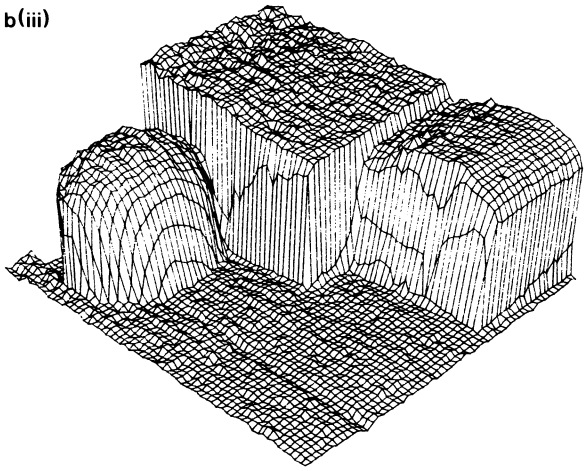


FIG. 7. (a) The reference object used in evaluating the depth measurements. (b) 3D perspective plots of the depth images produced by the three methods ((i) Method 1, (ii) Method 2, (iii) Method 3). (c) Error values of the three methods as grayscale images with brighter points depicting larger absolute errors ((i) Method 1, (ii) Method 2, (iii) Method 3). (d) Shaded surface plot constructed from the depth values derived using Method 2.

the depth translation for relative error determination. Figure 7d shows a shaded surface plot of the resulting depth values using Method 2. The surface is constructed as a mesh of polygonal elements where the vertices of the elements correspond to the pixel locations. The surface is then smoothly shaded using Gouraud [11] shading.

Figure 8a shows a photograph of a 3D scene with some familiar objects. Figures 8b and c show the views from the CCD camera using the wedge and constant filters respectively. Figure 8d shows the ratio image, which is represented using 12 bits of significance in a VICOM image processing system. The VICOM system has a processor capable of forming the ratio image (with 12 bits of significance) at essentially video rate. Using the Method 2 formula, depth values were calculated and





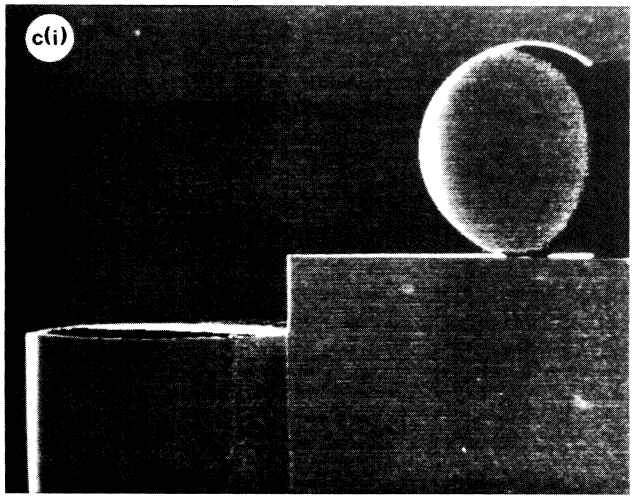


FIG. 7—*Continued*

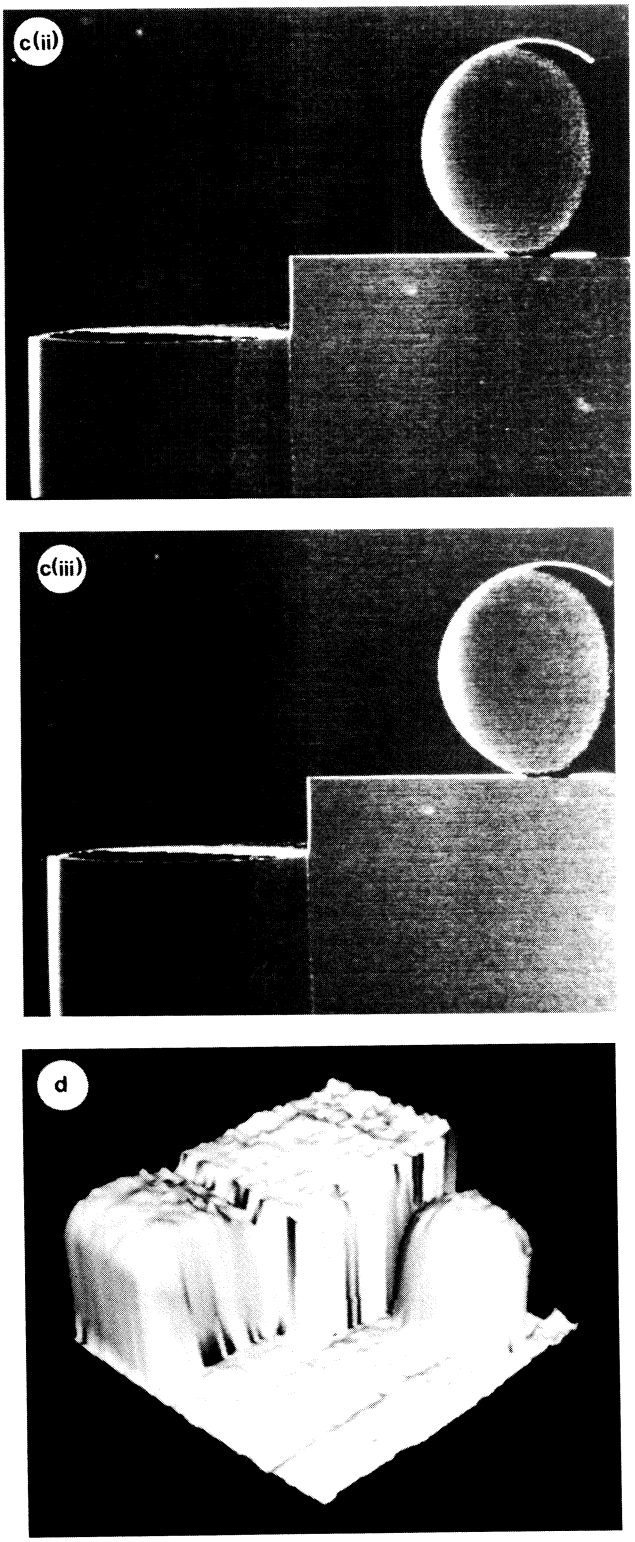


FIG. 7—*Continued*

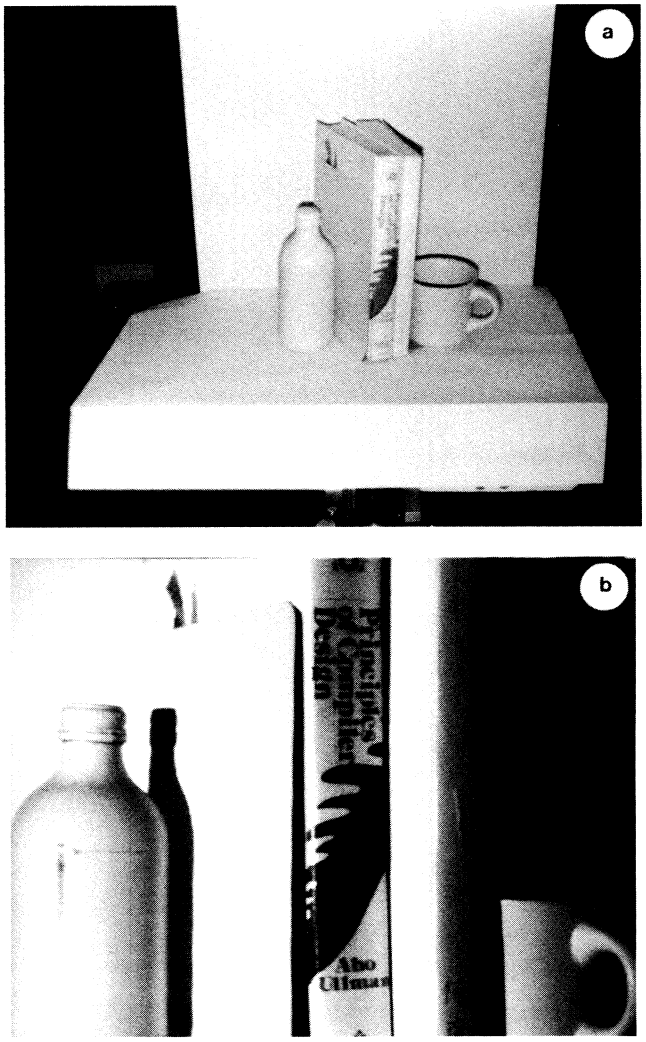


FIG. 8. (a) A scene containing familiar objects. (b) Wedge filter illuminated view of the scene. (c) Constant filter illuminated view of the scene. (d) Resulting grayscale representation of the ratio image. (e) Grayscale representation of the depth image formed using Method 2 (brighter points are closer to the camera). (f) Shaded surface plot constructed in the same manner as Fig. 7d.

displayed as a grayscale image in Fig. 8e. In applying the formula (Eq. (2)), $d(\rho)$ is implemented as a 12 bit in—12 bit out lookup table, which requires one frame time to apply to the entire ratio image, and the remaining arithmetic operations are computed with 12 bit precision using the VICOM pipeline processor. The result is that the complete depth computation requires only several frame times. The depth values are quantized to 4096 levels, and the scale was chosen so that the levels span only the usable workspace area (i.e, the first level is not $z = 0$). In Fig. 8e, the brightest pixels are the closest quantization depths, and more distant points are displayed with arithmetically decreasing intensity.

FIG. 8—*Continued*

Figure 8f shows a shaded surface plot constructed in the same manner as Fig. 7d. Note that the viewpoint is not the same as that of the camera as can be seen by comparison with Fig. 8c. Note also that all surface reflectivity variations have been eliminated by the transformations making use of the depth representation.

4. COMMENTS

The experiments in the previous section demonstrate the feasibility of the intensity ratio depth sensor. It is remarkable that reasonably accurate depth measurements can be made with a system involving only two views and eight bit digitizations of each view. Indeed, in our Fig. 8 experiment, the precision of all calculations is limited to 12 bits, and yet there is still considerable fidelity between the true and perceived shape of objects in the scene.

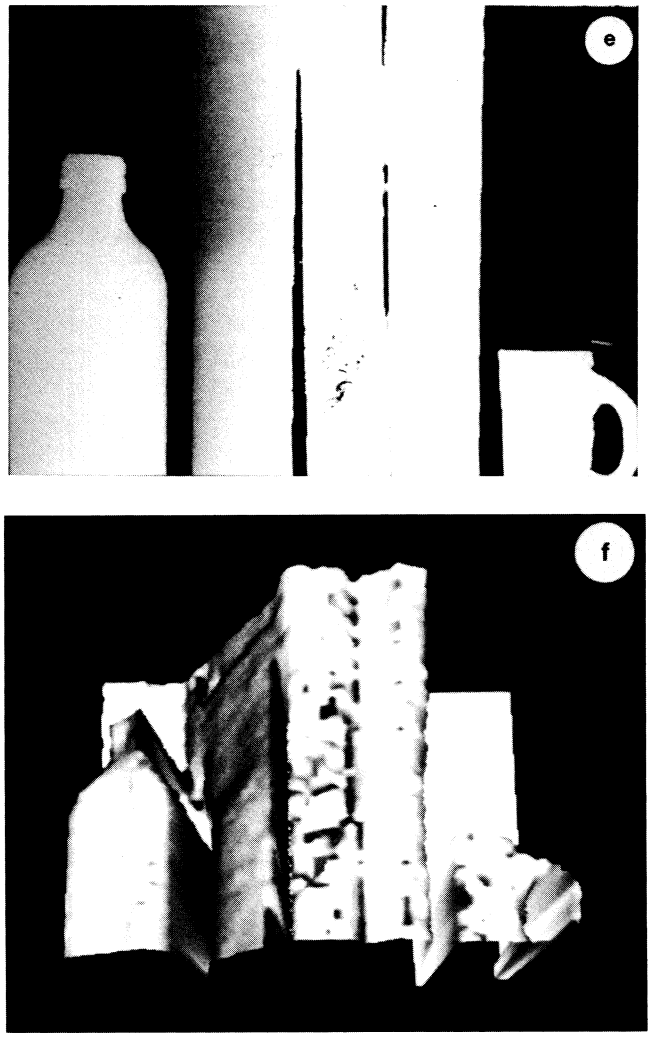


FIG. 8—*Continued*

There are many ways to improve the accuracy of the intensity ratio depth sensor. For example, the use of a third view allows considerable flexibility and greater accuracy. The ratio functions ($z_{u,v}(\rho)$) formed as a result of the additional view are no longer required to be strictly monotonic or single valued in terms of the ratio ρ as long as the pair of ratio functions together uniquely determine the depth. In particular, the use of a "sawtooth gradient" filter with a number of repeated, steep gradient intervals utilized for the third illumination makes effective use of the ratio value range by producing a finer quantization of the wedge of projected ratios. Further, for each extra bit of image intensity quantization, two additional significant bits will be realized in a two view system, resulting in a quadrupling of the depth accuracy. Finally, more carefully manufactured filters, lens projection system, and improved camera optics would yield more accurate results.

The relative position of the objects in the scene may give rise to shadow regions in the acquired intensity images. Shadows result when there are surfaces visible from the camera but are obscured from the projector focal point (due to the presence of occluding surfaces). This phenomenon exists in all structured light triangulation schemes where there is a displacement between the camera and projector focal points. Although the extent of the shadow regions in an image is scene dependent, in general, the larger the displacement the larger the shadow regions. There is a trade-off here between the triangulation calculation accuracy (which increases with the increase in displacement) and the extent of shadows. With the use of a projector in a second location, an additional ratio image (and its associated depth image) may be acquired with considerably different shadow regions. Since the camera position is fixed, the two depth images will be in registration and a simple merging step is sufficient to combine them. With two projector locations, intersecting shadow regions will generally be less in extent.

The most serious limitation of the intensity ratio depth sensor is that certain portions of the scene can reflect light and illuminate other portions of the scene. Although the same reflection properties apply to both the wedge and constant filter illuminations, the net result is that the observed ratio at the multiply illuminated portions of the scene will be some value in between the proper ratio (which results from the direct illumination from the projector), and ratio values from the reflected rays. Depth values are confused by being averaged with depths to points forming the source of the secondary illumination. It is possible that an iterative refinement procedure involving expensive ray tracing algorithms can disambiguate self-illumination intensity ratio errors. We have not pursued this here.

Our Method 1 results are not nearly as accurate as Methods 2 and 3, principally because the wedge filter is far from a constant gradient. The Method 2 table lookup and Eq. (2b) or (2c) computation method works quite well, validating the vertical plane assumption for the constant ratio surfaces. The main disadvantage of Method 2 is that in order to implement a global lookup table on a single-instruction multiple-processor grid array, it is necessary to store the entire table at each site, or conduct an exhaustive "case" statement over all possible input values. The former option is probably too demanding in memory, unless the lookup table is compressed into a few parameters; the latter option may well be too slow in comparison with a pipeline approach. It is certainly possible to apply the global lookup table quickly by processing pixels in raster scan format; although it would seem to limit Method 2 implementations to video rate sensors (and no faster). Method 3 is better suited for parallel processing with separate processors on-chip for each pixel. However, there is a trade-off in Method 3 of approximation accuracy versus computation time. In our experiments, the use of quadratic polynomials led to much more accurate results than using linear functions to calculate depth at each pixel.

Our use of three different methods for computing depth from ratio values suggests a number of variations are possible. In [10], for example, we will consider the formula analogous to Eqs. (2) that arises if the constant ratio surfaces are no longer assumed to lie in vertical planes. Methods for calibrating parameters in the system without assuming that x_0 and z_0 are known are also demonstrated. We chose here three simple methods involving a hierarchy of increasing numbers of parameters. In these experiments, it turns out that the simplifying assumptions leading to Eqs. (2) hold sufficiently well, relative to the inherent limitations in our setup, so that use of the intermediate number of parameters yields essentially peak performance.

ACKNOWLEDGMENTS

The ratio image depth sensor idea is due to Jacob Schwartz, whose encouragement and support is gratefully acknowledged. Development of the idea at NYU [12–13] has been greatly assisted by Marc Bastuscheck and Dayton Clark. Manuscript preparation was patiently performed by Connie Engle.

REFERENCES

1. P. M. Will and K. S. Pennington, Grid coding: A preprocessing technique for robot and machine vision, *Artif. Intell.* **2**, 1971, 319–329.
2. B. K. P. Horn, *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*, MIT, AI Tech. Report 79, Project MAC, 1970.
3. R. Bajcsy and L. Lieberman, Texture gradients as a depth cue, *Comput. Graphics Image Process.* **5**, 1976, 52–67.
4. K. Stevens, *Surface Perception from Local Analysis of Texture and Contour*, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, 1979.
5. D. C. Marr and T. Poggio, A computational theory of human stereo vision, *Proc. Roy. Soc. London B* **204**, 1979, 301–328.
6. R. Nevatia, Depth measurement by motion stereo, *Comput. Graphics Image Process.* **5**, 1976, 203–215.
7. R. A. Jarvis, A perspective on range finding techniques for computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-5**, No. 2, 1983, 122–139.
8. Y. Shirai and M. Suwa, Recognition of polyhedrons with a range finder, in *Proc. 2nd Int. Joint Conf. Artif. Intell.* 1971, 80–87.
9. J. T. Schwartz, *Structured Light Sensors for 3D Robot Vision*, Robotics Research Report No. 8, Courant Institute, New York Univ., 251 Mercer Street, New York, 1983.
10. B. Carrihill, *The Intensity Ratio Depth Sensor*, Ph.D. dissertation, New York Univ., in preparation.
11. H. Gouraud, Computer display of curved surfaces, *IEEE Trans. Comput.* **C-20**, No. 6, 1971, 623–628.
12. D. Clark and R. Hummel, *VSH User's Guide: A Software Environment for Image Processing*, Robotics Research Report 19, Courant Institute, New York Univ., 251 Mercer Street, New York, March 1984.
13. C. M. Bastuscheck and J. T. Schwartz, *Preliminary Implementation of a Ratio Depth Sensor*, Robotics Research Report 28, Courant Institute, New York Univ., 251 Mercer Street, New York, June 1984.