COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 35, 200– 219 (1986)

Calibration of a Camera and Light Source by Fitting to a Physical Model

PETER MANSBACH

Bldg. 220/Room B-124, National Bureau of Standards, Gaithersburg, Maryland 20899

Received December 5, 1984; revised April 1, 1986

This paper describes the calibration of a camera-and-plane-of-light ranging system. Equations are derived which relate the image coordinates in the camera to the external coordinate system. These equations contain coefficients which are functions of the geometrical parameters of the camera/light-source system (focal length, pixel spacing, camera-to-plane-of-light distance, etc.). Several pictures are taken of a test block at different distances, and the geometrical parameters in the equations are varied to achieve a best (least squares) fit to the data. The resulting equations have a remaining error of less than one pixel, and have been used successfully on a parts-acquisition robot. © 1986 Academic Press, Inc.

I. BACKGROUND

Camera calibration is the determination of the correspondence between points in the camera image and points in physical space. Only when the camera is properly calibrated can its 2-dimensional raster coordinates be translated into real-world locations for the objects in its field of view.

In this paper we examine specifically the configuration in which a light stripe is projected on an object, in order to determine the range (distance) to the object. The same calibration techniques may be applied to other geometries as well.

The technique of obtaining range by means of triangulation using a plane of light has been widely discussed (see, e. g., [1–7]). This technique has also been extended to multiple planes [8, 9]. The effort described in [9] in fact provided the impetus for the calibration procedure described in this paper.

As described there, a camera is mounted above the wrist of the robot. A light source is attached below the wrist, and is fitted with a slit and a cylindrical lens, so that the light emanating from it is confined to a plane region of space. An object in the field of this light source is thus illuminated by a narrow stripe or line. The position of this line in the camera's image can be used to obtain the distance to points on the object that are both illuminated by the light and visible to the camera. Knowing this distance (the range, or z-coordinate), the x and y coordinates of points on the object can then be obtained straightforwardly from a fully illuminated image.

First, however, it is necessary for the camera/lighting configuration to be calibrated; that is, for the precise relationship to be established between points on the image and their corresponding coordinates in the external world. The calibration must in principle be performed each time the camera system is demounted or modified. In the past a laborious procedure has been followed of manually ascertaining the entries in several 128-entry tables which were then used as look-up tables by the camera's software.

Other calibration procedures are mentioned in the literature. A straightforward means of calibrating a floodlit image is presented in [10]. In this case only one

0734-189X/86 $3.00

parameter, the scaling factor, was to be determined, and distortion and misalignment were neglected. In [11] a fitting technique was used to determine eight camera/robot parameters by fitting them to four data points. In this case range data were obtained from the phase shift of a modulated laser beam, so that no triangulation parameters were needed. Again, it appears that distortion and misalignment were not included.

In this paper we examine a number of camera parameters, including the location of the lens relative to the plane-of-light, the raster dimensions, lens distortion, raster misalignment, and light-source misalignment, and fit these to a number of data points (between 30 and 150).

In Section II we derive relatively simple expressions which represent the real-world coordinates in terms of the rows and columns of the camera pixels. These equations are expanded in Section III to take into account misalignment and lens distortion. In Section IV the various transformations are summarized, and given in terms of the physical parameters of the camera and its mounting, such as lens focal length, pixel spacing, and distance from the camera to the plane of light. The analogous equations for a fully illuminated object, retaining the distortion and misalignment terms, are also given.

These analytical expressions may be used directly in the run-time coordinate transformation routines, or they may be used offline to generate numerical tables, to allow more rapid computation.

In performing the calibration, rather than attempt to measure these camera parameters with the necessary accuracy, we have chosen to fit the parameterized expressions to a number of measured data points. This is described in Section V. This fitting procedure also assures that the real world data will be well represented by the model.

II. DERIVATION OF THE BASIC EQUATIONS

In this section we derive the basic equations for inferring, from the picture coordinates of an imaged object, the location of the object itself in an external coordinate system. Some assumptions are first made, to simplify the derivation. Next, the z' transformation is derived, and then the x'.

Assumptions

In the following derivation we first assume (1) that the lens behaves as an ideal, thin lens; (2) that the pixel matrix is aligned with the external coordinate system; (3) that the reflecting surface is perfectly diffuse; and (4) that the pixel raster is "ideal."

The ideal thin lens assumption states that the lens can be replaced by a pinhole, as far as determining where the image will be (see [12–15]). The location of this equivalent pinhole is called the optical center of the lens.

The alignment assumption states that the pixel rows must be parallel to the plane of light, and that the columns hence are perpendicular. The diffuse-surface assumption requires that the intensity of the reflected light be independent of reflection angle, and the assumption of an ideal pixel raster requires that the camera be noise-free and of infinite resolution. Effects of violating these assumptions are discussed subsequently, and appropriate corrections added.

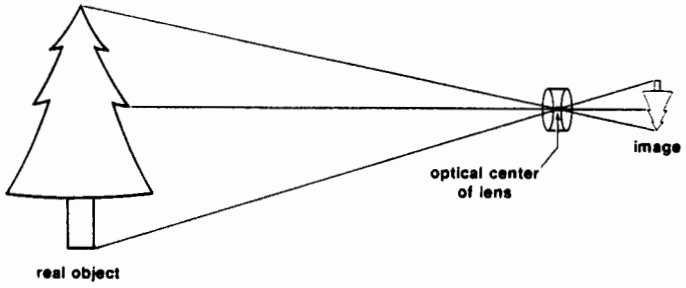

FIG. 1. Image formation for an ideal, thin lens.

Coordinate Systems

As indicated schematically in Fig. 1, all rays from a given field point are assumed to pass through the optical center and to continue in a straight line to the image plane. Figure 2 shows a similar scenario, with an arbitrary field point labelled (x, y, z), and its corresponding image point (g, h). The y values along the object are also indicated, and values of h in the image plane are shown. Note that since the lens inverts the image, h increases downward.

Image Coordinates. Points in the real image are identified by their distances (g, h) from the image center, where the g axis on the image is parallel to the x axis in the external world, and h is aligned with y. In the image, g distances are measured "horizontally," and h "vertically."

Pixel Coordinates. The pixels within the camera are discrete, and are identified by integer pairs (i, j) giving their column and row number, respectively, from the center of the raster. Define s to be the pixel spacing (the distance from the center of one pixel to the center of the adjacent pixel) in the g direction ("horizontally"), and t the spacing in the h direction ("vertically"). Then the image coordinates (g, h) of the center of a pixel are related to the integer pair (i, j) by $g = s * i$ and $h = t * j$. In the remainder of this section we derive expressions for determining the real-world coordinates of a point within the plane of light, given its position (g, h) in the image.

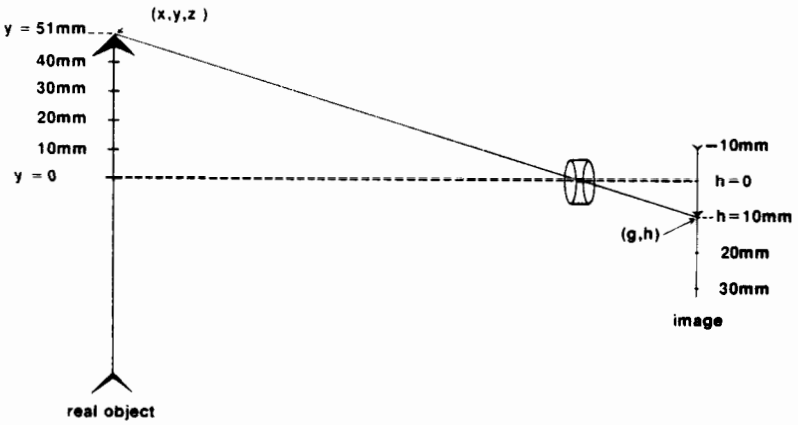

FIG. 2. Image formation, showing the relationship between (external) camera coordinates (x, y, z) and image coordinates (g, h).

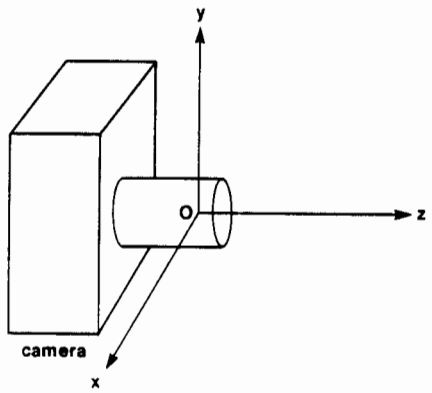

FIG. 3. Definition of camera coordinates.

Camera Coordinates. Two real-world (external) coordinate systems are used in the following derivation. The first is the natural coordinate system of the camera. In this system the z axis lies along the camera axis, with the origin at the optical center of the lens, and the positive direction outwards into the external world (Fig. 3). The x axis is chosen parallel to the pixel rows, and the y axis is perpendicular to the other two.

The positive direction of these axes is taken to be to the right (for x) and upwards (for y), when seen looking outward from the lens at the external world. It is noted that this choice of orientations results in a left-handed coordinate system.

Plane-of-light Coordinates. The second coordinate system is based on the projected-light source, and will be denoted with primes: (x', y', z') (Fig. 4). The z' axis is located within the plane of light, directly under the camera axis (the z axis defined

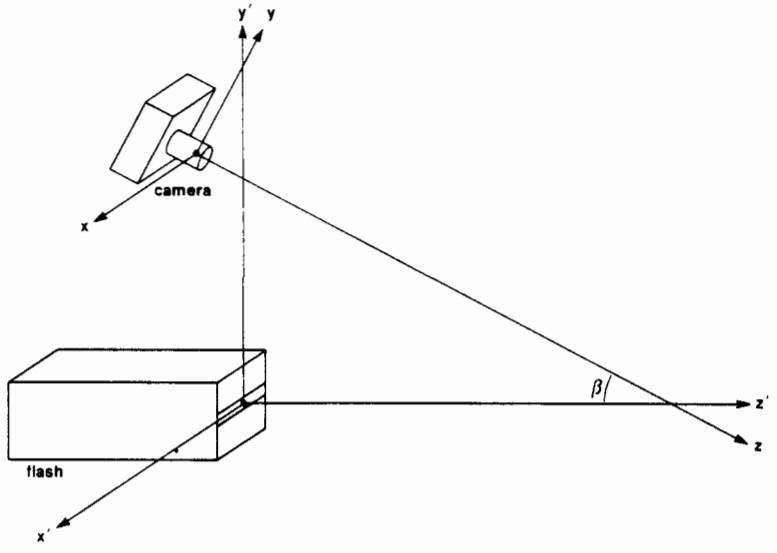

FIG. 4. Definition of plane-of-light coordinates and their relationship to camera coordinates.

above). The origin of the plane-of-light system has been chosen arbitrarily to lie at the face of the light projector. The x' axis also lies within the light plane, whereas the y' axis is perpendicular to it.

Since the pixel rows are parallel to the plane-of-light, the x and x' coordinates of any point will be the same. The z and z' axes intersect at a point called the "crossing point," and form an angle β (see Fig. 4). Note that one coordinate system can be obtained from the other by a translation and a rotation of angle β about the x axis.

z' Coordinate (Range)

In this subsection we derive an expression for the range z' as a function of image position (i, j) for points illuminated by the light. Because the light is confined (by the projector) to the $y' = 0$ plane, we know a priori that $y' = 0$ for any illuminated field point.

Figure 5 shows the camera and light source as seen from the side. The plane of the figure is described by $x' = 0$. The axis of the camera intersects the plane of light at the crossing point, labelled K. The crossing point, K, is at a distance C from the origin of the plane-of-light (primed) coordinate system. The optical center of the lens is labelled O, and is a distance d above the light plane. In this figure the field point $P(x', y', z')$ lies on the z' axis ($x' = 0$). (Because it is in the plane of light, $y' = 0$ also, as noted above.) For a distortion-free lens, properly aligned, z' is a function of j only and the $x' = 0$ result applies also for arbitrary x'.

Figure 6 shows an enlarged view of the camera portion of Fig. 5; f is nominally the focal length of the lens, but is actually the distance from the optical center of the lens to the image plane. (This distance is exactly the focal length only if the lens is focussed at infinity.) h is the distance of the image point from the center of the image in the y' direction. We have immediately

$$\tan \alpha = h/f. \tag{1}$$

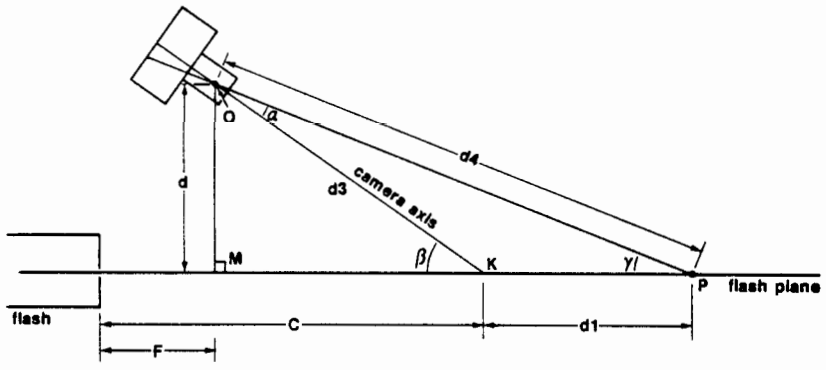

FIG. 5. Camera and plane-of-light geometry seen from the side, for an illuminated point P in the $x' = 0$ plane. OK is the optical axis of the camera. The plane of the drawing is defined by $x' = 0$.

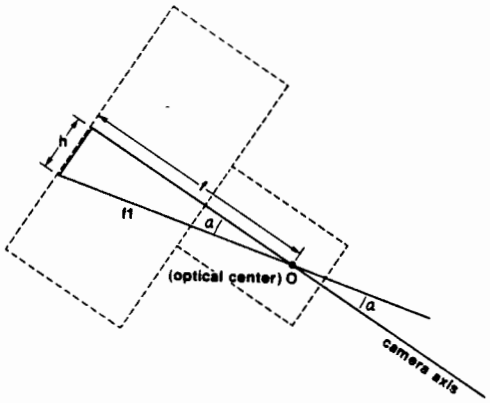

FIG. 6. Enlarged view of the camera geometry of Fig. 5.

Applying the law of sines to the exterior triangle OKP in Fig. 5, we have

$$\frac{\sin\alpha}{d_1} = \frac{\sin\gamma}{d_3}, \tag{2}$$

where γ is the angle made by the ray OP with the plane of light, and $d_3 = KO$ is the (fixed) distance of the crossing point from the optical center. $d_1 = KP$ is the unknown z' distance (positive or negative) of the illuminated object from the crossing point, and $d_1 + C$ is the object's z' coordinate. Note that $\gamma = \beta - \alpha$, so that

$$\sin\gamma = \sin\beta * \cos\alpha - \cos\beta * \sin\alpha. \tag{3}$$

Thus

$$\begin{aligned} d_1 &= \frac{d_3 * \sin\alpha}{\sin\beta * \cos\alpha - \cos\beta * \sin\alpha} \\ &= \frac{d_3 * \tan\alpha / \sin\beta}{1 - \operatorname{ctn}\beta * \tan\alpha} \\ &= \frac{(d_3 * \csc\beta / f) * h}{1 - (\operatorname{ctn}\beta / f) * h} \end{aligned} \tag{4}$$

In terms of raster units, where s and t are the horizontal and vertical pixel spacing, respectively, the image point (g, h) is given by the integer pair (i, j), where $g = s * i$ and $h = t * j$. Then

$$d_1 = \frac{A * j}{1 - B * j} \tag{5}$$

where

$$\begin{aligned} A &= d_3 * (t/f) * \csc\beta \\ &= d * (t/f) * (\csc\beta)^2 \\ B &= (t/f) * \mathrm{ctn}\,\beta. \end{aligned} \tag{6}$$

(We have defined d to be the height of the lens center above the plane of light.) Thus

$$z' = \frac{A * j}{1 - B * j} + C, \tag{7}$$

where C is the z' coordinate of the crossing point. Equation (7) above applies for any value of x'.

Note that the expression for z' is singular at some value of j for which $(1 - B * j)$ is zero. This corresponds to the vanishing point, the point in the image for which z' approaches infinity. In fact, it is possible to determine the constant B solely from observing the limiting pixel as z' becomes very large.

Roughly speaking, one may think of A as a measure of the magnification of the lens ($\Delta j/\Delta z'$) at the center of the field. B is the location of the vanishing point, and C is the z' coordinate of the crossing point.

x' Coordinate

In this subsection we examine points $P(x', y', z')$ for which x' need not be zero (although of course $y' = 0$ because the point lies within the plane of light); z' is still given by Eq. (7) above and we now derive an expression for x'.

Determination of the x' coordinate in terms of the pixel number i is conceptually simpler than the z' case, and is shown in Fig. 7. The two triangles in the figure are similar, and we have immediately $g/f_1 = x'/d_4$, or

$$x' = (d_4/f_1) * g; \tag{8}$$

d_4 is the distance from the origin of the camera coordinate system to the point P_0, which is the projection of the point P onto the z' axis. (Figure 5 also shows d_4 seen from the side, for the case $x' = 0$.) Also, f_1 is the distance from the optical center of the lens to the image point $(0, h)$, which is the projection of the arbitrary point (g, h) onto the h axis. Referring to Fig. 6, we have

$$f_1^2 = f^2 + h^2, \tag{9}$$

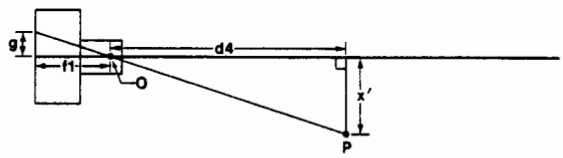

FIG. 7. Camera geometry seen from above, for an arbitrary illuminated point P. PO is the projection of P onto the z' axis. The plane of the drawing is defined by P, PO, and O.

so that in general f_1 differs from f [16]. In Fig. 5 the triangle OMP is a right triangle, so that

$$d_4^2 = d^2 + (z' - F)^2, \tag{10}$$

where F is the z' coordinate of the optical center O.

Combining these equations, and expressing the image distance g in raster coordinates as $s * i$, gives

$$x' = D * i * \sqrt{E + [z' - F]^2} \Big/ \sqrt{1 + [h/f]^2}, \tag{11}$$

where

$$\begin{aligned} D &= s/f \\ E &= d^2 \\ F &\quad \text{is the } z' \text{ coordinate of the optical center of the lens.} \end{aligned} \tag{12}$$

Note from Fig. 5 that $C = F + d * \operatorname{ctn} \beta$.

In physical terms, D, together with the square root factors, is the magnification in the horizontal (x') direction, and depends on z'. E is the (square of the) separation between camera and plane of light, and F results from the arbitrariness in the choice of origin of the z' coordinate. (If we required the origin of z' to be at the optical center of the lens, there would be no need for F).

III. DISCUSSION OF ASSUMPTIONS

The derivation above proceeded under several simplifying assumptions. In practice, the departures from these assumptions contribute measurably to the errors in the final data. In this section we develop corrections to be applied when the assumptions are not strictly valid. The deviations explored are (1) distortion, (2) nodal separation, (3) raster rotation, (4) raster translation, and (5) camera/light-source misalignment. Also discussed are raster misalignment, diffuseness, noise, and quantization. The complete transformation is then summarized in the following section.

1. Distortion

We assumed above that the lens was an ideal thin lens, so that the image of an object point could be found by drawing a straight line from the object, through the optical center, to the image plane. In practice the lens is neither ideal nor thin.

If a lens is not ideal, this means that not all points in space map through the same optical center. Thus rays converging on a point near the edge of the picture may pass through a point somewhat closer to (or further away from) the image plane, than rays converging on the center of the picture. It is this effect which is known as distortion.

In the following, we take (g, h) to be the coordinates of the actual image point, and (g_p, h_p) to be the image point coordinates that would obtain with a "perfect" (i. e., non-distorting) lens of the same focal length.

Clearly, since the lens is axisymmetric, the distortion is also axisymmetric. Expressed in terms of polar coordinates (r, θ) on the image, the distortion can

depend only on r, not on θ:

$$r = \text{function}\,(r_p),$$

where $r_p = g_p^2 + h_p^2$. The lowest-order distortion component present is the r^3 term [12]; its coefficient will be referred to as G.

The effect of this term on the image is

$$r = r_p + G * r_p^3$$

so that

$$\begin{aligned} g &= g_p * \left(1 + G * r_p^2\right) \\ h &= h_p * \left(1 + G * r_p^2\right). \end{aligned} \tag{13}$$

If G is negative, the distortion is called barrel distortion; if positive, pincushion distortion. The appearance of a rectangular grid, seen through a lens with pincushion distortion, is sketched in Fig. 8.

Consider a straight object lying parallel to the x axis. In the image, using a perfect lens, the object would again be straight, and parallel to the g axis. For a real lens, however, the object appears slightly curved. Its deviation from straightness is described by the value of its h coordinate near the edge of the image relative to its value at the h axis: h (at g) $-$ h (at center). This deviation is therefore

$$h - h\ (\text{at center}) = G * h * g^2. \tag{14}$$

Such distortion was observed in three of the lenses used in this study.

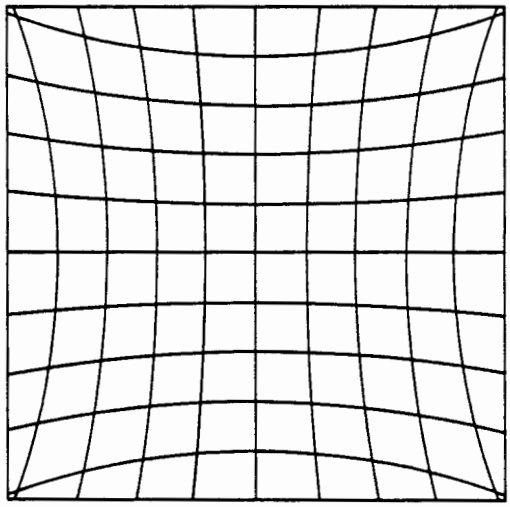

FIG. 8. Rectangular grid as seen through a lens with pincushion distortion.

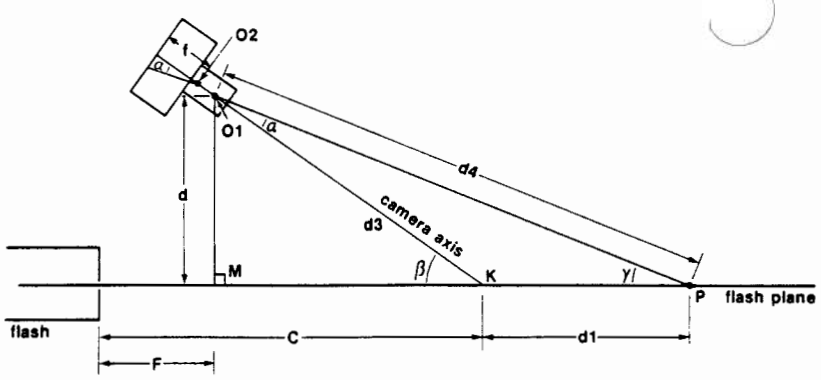

FIG. 9. Camera and plane-of-light geometry (see Fig. 5) for a lens which is not thin. O_1 and O_2 are respectively the front and rear nodal points of the lens.

2. Nodal Separation

If a lens is not thin, there are two optical centers, or nodal points, rather than one. Rays incident on the first nodal point emerge from the second at the same angle, as shown in Fig. 9. For many photographic lenses, the distance N (nodal separation) between the optical centers is considerably smaller than the physical thickness of the lens. Moreover, the effects of this separation on the image are proportional to N/z, and thus approach zero for points far from the camera.

For cases where the nodal separation N cannot be ignored, the derivation of Eqs. (7) and (11) proceeds much the same as before, and these equations are still valid. Now, however, all external distances (such as d and F) are defined relative to the front nodal point, whereas the lens' focal length f is measured from the rear nodal point. The following relationships apply:

$$d_{\text{front}} = d_{\text{rear}} - N * \sin\beta$$
$$F_{\text{front}} = F_{\text{rear}} + N * \cos\beta.$$

3. Raster Rotation

If the pixel raster is not aligned with the outside coordinate system, the image will appear to be rotated. For a rotation angle H,

$$\begin{aligned} g &= g_p * \cos H - h_p * \sin H \\ h &= g_p * \sin H + h_p * \cos H, \end{aligned} \tag{15}$$

where (g, h) are the actual image coordinates, and (g_p, h_p) are those one would have achieved with perfect alignment. For small angle H (rad) the first-order correction is

$$\begin{aligned} g &= g_p - H * h_p \\ h &= h_p + H * g_p. \end{aligned} \tag{16}$$

4. Raster Translation

In addition to the possibility of image rotation, there is the possibility of image translation, due to the center of the raster not coinciding with the lens axis. Writing

(k, l) for the translation distances in pixel units in the directions of g and h, respectively, the effect of such a translation on the "perfect setup" coordinates is simply

$$\begin{aligned} g &= g_p + k * s \\ h &= h_p + l * t. \end{aligned} \tag{17}$$

The pixel located on the optical centerline is thus offset by (k, l) from the center of the raster. The parameters (k, l) (offset from the center) have been re-expressed as (K, L), an offset from the lower left-hand corner of the raster. (Here $K = k + MIDI$, and $L = l + MIDJ$, where $(MIDI, MIDJ)$ are the pixel coordinates of the raster center relative to the lower left-hand corner.)

5. Camera / Light-Source Misalignment

The camera and light projector are held in their relative positions by a bracket, and this bracket forms the natural reference for the plane-of-light coordinate system. As a result of machining and assembly imperfections, the bracket we used was not perfectly aligned with the optical axis of the camera. Two additional parameters, I and J, were added to account for this misalignment. I is the x' coordinate of the lens center, while J is the angle between the x and x' axes, measured about the y' axis. These have no physical significance in the camera, but should be viewed as an artifact of the (plane-of-light) coordinate system, arising from choosing the light projector axis and the camera axis independently.

The true values z_t' and x_t' are given (to first order in J) by

$$\begin{aligned} x_t' &= x' + J * z' + I \\ z_t' &= z' - J * x'. \end{aligned} \tag{18}$$

6. Raster Misalignment

It has also been assumed that the pixel raster is perpendicular to the optical axis of the lens. Slight departures from perpendicularity would manifest themselves in yet additional correction terms to be applied. Much of this additional correction can be absorbed by other parameters already defined. Perhaps for this reason, a good fit with data was obtained in the configurations discussed in Section V, and these additional terms were not explored.

7. Departure from Surface Diffuseness

If a surface is perfectly diffuse, the intensity of reflected light is independent of reflection angle, i. e., light is reflected equally in all directions. When this is not the case, some parts of the object appear brighter than others. When the image is thresholded into black and white, the brighter parts of the object may appear larger than the dimmer parts, and the centroid may thus be incorrectly determined. This problem was sidestepped during calibration by using a small object, for which the reflection angle was essentially constant. For large objects, care in thresholding is required.

8. Camera Noise

Noise arising in the camera hardware appears in the image as occasional pixels being lighter or darker than the actual illumination would dictate. This is manifested

in the thresholded image by occasional white pixels in a black region, or vice versa. By rejecting all areas of single pixel size, the effect of remaining noise was kept small. Further, the noise appears to be random, so that it primarily affects the scatter of the data, and not their mean values.

9. Quantization

The above transformations were derived as though pixel coordinates could take on any value. In fact, only integer values are allowed, corresponding to the camera's discrete pixels in the image plane. Points in the image plane can only be reported to lie at one pixel or at the adjacent pixel, and never in between.

A proper analysis of the effects of quantization is beyond the scope of this paper. In general, however, for objects whose dimensions exceed several pixels in each direction, as ours did, quantization may be expected to increase the scatter of the data, while not introducing any appreciable systematic errors into the results.

IV. SUMMARY OF TRANSFORMATIONS

The full transformation equations, including the correction terms discussed above, are collected here. Given a pixel at column i, row j measured from the lower left pixel in the image plane, the following expressions determine the actual field coordinates x' and z'.

It is useful to recognize that there are three distinct physical "worlds" encountered in this calibration problem. These are the external or "real" world, the physical image, and the camera raster.

Coordinate systems are defined for each of these worlds as follows. The external world is described in terms of camera coordinates x, y, and z, and also in terms of flash (light-projector) coordinates x', y', and z' as described earlier in this paper. The physical image is described by the physical image coordinates g and h. These coordinates have their origin on the camera axis, and have physical dimensions (millimeters, in this paper).

The raster, by contrast, is taken to have no real physical distances. It is more like a template from which monitor displays of any size can be constructed. Thus, distances are expressed in terms of the number of picture elements (pixels) between two points. The raster coordinates are i and j, and their origin is at the lower left of the raster.

Thus there are really two transformations taking place. One is between the external world and the physical image, and is determined by the lens and by the camera and light projector geometry. The other is between the physical image and the pixel raster. This later includes effects of the raster geometry, and raster position and alignment relative to the lens. In this summary, care has been taken to keep these two transformations separate. The full transformation is:

Raster coordinates (i, j) to physical image (g, h):

Raster translation:

$$\begin{aligned} i_1 &= i - K \\ j_1 &= j - L. \end{aligned} \tag{19}$$

Raster rotation:

$$\begin{aligned} i_2 &= i_1 + H * j_1 \\ j_2 &= j_1 - H * i_1. \end{aligned} \tag{20}$$

Pixel spacing:

$$\begin{aligned} g &= i_2 * s \\ h &= j_2 * t. \end{aligned} \tag{21}$$

Physical image (g, h) to external world (x', z'):

Lens distortion:

$$\begin{aligned} r &= \sqrt{g^2 + h^2} \\ g_1 &= (1 - G * r^2) * g \\ h_1 &= (1 - G * r^2) * h. \end{aligned} \tag{22}$$

Basic transformations:

$$z_1' = \frac{A * h_1}{1 - B * h_1} + C, \tag{23}$$

$$x_1' = D * g_1 * \sqrt{E + [z_1' - F]^2} \Big/ \sqrt{1 + [h_1/f]^2}, \tag{24}$$

Axis misalignment:

$$\begin{aligned} x' &= x_1' - J * z_1' - I \\ z' &= z_1' + J * x_1', \end{aligned} \tag{25}$$

where x' and z' are the coordinates actually measured in the plane-of-light coordinate system. In the above,

$$\begin{aligned} A &= d_{\text{front}} * (\csc \beta)^2 / f. \\ B &= \operatorname{ctn} \beta / f. \\ C &= F_{\text{front}} + d_{\text{front}} * \operatorname{ctn} \beta \\ D &= 1/f \\ E &= d_{\text{front}}^2 \end{aligned} \tag{26}$$

F_{front} is the z' coordinate of the front nodal point of the lens.

The suffix "front," used above, means that the relevant parameter is measured from the front nodal point. Thus

$$\begin{aligned} d_{\text{front}} &= d_{\text{rear}} - N * \sin \beta \\ F_{\text{front}} &= F_{\text{rear}} + N * \cos \beta. \end{aligned} \tag{27}$$

Also

G is distortion (in units of inverse length squared),

H is the rotation angle of the pixel raster about the optical axis (radians),

I is the x' coordinate of the front nodal point of the lens center,

J is the angular misalignment between x' and x (radians),

K is the raster translation in the i direction (in pixels),

L is the raster translation in the j direction (in pixels),

and

N is the nodal separation of the lens, i. e., the distance from the frontmost to the rearmost principal nodal points.

In the above,

β is the angle between the optical axis and the light plane

d_{front} is the y' coordinate of the front nodal point of the lens (distance between the lens and the light plane)

F_{front} is the z' coordinate of the front nodal point of the lens.

s is the pixel spacing in the i direction (column spacing, in mm/pixel)

t is the pixel spacing in the j direction (row spacing, in mm/pixel)

f is the focal length of the lens (as used), i. e., the distance of the rear nodal point from the image plane.

Note that (I, d, F) are the (x', y', z') coordinates of the front nodal point of the lens in the external coordinate system of the light source. Also, (K, L) are the (i, j) coordinates of the lens centerline in pixel coordinates (K and L in pixels, and the origin $(i = 0, j = 0)$ at the lower left).

Many of the above parameters may be measured approximately by hand. More accurate determinations of these parameters may be obtained by fitting to data. This is discussed in the next section.

Note that the parameters determined above using the plane of light also determine the magnification (as a function of z) for fully illuminated images, so that no additional calibration is required. Thus, for an arbitrary point (x, y, z), in *camera* coordinates, and for a perfect lens and camera,

$$\begin{aligned} x &= (z/f) * s * i \\ y &= (z/f) * t * j. \end{aligned} \tag{28}$$

More generally, the raster (i, j) to physical image (g, h) transformation is the same as for the line flash, Eqs. (19)–(21). This is so because the size and alignment of the raster with the physical image in no way depends on the external lighting. Similarly, the correction for lens distortion, Eq. (22), is the same, since the lens has not changed. The remainder of the image-to-external-world transformation is simply the magnification due to the geometry, viz.

$$\begin{aligned} x &= (z/f) * g_1 \\ y &= (z/f) * h_1. \end{aligned} \tag{29}$$

Eq. (29) replaces Eqs. (23) and (24) of the projected light case above. Note that the corrections for axis misalignment, Eq. (25) in the projected light scenario, are absent here, since there is no additional light source coordinate system.

V. CALIBRATION OF THE CAMERA

In this section we summarize the data collection and parameter fitting techniques used on several camera/light-source configurations.

Data

The calibration procedure itself entails careful measurement of a number of field points together with the corresponding image points. To obtain such measurements, a small rectangular block was placed at a number of different positions in the camera's field of view, and was illuminated by the plane of light as described above. The location of the center of the block's face relative to the light projector was measured by hand. Its position in the image plane was taken to be the centroid of the largest blob in the thresholded image. This approach was found to yield consistent data insensitive to small amounts of noise in the image. Using the centroid reduces the effects of quantization mentioned earlier.

Parameters

The six basic physical parameters, defined in the derivation above, are

s—pixel spacing in i direction ("horizontally")

t—pixel spacing in j direction ("vertically")

f—lens focal length (as actually focussed)

β—angle between optical axis and plane of light

d_{front}—height of the front nodal point of the lens above light plane (y' coordinate of optical center)

F_{front}—z' coordinate of the front nodal point of the lens.

(Note that the transformation depends on the actual distance of the optical center of the lens from the image plane. Since changing the focus changes this distance, and hence the magnification of the lens, it is necessary to set the focus of the lens once at calibration time, and not to refocus later. In order to have a sharp picture of nearby objects, where accuracy is most important, we set the focus of the lens at its minimum distance—an easily repeatable setting.)

The six basic parameters can all be measured approximately. In particular, f is nominally the focal length engraved on the lens mount. The rear nodal point of the lens is located on the lens axis at a distance f from the pixel raster. F_{rear} and d_{rear} may be measured directly, and F_{front} and d_{front} obtained using Eq. (27). N may be initially taken to be zero (thin lens model); better values of N may be obtained by fitting to the data (below), or may be known from prior experience.

A set of coefficients A through F can be computed using Eqs. (26). In addition, there are seven small-correction parameters G through L, and N, described in Section III. These may be more difficult to measure directly, but they may be taken to be zero for the initial trial values. These initial values are close enough to the actual values that the fitting procedures described below converge dependably to the actual values of the parameters.

The trial values are not exact enough to be used in robot applications, however. The location of the optical center of the lens is known only approximately; the pixel spacing may be difficult to measure exactly; and the small departures from the perfect lens model nonetheless have to be included in the transformation equations. Thus it is generally necessary to perform a best fit calibration to real data.

It seems attractive at first to vary the coefficients A through F (rather than the parameters), since the fitting of (A, B, C) proceeds independently from that of (D, E, F). However, the best fit obtained in this manner may not correspond to realizable values of the physical parameters. This situation arises because f appears only as s/f or t/f, and so there are really only five independent parameters (s/f, t/f, d, F, and β). A fit performed on the six coefficients A through F independently results in the physical parameters being overspecified.

We therefore vary the five physical parameters, recomputing coefficients A through F for each selection of parameters. This has the additional advantage that bounds can be placed on the physical parameters, reflecting limits based on our measurements of the camera geometry. These bounds help prevent the fitting program from becoming bogged down in solutions that do not correspond with the physical system.

Measure of Fit

A set of values of the parameters must be selected, so that the transformation equations provide a description of the observed reality that is as accurate as possible. A measure of fit must be selected, so that different possible choices of parameters can be compared, and the "most accurate" selected.

Such a measure of fit was initially obtained by transforming the observed raster coordinates (i, j) of each data point, using Eqs. (23) and (24), to the corresponding real-world coordinates $(x'_{\text{comp}}, z'_{\text{comp}})$. These computed values were then compared with the coordinates (x', z') actually observed. However, this measure of fit on distances in the external world gives far more weight to distant points than to nearby points, since a one-row change in j corresponds to many millimeters in z', for large z'. It may even result in situations where the parameters are selected on the basis of one or two points, and are in fact selected on the basis of the quantization effects at those points, and not on the actual camera characteristics.

We therefore used the inverse transformation, from (x', z') to (i, j):

$$\begin{aligned} i &= \frac{x'/D}{\sqrt{E + [z' - F]^2} \Big/ \sqrt{1 + [j * t/f]^2}} \\ j &= \frac{z' - C}{A + B * (z' - C)}. \end{aligned} \tag{30}$$

Now, for each point (x', z'), its computed image location $(i_{\text{comp}}, j_{\text{comp}})$ (for a particular choice of parameters) is compared with the image location (i, j) actually observed. The measure of fit is now the sum of squares between computed and observed image coordinates:

$$\text{fit} = \sum \left[(i_{\text{comp}} - i)^2 + (j_{\text{comp}} - j)^2 \right]. \tag{31}$$

(The sum is taken over all the data points). The parameters were then selected to minimize the measure of fit.

It should be noted that in general the exact choice of parameters, by means of a fit to data, depends on the selection of the data points. Choosing most of the data points close to the camera will favor calibration accuracy close in, and conversely. We chose to concentrate our test points towards the center of the image, where we wanted the greatest accuracy for parts acquisition by the robot arm.

The Fit

The following recursive scheme was used to vary the parameters. For this scheme, the parameters are numbered from 1 to n. When the routine is called at some "level" (parameter number) **k**, it adjusts the parameters below **k** until a best fit is obtained. This is the best fit for the given values of the parameters at level **k** and above, varying only the parameters below level **k**. This routine is called for several values of the current parameter **k**. A 1-variable Newton's method is used to locate the particular value of parameter **k** that corresponds to the best fit. Each time parameter **k** takes on a new value, all parameters below **k** are again adjusted to assure best fit at the new value. This adjustment is accomplished simply by having the routine call itself, but with level **k** − 1.

Because the execution time of the fitting program was substantial, and increased exponentially as the number of parameters, it was necessary to allow only a subset of the parameters to vary during a single computer run. A more efficient optimization strategy is given in [17].

Three camera/light-projector configurations were calibrated in this way, varying the five physical parameters required by the model of a perfect camera. In each case, the least squares fit corresponded to an rms error of about one or two pixels per data point. Some error is, of course, expected. The quantization due to the finite pixel size is expected to give rise to an rms error of about 0.4 pixels per data point. Errors in the manual data collection would increase this, as would variations in the camera exposure (light output) without corresponding compensation in the thresholding.

Still, we felt that some of the error in the fit might be due to imperfect lenses or to geometric variables not yet considered. Inclusion of additional parameters in the recursive fitting program was straightforward.

Distortion due to the lens was added as a sixth parameter. The resulting transformations have been described above. For the first camera the inclusion of distortion allowed a sum-of-squares fit of parameters that was closer by nearly a factor of two. For the other two cases no distortion had been anticipated, and none was found.

However, the angle between the x and x' axes (camera/light-source misalignment) was introduced as an additional parameter. This parameter was to be zero by the definition of the coordinate system. But a small error in alignment of the camera and light projector would not be surprising. Introduction of this seventh parameter reduced the errors in the second and third camera configurations, again by nearly a factor of two. The error in alignment, represented by the fitted value of this seventh parameter, was about 0.5°. Final rms errors in fit were thus about one pixel per data point in all three cases, equivalent to about 2 mm of error at the center of the field of view (at a distance of about 350 mm).

TABLE 1
Comparison of Data from Seven calibration

ID	1	2	3	4	5	6	7
Date	2/82	7/82	7/82	12/82	10/83	5/84	11/84
Camera	2200	2500	2500	2500	2500	2500	2500
Lens (mm)	9	16	16	9	12.5	12.5	9
Flash bracket	*A*	*B*	*B'*[a]	*C*	*D*	*D*	*C*
bracket size	small	med	med	med	large	large	med
Narrow test object?	*N*	*N*	*N*	*N*	*Y*	*Y*	*Y*
No. of points	36	52	56	68	63	30	149
rms error	0.77	1.27	1.08	0.81	0.70	0.90[b]	0.57
Best fit values of parameters							
$1/s$ (*ll*/mm)	23.62	28.71	28.53	28.28	27.8	27.82	28.01
$1/t$ (*ll*/mm)	23.90	22.95	22.93	22.09	22.0	23.80[c]	24.24[c]
d (mm)[f]	97.0	158.5	130.7	162.2	281.	280.9	159.7
F (mm)[f]	−46.1	−41.6	−40.6	−13.3	123.	123.0	−13.2
Beta (deg)	22.1	23.6	23.1	25.6	16.6	16.6	24.1
Nodal separation (mm)	—[d]	—	—	—	44.	44.1	4.8
Distortion (inverse mm squared)	−.0008	0.	0.	0.	−0.00033	−0.00033	−0.00164
Rotation (deg)[e]	0.	−0.5	−0.5	0.9	−0.7	−0.7	1.4
I (mm)	0.	0.	0.	0.	0.	0.	−0.5
J (deg)[e]	0.	−0.5	−0.5	0.	1.2	1.2	−1.9
K (pixels)	—	—	—	104.3	119.7	128.2	108.8
L (pixels)	—	—	—	138.9	136.5	155.6[c]	164.4[c]

[a] Same bracket as *B* except for d_2.
[b] Some of the parameters were held fixed to their prior (10/83) value.
[c] These resolutions and offsets reflect a change in digitization hardware. The change affects t, K, and L, by increasing the number of apparent pixels in the final digitized video signal, and by changing the timing of VSYNC and HSYNC.
[d] "—" means this parameter had not been implemented yet.
[e] These angles must be converted to radians before being used in the equations as given.
[f] Measured from the front nodal point.

Subsequent experience, with additional cameras and light projectors, suggested the inclusion of the remaining parameters discussed in Section III. The results of these calibrations— seven in all— are summarized in Table 1, and an example of the parameters before and after fitting is given in Table 2.

Application

One of the camera/light-projector systems, calibrated as discussed in the previous subsection, was mounted on the wrist of a small industrial robot at NBS. This robot is used experimentally to retrieve objects in its field of view. Using the equations derived earlier in this paper, and the parameters obtained as described above, the robot can reliably ascertain the position of objects in its field of view, and also (using a double-line projector [9]) can determine the location and orientation of plane surfaces on the object, with sufficient accuracy to pick up the object perpendicularly to its face, or visually to explore it further.

TABLE 2
Comparison of Parameters before and after Fitting Procedure

	Initial params	Final fit
rms error	10.12	0.57
$1/s$ (ll/mm)	28.3	28.01
$1/t$ (ll/mm)	23.9	24.24
d (mm)	162.2	159.7
F (mm)	−13.3	−13.2
β (deg)	25.6	24.1
Nodal separation (mm)	0.	4.8
Distortion (inverse mm squared)	.00000	−.00164
Rotation (deg)	0.9	1.4
I (mm)	0.	−0.5
J (deg)	0.	−1.9
K (pixels)	118.3	108.8
L (pixels)	150.3	164.4

Notes. The data comes from calibration #7 (see Table 1). The initial parameters are taken from a previous fit of a similar bracket and camera, namely #4 (Table 1, adjusted per c).

Some small uncertainty was felt to exist in the fitted value of s, the horizontal pixel spacing. This was attributed to the light falloff which occurs within the plane of light, towards the edges of the picture. This light falloff affects the observed centroid of the light stripe in the image, moving it systematically towards the center of the picture. This in turn gives rise to a too small value for s. Using a narrower object in the data collection phase of calibrations #5, 6, and 7 (see Table 1) has in fact systematically increased the value of s (decreased $1/s$).

CONCLUSION

We have derived relatively simple transformation equations to convert from image coordinates to external downrange and crossrange coordinates for points illuminated by a plane of light. These equations, together with correction terms as needed, are used to calibrate a camera/light-projector system by varying a small number of parameters, using a least squares fit to a fairly small number of experimentally measured points. The resulting fitted parameters define an image-to-real-world transformation, which we have used satisfactorily on a parts-acquisition robot.

REFERENCES

1. G. J. Agin and T. O. Binford, Computer description of curved objects, in *Proceedings, 3rd Int. Joint Conf. Artif. Intell.*, Stanford, 1973, pp. 629–640.
2. G. J. Agin, *Real Time Control of a Robt with a Mobile Camera*, Technical Note 149, SRI International, Menlo Park, Calif., 1979.
3. R. C. Bolles and M. A. Fischler, A Ransac-based approach to model fitting and its application to finding cylinders in range data, in *Proceedings, 7th Int. Joint Conf. Artif. Intell.* 1981, pp. 637–643.
4. W. F. Clocksin, P. G. Davey, C. G. Morgan, and A. R. Vidler, Progress in visual feedback for robot arc-welding of thin sheet steel, in *Robot Visions*, (A. Pugh, Ed.), pp. 187–198, IFS Publications, Bedford, U.K., 1983.

5. J. Hill and W. T. Park, Real-time control of a robot with a mobile camera, in *Proceedings, 9th Int. Symp. Indus. Robots (SME)*, Washington, D.C., 1979.
6. R. N. Nagel, G. J. VanderBrug, J. S. Albus, and E. Lowenfeld, Experiments in parts acquisition using robot vision, in *Proceedings, Autofact II, Robots IV Conf.*, Detroit, Mich., 1979.
7. Y. Shirai and M. Suwa, Recognition of polyhedrons with a rangefinder, in *Proceedings, 2nd Int. Joint Conf. Artif. Intell.* 1971, pp. 80–87.
8. M. Oshima and Y. Shirai, Object recognition using three-dimensional information, in *Proceedings, 7th Int. Joint Conf. Artif. Intell.* 1981, pp. 601–606.
9. J. Albus, E. Kent, P. Mansbach, M. Nashman, L. Palombo, and M. Shneier, Six-dimensional vision system, in *Robot Vision, Proc. Soc. Photo-Opt. Instrum. Eng.* **336**, 1982.
10. B. Carlisle, J. Gleason, D. McGhie, and S. Roth, The PUMA/VS-100 robot vision system, in *Robot Vision*, (A. Pugh, Ed.), pp. 313–323, IFS Publications, Bedford, U.K., 1983.
11. D. Nitzan, A. E. Brain, and R. O. Duda, The measurement and use of registered reflectance and range data in scene analysis, *Proc. IEEE* **65**, 1977, 206–220.
12. M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford, 1965.
13. G. A. Boutry, *Instrumental Optics*, Interscience, New York, 1962.
14. E. B. Brown, *Modern Optics*, Reinhold, New York, 1965.
15. F. A. Jenkins and H. E. White, *Fundamentals of Optics*, McGraw–Hill, New York, 1976.
16. Y. Hung, private communication.
17. Y. Hung, P-S. Yeh, P. Mansbach, *Calibration of a Structured Light Vision System*, Technical Report CAR-TR-29, Univ. of Maryland, College Park, Md., 1983.