Simple Pinhole Camera Calibration

Kenneth M. Dawson-Howe and David Vernon
Department of Computer Science, Trinity College, Dublin, Ireland

ABSTRACT
A large number of cameras may be modeled quite accurately using the simple pinhole camera model, which may be defined either in terms of camera parameters or by the *c* matrix (which defines the mapping from 3D points to the image plane). We present formulations of the associated transformations between these two equivalent representations. We also introduce an inexpensive technique for calibrating a camera using a single two-plane calibration object, and employ a novel high-precision Hough transform technique for determining calibration grid lines. © 1994 John Wiley & Sons, Inc.

I. INTRODUCTION

A camera model is, as its name suggests, an attempt to model the imaging function of a camera. That is, it is a model of the relationship between the viewed image on the camera's sensor elements and "real" 3D world (which is the source of the image). The camera model is comprised of two distinct sets of parameters; the *intrinsic* parameters, which specify the internal geometry of the camera, and the *extrinsic* parameters, which specify the camera's position and orientation with respect to some frame of reference in the real world. Several models of varying complexity have been proposed, from simple orthographic projection to more complex perspective models which take into account possible lens distortion [1].

In certain instances use of a complex camera model is obviously preferable (e.g., if the camera exhibits significant distortions, or if the task requires extremely high precision). However, in many situations a simple camera model is quite sufficient. In this article we present the simple pinhole camera model (as described in [2] and [3]), and, more importantly, formulas for converting between the two different representations of the model. Additionally, an inexpensive* method of camera calibration is presented using a single view of a calibration object.

II. SIMPLE PINHOLE CAMERA MODEL

The simple pinhole camera model models perspective effects but does not model any form of distortion. It can be specified in one of two ways.

* The calibration technique is inexpensive from a financial point of view, in that it does not require the use of a *Z* stage.

Received 27 August 1991; revised manuscript received 22 December 1992

(1) In terms of the *c* matrix, which transforms 3D real-world points (x, y, z) to the image plane [i.e., as (I, J) points]:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} i \\ j \\ t \end{bmatrix}, \tag{1}$$

where $I = i/t$ and $J = j/t$.

(2) Explicitly in terms of the intrinsic and extrinsic camera parameters. The intrinsic parameters of the simple pinhole camera model may be considered in terms of the sensor resolutions s_i and s_j in the I and J directions, respectively, the focal length f, and the image center (i_c, j_c). The extrinsic camera parameters may be expressed in terms of a homogeneous transformation which specifies the position (or translation T) and the orientation (or rotation R) of the camera:

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Each form of representation may be computed from the other, and those transformations are detailed in the sections which follow.

A. Determining the *c* Matrix from Camera Parameters

The *c* matrix may be determined from the camera parameters, quite simply, as follows:

$$c = \begin{bmatrix} \frac{f}{s_i} & 0 & i_c & 0 \\ 0 & \frac{f}{s_j} & j_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}. \tag{2}$$

B. Determining Camera Parameters from the *c* Matrix

Calculating camera parameters from the *c* transformation matrix is not as straightforward, but may be performed as follows.

Firstly, it should be noted that the *c* transformation can be considered in terms of three plane equations (see fig. 1) by simply multiplying the two matrices on the left-hand side of Eq. (1):

CCC 0899-9457/94/010001-06

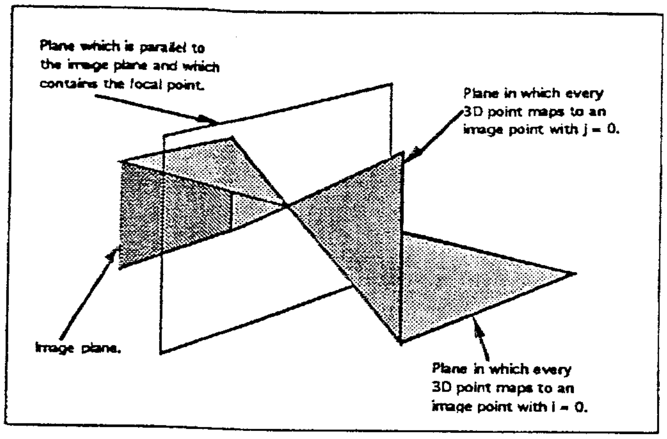

Figure 1. The three planes which are defined in the C camera transformation matrix.

$$\begin{bmatrix} c_{11}x + c_{12}y + c_{13}z + c_{14} \\ c_{21}x + c_{22}y + c_{23}z + c_{24} \\ c_{31}x + c_{32}y + c_{33}z + c_{34} \end{bmatrix} = \begin{bmatrix} i \\ j \\ t \end{bmatrix} . \tag{3}$$

If $t = 0$ then the point computed by the camera model is undefined. This only occurs when the 3D point [i.e., (x, y, z)] is in the plane parallel to the image plane through the focal point of the camera, and hence, from eq. (3), this plane is:

$$c_{31}x + c_{32}y + c_{33}z + c_{34} = 0 . \tag{4}$$

It is also quite simple to extract the plane in which $i = 0$. From Eq. (3) it is

$$c_{11}x + c_{12}y + c_{13}z + c_{14} = 0 . \tag{5}$$

Similarly, the plane in which $j = 0$ is

$$c_{21}x + c_{22}y + c_{23}z + c_{24} = 0 . \tag{6}$$

These three planes (i.e., that which is parallel to the image plane, through the focal point of the camera, that which includes all points that map to $i = 0$, and that which includes all points that map to $j = 0$) all include the focal point of the camera, so the focal point of the camera may be determined simply by calculating the intersection of the three planes.

Solving Eqs. (4)–(6) for x, y, and z we get*:

$$x = \frac{(c_{33}c_{12} - c_{32}c_{13})(c_{33}c_{24} - c_{34}c_{23}) - (c_{33}c_{22} - c_{32}c_{23})(c_{33}c_{14} - c_{34}c_{13})}{(c_{33}c_{22} - c_{32}c_{23})(c_{33}c_{11} - c_{11}c_{13}) - (c_{33}c_{12} - c_{32}c_{13})(c_{33}c_{21} - c_{31}c_{23})} , \tag{7}$$

$$y = \frac{(c_{31}c_{13} - c_{11}c_{33})(c_{31}c_{24} - c_{21}c_{34}) - (c_{31}c_{23} - c_{21}c_{33})(c_{31}c_{14} - c_{11}c_{34})}{(c_{31}c_{23} - c_{21}c_{33})(c_{31}c_{12} - c_{11}c_{32}) - (c_{31}c_{13} - c_{11}c_{33})(c_{31}c_{22} - c_{21}c_{32})} , \tag{8}$$

$$z = \frac{(c_{31}c_{22} - c_{21}c_{32})(c_{31}c_{14} - c_{11}c_{34}) - (c_{31}c_{12} - c_{11}c_{12})(c_{31}c_{24} - c_{21}c_{34})}{(c_{31}c_{12} - c_{11}c_{32})(c_{31}c_{23} - c_{21}c_{33}) - (c_{31}c_{22} - c_{21}c_{32})(c_{31}c_{13} - c_{11}c_{33})} . \tag{9}$$

As the focal point of the camera and a plane equation which is parallel to the image plane are both known, it is possible to calculate the point on the image plane which is directly behind the focal point. This is done by determining an arbitrary point directly in front of the focal point (by translating the focal point by an arbitrary distance in the direction perpendicular to the image plane) and mapping the point onto the image plane using the inverse perspective transformation (see Sec. II.C).

* It should be noted that the listed formulae are only one of three possibilities, depending on how the equations are solved. If any of these formulae fail (due to a division by zero), one of the two alternatives will always succeed.

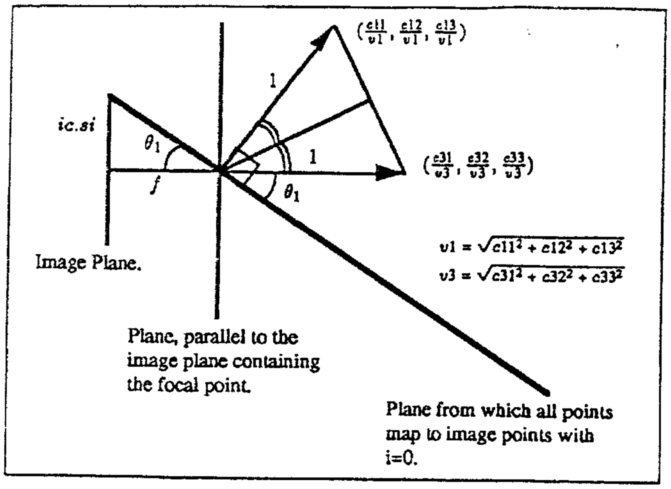

Figure 2. Geometry used in order to calculate the sensor resolutions to focal length ratios.

The orientation of the camera may also be readily calculated at this juncture. The plane which is parallel to the focal point indicates the direction of the focal axis and either of the other two plane equations (i.e., those from which all points map to image points with $i = 0$ or $j = 0$) may be used to determine the camera rotation around the focal axis.

The only remaining camera parameters, which have not yet been determined, are those of the focal length (f) and sensor resolution (s_i and s_j). In fact, only the ratios between them may be calculated (as it is only the ratio between them that is used in determining the simple pinhole camera model, and hence the c matrix). Consider the geometry shown in Fig. 2.

$$\theta_1 = \frac{\pi}{2} - 2\sin^{-1}\left[\frac{1}{2}\sqrt{\left(\frac{c_{31}}{v_3} - \frac{c_{11}}{v_1}\right)^2 + \left(\frac{c_{32}}{v_3} - \frac{c_{12}}{v_1}\right)^2 + \left(\frac{c_{33}}{v_3} - \frac{c_{13}}{v_1}\right)^2}\,\right] , \tag{10}$$

$$\frac{i_c s_i}{f} = \tan\theta_1 \tag{11}$$

$$\frac{s_i}{f} = \frac{\tan\theta_1}{i_c} . \tag{12}$$

Similarly,

$\theta_2 =$

$$\frac{\pi}{2} - 2\sin^{-1}\left[\frac{1}{2}\sqrt{\left(\frac{c_{31}}{v_3} - \frac{c_{21}}{v_2}\right)^2 + \left(\frac{c_{32}}{v_3} - \frac{c_{22}}{v_2}\right)^2 + \left(\frac{c_{33}}{v_3} - \frac{c_{23}}{v_2}\right)^2}\right], \quad (13)$$

$$\frac{j_c s_j}{f} = \tan\theta_2 \quad (14)$$

$$\frac{s_j}{f} = \frac{\tan\theta_2}{j_c}. \quad (15)$$

where v_1–v_3 are the lengths of the vectors pertaining to the various plane equations,

$$v_1 = \sqrt{c_{11}^2 + c_{12}^2 + c_{13}^2}, \quad (16)$$

$$v_2 = \sqrt{c_{21}^2 + c_{22}^2 + c_{23}^2}, \quad (17)$$

$$v_3 = \sqrt{c_{31}^2 + c_{32}^2 + c_{33}^2}. \quad (18)$$

Finally, the ratio between sensor resolutions in the I and J directions may be calculated by combining Eqs. (12) and (15), giving

$$\frac{s_i}{s_j} = \frac{j_c \tan\theta_1}{i_c \tan\theta_2}. \quad (19)$$

C. Inverse Perspective Transformation. For completeness we must also consider the mapping from points on the image plane into the real world (i.e., the inverse perspective transformation). A single view of a 3D point on an image plane is insufficient to allow the 3D coordinates of the point to be computed. This is caused by the loss of information inherent in the viewing process (which transforms 3D points to 2D points). However, it is possible for any image point (i, j) to generate a 3D vector (x_v, y_v, z_v) from the focal point of the camera along which the 3D point corresponding to (i, j) will lie. This may be done through the use of the inverse perspective transformation, which may be defined by a simple 4×3 matrix p as follows:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \\ p_{41} & p_{42} & p_{43} \end{bmatrix} \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} = \begin{bmatrix} x_v \\ y_v \\ y_v \\ w \end{bmatrix}. \quad (20)$$

The inverse perspective matrix p may be derived from the camera parameters:

$$p = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_i & 0 & -i_c s_i \\ 0 & -s_j & j_c s_j \\ 0 & 0 & f \\ 0 & 0 & 0 \end{bmatrix}. \quad (21)$$

III. CAMERA CALIBRATION

Camera calibration, in the context of 3D vision, is the process of determining both* the intrinsic and the extrinsic parameters of a given camera. In order to accurately calibrate *both* the intrinsic and the extrinsic parameters of a camera, points which are distributed in three dimensions (i.e., are not simply coplanar) must be viewed using the camera (e.g., see [1, 5]). The relationships between the viewed 2D points and the 3D points are then used to calibrate the camera.

The calibration points are typically obtained using a planar grid, and a Z stage, in order to view the points (which are already distributed in X and Y) at several different values of Z. This is done, rather than use a single calibration object with noncoplanar points, because the fabrication of such objects is difficult [6]. However, in any 3D vision applications it is important to be able to limit cost, and the additional expense of a Z stage may be prohibitive. Hence, the possibility of using a single calibration object with points distributed in 3D (i.e., not simply coplanar) is investigated in this research.

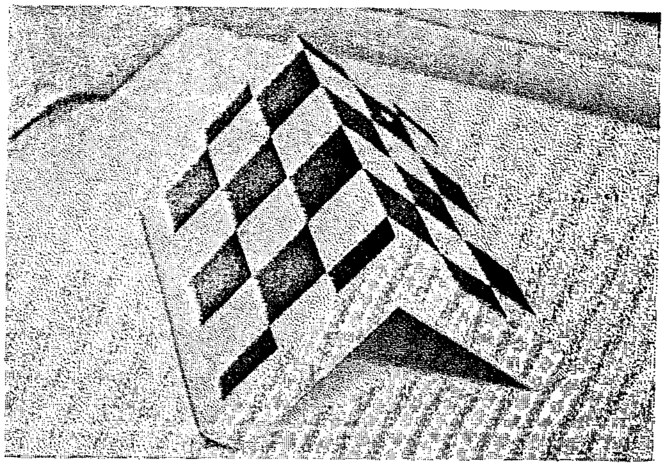

Figure 3. The calibration object employed.

A. Calibration Object. A calibration object was constructed, consisting of a checkerboard grid which was folded (at one of the grid lines) over a right angle (see Fig. 3). The calibration points were those at the corners of the grid boxes. The right-angle fold in the grid allows the points to be distributed in more than a single plane, but at the same time allows calculation of the relative 3D coordinates of the calibration points to remain trivial. Also, for the sake of automatic association of 3D point data with the viewed calibration points, a white dot was drawn on one of the black grid boxes.†

B. Determining grid points. Bearing in mind that the simple pinhole camera model is being assumed to be valid, it can also be assumed that the grid lines will appear straight in any image of the calibration object. Now, in order to make the determination of the grid points robust in the presence of possible marks on the grid or noise in the imaging process, the grid lines are located first, and then the relevant grid points are computed from the intersection of these lines. This search,

* In some instances it is necessary to calculate only the intrinsic or the extrinsic parameters (e.g., see [4]).

† An image of the calibration grid could be considered to be the grid in one of two different orientations, were the white dot not present.

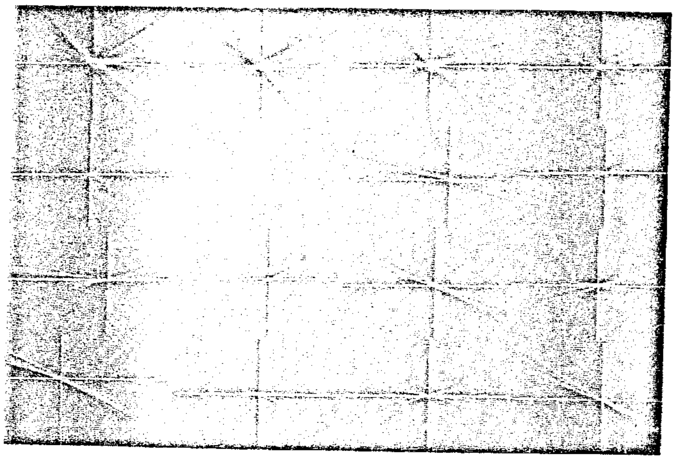

Figure 4. High precision determination of lines using the Hough transform. This figure shows fifteen different regions of Hough space, corresponding to the regions around the lines determined in the first stage of Hough transform (see text). In each case r is along the vertical axis, and ϕ is along the horizontal. The first five regions of Hough space shown (the four on the first line and the first on the second line) correspond to the grid lines which are parallel to the right angle edge of the calibration object. The other ten Hough spaces correspond to the lines which are orthogonal (in 3-D) to that edge.

for the grid lines, may be performed using the Hough transform for lines (e.g., see [7, 3]).

Unfortunately, in order for the lines to be identified with sufficient accuracy for the calibration task at hand, the Hough accumulator would have to be prohibitively large. To overcome this a two-stage Hough transform is used. In the first stage lines are roughly approximated, while in the second stage a high-precision Hough transform [which considers only a subspace of the Hough (r, ϕ) accumulator*] is employed in order to tune the values of each determined grid line. By constraining the search for edge points to a specific region of the image, for each grid line, the Hough accumulators shown in Fig. 4 are determined. The (r, ϕ) pair corresponding to the line is identified as the center of the narrowest ϕ column (i.e., the column with the thinnest section of active cells; see Fig. 4).

The grid lines and points determined are shown in Fig. 5. The white dot is also identified, and hence the 3D coordinates of the grid points (which are already known, as the calibration grid defines the frame of reference used) may be automatically associated with the viewed calibration points.

C. Mathematics of Calibration. Having determined a mapping from a number of 3D points to a number of 2D image points, the determination of the camera model may now be addressed.†

The transformation which maps any three-dimensional point (x, y, z) to the corresponding image point (I, J) may be

* Note that the coordinates in the Hough space, i.e., the line parameters r and ϕ, represent the normal distance of the line to the origin and the angle that this normal makes with the horizontal, respectively.

† The mathematics listed in this section is taken from the text *Machine Vision* by David Vernon, with permission of the publishers, Prentice-Hall International.

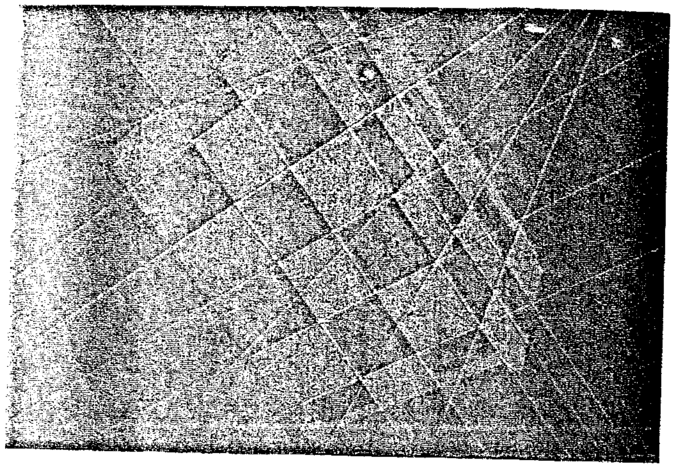

Figure 5. The grid lines and points determined for a sample calibration image. Note also that the white dot is identified, in order to determine the orientation of the grid.

defined as follows (see Sec. II):

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} i \\ j \\ t \end{bmatrix}, \qquad (22)$$

where

$$I = \frac{i}{t}, \qquad (23)$$

$$J = \frac{j}{t}. \qquad (24)$$

Expanding Eq. (22):

$$c_{11}x + c_{12}y + c_{13}z + c_{14} = i, \qquad (25)$$

$$c_{21}x + c_{22}y + c_{23}z + c_{24} = j, \qquad (26)$$

$$c_{31}x + c_{32}y + c_{33}z + c_{34} = t. \qquad (27)$$

But by rearranging Eq. (23) and (24) we know that

$$i = It, \qquad (28)$$

$$j = Jt, \qquad (29)$$

Hence

$$c_{11}x + c_{12}y + c_{13}z + c_{14} - Ic_{31}x - Ic_{32}y - Ic_{33}z - Ic_{34} = 0, \qquad (30)$$

$$c_{21}x + c_{22}y + c_{23}z + c_{24} - Jc_{31}x - Jc_{32}y - Jc_{33}z - Jc_{34} = 0. \qquad (31)$$

Overall scaling of the transformation matrix is irrelevant, so that c_{34} may arbitrarily be set to 1. At the same time completing Eq. (30) and (31), we get

$$c_{11}x + c_{12}y + c_{13}z + c_{14} + c_{21}0 + c_{22}0 + c_{23}0 + c_{24}0 - Ic_{31}x - Ic_{32}y - Ic_{33}z = I\,, \quad (32)$$

$$c_{21}x + c_{22}y + c_{23}z + c_{24} + c_{11}0 + c_{12}0 + c_{13}0 + c_{14}0 - Jc_{31}x - Jc_{32}y - Jc_{33}z = J\,. \quad (33)$$

We have 11 unknowns at this stage and in order to determine values for them we need at least 11 equations. For any three-dimensional point to two-dimensional image plane point mapping we obtain two equations [(32) and (33)], so at least six points are needed* in order to solve for the values of the transformation matrix. This allows us to reformulate as follows:

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -I_1x_1 & -I_1y_1 & -I_1z_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -J_1x_1 & -J_1y_1 & -J_1z_1 \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -I_2x_2 & -I_2y_2 & -I_2z_2 \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -J_2x_2 & -J_2y_2 & -J_2z_2 \\ x_3 & y_3 & z_3 & 1 & 0 & 0 & 0 & 0 & -I_3x_3 & -I_3y_3 & -I_3z_3 \\ 0 & 0 & 0 & 0 & x_3 & y_3 & z_3 & 1 & -J_3x_3 & -J_3y_3 & -J_3z_3 \\ x_4 & y_4 & z_4 & 1 & 0 & 0 & 0 & 0 & -I_4x_4 & -I_4y_4 & -I_4z_4 \\ 0 & 0 & 0 & 0 & x_4 & y_4 & z_4 & 1 & -J_4x_4 & -J_4y_4 & -J_4z_4 \\ x_5 & y_5 & z_5 & 1 & 0 & 0 & 0 & 0 & -I_5x_5 & -I_5y_5 & -I_5z_5 \\ 0 & 0 & 0 & 0 & x_5 & y_5 & z_5 & 1 & -J_5x_5 & -J_5y_5 & -J_5z_5 \\ x_6 & y_6 & z_6 & 1 & 0 & 0 & 0 & 0 & -I_6x_6 & -I_6y_6 & -I_6z_6 \\ 0 & 0 & 0 & 0 & x_6 & y_6 & z_6 & 1 & -J_6x_6 & -J_6y_6 & -J_6z_6 \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} = \begin{bmatrix} I_1 \\ J_1 \\ I_2 \\ J_2 \\ I_3 \\ J_3 \\ I_4 \\ J_4 \\ I_5 \\ J_5 \\ I_6 \\ J_6 \end{bmatrix}, \quad (34)$$

$$Xc = y\,. \quad (35)$$

Taking the pseudoinverse, we may then determine c:

$$c = (X^T X)^{-1} X^T y\,. \quad (36)$$

Finally, we note that Eq. (36) must be solved numerically, as it is not practical to solve it algebraically.

D. Determining the Accuracy of the Calibration. In order to evaluate the accuracy of the calibration several different measures are calculated:

(i) Variance in the internal camera parameters. These variances, to some extent, measure the repeatability of the calibration. See Table I.

(ii) Accuracy with which the calibration points are modeled. This measure is extremely important, as it is indicative of how accurately the simple pinhole camera model represents the imaging process. See Table II.

Table I

Parameter	Average value	Standard deviation (σ)
Image center (i_c, j_c)	(253.00, 258.23)	3.25 pixels
Aspect ratio s_i/s_j	0.684 42	0.000 35

Table II

Parameter	Standard deviation (σ)
Distance between imaged and modeled calibration points	0.2852 pixels

Table III

Measured distance	Average determined value	Standard deviation (σ)
60.0 mm	59.96 mm	0.40 mm

(iii) Length measurements. Finally, perhaps the most important measure of all is how accurately distances in the real world may be measured. In order to perform this experiment, two different cameras which viewed the same scene were used, both were calibrated and the three-dimensional point data was calculated using stereo vision techniques.[†] On the basis of 240 measurements the results were as shown in Table III.

Assuming a normal distribution, 95% of the values will be within 2 times the standard deviation, giving an approximate accuracy of 1 part in 75.

IV. SUMMARY

Transformations between the c matrix and the camera parameters of the simple pinhole camera model are presented. Additionally, a camera calibration technique using a single view of a two-plane calibration object is introduced and, experimentally, is found to be reasonably accurate (i.e., it allows determination of distances to within 1 part in 75). While more accurate camera models and calibration techniques exist (e.g., [1]), the accuracy achieved using this simple camera model and this inexpensive calibration technique would appear to be sufficient for a reasonably wide variety of applications.

REFERENCES

1. R. T. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lenses," *IEEE J. Robotics Autom.* **RA-3,** 323–344 (1987).

* It is possible to use more than six points when determining the c transformation matrix and, in fact, is advisable so to do, so that a good least-squares error may be obtained.

† For this test to be valid it was necessary to move the connected points within the imaged scene so that they appeared in many different positions and orientations.

2. D. H. Ballard and C. M. Brown, *Computer Vision* (Prentice-Hall, Englewood Cliffs, NJ, 1982).
3. D. Vernon, *Machine Vision*, (Prentice-Hall, Englewood Cliffs, NJ, 1991).
4. L. L. Wang and W. H. Tsai, "Camera calibrated by Vanishing Lines for 3-D Computer Vision," *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-13**, 370–376 (1991).
5. F. Chaumette and P. Rives, "Realisation et Calibration d'un Systeme Experimental de Vision Compose d'une camera mobile embarquee sur un robot-manipulateur." Institut National de Recherche en Informatique et en Automatique (INRIA), Rennes, Report No. 994, 1989.
6. R. K. Lenz and R. Y. Tsai, "Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology," *IEEE Trans. on Pattern Anal. Machine Intell.* **PAMI-10**, 713–720 (1988).
7. P. V. C. Hough, "Methods and Measures for Recognising Complex Patterns," US Patent No. 3069654, December 1962.