Pattern Recognition, Vol. 24, No. 6, pp. 533–541, 1991
Printed in Great Britain

0031-3203/91 \$3.00 + .00
Pergamon Press plc

DETERMINING CAMERA PARAMETERS FROM THE PERSPECTIVE PROJECTION OF A QUADRILATERAL

MICHAEL A. PENNA

Department of Computer and Information Science, Indiana University–Purdue University at Indianapolis, 1201 East 38th Street, Indianapolis, IN 46205-2868, U.S.A.

(*Received* 15 *March* 1990; *in revised form* 17 *August* 1990; *received for publication* 24 *September* 1990)

Abstract—In this paper we show that there is sufficient information in the two-dimensional perspective projection of an arbitrary quadrilateral of known shape and size in three-space to determine the exact three-dimensional coordinates of its vertices, generalizing known results for rectangles. Implementation results are also discussed.

Perspective projection Camera calibration Machine vision Computer vision Scene analysis

1. INTRODUCTION

In Haralick,[1] it is shown that there is sufficient information in the two-dimensional perspective projection of a rectangle of unknown size in three-space to determine the camera look-angle parameters, and that if the size of the rectangle is given then the exact three-dimensional coordinates of its vertices can be computed. In this paper we show that there is sufficient information in the two-dimensional perspective projection of an arbitrary quadrilateral (an arbitrary four-sided linear figure) of known shape and size in three-space to determine the exact three-dimensional coordinates of its vertices.

Our result generalizes the second result presented in Haralick[1] since a quadrilateral is more general than a rectangle. Our result applies, for example, to a quadrilateral determined by four sides of a polygonal face of an object, while the result presented in Haralick[1] does not. For a rectangle of unknown size, we can determined the camera look-angle parameters (thus giving us the first result presented in Haralick[1]) as well as the ratio of length to width of the rectangle. Our computations are simple, making our approach easy to implement. Further, our approach is line-based (as opposed to point-based) so it is numerically stable in the presence of noise.

This paper is organized as follows: in Section 2 we present an algorithm for determining, from the two-dimensional perspective projection of an arbitrary quadrilateral of known shape and size in three-space, its exact three-dimensional coordinates. Since this algorithm draws heavily on Penna and Patterson,[2] a brief introduction to projective geometry is provided in Appendix 1; in Appendix 2 we briefly discuss the formulas presented in Haralick[1] from this point of view. In Section 3, we discuss some implementation results. In Section 4 we show that for a rectangle of unknown size, we can determine the camera look-angle parameters as well as the ratio of length to width of the rectangle.

2. THE ALGORITHM

We assume that we are given an image of a quadrilateral taken with a camera whose focal length f is known. We use a pin-hole camera model, and we assume (see Fig. 1) that coordinates are chosen so that the focal point of the camera is the origin $O(0,0,0)$ and that the (positive) image plane Π is the plane whose equation is $z = d = (M + 1)f$ where M is the factor by which a negative image is magnified in creating a positive image. (The lens equation is $1/d + 1/d' = 1/f$ where d is the distance from the focal point of the lens to the positive image plane, and d' is the distance from the focal point of the lens to the negative image plane. Since $M = d/d'$, $d' = d/M$, so that $1/d + M/d = 1/f$. Thus $d = (M + 1)f$.) The x- and y-coordinate axes are chosen so that the

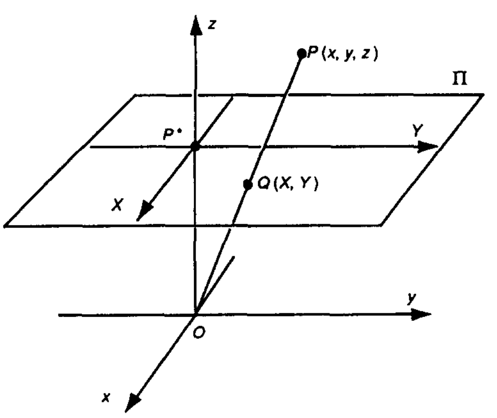


Fig. 1. The perspective viewing mechanism.

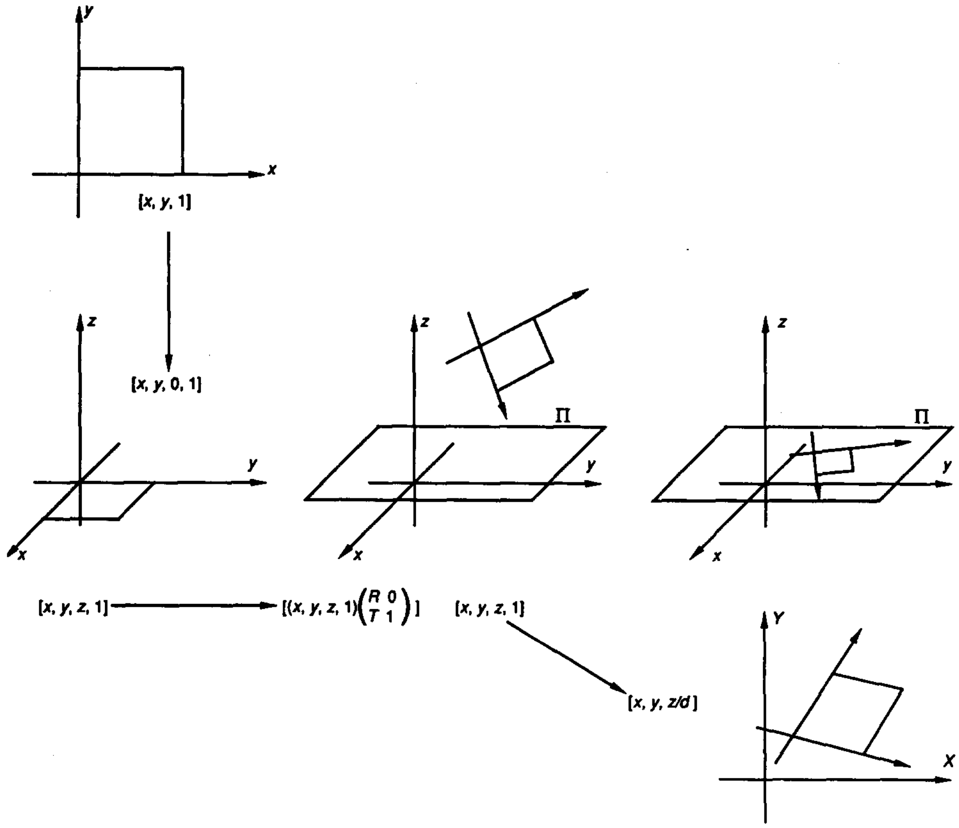


Fig. 2. The display transformation.

induced X- and Y-coordinate axes on Π (the lines of intersection of Π with the xz- and yz-coordinate planes) are the canonical image plane coordinate axes. Perspective projection π maps the point $P(x, y, z)$ to the point $Q(X, Y)$ in Π defined by intersecting the line determined by O and P with Π; it follows that

$$\pi(x, y, z) = (X, Y) = \left(\frac{dx}{z}, \frac{dy}{z}\right).$$

Let us now consider the process of displaying an image of a planar figure (see Fig. 2). If a (closed and bounded) figure is contained in a model xy-coordinate plane, then the model xy-coordinate plane can be mapped to the xy-coordinate plane in three-space by the transformation

$$[x, y, 1] \rightarrow \left[(x, y, 1)\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right] = [x, y, 0, 1].$$

(Throughout this paper we use parentheses (. . .) to denote Cartesian coordinates, and brackets [. . .] to denote homogeneous coordinates.[(2 or Appendix 1)]) The image of this figure can then be mapped into an arbitrary position in three-space by the rigid motion

$$[x, y, z, 1] \rightarrow \left[(x, y, z, 1)\begin{pmatrix} R & 0 \\ T & 1 \end{pmatrix}\right]$$

where the 3×3 orthogonal matrix

$$R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

represents rotation, and the 1×3 matrix $T = (h_1\, h_2\, h_3)$ represents translation. The image of this figure is mapped by perspective projection into Π by

$$[x, y, z, 1] \rightarrow \left[(x, y, z, 1)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/d \\ 0 & 0 & 0 \end{pmatrix}\right]$$

$$= \left[x, y, \frac{z}{d}\right] = \left[\frac{dx}{z}, \frac{dy}{z}, 1\right].$$

Thus the composition of mapping the model plane into three-space, performing a rigid motion of three-space, and perspectively projecting three-space to Π

is represented by the matrix

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & 0 \\ T & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/d \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13}/d \\ r_{21} & r_{22} & r_{23}/d \\ h_1 & h_2 & h_3/d \end{pmatrix}.$$

For the purpose of exposition we let $d = 1$ so that

$$M = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ h_1 & h_2 & h_3 \end{pmatrix}.$$

Observe that the first two rows of M are the first two rows $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ of R, and that the third row of M is T. Since the third row $\boldsymbol{r}_3$ of R is the cross product $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$ of the first two rows of R, we can determine both R and T if we can determine M.

$$R = \begin{pmatrix} \cos\alpha\cos\beta & \sin\alpha\cos\beta & -\sin\beta \\ \cos\alpha\sin\beta\sin\gamma - \sin\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma + \cos\alpha\cos\gamma & \cos\beta\sin\gamma \\ \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & \cos\beta\cos\gamma \end{pmatrix}.$$

To do this, we use the fact that M represents a projective transformation. Recall[2 or Appendix 1] that a projective transformation T of two-space is represented by an invertible 3×3 matrix A (which is unique, up to a scalar multiple)

$$T = T_A : [x, y, z] \to [(x, y, z)A]$$

and that this matrix can be determined by how it maps a quadrilateral. Indeed, if l_i', $i = 1, \ldots, 4$, are the lines determined by the edges of a quadrilateral in the model plane, and l_i'', $i = 1, \ldots, 4$, are the corresponding image lines in the image plane, then the projective transformation taking l_i', $i = 1, \ldots, 4$, to l_i'', $i = 1, \ldots, 4$, is represented by the matrix

$$A = (k_1'l_1' \; k_2'l_2' \; k_3'l_3')(k_1''l_1'' \; k_2''l_2'' \; k_3''l_3'')^{-1}$$

where k_1', k_2', k_3' and k_1'', k_2'', k_3'' are constants for which $l_4' = k_1'l_1' + k_2'l_2' + k_3'l_3'$ and $l_4'' = k_1''l_1'' + k_2''l_2'' + k_3''l_3''$, respectively. It follows that

$$M = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ h_1 & h_2 & h_3 \end{pmatrix} = k(k_1'l_1' \; k_2'l_2' \; k_3'l_3')(k_1''l_1'' \; k_2''l_2'' \; k_3''l_3'')^{-1}$$

for some non-zero constant k.

Thus we can determine R and T by using the following algorithm:

Input: The homogeneous coordinates l_i', $i, \ldots, 4$, of the lines l_i' determined by the sides of a quadrilateral in the image plane, and the homogeneous coordinates l_i'', $i = 1, \ldots, 4$, of the images of these lines under the composition of embedding into the xy-coordinate plane, a rigid motion, and perspective projection.

Output. The entries of rotation matrix R and the translation matrix T representing the rigid motion.

Step 1. Perform Gauss–Jordan elimination on the 3×4 matrix $(l_1' \; l_2' \; l_3' \; l_4')$; the entries in the fourth column of the resulting matrix are the constants k_1', k_2', k_3'. Let $A' = (k_1'l_1' \; k_2'l_2' \; k_3'l_3')$.

Step 2. Perform Gauss–Jordan elimination on the 3×4 matrix $(l_1'' \; l_2'' \; l_3'' \; l_4'')$; the entries in the fourth column of the resulting matrix are the constants k_1'', k_2'', k_3''. Let $A'' = (k_1''l_1'' \; k_2''l_2'' \; k_3''l_3'')^{-1}$.

Step 3. Let $A = (a_{ij}) = A'A''$ and let k be either $(a_{11}^2 + a_{12}^2 + a_{13}^2)^{1/2}$ or $(a_{21}^2 + a_{22}^2 + a_{23}^2)^{1/2}$ (both numbers are the same and non-zero). Then $\boldsymbol{r}_1 = (a_{11}, a_{12}, a_{13})/k$, $\boldsymbol{r}_2 = (a_{21}, a_{22}, a_{23})/k$, $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$, and $\boldsymbol{h} = (h_1, h_2, h_3) = (a_{31}, a_{32}, a_{33})/k$.

Having retrieved R, the camera look-angle parameters α, β and γ can be computed (in any number of ways) using the fact that [the matrix R displayed above].

For example, $\alpha = \tan^{-1}(r_{12}/r_{11})$, $\beta = \tan^{-1}(-r_{13}/(r_{11}^2 + r_{12}^2)^{1/2})$ and $\gamma = \tan^{-1}(r_{23}/r_{33})$.

3. IMPLEMENTATION RESULTS

Our algorithm was implemented on a system consisting of a Panasonic WV-BL200 CCTV Camera equipped with an MN516 16 mm lens and connected to an Apple Macintosh II via a DataTranslation QuickCapture Frame Grabber Board. Numerous experiments were performed under varying

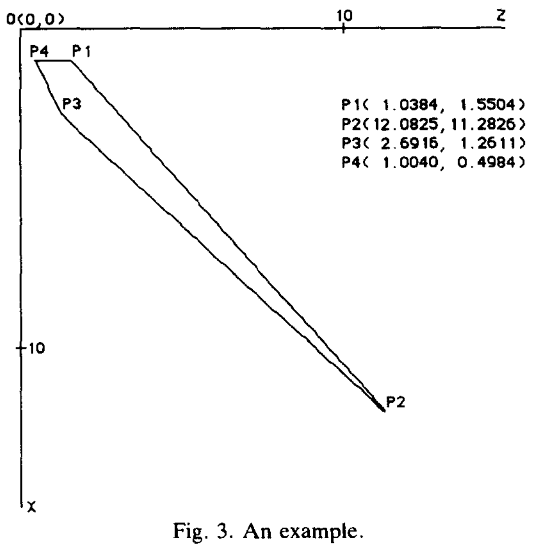


Fig. 3. An example.

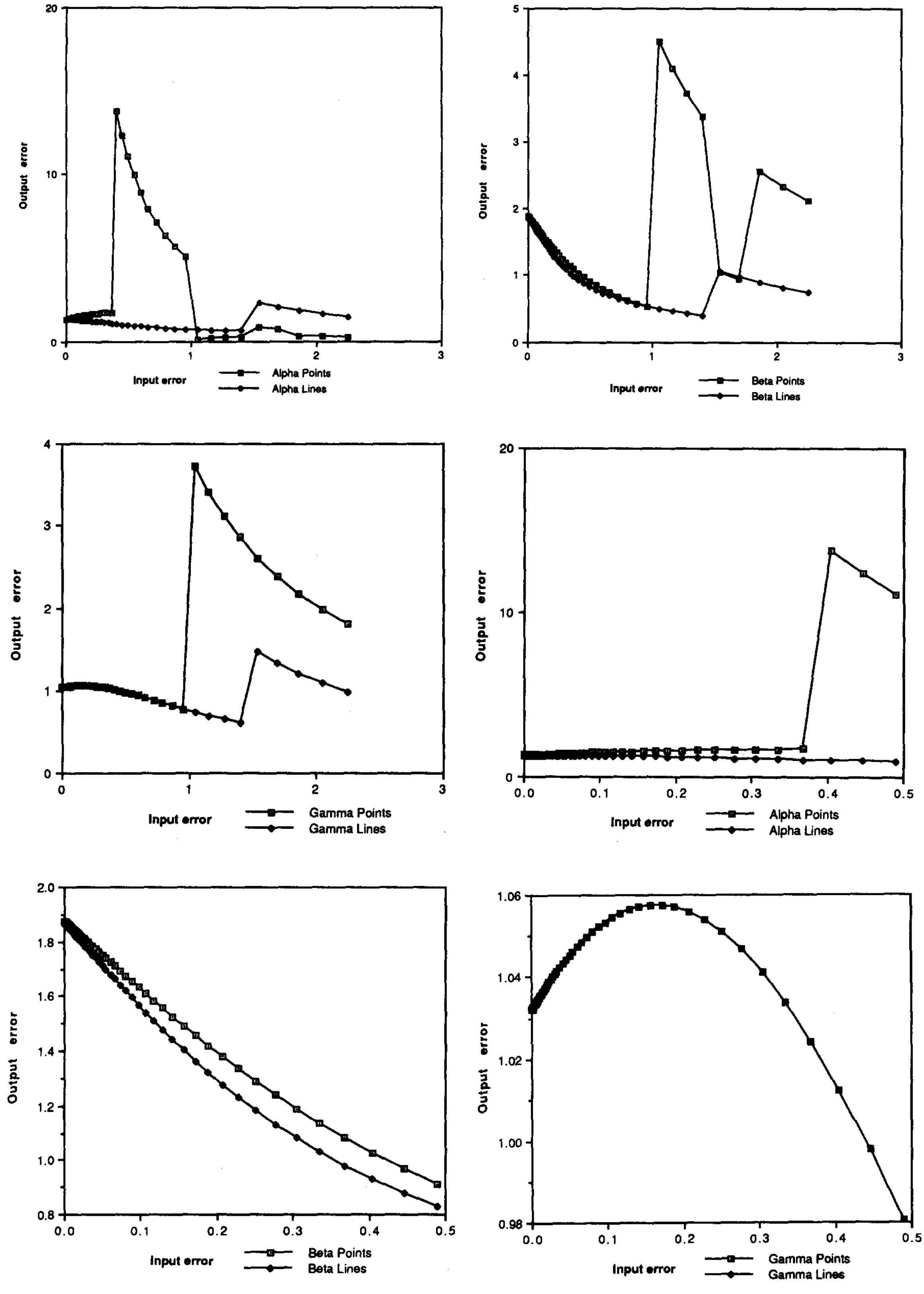


Fig. 4. Error analysis.

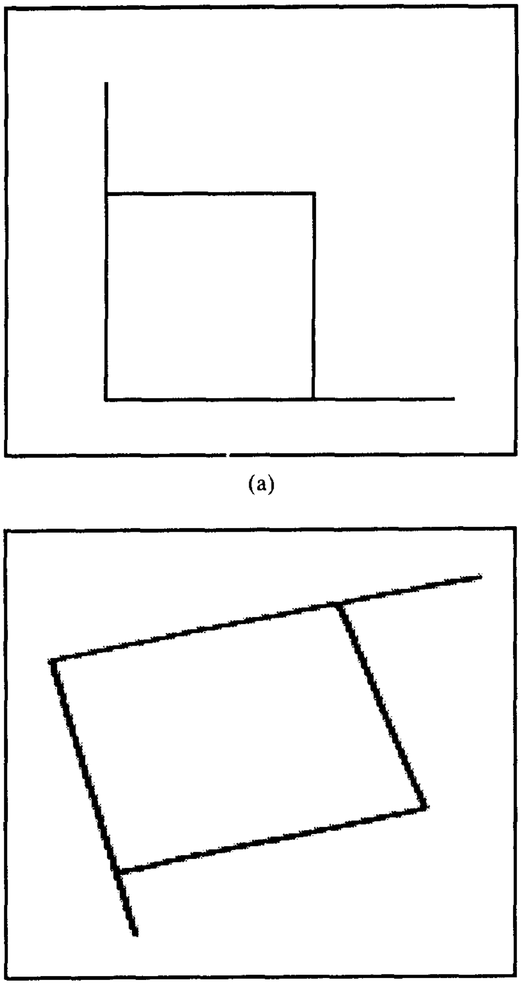
conditions, and in this section we discuss our results.

First we determined the intrinsic camera calibration parameters of our system.[3,4] In presenting our algorithm in the last section, we made several simplifying assumptions for the purpose of exposition. Determining the intrinsic camera calibration parameters is necessary both for implementing our algorithm (it is necessary to know the horizontal and vertical image scale factors, the center or principal point of the image, and d to implement it) and for correcting for real phenomena (such as spherical lens distortion) that we have suppressed in our discussion.

Next we implemented our algorithm and applied it to synthetic images to examine its response to noise. For rectangles, a comparison of the response of our algorithm to the response of the formulas presented in Haralick[1] depends on how the coordinates of vertices are (or the input is) computed for the latter. If the Hough method (or least squares) is used to determine the equations of the sides of the image quadrilateral, and these equations are used to determine the coordinates of the vertices, then the results obtained by applying the formulas presented in Haralick[1] would be the same as the results obtained by using our algorithm If the coordinates of the vertices are estimated directly, then the results obtained by using our algorithm. If the coordinates be more accurate than results obtained by applying the formulas presented in Haralick.[1] This occurs especially when the various image formation parameters are at extremes, and is probably simply because our algorithm is line-based (as opposed to point-based) and line-based techniques are generally more stable in the presence of noise than point-based techniques.

To illustrate, noise was randomly introduced both into the line data and into the vertex data of the rectangle image illustrated in Fig. 3. To the resulting line data we applied our algorithm, and to the resulting vertex data we applied the results of Haralick.[1] This was done 500 times at each maximum noise level of $\pm 2.25 * 1.1^{-n}$, $n = 1, \ldots$. Our results are presented in Fig. 4. For each Euler angle, the average ratio of output noise to input noise is plotted against maximum input noise. (The value of 1.68840880 at the noise level ± 0.27640344 in α, for example, represents the factor by which input noise in the range of ± 0.27640344 was modified in output noise in α.) Since jumps in these graphs occur at unstable input noise levels, the implication of these data is that, at least in this case, point-based formulas cannot tolerate noise over approximately 2.72%, but line-based results can easily tolerate noise up to approximately 10.00%.

Next, we applied our algorithm to real images. We created images of a known object, an object such as a piece of paper with lines drawn on it, that was resting on a table (see Fig. 5). We created images of the object in different positions, the camera being held fixed. Given that the planes of the (faces of the) objects being studied were constant, constancy of the computed normal vector was taken as a measure of success of our algorithm.



(a)

(b)

Fig. 5. A test image: (a) original test pattern; (b) actual image.

In general, we found that the accuracy of our algorithm depends on many factors. Least squares was used to determine the equations of lines, and the longer the line segments the more consistent was the output of our program. Also, the finer the line width, and the darker and more refined the image (the more contrast in the image), the more consistent was the output of our program. These properties of image lines were obviously greatly affected by the focus of the camera, the amount of light on the scene, and the type of material on which the lines appeared. Other factors that made major differences in the consistency of the output of our program were the degree of planarity of the surface on which the quadrilateral appeared, whether the lines were visual (e.g. drawn in ink) or structural (e.g. the edges of a polyhedron), the accuracy and optical properties of the real lines themselves, and whether the image being analyzed was obtained through simple thresholding or through some form of differentiation.

Studies were performed of quadrilaterals (including squares and rectangles) of different shapes and sizes. In general, there was no noticeable difference between working with squares, rectangles, and arbitrary quadrilaterals, except that if the dependence relation between the sides of a quadrilateral was close to being degenerate, then the accuracy of our algorithm began to suffer.

4. A SPECIAL CASE

The fact that there is sufficient information in the 2D perspective projection of an arbitrary quadrilateral of known shape and size in 3D space to determine its exact 3D coordinates, generalizes one of the primary results in Haralick.[1] We now show that the other primary result, namely that there is sufficient information in the 2D perspective projection of a rectangle of unknown size in 3D space to determine the camera look-angle parameters, is an immediate consequence of our first result.

Suppose that the (original) quadrilateral in the model plane is a rectangle, and, in particular, that the vertices of this rectangle are $V_1(0,0)$, $V_2(L,0)$, $V_3(L,W)$, and $V_4(0,W)$. If $l_1' = V_1V_2$, $l_2' = V_2V_3$, $l_3' = V_3V_4$, and $l_4' = V_4V_1$ ($l = PQ$ denoting the line determined by the points P and Q), then

$$A' = \begin{pmatrix} 0 & W & 0 \\ L & 0 & -L \\ 0 & -LW & LW \end{pmatrix}.$$

Thus, if $\boldsymbol{a}_i''$ is the ith row of A'', $i = 1, 2, 3$, then

$$A = A'A'' = \begin{pmatrix} 0 & W & 0 \\ L & 0 & -L \\ 0 & -LW & LW \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1'' \\ \boldsymbol{a}_2'' \\ \boldsymbol{a}_3'' \end{pmatrix}$$

$$= \begin{pmatrix} W\boldsymbol{a}_2'' \\ L(\boldsymbol{a}_1'' - \boldsymbol{a}_3'') \\ LW(\boldsymbol{a}_3'' - \boldsymbol{a}_2'') \end{pmatrix} = \frac{1}{k} \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \boldsymbol{h} \end{pmatrix}$$

so $\boldsymbol{r}_1 = \boldsymbol{a}_2''/|\boldsymbol{a}_2''|$, $\boldsymbol{r}_2 = (\boldsymbol{a}_1'' - \boldsymbol{a}_3'')/|\boldsymbol{a}_1'' - \boldsymbol{a}_3''|$, $\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2$, and $\boldsymbol{h}$ is a multiple of $(\boldsymbol{a}_3'' - \boldsymbol{a}_2'')/|\boldsymbol{a}_3'' - \boldsymbol{a}_2''|$. Since we can thus determine R, we can thus again completely determine the camera look-angle parameters (as well as the translation T up to a scalar multiple). Note also, in passing, that in this case we can actually say more: indeed, $L/W = |\boldsymbol{a}_2''|/|\boldsymbol{a}_1'' - \boldsymbol{a}_3''|$.

5. CONCLUSION

There are many more patterns in nature that have quadrilaterals in them than there are patterns that only have rectangles in them, so the algorithm presented in this paper significantly broadens the scope of the algorithms presented in Haralick.[1] Further, the algorithm presented in this paper is easy to implement and numerically stable in the presence of noise. The real issue, of course, is whether adequately accurate input data can be deduced from an arbitrary real image so that any of these algorithms can yield useful information. Our experiments confirm that under most conditions, accurate input data can be deduced from an image so that our algorithms can yield useful information. Under certain conditions, however, accurate input data cannot be deduced, and the best that our (or any other similar) algorithm can do is to yield coarse information: information that may not be useful in itself, but information which when combined with other knowledge is useful. This, however, is not too surprising.

SUMMARY

In a recent paper of Haralick appearing in this journal,[1] it is shown that there is sufficient information in the two-dimensional perspective projection of a rectangle of unknown size in three-space to determine the camera look-angle parameters, and that if the size of the rectangle is given then its exact three-dimensional coordinates can be computed. In this paper we show that there is sufficient information in the two-dimensional perspective projection of an arbitrary quadrilateral of known shape and size in three-space to determine its exact three-dimensional coordinates. This result generalizes Haralick's second result since a quadrilateral is more general than a rectangle. (Our result applies, for example, to a quadrilateral determined by four sides of a polygonal face of an object, while Haralick's does not.) For a rectangle of unknown size, we can determine the camera look-angle parameters, giving us Haralick's first result, as well as the ratio of length to width of the rectangle. Our computations are simple, making our approach easy to implement. Further, our approach is line-based (as opposed to point-based) so it is numerically stable in the presence of noise. Finally, in this paper we also discuss implementation results.

REFERENCES

1. R. Haralick, Determining camera parameters from the perspective projection of a rectangle, *Pattern Recognition* **22**, 225–230 (1989).
2. M. A. Penna and R. R. Patterson, *Projective Geometry and its Applications to Computer Graphics*. Prentice-Hall, Englewood Cliffs, NJ (1986).
3. 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. Pattern Anal. Mach. Intel.* **10**, 713–720 (1988).
4. R. Y. Tsai, Review of RAC-based camera calibration, Technical Report, IBM T. J. Watson Reseach Center, Yorktown Heights, NY (1988).

APPENDIX 1. A BRIEF INTRODUCTION TO PROJECTIVE GEOMETRY

In this section we provide a brief background on projective geometry. For more complete details, see Penna and Patterson.[2]

(a) *The projective plane and homogeneous coordinates*

The origin of homogeneous coordinates is in the quantitative analysis of perspective. Given, in three-space, a plane Π and a point O not in Π, the perspective image of an object is formed by associating to each point P of the object the point Q of intersection of Π and the line OP determined by O and P (see Fig. 1). The plane Π is the *image plane*, the point O is the *center of perspectivity*, and the map π from world space to Π is *perspective projection*. If coordinates are chosen so that *O is the origin and* Π is the plane whose equation is $z = 1$, then

$$\pi(x, y, z) = \left(\frac{x}{z}, \frac{y}{z}, 1\right).$$

Observe that π is a many-to-one mapping: indeed, for any point $Q(X, Y, 1)$ in Π and any non-zero real number z, $\pi(zX, zY, z) = (X, Y, 1)$.

In the study of perspective, one often thinks of (or identifies) each point $Q(X, Y, 1)$ in Π with one or more points $P(x, y, z)$ that project to it. (Rather than thinking of a point as a glob of paint on a canvas, one thinks of the point for what it artistically represents in the picture as a whole.) If we identify each point $Q(X, Y, 1)$ in Π with all points $P(x, y, z)$ in Euclidean three-space that project to it, then we are naturally associating to Q the set

$$[x, y, z] = \left\{(x, y, z) \mid X = \frac{x}{z}, Y = \frac{y}{z}\right\}.$$

These are the homogeneous coordinates of $Q(X, Y, 1)$; that is, the *homogeneous coordinates* of $Q(X, Y, 1)$ are the set $[x, y, z]$ of all triples (x, y, z) of real numbers for which $X = x/z$ and $Y = y/z$.

Note that the homogeneous coordinates of $Q(X, Y, 1)$ are not the entries of a triple of real numbers (as are Cartesian coordinates); rather, they are a set of triples (x, y, z) of real numbers. An element of this set is a *representative* of the homogeneous coordinates for $Q(X, Y, 1)$, and such representatives are uniquely defined only up to scalar multiples. That is, given two representatives (x_1, y_1, z_1) and (x_2, y_2, z_2) of the homogeneous coordinates $[x, y, z]$, there is a constant k for which $x_1 = kx_2$, $y_1 = ky_2$ and $z_1 = kz_2$. (For example, $(2, 3, 1)$ and $(10, 15, 5)$ are both representatives of the homogeneous coordinates $[2, 3, 1]$ of the point $Q(2, 3, 1)$; indeed, $[2, 3, 1] = [10, 15, 5]$.) Note also that the third Cartesian coordinate for each point $Q(X, Y, 1)$ in Π is 1; hence, when there is no room for ambiguity, we may use a phrase such as "... the point $Q(X, Y)$ in the plane whose homogeneous coordinates are $[x, y, z]$ ..." for short.

Now, consider the line l in Π whose direction numbers are $(a, b, 0)$ and which passes through the point whose Cartesian coordinates are $(x_0, y_0, 1)$. This line can be parametrized by the equations $x = at + x_0$, $y = bt + y_0$, $z = 1$ where $t \in R$. The limit of the points on l as $t \to \pm\infty$ is the *ideal point* of l; ideal points arise in addressing vanishing points in perspective images. Since

$$\lim_{t \to \pm\infty} [x, y, 1] = \lim_{t \to \pm\infty} [at + x_0, bt + y_0, 1]$$

$$= \lim_{t \to \pm\infty} \left[a + \frac{x_0}{t}, b + \frac{y_0}{t}, \frac{1}{t}\right]$$

$$= \left[\lim_{t \to \pm\infty} \left(a + \frac{x_0}{t}, b + \frac{y_0}{t}, \frac{1}{t}\right)\right] = [a, b, 0]$$

it follows that the homogeneous coordinates of the ideal point of l are $[a, b, 0]$. The set of Euclidean points $Q = Q(X, Y, 1)$ on Π together with the set of ideal points of lines in Π form the *projective plane* or *projective two-space*.

(b) *Projective transformations*

Among the most important transformations of the Euclidean plane are the linear transformations, those transformations represented by right matrix multiplication: $T(\boldsymbol{x}) = \boldsymbol{x}A$. Right multiplication by an invertible 3×3 matrix A

$$T = T_A : [x, y, z] \to [(x, y, z)A]$$

represents an important transformation of the projective plane known as a *projective transformation*. One can think of a projective transformation as an image transformation which is associated to (or induced by) the linear transformation of world space represented by multiplication by A. The study of projective transformations is important in image analysis where, for example, one might wish to say things about a rigid motion of an object in world space based on the induced projective transformation observed in the image plane. The matrix representing a projective transformation is unique up to a scalar multiple. Further, for any projective transformations $T_{A'}$ and $T_{A''}$, $T_{A'} \circ T_{A''} = T_{A'A''}$ ($\circ$ denoting composition), and thus the inverse of the projective transformation T_A represented by the matrix A is the projective transformation $T_{A^{-1}}$ represented by A^{-1}.

The most important results of projective geometry (the Fundamental Theorems of Projective Geometry) are statements about projective transformations. The Fundamental Theorem of Projective Geometry for points in the projective plane states that if $\{P'_1, P'_2, P'_3, P'_4\}$ and $\{P''_1, P''_2, P''_3, P''_4\}$ are two sets of points in the projective plane, no three points in either set being collinear, then there is a unique projective transformation T taking P'_i to P''_i, $i = 1, \ldots, 4$. The existence part of this result is important, it is by construction, and it goes as follows:

Suppose $\boldsymbol{p}'_i$ is a representative of the homogeneous coordinates of the point $\boldsymbol{P}'_i$, $i = 1, \ldots, 4$. Since the vectors $\boldsymbol{p}'_i$, $i = 1, \ldots, 4$ are linearly dependent, there are constants k'_i, $i = 1, \ldots, 3$ for which $\boldsymbol{p}'_4 = k'_1\boldsymbol{p}'_1 + k'_2\boldsymbol{p}'_2 + k'_3\boldsymbol{p}'_3$. Now if

$$A' = \begin{pmatrix} k'_1\boldsymbol{p}'_1 \\ k'_2\boldsymbol{p}'_2 \\ k'_3\boldsymbol{p}'_3 \end{pmatrix}$$

is the matrix whose ith row is $k'_i\boldsymbol{p}'_i$, $i = 1, \ldots, 3$, then the projective transformation $T_{A'}$ represented by A' maps the points $I_x[1, 0, 0]$, $I_y[0, 1, 0]$, $O[0, 0, 1]$ and $U[1, 1, 1]$ to the points P'_1, P'_2, P'_3 and P'_4, respectively:

$$T_{A'}[1, 0, 0] = [(1, 0, 0)A'] = [k'_1\boldsymbol{p}'_1] = [\boldsymbol{p}'_1]$$

$$T_{A'}[0, 1, 0] = [(0, 1, 0)A'] = [k'_2\boldsymbol{p}'_2] = [\boldsymbol{p}'_2]$$

$$T_{A'}[0, 0, 1] = [(0, 0, 1)A'] = [k'_3\boldsymbol{p}'_3] = [\boldsymbol{p}'_3]$$

$$T_{A'}[1, 1, 1] = [(1, 1, 1)A'] = [k'_1\boldsymbol{p}'_1 + k'_2\boldsymbol{p}'_2 + k'_3\boldsymbol{p}'_3] = [\boldsymbol{p}'_4].$$

Thus the projective transformation $T_{A'^{-1}}$ represented by A'^{-1} maps P'_1, P'_2, P'_3 and P'_4 to I_x, I_y, O and U. Consequently, if A'' is the matrix representing the projective transformation $T_{A''}$ that maps I_x, I_y, O and U to P''_1, P''_2, P''_3 and P''_4, then the composite $T_{A'^{-1}} \circ T_{A''} = T_{A'^{-1}A''}$ maps P'_1, P'_2, P'_3 and P'_4 to $(I_x, I_y, O$ and U, and then I_x, I_y, O and U back to) P''_1, P''_2, P''_3 and P''_4.

(c) *Generalizations*

We close this section by generalizing some of the ideas presented above.

First, although it may be difficult to visualize perspective projection from Euclidean four-space into a Euclidean three-plane, we can still formally define the homogeneous coordinates $[x, y, z, w]$ of points in Euclidean three-space, ideal points on lines in Euclidean three-space and finally projective three-space. Further, we can discuss projective transformations of projective three-space: these are transformations

$$T = T_A: [x, y, z, w] \to [(x, y, z, w)A]$$

represented by right multiplication by a 4×4 invertible matrix A. For such transformations there is a Fundamental Theorem of Projective Geometry which states that if $\{P'_1, P'_2, P'_3, P'_4, P'_5\}$ and $\{P''_1, P''_2, P''_3, P''_4, P''_5\}$ are two sets of points in the projective plane, no four points in either set being coplanar, then there is a unique projective transformation T taking P'_i to P''_i, $i = 1, \ldots, 5$.

Second, we can discuss projective transformations

$$T = T_A : [x, y, z] \to [(x, y, z, w)A]$$

from the projective plane to projective three-space, and projective tranformations

$$T = T_A : [x, y, z, w] \to [(x, y, z)A]$$

from projective three-space to the projective plane. These are transformations represented by 3×4 and 4×3 matrices A, respectively, of rank 3. Again, for any projective transformations $T_{A'}$ and $T_{A''}$, $T_{A'} \circ T_{A''} = T_{A'A''}$ ($\circ$ denoting composition).

Third, we can extend our statements about points in the projective plane to lines in the projective plane: as in the Euclidean plane, a line l in the projective plane can be characterized as the set of points $Q[x, y, z]$ for which $ax + by + cz = 0$ for some constants a, b, c. (This definition is independent of the representative (x, y, z) of $[x, y, z]$.) Since ka, kb, kc, for $k \neq 0$, define the same line as a, b, c, there is a natural association to l of $[a, b, c]$; $[a, b, c]$ are the homogeneous coordinates of l. Since $Q[x, y, z]$ lies on $l[a, b, c]$ if and only if

$$(x \quad y \quad z)\begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

and since

$$(x \quad y \quad z)\begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0 \text{ if and only if } (x \quad y \quad z)AA^{-1}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$$

it follows that the projective transformation T_A of the projective plane represented by the matrix A, maps lines by

$$T = T_A : \begin{bmatrix} a \\ b \\ c \end{bmatrix} \to \left[A^{-1}\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right].$$

Finally, there is a Fundamental Theorem of Projective Geometry for lines in the projective plane which states that if $\{l'_1, l'_2, l'_3, l'_4\}$ and $\{l''_1, l''_2, l''_3, l''_4\}$ are two sets of lines in the projective plane, no three lines in either set meeting at one point, then there is a unique projective transformation T of the projective plane taking l'_i to l''_i, $i = 1, \ldots, 4$: as in the case for points in the projective plane, if l'_i is a representative of the homogeneous coordinates of the line l'_i, $i = 1, \ldots, 4$, then there are constants k'_1, k'_2, k'_3 for which $l'_4 = k'_1 l'_4 + k'_2 l'_2 + k'_3 l'_3$. Thus if $A'^{-1} = (k'_1 l'_1 \; k'_2 l'_2 \; k'_3 l'_3)$ is the 3×3 matrix whose ith column is $k'_i l'_i = 1, 2, 3$, then the projective transformation represented by A' maps the lines $i_x[1, 0, 0]$, $i_y[0, 1, 0]$, $o[0, 0, 1]$ and $u[1, 1, 1]$ to the lines l'_1, l'_2, l'_3 and l'_4, respectively, and the projective transformation represented by A'^{-1} maps the lines l'_1, l'_2, l'_3 and l'_4 to i_x, i_y, o and u. Consequently, if A'' is the matrix representing the projective transformation that maps i_x, i_y, o and u to l''_1, l''_2, l''_3 and l''_4, then the composite transformation, which is represented by $A'^{-1}A''$, maps l'_1, l'_2, l'_3 and l'_4 to l''_1, l''_2, l''_3 and l''_4.

APPENDIX 2. HARALICK[1] REVISITED

The formulas presented in Haralick[1] are based on a different mechanism for modeling the perspective display of a rectangle than the mechanism used in the present paper. In this section we briefly discuss the formulas presented in Haralick[1] from the point of view of this paper. For simplicity we assume throughout this section that the focal length $f = 1$.

A rigid motion of three-space may be written in Cartesian coordinates

$$(x, y, z) \to (x, y, z)R + (h_1, h_2, h_3) = (x + h'_1, y + h'_2, z + h'_3)R,$$

where $(h'_1, h'_2, h'_3) = (h_1, h_2, h_3)R^{-1}$. If $(h_1, h_2, h_3) \neq (0, 0, 0)$, then this map does not induce a projective transformation of the projective plane. However, right multiplication $(x, y, z) \to (x, y, z)R$ by R does induce a projective tranformation $[x, y, z] \to [(x, y, z)R]$ of the projective plane, and we have the following result.

Theorem. If (x^*_i, z^*_i), $i = 1, \ldots, 4$, are the vertices of the image of a rectangle under perspective projection into the plane whose equation is $y = 1$, $(x_2, y_2, z_2) = (x_1 + W, y_1, z_1)$, $(x_3, y_3, z_3) = (x_1, y_1 + L, z_1)$, $(x_4, y_4, z_4) = (x_1 + W, y_1 + L, z_1)$ and

$$R = \begin{pmatrix} \cos\theta\cos\xi + \sin\theta\sin\phi\sin\xi & -\sin\theta\cos\phi & -\cos\theta\sin\xi + \sin\theta\sin\phi\cos\xi \\ \sin\theta\cos\xi - \cos\theta\sin\phi\sin\xi & \cos\theta\cos\phi & -\sin\theta\sin\xi - \cos\theta\sin\phi\cos\xi \\ \cos\phi\sin\xi & \sin\phi & \cos\phi\cos\xi \end{pmatrix}$$

then the system

$$[x^*_i, 1, z^*_i] = [(x_i, y_i, z_i)R]$$

represents a system

$$(x^*_i, 1, z^*_i) = \lambda_i(x_i, y_i, z_i)R \qquad i = 1, \ldots, 4$$

of 12 real equations in 12 unknowns—namely x_1, y_1, z_1, L, W, θ, ϕ, ξ, λ_1, λ_2, λ_3, λ_4—which can be solved explicitly for θ, ϕ and ξ. Further, if L and W are known, then this system can be solved explicitly for x_1, y_1, z_1 as well.

(Note that the form of the rotation matrix R in this result is somewhat different than the form of the rotation matrix used in Section 2.)

This result is proved in Haralick[1] by explicitly solving the system of equations. What is perhaps most interesting about this result is the way in which it handles a rigid motion by absorbing the translational part into data, and treating the rotational part as a linear transformation of Euclidean three-space that induces a projective transformation of the projective plane.

About the Author—MICHAEL A. PENNA received a B.S. degree in Mathematics from Union College, Schenectady, NY, in 1967, and M.S. and Ph.D. degrees in Mathematics from the University of Illinois, Urbana-Champaign, IL, in 1968 and 1973, respectively.

Dr Penna joined the faculty of Indiana University–Purdue University at Indianapolis in 1973. He has been a faculty member there since then, and currently holds the rank of Profesor of Mathematical Sciences and Professor of Computer and Information Science. Dr Penna's research interests include computer vision, image processing and pattern recognition.