Reprinted from Journal of the Optical Society of America A, Vol. 4, page 629, April 1987

Closed-form solution of absolute orientation using unit quaternions

Berthold K. P. Horn

Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96720

Received August 6, 1986; accepted November 25, 1986

Finding the relationship between two coordinate systems using pairs of measurements of the coordinates of a number of points in both systems is a classic photogrammetric task. It finds applications in stereophotogrammetry and in robotics. I present here a closed-form solution to the least-squares problem for three or more points. Currently various empirical, graphical, and numerical iterative methods are in use. Derivation of the solution is simplified by use of unit quaternions to represent rotation. I emphasize a symmetry property that a solution to this problem ought to possess. The best translational offset is the difference between the centroid of the coordinates in one system and the rotated and scaled centroid of the coordinates in the other system. The best scale is equal to the ratio of the root-mean-square deviations of the coordinates in the two systems from their respective centroids. These exact results are to be preferred to approximate methods based on measurements of a few selected points. The unit quaternion representing the best rotation is the eigenvector associated with the most positive eigenvalue of a symmetric 4 × 4 matrix. The elements of this matrix are combinations of sums of products of corresponding coordinates of the points.

1. INTRODUCTION

Suppose that we are given the coordinates of a number of points as measured in two different Cartesian coordinate systems (Fig. 1). The photogrammetric problem of recovering the transformation between the two systems from these measurements is referred to as that of *absolute orientation*.[1] It occurs in several contexts, foremost in relating a stereo model developed from pairs of aerial photographs to a geodetic coordinate system. It also is of importance in robotics, in which measurements in a camera coordinate system must be related to coordinates in a system attached to a mechanical manipulator. Here one speaks of the determination of the hand–eye transform.[2]

A. Previous Work

The problem of absolute orientation is usually treated in an empirical, graphical, or numerical iterative fashion.[1,3,4] Thompson[5] gives a solution to this problem when three points are measured. His method, as well as the simpler one of Schut,[6] depends on selective neglect of the extra constraints available when all coordinates of three points are known. Schut uses unit quaternions and arrives at a set of linear equations. I present a simpler solution to this special case in Subsection 2.A that does not require solution of a system of linear equations. These methods all suffer from the defect that they cannot handle more than three points. Perhaps more importantly, they do not even use all the information available from the three points.

Oswal and Balasubramanian[7] developed a least-squares method that can handle more than three points, *but* their method does not enforce the orthonormality of the rotation matrix. An iterative method is then used to square up the result—bringing it closer to being orthonormal. The method for doing this is iterative, and the result is *not* the solution of the original least-squares problem.

I present a closed-form solution to the least-squares problem in Sections 2 and 4 and show in Section 5 that it simplifies greatly when only three points are used. This is important, since at times only three points may be available. The solution is different from those described at the beginning of this section because it does not selectively neglect information provided by the measurements—it uses all of it.

The groundwork for the application of quaternions in photogrammetry was laid by Schut[8] and Thompson.[9] In robotics, Salamin[10] and Taylor[11] have been the main propagandists. The use of unit quaternions to represent rotation is reviewed in Section 3 and the appendixes (see also Ref. 2).

B. Minimum Number of Points

The transformation between two Cartesian coordinate systems can be thought of as the result of a rigid-body motion and can thus be decomposed into a rotation and a translation. In stereophotogrammetry, in addition, the scale may not be known. There are obviously three degrees of freedom to translation. Rotation has another three (direction of the axis about which the rotation takes place plus the angle of rotation about this axis). Scaling adds one more degree of freedom. Three points known in both coordinate systems provide nine constraints (three coordinates each), more than enough to permit determination of the seven unknowns.

By discarding two of the constraints, seven equations in seven unknowns can be developed that allow one to recover the parameters. I show in Subsection 2.A how to find the rotation in a similar fashion, provided that the three points are not collinear. Two points clearly do not provide enough constraint.

C. Least Sum of Squares of Errors

In practice, measurements are not exact, and so greater accuracy in determining the transformation parameters will be sought for by using more than three points. We no longer

0740-3232/87/040629-14$02.00

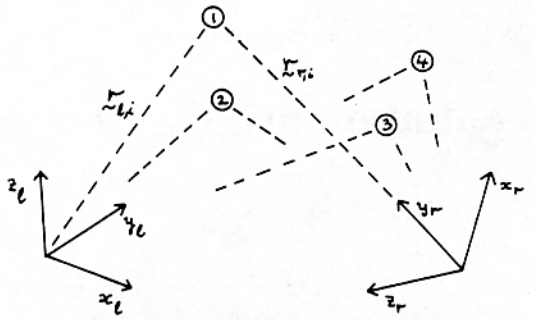

Fig. 1. The coordinates of a number of points is measured in two different coordinate systems. The transformation between the two systems is to be found.

expect to be able to find a transformation that maps the measured coordinates of points in one system exactly into the measured coordinates of these points in the other. Instead, we minimize the sum of squares of residual errors. Finding the best set of transformation parameters is not easy. In practice, various empirical, graphical, and numerical procedures are in use. These are iterative in nature. That is, given an approximate solution, such a method leads to a better, but still imperfect, answer. The iterative method is applied repeatedly until the remaining error is negligible.

At times, information is available that permits one to obtain so good an initial guess of the transformation parameters that a single step of the iteration brings one close enough to the true solution of the least-squares problem to eliminate the need for further iteration in a practical situation.

D. Closed-Form Solution

In this paper I present a closed-form solution to the least-squares problem of absolute orientation, one that does not require iteration. One advantage of a closed-form solution is that it provides one in a single step with the best possible transformation, given the measurements of the points in the two coordinate systems. Another advantage is that one need not find a good initial guess, as one does when an iterative method is used.

I give the solution in a form in which unit quaternions are used to represent rotations. The solution for the desired quaternion is shown to be the eigenvector of a symmetric 4 × 4 matrix associated with the most positive eigenvalue. The elements of this matrix are simple combinations of sums of products of corresponding coordinates of the points. To find the eigenvalues, a quartic equation has to be solved whose coefficients are sums of products of elements of the matrix. It is shown that this quartic is particularly simple, since one of its coefficients is zero. It simplifies even more when one or the other of the sets of measurements is coplanar.

E. Orthonormal Matrices

While unit quaternions constitute an elegant representation for rotation, most of us are more familiar with orthonormal matrices with positive determinant. Fortunately, the appropriate 3 × 3 rotation matrix can be easily constructed from the four components of the unit quaternion, as is shown in Subsection 3.E. Working directly with matrices is difficult because of the need to deal with six nonlinear constraints that ensure that the matrix is orthonormal. A closed-form solution for the rotation matrix using manipulations of matrices will be presented in a subsequent paper. This closed-form solution requires the computation of the square root of a symmetric 3 × 3 matrix.

F. Symmetry of the Solution

Let us call the two coordinate systems "left" and "right." A desirable property of a solution method is that, when applied to the problem of finding the best transformation from the right to the left system, it gives the exact inverse of the best transformation from the left to the right system. I show in Subsection 2.D that the scale factor has to be treated in a particular way to guarantee that this will happen. Symmetry is guaranteed when one uses unit quaternions to represent rotation.

2. SOLUTION METHODS

As we shall see, the translation and the scale factor are easy to determine once the rotation is known. The difficult part of the problem is finding the rotation. Given three noncollinear points, we can easily construct a useful triad in each of the left and the right coordinate systems (Fig. 2). Let the origin be at the first point. Take the line from the first to the second point to be the direction of the new x axis. Place the new y axis at right angles to the new x axis in the plane formed by the three points. The new z axis is then made to be orthogonal to the x and y axes, with orientation chosen to satisfy the right-hand rule. This construction is carried out in both left and right systems. The rotation that takes one of these constructed triads into the other is also the rotation that relates the two underlying Cartesian coordinate systems. This rotation is easy to find, as we show below.

A. Selective Discarding Constraints

Let the coordinates of the three points in each of the two coordinate systems be $\mathbf{r}_{l,1}$, $\mathbf{r}_{l,2}$, $\mathbf{r}_{l,3}$ and $\mathbf{r}_{r,1}$, $\mathbf{r}_{r,2}$, $\mathbf{r}_{r,3}$, respectively. Construct

$$\mathbf{x}_l = \mathbf{r}_{l,2} - \mathbf{r}_{l,1}.$$

Then

$$\hat{x}_l = \mathbf{x}_l / \|\mathbf{x}_l\|$$

is a unit vector in the direction of the new x axis in the left-hand-system. Now let

$$\mathbf{y}_l = (\mathbf{r}_{l,3} - \mathbf{r}_{l,1}) - [(\mathbf{r}_{l,3} - \mathbf{r}_{l,1}) \cdot \hat{x}_l]\hat{x}_l$$

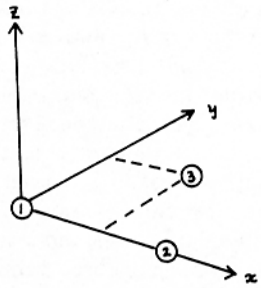

Fig. 2. Three points define a triad. Such a triad can be constructed by using the left measurements. A second triad is then constructed from the right measurements. The required coordinate transformation can be estimated by finding the transformation that maps one triad into the other. This method does not use the information about each of the three points equally.

be the component of $(\mathbf{r}_{l,3} - \mathbf{r}_{l,1})$ perpendicular to $\hat{x}$. The unit vector

$$\hat{y}_l = \mathbf{y}_l / \|\mathbf{y}_l\|$$

is in the direction of the new y axis in the left-hand system. To complete the triad, we use the cross product

$$\hat{z}_l = \hat{x}_l \times \hat{y}_l.$$

This construction is now repeated in the right-hand system to obtain $\hat{x}_r$, $\hat{y}_r$, and $\hat{z}_r$. The rotation that we are looking for takes $\hat{x}_l$ into $\hat{x}_r$, $\hat{y}_l$ into $\hat{y}_r$, and $\hat{z}_l$ into $\hat{z}_r$.

Now adjoin column vectors to form the matrices M_l and M_r as follows:

$$M = |\hat{x}_l \hat{y}_l \hat{z}_l|, \qquad M_r = |\hat{x}_r \hat{y}_r \hat{z}_r|.$$

Given a vector $\mathbf{r}_l$ in the left coordinate system, we see that

$$M_l{}^T \mathbf{r}_l$$

gives us the components of the vector $\mathbf{r}_l$ along the axes of the constructed triad. Multiplication by M_r then maps these into the right-hand coordinate system, so

$$\mathbf{r}_r = M_r M_l{}^T \mathbf{r}_l.$$

The sought-after rotation is given by

$$R = M_r M_l{}^T.$$

The result is orthonormal since M_r and M_l are orthonormal, by construction. The above constitutes a closed-form solution for finding the rotation, given three points. Note that it uses the information from the three points selectively. Indeed, if we renumber the points, we get a different rotation matrix, unless the data happen to be perfect. Also note that the method cannot be extended to deal with more than three points.

Even with just three points we should really attack this problem by using a least-squares method, since there are more constraints than unknown parameters. The least-squares solution for translation and scale will be given in Subsections 2.C and 2.E. The optimum rotation is found in Section 4.

B. Finding the Translation

Let there be n points. The measured coordinates in the left and right coordinate system will be denoted by

$$\{\mathbf{r}_{l,i}\} \quad \text{and} \quad \{\mathbf{r}_{r,i}\},$$

respectively, where i ranges from 1 to n. We are looking for a transformation of the form

$$\mathbf{r}_r = sR(\mathbf{r}_l) + \mathbf{r}_0$$

from the left to the right coordinate system. Here s is a scale factor, $\mathbf{r}_0$ is the translational offset, and $R(\mathbf{r}_l)$ denotes the rotated version of the vector $\mathbf{r}_l$. We do not, for the moment, use any particular notation for rotation. We use only the facts that rotation is a linear operation and that it preserves lengths so that

$$\|R(\mathbf{r}_l)\|^2 = \|\mathbf{r}_l\|^2,$$

where $\|\mathbf{r}\|^2 = \mathbf{r} \cdot \mathbf{r}$ is the square of the length of the vector $\mathbf{r}$.

Unless the data are perfect, we will not be able to find a scale factor, a translation, and a rotation such that the transformation equation above is satisfied for each point. Instead there will be a residual error

$$\mathbf{e}_i = \mathbf{r}_{r,i} - sR(\mathbf{r}_{l,i}) - \mathbf{r}_0.$$

We will minimize the sum of squares of these errors

$$\sum_{i=1}^{n} \|\mathbf{e}_i\|^2.$$

(I show in Appendix A that the measurements can be weighted without changing the basic solution method.)

We consider the variation of the total error first with translation, then with scale, and finally with respect to rotation.

C. Centroids of the Sets of Measurements

It turns out to be useful to refer all measurements to the centroids defined by

$$\bar{\mathbf{r}}_l = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}_{l,i}, \qquad \bar{\mathbf{r}}_r = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}_{r,i}.$$

Let us denote the new coordinates by

$$\mathbf{r}'_{l,i} = \mathbf{r}_{l,i} - \bar{\mathbf{r}}_l, \qquad \mathbf{r}'_{r,i} = \mathbf{r}_{r,i} - \bar{\mathbf{r}}_r.$$

Note that

$$\sum_{i=1}^{n} \mathbf{r}'_{l,i} = \mathbf{0}, \qquad \sum_{i=1}^{n} \mathbf{r}'_{r,i} = \mathbf{0}.$$

Now the error term can be rewritten as

$$\mathbf{e}_i = \mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i}) - \mathbf{r}'_0,$$

where

$$\mathbf{r}'_0 = \mathbf{r}_0 - \bar{\mathbf{r}}_r + sR(\bar{\mathbf{r}}_l).$$

The sum of squares of errors becomes

$$\sum_{i=1}^{n} \|\mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i}) - \mathbf{r}'_0\|^2$$

or

$$\sum_{i=1}^{n} \|\mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i})\|^2 - 2\mathbf{r}'_0 \cdot \sum_{i=1}^{n} [\mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i})] + n\|\mathbf{r}'_0\|^2.$$

Now the sum in the middle of this expression is zero since the measurements are referred to the centroid. So we are left with the first and third terms. The first does not depend on $\mathbf{r}'_0$, and the last cannot be negative. The total error is obviously minimized with $\mathbf{r}'_0 = 0$ or

$$\mathbf{r}_0 = \bar{\mathbf{r}}_r - sR(\bar{\mathbf{r}}_l).$$

That is, the translation is just the difference of the right centroid and the scaled and rotated left centroid. We return to this equation to find the translational offset once we have found the scale and rotation.

This method, based on all available information, is to be preferred to one that uses only measurements of one or a few selected points to estimate the translation.

At this point we note that the error term can be written as

$$\mathbf{e}_i = \mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i})$$

since $\mathbf{r}'_0 = 0$. So the total error to be minimized is just

$$\sum_{i=1}^{n} \|\mathbf{r}'_{r,i} - sR(\mathbf{r}'_{l,i})\|^2.$$

D. Finding the Scale

Expanding the total error and noting that

$$\|R(\mathbf{r}'_{l,i})\|^2 = \|\mathbf{r}'_{l,i}\|^2,$$

we obtain

$$\sum_{i=1}^{n} \|r'_{r,i}\|^2 - 2s \sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot R(\mathbf{r}'_{l,i}) + s^2 \sum_{i=1}^{n} \|\mathbf{r}'_{l,i}\|^2,$$

which can be written in the form

$$S_r - 2sD + s^2 S_l,$$

where S_r and S_l are the sums of the squares of the measurement vectors (relative to their centroids), while D is the sum of the dot products of corresponding coordinates in the right system with the rotated coordinates in the left system. Completing the square in s, we get

$$(s\sqrt{S_l} - D/\sqrt{S_l})^2 + (S_r S_l - D^2)/S_l.$$

This is minimized with respect to scale s when the first term is zero or $s = D/S_l$, that is,

$$s = \sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot R(\mathbf{r}'_{l,i}) \Bigg/ \sum_{i=1}^{n} \|\mathbf{r}'_{l,i}\|^2.$$

E. Symmetry in Scale

If, instead of finding the best fit to the transformation,

$$\mathbf{r}_r = sR(\mathbf{r}_l) + \mathbf{r}_0,$$

we try to find the best fit to the inverse transformation,

$$\mathbf{r}_l = \bar{s}\bar{R}(\mathbf{r}_r) + \bar{\mathbf{r}}_0,$$

we might hope to get the exact inverse:

$$\bar{s} = 1/s, \qquad \bar{\mathbf{r}}_0 = -\frac{1}{s} R^{-1}(\mathbf{r}_0), \qquad \bar{R} = R^{-1}.$$

This does not happen with the above formulation. By exchanging left and right, we find instead that $\bar{s} = \bar{D}/S_r$ or

$$\bar{s} = \sum_{i=1}^{n} \mathbf{r}'_{l,i} \cdot \bar{R}(\mathbf{r}'_{r,i}) \Bigg/ \sum_{i=1}^{n} \|\mathbf{r}'_{r,i}\|^2,$$

which in general will not equal $1/s$, as determined above. (This is illustrated in an example in Appendix A1.)

One of the two asymmetrical results shown above may be appropriate when the coordinates in one of the two systems are known with much greater precision than those in the other. If the errors in both sets of measurements are similar, it is more reasonable to use a symmetrical expression for the error term:

$$\mathbf{e}_i = \frac{1}{\sqrt{s}} \mathbf{r}'_{r,i} - \sqrt{s} R(\mathbf{r}'_{l,i}).$$

Then the total error becomes

$$\frac{1}{s} \sum_{i=1}^{n} \|r'_{r,i}\|^2 - 2 \sum_{i=1}^{n} r'_{r,i} \cdot R(\mathbf{r}'_{l,i}) + s \sum_{i=1}^{n} \|\mathbf{r}'_{l,i}\|^2$$

or

$$\frac{1}{s} S_r - 2D + sS_l.$$

Completing the square in s, we get

$$\left(\sqrt{s}S_l - \frac{1}{\sqrt{s}} S_r\right)^2 + 2(S_l S_r - D).$$

This is minimized with respect to scale s when the first term is zero or $s = S_r/S_l$, that is,

$$s = \left(\sum_{i=1}^{n} \|\mathbf{r}'_{r,i}\|^2 \Bigg/ \sum_{i=1}^{n} \|\mathbf{r}'_{l,i}\|^2 \right)^{1/2}.$$

One advantage of this symmetrical result is that it allows one to determine the scale without the need to know the rotation. Importantly, the determination of the rotation is *not* affected by our choice of one of the three values of the scale factor. In each case the remaining error is minimized when D is as large as possible. That is, we have to choose the rotation that makes

$$\sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot R(\mathbf{r}'_{l,i})$$

as large as possible.

3. REPRESENTATION OF ROTATION

There are many ways to represent rotation, including the following: Euler angles, Gibbs vector, Cayley–Klein parameters, Pauli spin matrices, axis and angle, orthonormal matrices, and Hamilton's quaternions.[12,13] Of these representations, orthonormal matrices have been used most often in photogrammetry and robotics. There are a number of advantages, however, to the unit-quaternion notation. One of these is that it is much simpler to enforce the constraint that a quaternion have unit magnitude than it is to ensure that a matrix is orthonormal. Also, unit quaternions are closely allied to the geometrically intuitive axis and angle notation.

Here I solve the problem of finding the rotation that maximizes

$$\sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot R(\mathbf{r}'_{l,i})$$

by using unit quaternions. If desired, an orthonormal matrix can be constructed from the components of the resulting unit quaternion, as is shown in Subsection 3.E. We start here by reviewing the use of unit quaternions in representing rotations. Further details may be found in Appendixes A6–A8. The reader familiar with this material may wish to skip ahead to Section 4.

A. Quaternions

A quaternion can be thought of as a vector with four components, as a composite of a scalar and an ordinary vector, or as a complex number with three different imaginary parts.

Quaternions will be denoted here by using symbols with circles above them. Thus, using complex number rotation, we have

$$\dot{q} = q_0 + iq_x + jq_y + kq_z,$$

a quaternion with real part q_0 and three imaginary parts, q_x, q_y, and q_z.

Multiplication of quaternions can be defined in terms of the products of their components. Suppose that we let

$$i^2 = -1, \qquad j^2 = -1, \qquad k^2 = -1;$$
$$ij = k, \qquad jk = i, \qquad ki = j;$$

and

$$ji = -k, \qquad kj = -i, \qquad ik = -j.$$

Then, if

$$\dot{r} = r_0 + ir_x + jr_y + kr_z,$$

we get

$$\begin{aligned}\dot{r}\dot{q} = {} & (r_0q_0 - r_xq_x - r_yq_y - r_zq_z) \\ & + i(r_0q_x + r_xq_0 + r_yq_z - r_zq_y) \\ & + j(r_0q_y - r_xq_z + r_yq_0 + r_zq_x) \\ & + k(r_0q_z + r_xq_y - r_yq_x + r_zq_0).\end{aligned}$$

The product $\dot{q}\dot{r}$ has a similar form, but six of the signs are changed, as can readily be verified. Thus $\dot{r}\dot{q} \neq \dot{q}\dot{r}$, in general. In Subsection 3.B we think of quaternions as column vectors with four components.

B. Products of Quaternions

The product of two quaternions can also be conveniently expressed in terms of the product of an orthogonal 4×4 matrix and a vector with four components. One may choose to expand either the first or the second quaternion in a product into an orthogonal 4×4 matrix as follows:

$$\dot{r}\dot{q} = \begin{bmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & -r_z & r_y \\ r_y & r_z & r_0 & -r_x \\ r_z & -r_y & r_x & r_0 \end{bmatrix} \dot{q} = \mathbb{R}\dot{q}$$

or

$$\dot{q}\dot{r} = \begin{bmatrix} r_0 & -r_x & -r_y & -r_z \\ r_x & r_0 & r_z & -r_y \\ r_y & -r_z & r_0 & r_x \\ r_z & r_y & -r_x & r_0 \end{bmatrix} \dot{q} = \bar{\mathbb{R}}\dot{q}.$$

Note that $\bar{\mathbb{R}}$ differs from $\mathbb{R}$ in that the lower-right-hand 3×3 submatrix is transposed. (This again illustrates the noncommutative nature of multiplication of quaternions.) Note that the sum of squares of elements of each column (or row) of $\mathbb{R}$ and $\bar{\mathbb{R}}$ equals

$$r_0^2 + r_x^2 + r_y^2 + r_z^2,$$

which is just the dot product $\dot{r} \cdot \dot{r}$, as we see below.

C. Dot Products of Quaternions

The dot product of two quaternions is the sum of products of corresponding components:

$$\dot{p} \cdot \dot{q} = p_0q_0 + p_xq_x + p_yq_y + p_zq_z.$$

The square of the magnitude of a quaternion is the dot product of the quaternion with itself:

$$\| \dot{q} \|^2 = \dot{q} \cdot \dot{q}.$$

A unit quaternion is a quaternion whose magnitude equals 1. Taking the conjugate of a quaternion negates its imaginary part; thus

$$\dot{q}^* = q_0 - iq_x - jq_y - kq_z.$$

The 4×4 matrices associated with the conjugate of a quaternion are just the transposes of the matrices associated with the quaternion itself. Since these matrices are orthogonal, the products with their transposes are diagonal; that is, $QQ^T = \dot{q} \cdot \dot{q} I$, where I is the 4×4 identity matrix. Correspondingly, the product of $\dot{q}$ and $\dot{q}^*$ is real:

$$\dot{q}\dot{q}^* = (q_0^2 + q_x^2 + q_y^2 + q_z^2) = \dot{q} \cdot \dot{q}.$$

We immediately conclude that a nonzero quaternion has an inverse

$$\dot{q}^{-1} = (1/\dot{q} \cdot \dot{q})\dot{q}^*.$$

In the case of a unit quaternion, the inverse is just the conjugate.

D. Useful Properties of Products

Dot products are preserved because the matrices associated with quaternions are orthogonal; thus

$$(\dot{q}\dot{p}) \cdot (\dot{q}\dot{r}) = (Q\dot{p}) \cdot (Q\dot{r}) = (Q\dot{p})^T(Q\dot{r})$$

and

$$(Q\dot{p})^T(Q\dot{r}) = \dot{p}^TQ^TQ\dot{r} = \dot{p}^T(\dot{q} \cdot \dot{q})I\dot{r}.$$

We conclude that

$$(\dot{q}\dot{p}) \cdot (\dot{q}\dot{r}) = (\dot{q} \cdot \dot{q})(\dot{p} \cdot \dot{r}),$$

which, in the case when $\dot{q}$ is a unit quaternion, is just $\dot{p} \cdot \dot{r}$. A special case follows immediately:

$$(\dot{p}\dot{q}) \cdot (\dot{p}\dot{q}) = (\dot{p} \cdot \dot{p})(\dot{q} \cdot \dot{q});$$

that is, the magnitude of a product is just the product of the magnitudes. It also follows that

$$(\dot{p}\dot{q}) \cdot \dot{r} = \dot{p} \cdot (\dot{r}\dot{q}^*),$$

a result that we will use later.

Vectors can be represented by purely imaginary quaternions. If $\mathbf{r} = (x, y, z)^T$, we can use the quaternion

$$\dot{r} = 0 + ix + jy + kz.$$

(Similarly, scalars can be represented by using real quaternions.) Note that the matrices $\mathbb{R}$ and $\bar{\mathbb{R}}$ associated with a purely imaginary quaternion are skew symmetric. That is, in this special case,

$$\mathbb{R}^T = -\mathbb{R} \quad \text{and} \quad \bar{\mathbb{R}}^T = -\bar{\mathbb{R}}.$$

E. Unit Quaternions and Rotation

The length of a vector is not changed by rotation, nor is the angle between vectors. Thus rotation preserves dot products. Reflection also preserves dot products but changes the sense of a cross product—a right-hand triad of vectors is changed into a left-hand triad. Only rotations and reflections preserve dot products. Thus we can represent rotations by using unit quaternions if we can find a way of mapping purely imaginary quaternions into purely imaginary quaternions in such a way that dot products are preserved, as is the sense of cross products. (The purely imaginary quaternions represent vectors, of course.)

Now we have already established that multiplication by a unit quaternion preserves dot products between two quaternions. That is,

$$(\dot{q}\dot{p}) \cdot (\dot{q}\dot{r}) = \dot{p} \cdot \dot{r},$$

provided that $\dot{q} \cdot \dot{q} = 1$. We cannot use simple multiplication to represent rotation, however, since the product of a unit quaternion and a purely imaginary quaternion is generally not purely imaginary.

What we can use, however, is the composite product

$$\dot{r}' = \dot{q}\dot{r}\dot{q}^*,$$

which *is* purely imaginary. We show this by expanding

$$\dot{q}\dot{r}\dot{q}^* = (Q\dot{r})\dot{q}^* = \bar{Q}^T(Q\dot{r}) = (\bar{Q}^TQ)\dot{r},$$

where Q and $\bar{Q}$ are the 4×4 matrices corresponding to $\dot{q}$. Then we note that

$$\bar{Q}^TQ = \begin{bmatrix} \dot{q} \cdot \dot{q} & 0 & 0 & 0 \\ 0 & (q_0^2 + q_x^2 - q_y^2 - q_z^2) & 2(q_xq_y - q_0q_z) & 2(q_xq_z + q_0q_y) \\ 0 & 2(q_yq_x + q_0q_z) & (q_0^2 - q_x^2 + q_y^2 - q_z^2) & 2(q_yq_z - q_0q_x) \\ 0 & 2(q_zq_x - q_0q_y) & 2(q_zq_y + q_0q_x) & (q_0^2 - q_x^2 - q_y^2 + q_z^2) \end{bmatrix}.$$

So $\dot{r}'$ is purely imaginary if $\dot{r}$ is. Now Q and $\bar{Q}$ are orthonormal if $\dot{q}$ is a unit quaternion. Then $\dot{q} \cdot \dot{q} = 1$, and so the lower-right-hand 3×3 submatrix of $\bar{Q}^TQ$ must also be orthonormal. Indeed, it is the familiar rotation matrix R that takes r into r':

$$\mathbf{r}' = R\mathbf{r}.$$

In Appendix A6 it is shown that cross products are also preserved by the composite product $\dot{q}\dot{r}\dot{q}^*$, so that we are dealing with a rotation, not a reflection (equivalently, we can demonstrate that the determinant of the matrix above is $+1$).

Note that

$$(-\dot{q})\dot{r}(-\dot{q}^*) = \dot{q}\dot{r}\dot{q}^*$$

so that $-\dot{q}$ represents the same rotation as does $\dot{q}$.

F. Relationship to Other Representations

The expansion of $\bar{Q}^TQ$ given above provides an explicit method for computing the orthonormal rotation matrix R from the components of the unit quaternion $\dot{q}$. I give in Appendix A8 a method to recover the unit quaternion from an orthonormal matrix. It may be helpful to relate the quaternion notation to others with which the reader may be familiar. It is shown in Appendix A7 that a rotation by an angle θ about the axis defined by the unit vector $\hat{\omega} = (\omega_x, \omega_y, \omega_z)^T$ can be represented by the unit quaternion

$$\dot{q} = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}(i\omega_x + j\omega_y + k\omega_z).$$

Thus the imaginary part of the unit quaternion gives the direction of the axis of rotation, whereas the angle of rotation can be recovered from the real part and the magnitude of the imaginary part.

G. Composition of Rotations

Consider again the rotation $\dot{r}' = \dot{q}\dot{r}\dot{q}^*$. Suppose that we now apply a second rotation, represented by the unit quaternion $\dot{p}$. We have

$$\dot{r}'' = \dot{p}\dot{r}'\dot{p}^* = \dot{p}(\dot{q}\dot{r}\dot{q}^*)\dot{p}^*.$$

It is easy to verify that $(q^*p^*) = (pq)^*$. So we can write

$$\dot{r}'' = (\dot{p}\dot{q})\dot{r}(\dot{p}\dot{q})^*.$$

The overall rotation is represented by the unit quaternion $\dot{p}\dot{q}$. Thus composition of rotations corresponds to multiplication of quaternions.

It may be of interest to note that it takes fewer arithmetic operations to multiply two quaternions than it does to multiply two 3×3 matrices. Also, calculations are not carried out with infinite precision. The product of many orthonormal matrices may no longer be orthonormal, just as the product of many unit quaternions may no longer be a unit quaternion because of the limitation of the arithmetic used. It is, however, trivial to find the nearest unit quaternion, whereas it is quite difficult to find the nearest orthonormal matrix.

4. FINDING THE BEST ROTATION

We now return to the problem of absolute orientation. We have to find the unit quaternion $\dot{q}$ that maximizes

$$\sum_{i=1}^{n} (\dot{q}\dot{r}'_{l,i}\dot{q}^*) \cdot \dot{r}'_{r,i}.$$

Using one of the results derived above, we can rewrite this in the form

$$\sum_{i=1}^{n} (\dot{q}\dot{r}'_{l,i}) \cdot (\dot{r}'_{r,i}\dot{q}).$$

Suppose that $\mathbf{r}'_{l,i} = (x'_{l,i}, y'_{l,i}, z'_{l,i})^T$ while $\mathbf{r}'_{r,i} = (x'_{r,i}, y'_{r,i}, z'_{r,i})^T$; then

$$\dot{q}\dot{r}'_{l,i} = \begin{bmatrix} 0 & -x'_{l,i} & -y'_{l,i} & -z'_{l,i} \\ x'_{l,i} & 0 & z'_{l,i} & -y'_{l,i} \\ y'_{l,i} & -z'_{l,i} & 0 & x'_{l,i} \\ z'_{l,i} & y'_{l,i} & -x'_{l,i} & 0 \end{bmatrix} \dot{q} = \mathbb{R}_{l,i}\dot{q},$$

while

$$\dot{r}'_{r,i}\dot{q} = \begin{bmatrix} 0 & -x'_{r,i} & -y'_{r,i} & -z'_{r,i} \\ x'_{r,i} & 0 & -z'_{r,i} & y'_{r,i} \\ y'_{r,i} & z'_{r,i} & 0 & -x'_{r,i} \\ z'_{r,i} & -y'_{r,i} & x'_{r,i} & 0 \end{bmatrix} \dot{q} = \mathbb{R}_{r,i}\dot{q}.$$

Note that $\mathbb{R}_{r,i}$ and $*\mathbb{R}_{l,i}$ are skew symmetric as well as orthogonal. The sum that we have to maximize can now be written in the form

$$\sum_{i=1}^{n} (\bar{\mathbb{R}}_{l,i}\dot{q}) \cdot (\mathbb{R}_{r,i}\dot{q})$$

or

$$\sum_{i=1}^{n} \dot{q}^T \bar{\mathbb{R}}_{l,i}{}^T \mathbb{R}_{r,i}\dot{q};$$

that is,

$$\dot{q}^T \left(\sum_{i=1}^{n} \bar{\mathbb{R}}_{l,i}{}^T \mathbb{R}_{r,i} \right) \dot{q}$$

or

$$\dot{q}^T \left(\sum_{i=1}^{n} N_i \right) \dot{q} = \dot{q}^T N \dot{q},$$

where $N_i = *\mathbb{R}_{l,i}{}^T \mathbb{R}_{r,i}$ and $N = \sum_{i=1}^{n} N_i$. It is easy to verify that each of the N_i matrices is symmetric, so N must also be symmetric.

A. Matrix of Sums of Products

It is convenient at this point to introduce the 3×3 matrix

$$M = \sum_{i=1}^{n} \mathbf{r}'_{l,i} \mathbf{r}'^T_{r,i}$$

whose elements are sums of products of coordinates measured in the left system with coordinates measured in the right system. It turns out that this matrix contains *all* the information required to solve the least-squares problem for rotation. We may identify the individual elements by writing M in the form

$$M = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix},$$

where

$$S_{xx} = \sum_{i=1}^{n} x'_{l,i} x'_{r,i}, \qquad S_{xy} = \sum_{i=1}^{n} x'_{l,i} y'_{r,i},$$

and so on. Then

$$N = \begin{bmatrix} (S_{xx} + S_{yy} + S_{zz}) & S_{yz} - S_{zy} & S_{zx} - S_{xz} & S_{xy} - S_{yx} \\ S_{yz} - S_{zy} & (S_{xx} - S_{yy} - S_{zz}) & S_{xy} + S_{yx} & S_{zx} + S_{xz} \\ S_{zx} - S_{xz} & S_{xy} + S_{yx} & (-S_{xx} + S_{yy} - S_{zz}) & S_{yz} + S_{zy} \\ S_{xy} - S_{yx} & S_{zx} + S_{xz} & S_{yz} + S_{zy} & (-S_{xx} - S_{yy} + S_{zz}) \end{bmatrix}.$$

Thus the 10 independent elements of the real symmetric 4×4 matrix N are sums and differences of the nine elements of the 3×3 matrix M. Note that the trace, sum of diagonal elements of the matrix N, is zero. (This takes care of the 10th degree of freedom.)

B. Eigenvector Maximizes Matrix Product

It is shown in Appendix A3 that the unit quaternion that maximizes

$$\dot{q}^T N \dot{q}$$

is the eigenvector corresponding to the most positive eigenvalue of the matrix N.

The eigenvalues are the solutions of the fourth-order polynomial in λ that we obtain when we expand the equation

$$\det(N - \lambda I) = 0,$$

where I is the 4×4 identity matrix. We show in what follows how the coefficients of this polynomial can be computed from the elements of the matrix N.

Once we have selected the largest positive eigenvalue, say, λ_m, we find the corresponding eigenvector $\dot{e}_m$ by solving the homogeneous equation

$$[N - \lambda_m I]\dot{e}_m = 0.$$

I show in Appendix A5 how the components of $\dot{e}_m$ can be found by using the determinants of submatrices obtained from $(N - \lambda_m I)$ by deleting one row and one column at a time.

C. Nature of the Closed-Form Solution

We can now summarize the algorithm. We first find the centroids $\bar{\mathbf{r}}_l$ and $\bar{\mathbf{r}}_r$ of the two sets of measurements in the left and the right coordinate system. The centroids are subtracted from all measurements so that, from now on, we deal only with measurements relative to the centroids. For each pair of coordinates we compute the nine possible products $x'_l x'_r$, $x'_l y'_r, \ldots z'_l z'_r$ of the components of the two vectors. These are added up to obtain $S_{xx}, S_{xy}, \ldots, S_{zz}$. These nine totals contain all the information that is required to find the solution.

We now compute the 10 independent elements of the 4×4 symmetric matrix N by combining the sums obtained above. From these elements, in turn, we calculate the coefficients of the fourth-order polynomial that has to be solved to obtain the eigenvalues of N. There are closed-form methods for solving quartics, such as Ferrari's method, which splits the quartic into two quadratics.[12,14,15] We pick the most positive root and use it to solve the four linear homogeneous equations to obtain the corresponding eigenvector. The quaternion representing the rotation is a unit vector in the same direction. (If desired, an orthonormal 3×3 matrix can now be constructed from the unit quaternion, using the result in Subsection 3.E.)

At this point, we compute the scale by using one of the

three formulas given for that purpose in Subsections 2.D and 2.E. If we choose the symmetric form, we actually do not need to know the rotation for this step. The scale is then just the ratio of the root-mean-square deviations of the two sets of measurements from their respective centroids.

Finally, we compute the translation as the difference between the centroid of the right measurements and the scaled and rotated centroid of the left measurements.

While these calculations may be nontrivial, there is no approximation involved—no need for iterative correction. (Also, the basic data extracted from the measurements are the same as those used by existing iterative schemes.)

D. Coefficients of the Quartic—Some Details

Suppose that we write the matrix N in the form

$$N = \begin{bmatrix} a & e & h & j \\ e & b & f & i \\ h & f & c & g \\ j & i & g & d \end{bmatrix},$$

where $a = S_{xx} + S_{yy} + S_{zz}$, $e = S_{yz} - S_{zy}$, $h = S_{zx} - S_{xz}$, and so on. Then

$$\det(N - \lambda I) = 0$$

can be expanded in the form

$$\lambda^4 + c_3\lambda^3 + c_2\lambda^2 + c_1\lambda + c_0 = 0,$$

where

$$\begin{aligned} c_3 &= a + b + c + d, \\ c_2 &= (ac - h^2) + (bc - f^2) + (ad - j^2) + (bd - i^2) \\ &\quad + (cd + g^2) + (ab - e^2), \\ c_1 &= [-b(cd - g^2) + f(df - gi) - i(fg - ci) \\ &\quad - a(cd - g^2) + h(dh - gj) - j(gh - cj) \\ &\quad - a(bd - i^2) + e(de - ij) - j(ei - bj) \\ &\quad - a(bc - f^2) + e(ce - fh) - h(ef - bh)], \\ c_0 &= (ab - e^2)(cd - g^2) + (eh - af)(fd - gi) \\ &\quad + (ai - ej)(fg - ci) + (ef - bh)(hd - gj) \\ &\quad + (bj - ei)(hg - cj) + (hi - fj)^2. \end{aligned}$$

(These expressions may be rewritten and simplified somewhat.) It is easy to see that $c_3 = 0$ since it is the trace of the matrix N. (This makes it easier to solve the quartic, since the sum of the four roots is zero.) Furthermore,

$$\begin{aligned} c_2 = -2(&S_{xx}^2 + S_{xy}^2 + S_{xz}^2 + S_{yx}^2 \\ &+ S_{yy}^2 + S_{yz}^2 + S_{zx}^2 + S_{zy}^2 + S_{zz}^2), \end{aligned}$$

and so c_2 is always negative. (This means that some of the roots must be negative while others are positive.) Note that $c_2 = -2\,\mathrm{Tr}(M^TM)$. Next we find that

$$\begin{aligned} c_1 = 8&(S_{xx}S_{yz}S_{zy} + S_{yy}S_{zx}S_{xz} + S_{zz}S_{xy}S_{yx}) \\ &- 8(S_{xx}S_{yy}S_{zz} + S_{yz}S_{zx}S_{xy} + S_{zy}S_{yx}S_{xz}). \end{aligned}$$

This may be either positive or negative. It turns out to be zero when the points are coplanar, as we shall see below. This always happens when there are only three points. Note that $c_1 = -8\det(M)$, while $c_0 = \det(N)$.

5. POINTS IN A PLANE

If one of the two sets of measured points lies in a plane, the corresponding centroid must also lie in the plane. This means that the coordinates taken relative to the centroid lie in a plane that passes through the origin. The components of the measurements then are linearly related. Suppose, for concreteness, that it is the right set of measurements that is so affected. Let $\mathbf{n}_r$ be a vector normal to the plane. Then

$$\mathbf{r}'_{r,i} \cdot \mathbf{n}_r = 0$$

or

$$(x'_{r,i}, y'_{r,i}, z'_{r,i})^T \cdot \mathbf{n}_r = 0.$$

This implies that

$$\begin{aligned} (S_{xx}, S_{xy}, S_{xz})^T \cdot \mathbf{n}_r &= 0, \\ (S_{yx}, S_{yy}, S_{yz})^T \cdot \mathbf{n}_r &= 0, \\ (S_{zx}, S_{zy}, S_{zz})^T \cdot \mathbf{n}_r &= 0, \end{aligned}$$

or

$$M\mathbf{n}_r = \mathbf{0}.$$

We conclude that M is singular. Now $c_1 = -8\det(M)$, and hence $c_1 = 0$. The same reasoning applies when all the points in the left set of measurements are coplanar.

This means that the quartic is particularly easy to solve. It has the simple form

$$\lambda^4 + c_2\lambda^2 + c_0 = 0.$$

If we let $\mu = \lambda^2$, we obtain

$$\mu^2 + c_2\mu + c_0 = 0,$$

and the roots are just

$$\mu = \pm\tfrac{1}{2}(c_2^2 - 4c_0)^{1/2} - c_2.$$

Now c_2 is negative, so we get two positive roots, say, μ_+ and μ_-. The solutions for λ are then

$$\lambda = \pm\sqrt{\mu_+}, \qquad \lambda = \pm\sqrt{\mu_-}.$$

The most positive of the four solutions is

$$\lambda_m = [\tfrac{1}{2}(c_2^2 - 4c_0)^{1/2} - c_2]^{1/2}.$$

This simple method applies when the points all lie in a plane. In particular, if we are given three points this will always be the case. So we find a simple least-squares solution to this case that is normally treated in an approximate fashion by selectively discarding information, as we did in Subsection 2.A.

A. Special Case—Coplanar Points

When *both* sets of measurements are exactly coplanar (as always happens when there are only three) the general solution simplifies further. This suggests that this case may be dealt with more directly. We could attack it by using unit quaternions or orthonormal matrices. For variety we argue

this case geometrically. (Also, we make use of a dual interpretation: we consider the measurements to represent two sets of points in one coordinate system.)

First, the plane containing the left measurements has to be rotated to bring it into coincidence with the plane containing the right measurements (Fig. 3). We do this by rotating the left measurements about the line of intersection of the two planes. The direction of the line of intersection is given by the cross product of the normals of the two planes. The angle of rotation is that required to bring the left normal to coincide with the right normal, that is, the angle between the two normals.

At this point, the remaining task is the solution of a least-squares problem in a plane. We have to find the rotation that minimizes the sum of squares of distances between corresponding left and (rotated) right measurements. This second rotation is about the normal to the plane. The angle is determined by the solution of the least-squares problem.

The overall rotation is just the composition of the two rotations found above. It can be found by multiplication of unit quaternions or multiplication of orthonormal vectors, as desired. The solution is simpler than the general one because the rotation to be found by the least-squares method is in a *plane*, so it depends on only one parameter, not three, as does rotation in the general case (see Fig. 4).

We start by finding normals to the two planes. We can use the cross products of any pair of nonparallel vectors in the plane:

$$\mathbf{n}_l = \mathbf{r}'_{l2} \times \mathbf{r}'_{l1}, \qquad \mathbf{n}_r = \mathbf{r}'_{r2} \times \mathbf{r}'_{r1}.$$

(The normals can also readily be expressed in terms of three cross products of three of the original measurements rather than two measurements relative to the centroid.) We next construct unit normals $\hat{n}_l$ and $\hat{n}_r$ by dividing $\mathbf{n}_l$ and $\mathbf{n}_r$ by their magnitudes. The line of intersection of the two planes lies in both planes, so it is perpendicular to both normals. It is thus parallel to the cross product of the two normals. Let

$$\mathbf{a} = \mathbf{n}_l \times \mathbf{n}_r.$$

(This can be expanded either as a weighted sum of the two vectors $\mathbf{r}'_{l1}$ and $\mathbf{r}'_{l2}$ or as a weighted sum of the two vectors $\mathbf{r}'_{r1}$ and $\mathbf{r}'_{r2}$.) We find a unit vector $\hat{a}$ in the direction of the line of intersection by dividing $\mathbf{a}$ by its magnitude.

The angle ϕ through which we have to rotate is the angle between the two normals. So

$$\cos\phi = \hat{n}_l \cdot \hat{n}_r, \qquad \sin\phi = \|\hat{n}_l \times \hat{n}_r\|.$$

(Note that $0 \le \phi \le \pi$.)

We now rotate the left measurements into the plane containing the right measurements. Let $\mathbf{r}''_{l,i}$ be the rotated version of $\mathbf{r}'_{l,i}$. The rotation can be accomplished by using Rodrigues's formula, the unit quaternion

$$\dot{q}_a = \cos\frac{\phi}{2} + \sin\frac{\phi}{2}(ia_x + ja_y + ka_z),$$

or the corresponding orthonormal matrix, say, R_a.

B. Rotation in the Plane

We now have to find the rotation in the plane of the right-hand measurements that minimizes the sum of squares of distances between corresponding measurements. That is, we wish to minimize

$$\sum_{i=1}^{n} \|\mathbf{r}'_{r,i} - \mathbf{r}''_{l,i}\|^2.$$

Let α_i be the angle between corresponding measurements. That is,

$$r'_{r,i} r''_{l,i} \cos\alpha_i = \mathbf{r}'_{r,i} \cdot \mathbf{r}''_{l,i}$$

and

$$r'_{r,i} r''_{l,i} \sin\alpha_i = (\mathbf{r}'_{r,i} \times \mathbf{r}''_{l,i}) \cdot \hat{n}_r,$$

where

$$r'_{r,i} = \|\mathbf{r}'_{r,i}\|, \qquad r''_{l,i} = \|\mathbf{r}''_{l,i}\| = \|\mathbf{r}'_{l,i}\|.$$

Note that $\mathbf{r}'_{r,i} \times \mathbf{r}''_{l,i}$ is parallel to $\hat{n}_r$, so the dot product with $\hat{n}_r$ has a magnitude

$$\|\mathbf{r}'_{r,i} \times \mathbf{r}''_{l,i}\|$$

but a *sign* that depends on the order of $\mathbf{r}'_{r,i}$ and $\mathbf{r}''_{l,i}$ in the plane.

Using the cosine rule for triangles, or by expanding the term above, we find that the square of the distance between corresponding measurements is

$$(r'_{r,i})^2 + (r''_{l,i})^2 - 2r'_{r,i} r''_{l,i} \cos\alpha_i.$$

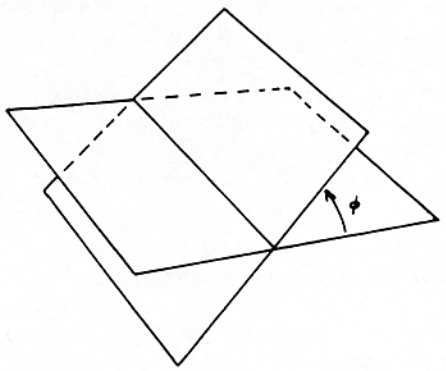

Fig. 3. When the two sets of measurements each lie in a plane, the rotation decomposes conveniently into rotation about the line of intersection of the two planes and rotation about a normal of one of the planes. (In this figure the two coordinate systems have been aligned and superimposed.)

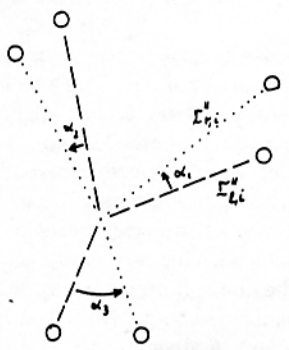

Fig. 4. The second rotation, when both sets of measurements are coplanar, is about the normal of one of the planes. The rotation that minimizes the sum of errors is found easily in this special case since there is only one degree of freedom.

When the left measurements are rotated in the plane through an angle θ, the angles α_i are reduced by θ. So to minimize the sum of squares of distances we need to maximize

$$\sum_{i=1}^{n} r'_{r,i} r''_{l,i} \cos(\alpha_i - \theta)$$

or

$$C \cos\theta + S \sin\theta,$$

where

$$C = \sum_{i=1}^{n} r'_{r,i} r''_{l,i} \cos\alpha_i = \sum_{i=1}^{n} (\mathbf{r}'_{r,i} \cdot \mathbf{r}''_{l,i})$$

and

$$S = \sum_{i=1}^{n} r'_{r,i} r''_{l,i} \sin\alpha_i = \left(\sum_{i=1}^{n} \mathbf{r}'_{r,i} \times \mathbf{r}''_{l,i} \right) \cdot \hat{n}_r.$$

This expression has extrema where

$$C \sin\theta = S \cos\theta$$

or

$$\sin\theta = \pm \frac{S}{\sqrt{S^2 + C^2}}, \qquad \cos\theta = \pm \frac{C}{\sqrt{S^2 + C^2}}.$$

The extreme values are $\pm\sqrt{S^2 + C^2}$. So for a maximum we chose the pluses. (Note that S and C may be positive or negative, however.)

The second rotation is then about the axis $\hat{n}_r$ by an angle θ. It can be represented by the unit quaternion

$$\dot{q}_p = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}(in_x + jn_y + kn_z)$$

or the corresponding orthonormal matrix, say, R_p.

The overall rotation is the composition of the two rotations; that is,

$$\dot{q} = \dot{q}_p \dot{q}_a \quad \text{or} \quad R = R_p R_a.$$

This completes the least-squares solution for the case in which both the left and the right sets of measurements are coplanar. Note that, in practice, it is unlikely that the measurements would be exactly coplanar even when the underlying points are, because of small measurement errors, *unless* $n = 3$.

The above provides a closed-form solution of the least-squares problem for three points. The special case of three points is usually treated by selectively neglecting some of the information, as discussed earlier. The answer then depends on which pieces of information were discarded. The method described above, while more complex, uses all the information. (The number of arithmetic operations is, of course, not a significant issue with modern computational tools.)

By the way, in the above it may have appeared that trigonometric functions needed to be evaluated in order for the components of the unit quaternions to be obtained. In fact, we only need to be able to compute sines and cosines of half-angles given sines and cosines of the angles. But for $-\pi \leq \theta \leq \pi$

$$\cos\theta/2 = [(1 + \cos\theta)/2]^{1/2},$$

while

$$\sin\theta/2 = \sin\theta/[2(1 + \cos\theta)]^{1/2}.$$

(As usual, it is best to know both the sine and the cosine of an angle.)

While unit quaternions constitute an elegant notation for representing rotation, much of the previous work has been based on orthonormal matrices. An orthonormal rotation matrix can be constructed easily from the components of a given unit quaternion, as shown in Subsection 3.E.

6. CONCLUSION

I have presented a closed-form solution for the least-squares problem of absolute orientation. It provides the best rigid-body transformation between two coordinate systems given measurements of the coordinates of a set of points that are not collinear. The solution given uses unit quaternions to represent rotations. The solution simplifies when there are only three points. I also gave an alternative method for this special case.

I showed that the best scale is the ratio of the root-mean-square deviations of the measurements from their respective centroids. The best translation is the difference between the centroid of one set of measurements and the scaled and rotated centroid of the other measurements. These exact results are to be preferred to ones based only on measurements of one or two points. The unit quaternion representing the rotation is the eigenvector associated with the most positive eigenvalue of a symmetric 4×4 matrix. The corresponding orthonormal matrix can be easily found.

The solutions presented here may seem relatively complex. The ready availability of program packages for solving polynomial equations and finding eigenvalues and eigenvectors of symmetric matrices makes implementation straightforward, however.

APPENDIXES

A1. The Optimum Scale Factor Depends on the Choice of the Error Term

Consider a case in which there is no translation or rotation (Fig. 5). We are to find the best scale factor relating the sets of measurements in the left coordinate system to those in the right. If we decide to scale the left measurements before matching them to the right ones, we get an error term of the form

$$E_{rl} = 2(c - sa)^2 + 2(d - sb)^2,$$

where s is the scale factor. This error term is minimized when

$$s = s_{rl} = \frac{ac + bd}{a^2 + b^2}.$$

If, instead, we decide to scale the right measurements before matching them to the left ones, we get an error term of the form

$$E = 2(a - \bar{s}c)^2 + 2(b - \bar{s}d)^2,$$

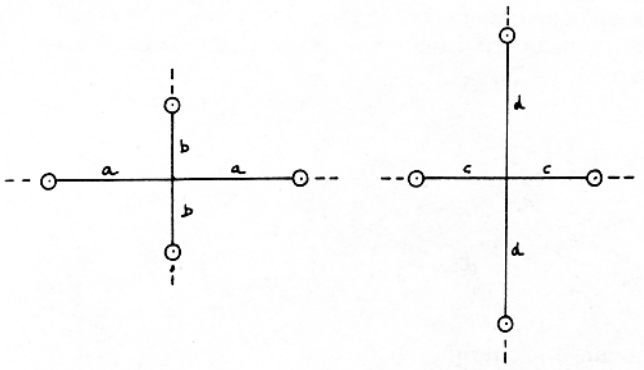

Fig. 5. Three different optimal scale factors can be obtained by choosing three different forms for the error term. In general, a symmetric form for the error is to be preferred.

where $\bar{s}$ is the new scale factor. This error term is minimized when

$$\bar{s} = \frac{ac + bd}{c^2 + d^2}.$$

The equivalent scaling of the left coordinate system is $1/\bar{s}$ or

$$s = s_{lr} = \frac{c^2 + d^2}{ac + bd}.$$

In general, $s_{lr} \neq s_{rl}$, so these two methods violate the symmetry property sought for. If we use the symmetric form

$$E_{sy} = 2\left(\frac{1}{\sqrt{s}}\, c - \sqrt{s}a\right)^2 + 2\left(\frac{1}{\sqrt{s}}\, d - \sqrt{s}c\right)^2,$$

we obtain instead

$$S = S_{sy} = \left(\frac{c^2 + d^2}{a^2 + b^2}\right)^{1/2}.$$

Note that S_{sy} is the geometric mean of S_{lr} and S_{rl} and that the expression for S_{sy} does not depend on products of measurements in the left coordinate system with measurement in the right coordinate system. If there is no strong reason to do otherwise, the symmetric form is to be preferred, although the method presented in this paper for determining the best rotation applies no matter how the scale factor is chosen.

A2. Weighting the Errors

The expected errors in measurements are not always equal. In this case it is convenient to introduce weights into the sum of errors so that one minimizes

$$\sum_{i=1}^{n} w_i \|\mathbf{e}_i\|^2,$$

where w_i is some measure of confidence in the measurements of the ith point. The method presented in this paper can easily accommodate this elaboration. The centroids become *weighted* centroids:

$$\bar{\mathbf{r}}_l = \sum_{i=1}^{n} w_i \mathbf{r}_{l,i} \Big/ \sum_{i=1}^{n} w_i$$

and

$$\bar{r}_r = \sum_{i=1}^{n} w_i \mathbf{r}_{r,i} \Big/ \sum_{i=1}^{n} w_i.$$

The translation is computed by using these centroids, as before. Computation of the best scale is also changed only slightly by the inclusion of weighting factors:

$$S = \left(\sum_{i=1}^{n} w_i \|\mathbf{r}'_{r,i}\|^2 \Big/ \sum_{i=1}^{n} w_i \|\mathbf{r}'_{l,i}\|^2\right)^{1/2}.$$

The only change in the method for recovering rotation is that the products in the sums are weighted, that is,

$$M = \sum_{i=1}^{n} w_i \mathbf{r}'_{l,i} \mathbf{r}'^{T}_{r,i},$$

so

$$S_{xx} = \sum_{i=1}^{n} w_i x'_{l,i} x'_{r,i},$$

and so on. Once the elements of the matrix M are found, the solution proceeds exactly as it did before the introduction of weights.

A3. Maximizing the Quadratic Form $\dot{q}^T N \dot{q}$

To find the rotation that minimizes the sum of squares of errors, we have to find the quaternion $\dot{q}$ that maximizes $\dot{q}^T N \dot{q}$ subject to the constraint that $\dot{q} \cdot \dot{q} = 1$. The symmetric 4×4 matrix N will have four real eigenvalues, say, $\lambda_1, \lambda_2, \ldots, \lambda_4$. A corresponding set of orthogonal unit eigenvectors $\dot{e}_1, \dot{e}_2, \ldots, \dot{e}_4$ can be constructed such that

$$N\dot{e}_i = \lambda_i \dot{e}_i \qquad \text{for} \quad i = 1, 2, \ldots, 4.$$

The eigenvectors span the four-dimensional space, so an arbitrary quaternion $\dot{q}$ can be written as a linear combination in the form

$$\dot{q} = \alpha_1 \dot{e}_1 + \alpha_2 \dot{e}_2 + \alpha_3 \dot{e}_3 + \alpha_4 \dot{e}_4.$$

Since the eigenvectors are orthogonal, we find that

$$\dot{q} \cdot \dot{q} = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2.$$

We know that this has to equal one, since $\dot{q}$ is a unit quaternion. Next, note that

$$N\dot{q} = \alpha_1 \lambda_1 \dot{e}_1 + \alpha_2 \lambda_2 \dot{e}_2 + \alpha_3 \lambda_3 \dot{e}_3 + \alpha_4 \lambda_4 \dot{e}_4$$

since $\dot{e}_1, \dot{e}_2, \ldots, \dot{e}_4$ are eigenvectors of N. We conclude that

$$\dot{q}^T N \dot{q} = \dot{q} \cdot (N\dot{q}) = \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_2 + \alpha_3^2 \lambda_3 + \alpha_4^2 \lambda_4.$$

Now suppose that we have arranged the eigenvalues in order, so that

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4.$$

Then we see that

$$\dot{q}^T N \dot{q} \leq \alpha_1^2 \lambda_1 + \alpha_2^2 \lambda_1 + \alpha_3^2 \lambda_1 + \alpha_4^2 \lambda_1 = \lambda_1$$

so that the quadratic form cannot become larger than the most positive eigenvalue. Also, this maximum is attained when we chose $\alpha_1 = 1$ and $\alpha_2 = \alpha_3 = \alpha_4 = 0$, that is, $\dot{q} = \dot{e}_1$. We conclude that the unit eigenvector corresponding to the

most positive eigenvalue maximizes the quadratic form $\dot{q}^T N \dot{q}$. (The solution is unique if the most positive eigenvalue is distinct.)

A4. Finding the Eigenvector

There are iterative methods available for finding eigenvalues and eigenvectors of arbitrary symmetric matrices. One can certainly apply these algorithms to the problem presented here. We are, however, dealing only with a 4×4 matrix, for which noniterative methods are available. We have already mentioned that the eigenvalues may be found by applying a method such as that of Ferrari to the quartic equation obtained by expanding $\det(N - \lambda I) = 0$.

Once the most positive eigenvalue, say, λ_m, has been determined, we obtain the corresponding eigenvector $\dot{e}_m$ by solving the homogeneous equation

$$(N - \lambda_m I)\dot{e}_m = 0.$$

This can be accomplished by arbitrarily setting one component of $\dot{e}_m$ to 1, solving the resulting inhomogeneous set of three equations in three unknowns, and then scaling the result.

If the component of $\dot{e}_m$ chosen happens to be small, numerical inaccuracies will tend to corrupt the answer. For best results, this process is repeated by setting each of the components of $\dot{e}_m$ to 1 in turn and choosing the largest of the four results so obtained. Alternatively, the four results can be added up, after changing signs if necessary, to ensure that they all have the same sense. In either case, the result may be scaled so that its magnitude is 1. The inhomogeneous equations can be solved by using Gaussian elimination (or even Cramer's rule).

A5. Using the Matrix of Cofactors

The method for finding an eigenvector (given an eigenvalue) described in the previous appendix can be implemented in an elegant way by using the cofactors of the matrix $(N - \lambda I)$. A cofactor of a matrix is the determinant of the submatrix obtained by striking a given row and a given column. There are as many cofactors as there are elements in a matrix since there are as many ways of picking rows and columns to strike. The cofactors can be arranged into a matrix. The element in the ith row and jth column of this new matrix is the determinant of the submatrix obtained by striking the ith row and jth column of the original matrix.

The matrix of cofactors has an intimate connection with the inverse of a matrix: the inverse is the transpose of the matrix of cofactors, divided by the determinant. Let

$$N_\lambda = N - \lambda I;$$

then

$$\det(N_\lambda) N_\lambda^{-1} = (\bar{N}_\lambda)^T,$$

where $\bar{N}_\lambda$ is the matrix of cofactors of N_λ. Next, we note that the determinant is the product of the eigenvalues and that the eigenvalues of $(N - \lambda I)$ are equal to the eigenvalues of N minus λ. So

$$\det(N_\lambda) = \prod_{j=1}^{n} (\lambda_j - \lambda).$$

We will also need the result that the eigenvalues of the inverse of a matrix are the (algebraic) inverses of the eigenvalues of the matrix itself. So, if we write for an arbitrary quaternion

$$\dot{q} = \sum_{i=1}^{n} \lambda_i \dot{e}_i,$$

we obtain

$$N_\lambda^{-1} \dot{q} = \sum_{i=1}^{n} \frac{\alpha_i}{(\lambda_i - \lambda)} \dot{e}_i.$$

Combining these results, we see that

$$\sum_{i=1}^{n} \prod_{j \neq 1}^{n} (\lambda_j - \lambda) \alpha_i \dot{e}_i = (\bar{N}_\lambda)^T \dot{q}.$$

We remove all terms but one from the sum on the left-hand side by setting $\lambda = \lambda_m$. This gives us

$$\prod_{j \neq m}^{n} (\lambda_j - \lambda_m) \alpha_m \dot{e}_m = (\bar{N}_{\lambda_m})^T \dot{q}.$$

The left-hand side is a multiple of the sought after eigenvector $\dot{e}_m$. This will be nonzero as long as the arbitrary quaternion $\dot{q}$ has some component in the direction $\dot{e}_m$, that is, $\alpha_m \neq 0$ (and provided that the most positive eigenvalue is distinct). It follows from consideration of the right-hand side of the expression above that each column of $(\bar{N}_{\lambda_m})^T$, and hence each each row of $\bar{N}_{\lambda_m}$, must be parallel to $\dot{e}_m$.

All that we need to do then is to compute the matrix of cofactors of $(N - \lambda_m I)$. Each row of the result is parallel to the desired solution. We may pick the largest row or add up the rows in the fashion described at the end of the previous appendix. Note that we could have saved ourselves some computation by generating only *one* of the rows. It may happen, however, that this row is zero or very small.

The matrix of cofactors, by the way, can be computed easily by using a modified Gaussian elimination algorithm for inverting a matrix.

A6. Quaternions as Sums of Scalar and Vector Parts

We found it convenient to treat quaternions as quantities with three different imaginary parts (and occasionally as vectors with four components). At other times it is useful to think of quaternions as composed of a scalar and a vector (with three components).

Thus

$$\dot{q} = q_0 + i q_x + j q_y + k q_z$$

can be represented alternatively in the form

$$\dot{q} = q + \mathbf{q},$$

where $q = q_0$ and $\mathbf{q} = (q_x, q_y, q_z)^T$. The rules for multiplication of quaternions given earlier can then be written in the more compact form

$$p = rs - \mathbf{r} \cdot \mathbf{s}, \qquad \mathbf{p} = r\mathbf{s} + s\mathbf{r} + \mathbf{r} \times \mathbf{s},$$

when $\dot{p} = \dot{r}\dot{s}$ and

$$\dot{r} = r + \mathbf{r}, \qquad \dot{s} = s + \mathbf{s}, \qquad \dot{p} = p + \mathbf{p}.$$

The results simplify if $\dot{r}$ and $\dot{s}$ are purely imaginary, that is, if $r = 0$ and $s = 0$. In this special case

$$p = -\mathbf{r} \cdot \mathbf{s}, \qquad \mathbf{p} = \mathbf{r} \times \mathbf{s}.$$

Now let us apply the composite product with a unit quaternion $\dot{q}$ to $\mathbf{r}$, $\mathbf{s}$, and $\mathbf{p}$. That is, let

$$\dot{r}' = \dot{q}\dot{r}\dot{q}^*, \qquad \dot{s}' = \dot{q}\dot{s}\dot{q}^*, \qquad \dot{p}' = \dot{q}\dot{p}\dot{q}^*.$$

Clearly

$$\dot{r}'\dot{s}' = (\dot{q}\dot{r}\dot{q}^*)(\dot{q}\dot{s}\dot{q}^*) = (\dot{q}\dot{r})(\dot{q}^*\dot{q})(\dot{s}\dot{q}^*) = \dot{q}(\dot{r}\dot{s})\dot{q}^*,$$

so consequently

$$-\mathbf{r}' \cdot \mathbf{s}' = -\mathbf{r} \cdot \mathbf{s}.$$

Also, $\mathbf{r}' \times \mathbf{s}'$ is the result of applying the composite product with the unit quaternion to $\mathbf{r} \times \mathbf{s}$. Thus dot products and cross products are preserved. Consequently, the composite product with a unit quaternion can be used to represent rotation.

A7. Unit Quaternion from Axis and Angle

Suppose that a vector $\mathbf{r}$ is taken into $\mathbf{r}'$ by a rotation of angle θ about an axis (through the origin) given by the unit vector $\hat{\omega} = (\omega_x, \omega_y, \omega_z)^T$. Analysis of the diagram (Fig. 6) leads to the celebrated formula of Rodrigues:

$$\mathbf{r}' = \cos\theta\mathbf{r} + \sin\theta\hat{\omega} \times \mathbf{r} + (1 - \cos\theta)(\hat{\omega} \cdot x)\hat{\omega}.$$

We would like to show that this transformation can also be written in the form

$$\dot{r}' = \dot{q}\dot{r}\dot{q}^*,$$

where the quaternions, expressed in terms of scalar and vector parts, are

$$\dot{r} = 0 + \mathbf{r}, \qquad \dot{r}' = 0 + \mathbf{r}',$$

and

$$\dot{q} = \cos(\theta/2) + \sin(\theta/2)\hat{\omega}.$$

Straightforward manipulation shows that, if the scalar part of $\dot{r}$ is zero and $\dot{q}$ is a unit quaternion, then the composite product $\dot{q}\dot{r}\dot{q}^*$ has zero scalar part and vector part

$$\mathbf{r}' = (q^2 - \mathbf{q} \cdot \mathbf{q})\mathbf{r} + 2q\mathbf{q} \times \mathbf{r} + 2(\mathbf{q} \cdot \mathbf{r})\mathbf{q}.$$

If we now substitute

$$q = \cos(\theta/2), \qquad \mathbf{q} = \sin(\theta/2)\hat{\omega}$$

and use the identities

$$2\sin(\theta/2)\cos(\theta/2) = \sin\theta,$$

$$\cos^2(\theta/2) - \sin^2(\theta/2) = \cos\theta,$$

we obtain the formula of Rodrigues.

The relationships presented here allow us to convert easily between axis and angle and unit-quaternion notations for rotations. We have

$$\hat{\omega} = \mathbf{q}/\|\mathbf{q}\|$$

and

$$\sin\theta = 2q\|\mathbf{q}\|, \qquad \cos\theta = (q^2 - \|\mathbf{q}\|^2).$$

A8. Unit Quaternion from Orthonormal Matrix

We have shown in Subsection 3.E that the orthonormal matrix corresponding to the unit quaternion

$$\dot{q} = q_0 + iq_x + jq_y + kq_z$$

is

$$R = \begin{bmatrix} (q_0^2 + q_x^2 - q_y^2 - q_z^2) & 2(q_xq_y - q_0q_z) & 2(q_xq_z + q_0q_y) \\ 2(q_yq_x + q_0q_z) & (q_0^2 - q_x^2 + q_y^2 - q_z^2) & 2(q_yq_z - q_0q_x) \\ 2(q_zq_x - q_0q_y) & 2(q_zq_y + q_0q_z) & (q_0^2 - q_x^2 - q_y^2 + q_z^2) \end{bmatrix}.$$

At times it is necessary to obtain the unit quaternion corresponding to a given orthonormal matrix. Let r_{ij} be the element of the matrix R in the ith row and the jth column.

The following combinations of the diagonal terms prove useful:

$$1 + r_{11} + r_{22} + r_{33} = 4q_0^2,$$

$$1 + r_{11} - r_{22} - r_{33} = 4q_x^2,$$

$$1 - r_{11} + r_{22} - r_{33} = 4q_y^2,$$

$$1 - r_{11} - r_{22} + r_{33} = 4q_z^2.$$

We evaluate these four expressions to ascertain which of $q_0, q_x, \ldots, q_z$ is the largest. This component of the quaternion is then extracted by taking the square root (and dividing by 2). We may choose either sign for the square root since $-\dot{q}$ represents the same rotation as $+\dot{q}$.

Next, consider the three differences of corresponding off-diagonal elements

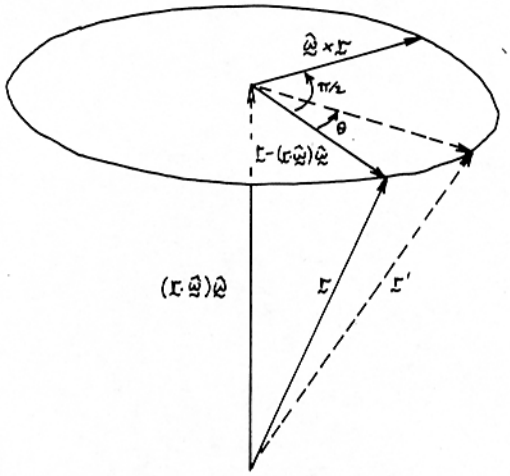

Fig. 6. The rotation of a vector $\mathbf{r}$ to a vector $\mathbf{r}'$ can be understood in terms of the quantities appearing in this diagram. Rodrigues's formula follows.

$$r_{32} - r_{23} = 4q_0q_x,$$

$$r_{13} - r_{31} = 4q_0q_y,$$

$$r_{21} - r_{12} = 4q_0q_z.$$

and the three sums of corresponding off-diagonal elements

$$r_{21} + r_{12} = 4q_xq_y,$$

$$r_{32} + r_{23} = 4q_yq_z,$$

$$r_{13} + r_{31} = 4q_zq_x.$$

Three of these are used to solve for the remaining three components of the quaternion using the one component already found. For example, if q_0 turned out to be the largest, then q_x, q_y, and q_z can be found from the three differences yielding $4q_0q_x$, $4q_0q_y$, and $4q_0q_z$.

We select the largest component to start with in order to ensure numerical accuracy. If we chose to solve first for q_0, for example, and it turned out that q_0 were very small, then q_x, q_y, and q_z would not be found with high precision.

ACKNOWLEDGMENTS

I would like to thank Shahriar Negahdaripour and Hugh M. Hilden, who helped me with the least-squares solution using orthonormal matrices that will be presented in a subsequent joint paper. Jan T. Galkowski provided me with a considerable number of papers on related subjects. Thomas Poiker and C. J. Standish made several helpful suggestions. Elaine Shimabukuro uncomplainingly typed several drafts of this paper. This research was supported in part by the National Science Foundation under grant no. DMC85-11966.

B. K. P. Horn is on leave from the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

REFERENCES

1. C. C. Slama, C. Theurer, and S. W. Henrikson, eds., *Manual of Photogrammetry* (American Society of Photogrammetry, Falls Church, Va., 1980).
2. B. K. P. Horn, *Robot Vision* (MIT/McGraw-Hill, New York, 1986).
3. P. R. Wolf, *Elements of Photogrammetry* (McGraw Hill, New York, 1974).
4. S. K. Gosh, *Theory of Stereophotogrammetry* (Ohio U. Bookstores, Columbus, Ohio, 1972).
5. E. H. Thompson, "A method for the construction of orthogonal matrices," Photogramm. Record **3,** 55–59 (1958).
6. G. H. Schut, "On exact linear equations for the computation of the rotational elements of absolute orientation," Photogrammetria **16,** 34–37 (1960).
7. H. L. Oswal and S. Balasubramanian, "An exact solution of absolute orientation," Photogramm. Eng. **34,** 1079–1083 (1968).
8. G. H. Schut, "Construction of orthogonal matrices and their application in analytical photogrammetry," Photogrammetria **15,** 149–162 (1959).
9. E. H. Thompson, "On exact linear solution of the problem of absolute orientation," Photogrammetria **15,** 163–179 (1959).
10. E. Salamin, "Application of quaternions to computation with rotations," Internal Rep. (Stanford University, Stanford, California, 1979).
11. R. H. Taylor, "Planning and execution of straight line manipulator trajectories," in *Robot Motion: Planning and Control,* M. Bradey, J. M. Mollerbach, T. L. Johnson, T. Lozano-Pérez, and M. T. Mason, eds. (MIT, Cambridge, Mass., 1982).
12. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw Hill, New York, 1968).
13. J. H. Stuelpnagle, "On the parameterization of the three-dimensional rotation group," SIAM Rev. **6,** 422–430 (1964).
14. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra* (Macmillan, New York, 1953).
15. P. H. Winston and B. K. P. Horn, *LISP* (Addison-Wesley, Reading, Mass., 1984).