Reprinted from Journal of the Optical Society of America A, Vol. 5, page 1127, July 1988
Copyright © 1988 by the Optical Society of America and reprinted by permission of the copyright owner.

Closed-form solution of absolute orientation using orthonormal matrices

Berthold K. P. Horn*

Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822

Hugh M. Hilden

Department of Mathematics, University of Hawaii at Manoa, Honolulu, Hawaii 96822

Shahriar Negahdaripour†

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received June 2, 1987; accepted March 25, 1988

Finding the relationship between two coordinate systems by using pairs of measurements of the coordinates of a number of points in both systems is a classic photogrammetric task. The solution has applications in stereophotogrammetry and in robotics. We 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 a closed-form solution can be simplified by using unit quaternions to represent rotation, as was shown in an earlier paper [J. Opt. Soc. Am. A **4,** 629 (1987)]. Since orthonormal matrices are used more widely to represent rotation, we now present a solution in which 3×3 matrices are used. Our method requires the computation of the square root of a symmetric matrix. We compare the new result with that obtained by an alternative method in which orthonormality is not directly enforced. In this other method a best-fit linear transformation is found, and then the nearest orthonormal matrix is chosen for the rotation. We note that 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.

1. ORIGIN OF THE PROBLEM

Suppose that we are given the coordinates of a number of points as measured in two different Cartesian coordinate systems (see Fig. 1). The photogrammetric problem of recovering the transformation between the two systems from these measurements is referred to as that of absolute orientation.[1–3] 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 we speak of the determination of the hand–eye transform.[4]

A. Previous Work

The problem of absolute orientation is usually treated in an empirical, graphic, or numerical, iterative fashion.[1–3] Thompson[5] gave a solution to this problem for the case in which exactly 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, as is discussed in Subsection 1.B. Schut used unit quaternions and arrived at a set of linear equations. A simpler solution that does not require the solution of a system of linear equations was presented in a precursor of this paper.[7] These three 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[8] developed a least-squares method that can handle more than three points, but their method does not enforce the orthonormality of the rotation matrix. Instead, they simply find the best-fit linear transform. An iterative method is then used to square up the result, bringing it closer to being orthonormal. Their method for doing this is iterative (and without mathematical justification). In addition, the result obtained is not the solution of the original least-squares problem.

We study their approach in Section 4 by using a closed-form solution for the nearest orthonormal matrix derived in Subsection 3.F. This is apparently not entirely novel, since an equivalent problem was treated in the psychological literature.[9–15] The existing methods, however, cannot deal with a singular matrix. We extend our method to deal with the case in which the rank deficiency of the matrix is 1. This is an important extension, since the matrix is singular when either of the sets of measurements is coplanar, as is always the case when there are only three measurements.

The main result presented here, however, is the closed-form solution to the least-squares problem of absolute orientation. Our new result can be applied in the special case when one or the other of the sets of measurements happens to be coplanar. This is important because sometimes only

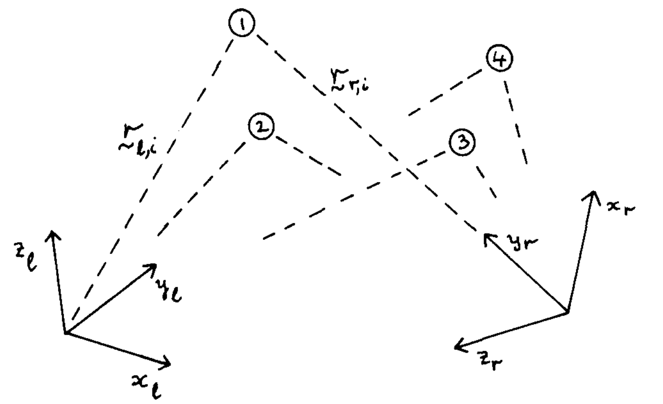

Fig. 1. The coordinates of a number of points are measured in two coordinate systems. The transformation between the two systems can be found by using these measurements.

three points are available, and three points are, of course, always coplanar. The solution that we present differs from the schemes discussed at the beginning of this section in that it does not selectively neglect information provided by the measurements: it uses all the information.

We should point out that a version of this problem was solved by Farrel and Stuelpnagel.[16] However, their solution applies only when neither of the sets of measurements is coplanar. We also learned recently that Arun *et al.*[17] independently developed a solution to an equivalent problem. They used a singular-value decomposition of an arbitrary matrix instead of the eigenvalue–eigenvector decomposition of a symmetric matrix inherent in our approach.

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 (the 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 allow determination of the seven unknowns, as shown, for example, in Ref. 7. By discarding two of the constraints, we can develop seven equations in seven unknowns that permit us to recover the parameters.

C. Least Sum of Squares of Errors

In practice, measurements are not exact, and so greater accuracy in determining the transformation parameters is sought by using more than three points. We no longer 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 the squares of the residual errors. Finding the best set of transformation parameters is not easy. Various empirical, graphic, and numerical procedures are currently in use.[1–3] These are all iterative in nature; that is, given an approximate solution, such a method is applied repeatedly until the remaining error becomes negligible.

At times, information is available that permits us to obtain so good an initial guess of the transformation parameters that a single step of the iteration brings us close enough to the true solution of the least-squares problem for all practical purposes, but this is rare.

D. Closed-Form Solution

In this paper we 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 us in one step with the best possible transformation, given the measurements of the points in the two coordinate systems. Another advantage is that it is not necessary to find a good initial guess, as it is when an iterative method is used.

A solution to this problem was presented previously in which unit quaternions are used to represent rotations.[7] The solutions for the desired quaternion was shown to be the eigenvector of a symmetric 4×4 matrix associated with the largest positive eigenvalue. The elements of this matrix are simple combinations of sums of products of coordinates of the points. To find the eigenvalues, one must solve a quartic equation whose coefficients are sums of products of elements of the matrix. It was shown that this quartic equation is particularly simple, since one of its coefficients is zero. It simplifies even more when only three points are used.

E. Orthonormal Matrices

Whereas unit quaternions constitute an elegant representation for rotation, most of us are more familiar with the use of proper orthonormal matrices for this purpose. Working directly with matrices is difficult because of the need to deal with six nonlinear constraints that ensure that the matrix is orthonormal. We nevertheless are able to derive a solution for the rotation matrix by using direct manipulation of 3×3 matrices. This closed-form solution requires the computation of the positive semidefinite square root of a positive semidefinite matrix. We show in Subsection 3.C how such a square root can be found once the eigenvalues and eigenvectors of the matrix are available. Finding the eigenvalues requires the solution of a cubic equation.

The method discussed here finds the same solution as does the method presented previously,[7] which uses unit quaternions to represent rotation, since it minimizes the same error sum. We present the new method only because the use of orthonormal matrices is so widespread. We actually consider the method using unit quaternions to be more elegant.

F. Symmetry of the Solution

Let us call the two coordinate systems "left" and "right." A desired property of a solution method is that, when it is applied to the problem of finding the best transformation from the left to the right system, it gives the exact inverse of the best transformation from the left system to the right system. It was shown in Ref. 7 that the scale factor must be treated in a particular way to guarantee that this happens. The method that we develop here for directly computing the rotation matrix gives two apparently different results when it is applied to the problem of finding the best transformation from left to right and the problem of finding the best transformation from right to left. We show that these two results are in fact different forms of the same solution and that our method does indeed have the desired symmetry property.

G. Nearest Orthonormal Matrix

Since the constraint of orthonormality leads to difficulties, some authors chose to find a 3×3 matrix that fits the data best in a least-squares sense without a constraint on its element.[8] The result is typically not orthonormal. If the data are fairly accurate, the matrix may be almost orthonormal. In this case, we wish to find the nearest orthonormal matrix; that is, we wish to minimize the sum of the squares of differences between the elements of the matrix obtained from the measurements and an ideal orthonormal matrix. Iterative methods exist for finding the nearest orthonormal matrix.

A closed-form solution, shown in Subsection 3.F, again involves square roots of 3×3 symmetric matrices. The answer obtained by this method is different, however, from that obtained by the solution that minimizes the sum of the squares of the residual errors. In particular, it does not have the highly desirable symmetry property discussed above, and it requires the accumulation of a larger number of sums of products of coordinates of measured points.

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 right coordinate systems[7] (see Fig. 2). 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 a right angle to the new x axis in the plane formed by the three points. The new z axis is then made orthogonal to the x and y axes with an orientation chosen to satisfy the right-hand rule. This construction is carried out in both the left and the 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 is shown in Ref. 7.

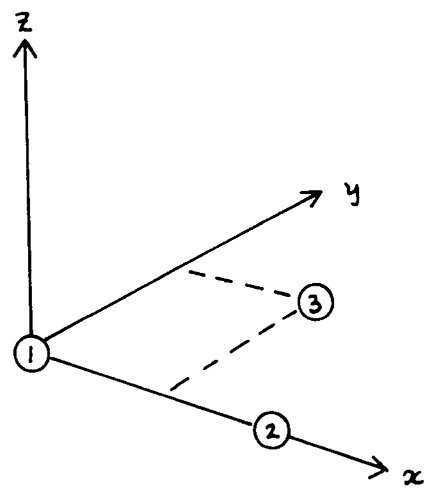

Fig. 2. Three points can be used to define a triad. Such a triad can be constructed by using the left measurements. A second triad is then constructed from the right measurements. The required coordinate transformation can be estimated by finding the transformation that maps one triad into the other.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

This *ad hoc* method constitutes a closed-form solution for finding the rotation from three points. Note that it uses the information from the three points selectively. Indeed, if we renumber the points, we obtain 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 only three points we should really attack this problem by means of a least-squares method, since there are more constraints than unknown parameters. The least-squares solutions for translation and scale are given in Subsections 2.B and 2.C. The optimum rotation is found in Section 4.

A. Finding the Translation

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

$$\{\mathbf{r}_{l,i}\}, \qquad \{\mathbf{r}_{r,i}\},$$

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 fact 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 cannot find a scale factor, a translation, and a rotation such that the transformation equation above is satisfied for each point. Instead there is a residual error,

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

We minimize the sum of the squares of these errors,

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

(It was shown in Ref. 7 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.

B. Centroids of the Sets of Measurements

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

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

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

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

The error term can now be rewritten in the form

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

where

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

The sum of the squares of the 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 sum of the vectors $\{\mathbf{r}'_{l,i}\}$ and the sum of the vectors $\{\mathbf{r}'_{r,i}\}$ are zero, as mentioned above. As a result, we are left with the first and the third terms. The first term does not depend on $\mathbf{r}'_0$, and the last cannot be negative. The total error obviously is 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, which uses all the 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 simplified to read as

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

since $\mathbf{r}'_0 = 0$, and so the total error to be minimized is just

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

C. Symmetry in Scale

It was shown in Ref. 7 that the formulation of the error term given in Subsection 2.B leads to an asymmetry in the determination of the optimal scale factor; that is, the optimal transformation from the left to the right coordinate system is then not the exact inverse of the optimal transformation from the right to the left coordinate system. The latter corresponds to use of the error term

$$\mathbf{e}_i = \mathbf{r}'_{l,i} - (1/s)R^T(\mathbf{r}'_{r,i}),$$

or

$$\mathbf{e}_i = -[(1/s)(\mathbf{r}'_{r,i}) - R(\mathbf{r}'_{l,i})],$$

and leads to a total error to be minimized of

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

If the errors in both sets of measurements are similar, it is more reasonable to use a symmetrical expression for the error term:

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

The total error then becomes

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

or

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

where

$$S_l = \sum_{i=1}^{n} \|\mathbf{r}'_{r,l}\|^2, \qquad D = \sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot [R(\mathbf{r}'_{l,i})], \qquad S_r = \sum_{i=1}^{n} \|\mathbf{r}'_{r,i}\|^2.$$

Completing the square in s, we obtain

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

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

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

One advantage of this symmetrical result is that it permits us to determine the scale without knowledge of the rotation. Importantly, however, the determination of the rotation is not affected by the choice of the value of the scale factor. In each case the remaining error is minimized when D is as large as possible; that is, we must choose the rotation that makes

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

as large as possible.

3. DEALING WITH ROTATION

There are many ways to represent rotation, including Euler angles, the Gibbs vector, Cayley–Klein parameters, Pauli spin matrices, axis-and-angle systems, orthonormal matrices, and Hamilton's quaternions.[18,19] Of these representations, orthonormal matrices are used most often in photogrammetry, graphics, and robotics. Although unit quaternions have many advantages when used to represent rotation, few investigators are familiar with their properties. We therefore present here a closed-form solution that uses orthonormal matrices and is similar to the closed-form solution obtained earlier that uses unit quaternions.[7]

The new method, which we present in this section, depends on the eigenvalue–eigenvector decomposition of a 3 × 3 matrix and so requires the solution of a cubic equation. Well-known methods such as Ferrari's solution can be used.[19–21] When one or the other (left or right) set of measurements is coplanar, the method simplifies, in that only a quadratic equation needs to be solved. It turns out, however, that much of the complexity of this approach stems from the need to deal with this and other special cases.

A. Best-Fit Orthonormal Matrix

We must find a matrix R that maximizes

$$\sum_{i=1}^{n} \mathbf{r}'_{r,i} \cdot [R(\mathbf{r}'_{l,i})] = \sum_{i=1}^{n} (\mathbf{r}'_{r,i})^T R(\mathbf{r}'_{l,i}).$$

Now

$$\mathbf{a}^T R\mathbf{b} = \mathrm{Tr}(R^T \mathbf{a}\mathbf{b}^T),$$

so we can rewrite the above expression in the form

$$\mathrm{Tr}\left[R^T \sum_{i=1}^{n} \mathbf{r}'_{r,i}(\mathbf{r}'_{l,i})^T\right] = \mathrm{Tr}(R^T M),$$

where

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

that is,

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

with

$$S_{xx} = \sum_{i=1}^{n} x_{r,i} x_{l,i}, \qquad S_{xy} = \sum_{i=1}^{n} x_{r,i} y_{l,i}, \ldots,$$

and so on. (We denote the elements of the matrix by S_{xx}, $S_{xy}, \ldots$ rather than by $M_{xx}, M_{xy}, \ldots$ in order to be consistent with Ref. 7.)

To find the rotation that minimizes the residual error, we must find the orthonormal matrix R that maximizes

$$\mathrm{Tr}(R^T M).$$

B. Product of Orthonormal and Symmetric Matrices

It follows from Theorem 1 on p. 169 of Ref. 22, that a square matrix M can be decomposed into the product of an orthonormal matrix U and a positive semidefinite matrix S. The matrix S is always uniquely determined. The matrix U is uniquely determined when M is nonsingular. (We show in Subsection 3.D that U can also be determined up to a two-way ambiguity when M is singular with a rank deficiency of 1.) When M is nonsingular, we can actually write directly

$$M = US,$$

where

$$S = (M^T M)^{1/2}$$

is the positive definite square root of the symmetric matrix $M^T M$, while

$$U = M(M^T M)^{-1/2}$$

is an orthonormal matrix. It is easy to verify that $M = US$, $S^T = S$, and $U^T U = I$.

C. Positive Definite Square Root of Positive Definite Matrix

The matrix $M^T M$ can be written in terms of the set of its eigenvalues $\{\lambda_i\}$ and the corresponding orthogonal set of unit eigenvectors $\{\hat{\mathbf{u}}_i\}$ as follows:

$$M^T M = \lambda_1 \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \lambda_2 \hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \lambda_3 \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T.$$

(This can be seen by checking that the expression on the right-hand side has eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\hat{\mathbf{u}}_i\}$.)

Now $M^T M$ is positive definite, so the eigenvalues are positive. Consequently, the square roots of the eigenvalues are real, and we can construct the symmetric matrix

$$S = \sqrt{\lambda_1}\hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \sqrt{\lambda_2}\hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \sqrt{\lambda_3}\hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T.$$

It is easy to show that

$$S^2 = \lambda_1 \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \lambda_2 \hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \lambda_3 \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T = M^T M,$$

using the fact that the eigenvectors are orthogonal. Also, for any nonzero vector $\mathbf{x}$,

$$\mathbf{x}^T S \mathbf{x} = \lambda_1(\hat{\mathbf{u}}_1 \cdot \mathbf{x})^2 + \lambda_2(\hat{\mathbf{u}}_2 \cdot \mathbf{x})^2 + \lambda_3(\hat{\mathbf{u}}_3 \cdot \mathbf{x})^2 > 0.$$

We see that S is positive definite, since $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_3 > 0$. This construction of $S = (M^T M)^{1/2}$ applies even when some of the eigenvalues are zero; the result then is positive semidefinite (rather than positive definite).

D. Orthonormal Matrix in the Decomposition

If all the eigenvalues are positive, then

$$S^{-1} = (M^T M)^{-1/2} = \frac{1}{\sqrt{\lambda_1}}\hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \frac{1}{\sqrt{\lambda_2}}\hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T + \frac{1}{\sqrt{\lambda_3}}\hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T,$$

as can be verified by multiplying by S. This expansion can be used to calculate the orthonormal matrix

$$U = MS^{-1} = M(M^T M)^{-1/2}.$$

The sign of $\det(U)$ is the same as the sign of $\det(M)$, because

$$\det(U) = \det(MS^{-1}) = \det(M)\det(S^{-1}),$$

and $\det(S^{-1})$ is positive, as all its eigenvalues are positive. Thus U represents a rotation when $\det(M) > 0$ and a reflection when $\det(M) < 0$. (We expect always to obtain a rotation in our case. Only if the data are severely corrupted may a reflection provide a better fit).

When M has a rank of only 2, the above method for constructing the orthonormal matrix breaks down. Instead, we use

$$U = M\left(\frac{1}{\lambda_1}\hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^T + \frac{1}{\lambda_2}\hat{\mathbf{u}}_2 \hat{\mathbf{u}}_2^T\right) \pm \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T,$$

or

$$U = MS^+ \pm \hat{\mathbf{u}}_3 \hat{\mathbf{u}}_3^T,$$

where S^+ is the pseudoinverse of S, that is,

$$S^+ = \frac{1}{\sqrt{\lambda_1}}\hat{u}_1 \hat{u}_1^T + \frac{1}{\sqrt{\lambda_2}}\hat{u}_2 \hat{u}_2^T,$$

and $\hat{\mathbf{u}}_3$ is the eigenvector with zero eigenvalue. The sign of the last term in the expression for U above is chosen to make the determinant of U positive. It is easy to show that the matrix constructed in this fashion is orthonormal and provides the desired decomposition $M = US$.

E. Maximizing the Trace

We must maximize

$$\mathrm{Tr}(R^T M) = \mathrm{Tr}(R^T US),$$

where $M = US$ is the decomposition of M discussed above. From the expression for S in Subsection 3.C, we see that

$$\mathrm{Tr}(R^T US) = \sqrt{\lambda_1}\,\mathrm{Tr}(R^T U\hat{\mathbf{u}}_1\hat{\mathbf{u}}_1^T) + \sqrt{\lambda_2}\,\mathrm{Tr}(R^T U\hat{\mathbf{u}}_2\hat{\mathbf{u}}_2^T) + \sqrt{\lambda_3}\,\mathrm{Tr}(R^T U\hat{\mathbf{u}}_3\hat{\mathbf{u}}_3^T).$$

For any matrices X and Y, such that XY and YX are square, $\mathrm{Tr}(XY) = \mathrm{Tr}(YX)$. Therefore

$$\mathrm{Tr}(R^T U\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T) = \mathrm{Tr}(\hat{\mathbf{u}}_i^T R^T U\hat{\mathbf{u}}_i) = \mathrm{Tr}(R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i) = (R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i).$$

Since $\hat{\mathbf{u}}_i$ is a unit vector and both U and R are orthonormal transformations, we have $(R\hat{\mathbf{u}}_i \cdot U\hat{\mathbf{u}}_i) \le 1$, with equality if and only if $R\hat{\mathbf{u}}_i = U\hat{\mathbf{u}}_i$. It follows that

$$\mathrm{Tr}(R^T US) \le \sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3} = \mathrm{Tr}(S),$$

and the maximum value of $\mathrm{Tr}(R^T US)$ is attained when $R^T U = I$, or $R = U$. Thus the orthonormal matrix that we seek is the one that occurs in the decomposition of M into the product of an orthonormal matrix and a symmetric one. If M is not singular, then

$$R = M(M^T M)^{-1/2}.$$

If M has a rank of only 2, however, we must resort to the second method discussed in Subsection 3.D to find R.

F. Nearest Orthonormal Matrix

We can now show that the nearest orthonormal matrix R to a given nonsingular matrix M is the matrix U that occurs in the decomposition of M into the product of an orthonormal matrix and a positive definite matrix; that is,

$$U = M(M^T M)^{-1/2}.$$

We wish to find the matrix R that minimizes

$$\sum_{i=1}^{3}\sum_{j=1}^{3}(m_{i,j} - r_{i,j})^2 = \mathrm{Tr}[(M - R)^T(M - R)],$$

subject to the condition that $R^T R = \mathrm{I}$; that is, we must minimize

$$\mathrm{Tr}(M^T M) - 2\mathrm{Tr}(R^T M) + \mathrm{Tr}(R^T R).$$

Now $R^T R = I$, so we conclude that the first and third terms do not depend on R. The problem then is to maximize

$$\mathrm{Tr}(R^T M).$$

We conclude immediately, using the result of Subsection 3.E, that the nearest orthonormal matrix to the matrix M is the orthonormal matrix that occurs in the decomposition of M into the product of an orthonormal matrix and a symmetric matrix.

Thus the orthonormal matrix that maximizes the residual error in our original least-squares problem is the orthonormal matrix nearest to the matrix

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

We note here that this orthonormal matrix can be found once an eigenvalue–eigenvector decomposition of the symmetric 3×3 matrix $M^T M$ has been obtained.

G. Rank of the Matrix M

It is clear that the rank of $M^T M$ is the same as the rank of M, since the two matrices have exactly the same eigenvectors with zero eigenvalue. The first method for finding the desired orthonormal matrix applies only when M, and hence $M^T M$, is nonsingular.

If, on the other hand,

$$M\mathbf{n}_l = 0$$

for any nonzero vector $\mathbf{n}_l$, then the matrix M, and hence $M^T M$, is singular. This happens when all the left measurements lie in the same plane, that is, when

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

for $i = 1, 2, \ldots, n$, where $\mathbf{n}_l$ is normal to the plane, since

$$M\mathbf{n}_l = \left(\sum_{i=1}^{n} \mathbf{r}'_{r,i}(\mathbf{r}'_{l,i})^T\right)\mathbf{n}_l = \sum_{i=1}^{n} \mathbf{r}'_{r,i}(\mathbf{r}'_{l,i} \cdot \mathbf{n}_l) = 0.$$

Similarly, if all the right measurements lie in the same plane, then

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

where $\mathbf{n}_r$ is a normal to the plane, and so $M^T\mathbf{n}_r = 0$. Now $\det(M^T) = \det(M)$, so this implies that M is singular also. As a result, we cannot use the simple expression

$$U = M(M^T M)^{-1/2}$$

to find the orthonormal matrix when either of the two sets of measurements (left or right) is coplanar. This happens, for example, when there are only three points.

If one or both sets of measurements are coplanar, we must use the second method for constructing U, which is given in Subsection 3.D. This method requires that the matrix M have a rank of 2 (which is the case unless the measurements happen to be collinear, in which case the absolute orientation problem does not have a unique solution). Note that the second method requires the solution of a quadratic equation to find the eigenvalues, whereas a cubic equation must be solved in the general case. We might, by the way, anticipate possible numerical problems when the matrix M is ill conditioned, that is, when one of the eigenvalues is nearly zero. This happens when one of the sets of measurements lies almost in a plane.

H. Symmetry in the Transformation

If, instead of finding the best transformation from the left to the right coordinate system, we decided to find the best transformation from the right to the left, then we would have to maximize

$$\sum_{i=1}^{n} (\mathbf{r}'_{l,i})^T \overline{R}\mathbf{r}'_{r,i}$$

by choosing an orthonormal matrix $\bar{R}$. We can immediately write down the solution

$$\bar{R} = M^T(MM^T)^{-1/2},$$

since M becomes M^T when we interchange the left and right systems. We would expect $\bar{R}^T$ to be equal to R, but, much to our surprise,

$$\bar{R}^T = (MM^T)^{-1/2}M.$$

This appears to be different from

$$R = M(M^TM)^{-1/2},$$

but, in fact, they are equal. This is so because

$$[M^{-1}(MM^T)^{1/2}M]^2 = M^{-1}(MM^T)M = M^TM.$$

Taking inverses and square roots, we obtain

$$M^{-1}(MM^T)^{-1/2}M = (M^TM)^{-1/2},$$

and, premultiplying by M, we find that

$$\bar{R}^T = (MM^T)^{-1/2}M = M(M^TM)^{-1/2} = R.$$

I. Finding the Eigenvalues and Eigenvectors

We must find the roots of the cubic equation in λ obtained by expanding

$$\det(M^TM - \lambda I) = 0,$$

where M^TM is

$$\begin{bmatrix} S_{xx}^2 + S_{yx}^2 + S_{zx}^2 & S_{xx}S_{xy} + S_{yx}S_{xy} + S_{zx}S_{zy} & S_{xx}S_{xz} + S_{yx}S_{yz} + S_{zx}S_{zz} \\ S_{xy}S_{xx} + S_{yy}S_{yx} + S_{zy}S_{zx} & S_{xy}^2 + S_{yy}^2 + S_{zy}^2 & S_{xy}S_{xz} + S_{yy}S_{yz} + S_{zy}S_{zz} \\ S_{xz}S_{xx} + S_{yz}S_{yx} + S_{zz}S_{zx} & S_{xz}S_{xy} + S_{yz}S_{yy} + S_{zz}S_{zy} & S_{xz}^2 + S_{yz}^2 + S_{zz}^2 \end{bmatrix}.$$

Having found the three solutions of the cubic equation, λ_i for $i = 1, 2, 3$ (all real and, in fact, positive), we then solve the homogeneous equations

$$(M^TM - \lambda_i I)\hat{\mathbf{u}}_i = 0$$

to find the three orthogonal eigenvectors $\hat{\mathbf{u}}_i$ for $i = 1, 2, 3$.

J. Coefficients of the Cubic Equation

Suppose that we write the matrix M^TM in the form

$$M^TM = \begin{bmatrix} a & d & f \\ d & b & e \\ f & e & c \end{bmatrix},$$

where $a = (S_{xx}^2 + S_{yy}^2 + S_{zz}^2)$ and so on; then

$$\det(M^TM - \lambda I) = 0$$

can be expanded as

$$-\lambda^3 + d_2\lambda^2 + d_1\lambda + d_0 + 0,$$

where

$$d_2 = a + b + c,$$
$$d_1 = (e^2 - bc) + (f^2 - ac) + (d^2 - ab),$$
$$d_0 = abc + 2def - (ae^2 + bf^2 + cd^2).$$

We may note at this point that

$$d_2 = \mathrm{Tr}(M^TM),$$

so

$$d_2 = (S_{xx}^2 + S_{xy}^2 + S_{xz}^2) + (S_{yx}^2 + S_{yy}^2 + S_{yz}^2) + (S_{zx}^2 + S_{zy}^2 + S_{zz}^2),$$

while

$$d_0 = \det(M^TM) = [\det(M)]^2$$

or

$$d_0 = [(S_{xx}S_{yy}S_{zz} + S_{xy}S_{yz}S_{zx} + S_{yx}S_{xz}S_{zy}) - (S_{xx}S_{yz}S_{zy} + S_{yy}S_{zx}S_{xz} + S_{zz}S_{xy}S_{yx})]^2.$$

4. IGNORING THE ORTHONORMALITY

Since it is so difficult to enforce the six nonlinear constraints that ensure that the matrix R is orthonormal, it is tempting just to find the best-fit linear transformation from the left to the right coordinate system. This is a straightforward least-squares problem. We can then try to find the nearest orthonormal matrix to the one obtained in this fashion. We show that this approach actually involves more work and does not produce the solution to the original least-squares problem. In fact, the result is asymmetric, in that the best-fit linear transform from left to right is not the inverse of the best-fit linear transform from right to left. Furthermore, at least four points must be measured, whereas the method that enforces orthonormality requires only three. We discuss this approach in this section.

A. Best-Fit Linear Transformation

We must find the matrix X that minimizes

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

or

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

Since X is not necessarily orthonormal, we cannot simply replace $\|X\mathbf{r}'_{l,i}\|^2$ by $\|\mathbf{r}'_{l,i}\|^2$. Note that $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$ and that $\mathbf{x} \cdot \mathbf{y} = \mathrm{Tr}(\mathbf{x}\mathbf{y}^T)$. The sum above can be rewritten in the form

$$\sum_{i=1}^{n} \mathrm{Tr}[\mathbf{r}'_{r,i}(\mathbf{r}'_{r,i})^T - 2\mathbf{r}'_{r,i}(\mathbf{r}'_{l,i})^T X^T + X\mathbf{r}'_{l,i}(\mathbf{r}'_{l,i})^T X^T] = \mathrm{Tr}(XA_lX^T - 2MX^T + A_r),$$

where

$$A_l = \sum_{i=1}^{n} \mathbf{r}'_{l,i}(\mathbf{r}'_{l,i})^T, \qquad A_r = \sum_{i=1}^{n} \mathbf{r}'_{r,i}(\mathbf{r}'_{r,i})^T$$

are symmetric 3 × 3 matrices obtained from the left and right sets of measurements, respectively.

We can find the minimum essentially by completing the square. First, we use the fact that $\mathrm{Tr}(MX^T) = \mathrm{Tr}(X^TM)$ to rewrite the above expression in the form

$$\mathrm{Tr}(XA_lX^T - MX^T - XM^T + MA_l^{-1}M^T) + \mathrm{Tr}(A_r - MA_l^{-1}M^T).$$

The second term does not depend on X, and the first term can be written as the trace of a product:

$$\mathrm{Tr}[(XA_l - M)(X - MA_l^{-1})^T].$$

Now it is easy to see that A_l is positive semidefinite. In fact, the matrix A_l is positive definite, provided that at least four measurements are available that are not collinear. This means that A_l has a positive definite square root and that this square root has an inverse. As a result, we can then rewrite the above expression in the form

$$\mathrm{Tr}[(XA_l^{1/2} - MA_l^{-1/2})(XA_l^{1/2} - MA_l^{-1/2})^T] = \|XA_l^{1/2} - MA_l^{-1/2}\|^2.$$

This is zero when

$$XA_l^{1/2} = MA_l^{-1/2}$$

or

$$X = MA_l^{-1}.$$

B. Asymmetry in the Simple Linear Solution

It is easy to find X by multiplying M by the inverse of A_l. Note, however, that we are using more information here than before. The method that does enforce orthonormality requires only the matrix M. Also note that A_l depends on the left measurements alone. This suggests an asymmetry. Indeed, if we minimize instead

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

we obtain

$$\bar{X} = M^TA_r^{-1}.$$

In general, $\bar{X}$ is not equal to X^{-1}, as one might expect.

Neither X nor $\bar{X}$ need be orthonormal. The nearest orthonormal matrix to X is shown in Subsection 3.F to be equal to

$$R = X(X^TX)^{-1/2} = (XX^T)^{-1/2}X,$$

whereas the matrix nearest to $\bar{X}$ is

$$\bar{R} = \bar{X}(\bar{X}^T\bar{X})^{-1/2} = (\bar{X}\bar{X}^T)^{-1/2}\bar{X}.$$

Typically $\bar{R}^T \neq R$.

C. Relationship of Simple Linear Solution to Exact Solution

We know from Subsection 3.F that the solution of the original least-squares problem is the orthonormal matrix closest to M. The simple best-fit linear solution instead leads to the matrix MA_l^{-1}. The closest orthonormal matrix to MA_l^{-1} in general is not equal to that closest to M. To see this, suppose that

$$M = US, \qquad MA_l^{-1} = U'S'$$

are the decompositions of M and MA_l^{-1} into orthonormal and positive definite matrices; then

$$US = U'(S'A_l).$$

For the solutions to be identical (that is, $U = U'$), we would need to have

$$S = S'A_l,$$

but the product of two symmetric matrices is, in general, not symmetric; so, in general, $U' \neq U$.

D. Disadvantage of the Simple Linear Method

The simple linear method does not lead to an orthonormal matrix. The closest orthonormal matrix can be found, but that requires just as much work as that required for the exact solution of the original least-squares problem. In addition, the simple linear method requires that twice as many data be accumulated (A_l or A_r in addition to M). Furthermore, the linear transformation has more degrees of freedom (nine independent matrix elements) than does an orthonormal matrix, so more constraint is required. Indeed, for A_l or A_r to be nonsingular, at least four points must be measured. This is a result of the fact that the vectors are taken relative to the centroid, and so three measurements do not provide three independent vectors. More seriously, this method does not produce the solution to the original least-squares problem.

5. CONCLUSION

We presented here a closed-form solution of the least-squares problem of absolute orientation, using orthonormal matrices to represent rotation. The method provides the best rigid-body transformation between two coordinate systems, given measurements of the coordinates of a set of points that are not collinear. A closed-form solution using unit quaternions to represent rotation was given previously.[7] In this paper we derive an alternative method that uses manipulation of matrices and their eigenvalue–eigenvector decomposition. The description of this method may perhaps appear to be rather lengthy. This is the result of the need to deal with various special cases, such as that of coplanar sets of measurements.

We show here 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 on measurements of one or two points only.

We contrast the exact solution of the absolute orientation problem to various approaches advocated in the past. The exact solution turns out to be easier to compute than one of these alternatives. The solution presented here may seem relatively complex. The ready availability of program packages for solving algebraic equations and finding eigenvalues and eigenvectors of symmetric matrices makes implementation straightforward, however. Methods for finding the eigenvectors efficiently were discussed in Ref. 7. It should also be noted that we deal only with 3×3 matrices here.

ACKNOWLEDGMENTS

Jan T. Galkowski provided us with copies of a number of interesting papers on subjects relevant to our endeavor. He also put us in touch with C. J. Standish, who discovered a paper by Farrel and Stuelpnagel[16] in which a closed-form solution to the least-squares problem of absolute orientation was presented that applies when neither of the sets of measurements is coplanar. Thomas Poiker made several helpful suggestions. Takeo Kanade drew our attention to some references in the psychological literature that are relevant to the task of finding the nearest orthonormal matrix.[9–15] After we had arrived at our solution, Thomas Huang brought to our attention a solution of this problem that he, S. Arun, and S. D. Blostein had developed,[17] using the singular-value decomposition of an arbitrary matrix instead of the eigenvalue–eigenvector decomposition of a symmetric matrix inherent in our approach. Elaine Shimabukuro uncomplainingly typed a number of drafts of this paper. We also wish to thank the reviewers for several helpful comments.

* On leave from the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

† Present address, Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822.

REFERENCES

1. C. C. Slama, C. Theurer, and S. W. Henrikson, eds., *Manual of Photogrammetry* (American Society of Photogrammetry, Falls Church, Va., 1980).
2. P. R. Wolf, *Elements of Photogrammetry* (McGraw-Hill, New York, 1974).

3. S. K. Gosh, *Theory of Stereophotogrammetry* (Ohio State U. Bookstores, Columbus, Ohio, 1972).
4. B. K. P. Horn, *Robot Vision* (MIT Press, Cambridge, Mass.; McGraw-Hill, New York, 1986).
5. E. H. Thompson, "An exact linear solution of the problem of absolute orientation," Photogrammetria 15(4), 163–179 (1958).
6. G. H. Schut, "On exact linear equations for the computation of the rotational elements of absolute orientation," Photogrammetria **16**(1), 34–37 (1960).
7. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," J. Opt. Soc. A **4**, 629–642 (1987).
8. H. L. Oswal and S. Balasubramanian, "An exact solution of absolute orientation," Photogramm. Eng. **34**, 1079–1083 (1968).
9. C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," Psychometrika **1**, 211–218 (1936).
10. B. F. Green, "The orthogonal approximation of an oblique structure in factor analysis," Psychometrika **1**, 429–440 (1952).
11. W. A. Gibson, "On the least-squares orthogonalization of an oblique transformation," Psychometrika **27**, 193–195 (1962).
12. R. M. Johnson, "On a theorem stated by Eckart and Young," Psychometrika **28**, 259–263 (1963).
13. P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," Psychometrika **31**, 1–10 (1966).
14. N. Cliff, "Orthogonal rotation to congruence," Psychometrika **31**, 33–42 (1966).
15. P. H. Schönemann and Robert M. Carroll, "Fitting one matrix to another under choice of a central dilation and a rigid motion," Psychometrika **35**, 169–183 (1970).
16. J. L. Farrell and J. C. Stuelpnagel, "Problem 65-1: a least squares estimated of satellite attitude," SIAM Rev. **8**, 384–386 (1966).
17. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," Tech. Note ISP-81 (Coordinated Science Laboratory, University of Illinois, Urbana, Ill., 1986); submitted to IEEE Trans. Patt. Anal. Machine Intell.
18. J. H. Stuelpnagel, "On the parameterization of the three-dimensional rotation group," SIAM Rev. **6**, 422–430 (1964).
19. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968).
20. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra* (Macmillan, New York 1953).
21. P. H. Winston and B. K. P. Horn, *LISP* (Addison-Wesley, Reading, Mass., 1984).
22. P. R. Halmos, *Finite Dimensional Vector Spaces*, 2nd ed. (Van Nostrand, Princeton, N.J., 1958).