G. H. Schut
National Research Council of Canada
Ottawa K1A 0S1, Canada

Similarity Transformation and Least Squares

Parameters based on the space-time formulation of special relativity are indistinguishable from those obtained if the parameters are computed by least-squares minimization of the residuals at the control points.

Introduction

A similarity transformation is a transformation of coordinates which can comprise no more than a change of scale, a rotation, and a translation. One can express this in somewhat more sophisticated language by stating[3] that it is a single-valued mapping of space onto itself which satisfies the requirement that it multiplies the distance between any two points by the same real positive number.

Abstract: *This article investigates three published methods in which the computation of the rotation matrix of a similarity transformation does not require initial estimates of the parameters. It is found that one method, based on the space-time transformation of special relativity, gives results that are virtually indistinguishable from those of the minimization of the residuals at the control points by the method of least squares. The other two are based on the initial computation of an affine transformation and in topographic applications of photogrammetry they are found to give at best appreciably inferior results.*

In matrix form, the transformation can be written as a function of the three component transformations:

$$x' = \lambda R x + c \tag{1}$$

Here, x and x' are vectors of coordinates (position vectors) before and after transformation, respectively, λ is a scale factor, R is an orthogonal matrix which applies the rotation, and c is a translation vector.

The computation of the parameters in this transformation can be performed by the method of least squares. The coordinates before transformation are here treated as observations or as pseudo-observations, and the weighted sum of squares of the residuals at the control points is minimized. This least-squares adjustment requires estimates of the parameters for use as initial approximations.

Good estimates for the parameters of the rotation cannot in all cases be easily determined. This is a problem that has led to the development of alternate methods for computing the matrix R of the similarity transformation and to the abandonment of the least squares adjustment.

Schut[4,5] developed a method by which four parameters which define the rotation matrix can be computed in a very simple way. This method does not place any restriction on the size of the rotation; it requires only that the coordinates of three non-collinear points be known in the two coordinate systems. The method seems to have received less attention than it deserves, perhaps because the two publications on the method deal only with special applications in which an exact fit was made at one of the control points. However, this is not a limiting feature of the method.

More recently, Oswal and Balasubramanian[2] and Blais[1] have given different methods, both based on the initial computation of an affine transformation. The orthogonal matrix is then derived from the matrix of that transformation. The second one of these two methods requires that the rotation be not equal to, and not very close to, a half-turn and both require a minimum of four control points. These points must not lie in a plane.

All three of these methods make a formal use of the method of least squares. However, they cannot be properly called least-squares adjustments because they do not minimize the weighted sum of squares of the residuals of observations. Schut minimizes the sum of squares of closures in certain condition equations. Oswal and Blais do not minimize any function in their respective derivations of an orthogonal matrix from the matrix of an affine transformation. Consequently, the results obtained with these three methods can be expected to differ from those obtained by the least squares minimization of residuals. No information on the possible size of these differences has been supplied by the authors.

The present article serves to compare the results that can be obtained with these three methods with those of the least squares adjustment. First, a short description of the methods is given. Use is made of this opportunity to show that Schut's formulation is based directly on the space-time transformation of special relativity. Also, three sometimes less desirable features of the formulation which served their purpose in the special applications have been changed. Oswal's method has been given its theoretical basis This has provided also a method to improve its orthogonalization procedure which is not exact. The methods are applied to the connection of independent models of two strips and to the absolute orientations of these models.

Least-Squares Minimization of Residuals

Because the results of the least-squares adjustment are of basic importance in the following, a short account of it must be given here first.

For the purpose of this adjustment, Equation 1 can be written:

$$x' = (1+\lambda_1)\lambda_0 \boldsymbol{R}_1 \boldsymbol{R}_0 x + c_0 + c_1 \qquad (2)$$

Here, the subscripts *0* and *1* indicate approximate values and corrections, respectively. Both $\boldsymbol{R}_1$ and $\boldsymbol{R}_0$ are exactly orthogonal matrices.

Correction equations, which are linear with respect to λ_1, the parameters of $\boldsymbol{R}_1$, and the components of c_1 are derived from this equation. First, $\boldsymbol{R}_1$ is written as the sum of the unit matrix and a skew-symmetric matrix:

$$\boldsymbol{R}_1 - \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix} \qquad (3)$$

Then, products of λ_1 and the parameters of $\boldsymbol{R}_1$ are ommitted. The equation can now be written in the form:

$$\begin{bmatrix} \lambda_1 & -\gamma & \beta \\ \gamma & \lambda_1 & -\alpha \\ -\beta & \alpha & \lambda_1 \end{bmatrix} \lambda_0 \boldsymbol{R}_0 x + c_1 = x' - (\lambda_0 \boldsymbol{R}_0 x + c_0) \quad . \qquad (4)$$

This vector equation represents three scalar correction equations with seven parameters. Normal equations are derived from these equations in the conventional way and are solved for the parameters.

The normal equations have the following interesting feature. If the origin of the unprimed system is chosen in the centroid of the control points, they fall apart into one equation for the scale correction λ_1, three separate equations for the three components of c_1, and a set of three equations for the parameters of $\boldsymbol{R}_1$.

The equation for the scale correction becomes:

$$\lambda_1 = \frac{\Sigma\ x_0 \bullet w}{\Sigma\ x_0 \bullet x_0} \qquad (5)$$

in which $x_0 = \lambda_0 \boldsymbol{R}_0 x$, w is the vector of residuals in the second part of Equation 4, the dot signifies the scalar product, and Σ indicates the summation over all control points.

The three separate equations for the translation corrections now specify that the origin of the unprimed system shall be transformed into the centroid of the control points in the primed system. In other words, the similarity transformation has the property that it transforms the centroid of the control points in one system into the centroid in the other system. Consequently, one should first shift in both systems the origin to the centroid. This leaves only the scale factor and the rotation matrix to be determined.

A good approximate value of the scale factor can be easily determined, for instance from distances in both systems. Good estimates of the parameters of the rotation matrix are often less readily available.

Derivation of R from the Space-Time Transformation

The elements of an orthogonal matrix can be written as rational functions of three or four parameters. This formulation is related to the Lorentz transformation, the space-time transformation of special relativity, and in particular to its quaternion formulation. To derive the elements in this form, the quaternions can be conveniently replaced by certain 4×4 matrices.

Let a matrix $\boldsymbol{T}$ be constructed from the space coordinates x, y, z and from a fourth parameter which shall represent the time as follows:

$$\boldsymbol{T} = \begin{matrix} t & -z & y & x \\ z & t & -x & y \\ -y & x & t & z \\ -x & -y & -z & t \end{matrix} \quad . \tag{6}$$

The set of all matrices which can be constructed in this way from four parameters form a system called a ring. This means that they obey the same rules with respect to addition, multiplication, and division as the set of all square nonsingular matrices of a given order. For instance, sums and products of matrices of this system are again matrices of the system.

The Lorentz transformation is effected by a matrix $\boldsymbol{D}$ of the ring whose parameters a, b, c, and d have the property:

$$a^2 + b^2 + c^2 + d^2 = 1 \quad . \tag{7}$$

The determinant of $\boldsymbol{D}$, which is the square of this expression, is unity. The inverse $\boldsymbol{D}^{-1}$ equals the transpose $\boldsymbol{D}^T$.

In photogrammetry, only a space transformation is of interest. In this special application, the Lorentz transformation becomes a space rotation about the origin, the four parameters of $\boldsymbol{D}$ become real and the transformation can be written

$$\boldsymbol{T}' = \boldsymbol{D}\boldsymbol{T}\boldsymbol{D}^T \quad . \tag{8}$$

Because the parameter t is now a constant, one can conveniently specify $t' = t = 0$. Also, a change of sign of all four parameters of $\boldsymbol{D}$ does not affect the transformation and, therefore, one may specify that d shall be positive.

The relations between the transformed coordinates x', y', z' and the coordinates x, y, z can now be found in a simple but laborious way by performing the matrix multiplications in Equation 8. This gives $x' = \boldsymbol{R}x$, where

$$\boldsymbol{R} = \begin{bmatrix} d^2+a^2-b^2-c^2 & 2ab-2cd & 2ac+2bd \\ 2ab+2cd & d^2-a^2+b^2-c^2 & 2bc-2ad \\ 2ac-2bd & 2bc+2ad & d^2-a^2-b^2+c^2 \end{bmatrix} \tag{9}$$

$$d^2+a^2+b^2+c^2 = 1 \text{ and } d > 0 \quad .$$

A more elegant derivation of Equation 9 is obtained by a generalization of the above matrix algebra which has no counterpart in quaternion algebra. Let the matrix obtained by transposing the fourth row and column of $\boldsymbol{D}$ be denoted by $\boldsymbol{D}$. This matrix can be obtained from $\boldsymbol{D}^T$ by transposing the first three columns and rows. The set of matrices constructed in the same way as $\boldsymbol{D}$ forms a second ring. The matrices of this ring commute with the matrices of the first ring.

The matrix $\boldsymbol{D}$ has, by virtue of its construction, the same fourth column as the matrix $\boldsymbol{D}^T$. It follows now from the rules of matrix multiplication that the matrices $\boldsymbol{T}' = \boldsymbol{D}\boldsymbol{T}\boldsymbol{D}^T$ and $\boldsymbol{D}\boldsymbol{D}\boldsymbol{T} = \boldsymbol{D}\boldsymbol{T}\boldsymbol{D}$ have the same fourth column. Accordingly, denoting the fourth column of $\boldsymbol{T}$ by $\boldsymbol{t}$, the transformation of Equation 8 can be written as the vector transformation

$$\boldsymbol{t}' = \boldsymbol{D}\boldsymbol{D}\boldsymbol{t} = \boldsymbol{D}\boldsymbol{D}\boldsymbol{t} \quad . \tag{10}$$

Forming the product $\boldsymbol{D}\boldsymbol{D}$, one finds that it has the matrix $\boldsymbol{R}$ of Equation 9 as the submatrix in the first three columns and rows. The element in the fourth column and row is unity, and the remaining six elements are zero. From this, Equation 9 follows directly.

The computation of the parameters of the matrix $\boldsymbol{R}$ by a direct procedure is based upon the four-dimensional form of the rotation, given by Equation 8. Postmultiplying both sides of the equation by $\boldsymbol{D}$, one obtains:

$$\boldsymbol{T}'\boldsymbol{D} = \boldsymbol{D}\boldsymbol{T} \quad . \tag{11}$$

This equation represents four scalar equations which can be obtained by equating the elements in the fourth column of the two products. Placing all terms in these equations in the first part and collecting terms with the same parameter, this gives

$$\begin{aligned} -(z'+z)b + (y'+y)c + (x'-x)d &= 0 \\ +(z'+z)a - (x'+x)c + (y'-y)d &= 0 \\ -(y'+y)a + (x'+x)b + (z'-z)d &= 0 \\ +(x'-x)a + (y'-y)b + (z'-z)c &= 0 \quad . \end{aligned} \tag{12}$$

Each complete control point gives rise to four Equations 12, but these equations are not all independent. The parameters of $\boldsymbol{R}$ can be computed from the equations of at least three non-collinear points. This can be done most conveniently by the method of least squares, minimizing the (weighted) sum of squares of the closures in these equations. A simple algorithm for this computation is given by Schut[5].

The special applications of this method contained two features, not related to the derivation of R, which differ from the corresponding ones in the least-squares adjustment:

a. Before computing the rotation matrix, the origins of the coordinate systems were shifted to one of the control points.
b. After the translation, the scale factor was computed from the mean of the lengths of the position vectors of the control points in the two systems.

Because Oswal's and Blais' method use the least-squares adjustment for the computation of translations and scale, a proper comparison of the methods requires that the same be done here.

Further, in the special applications all position vectors were reduced to unit length before their components were entered in Equations 12. This feature was copied from an earlier application of the method to the calibration of a camera by means of star photographs. It is in fact a weighting that gives a control point a greater weight the closer it lies to the centroid. If a point lies very close to the centroid, the adjustment can be adversely affected. Instead, each position vector in the x, y, z system must now be scaled to the x', y', z' system.

Derivation of R from an Affine Transformation

The affine transformation

The affine transformation can be written, similar to Equation 1, as:

$$x' = \lambda A x + c \quad .$$

Here, λA is a 3×3 matrix between whose elements no conditions are enforced. The factor λ serves to make A a matrix whose determinant equals unity. The elements of λA and the components of c can be computed directly from this equation. This requires the use of at least four complete control points. If more points are available, the method of least squares can be used. The computation breaks down if all control points lie in one plane. From the matrix A obtained in this way, Oswal and Blais derive the orthogonal matrix R of Equation 1.

Oswal's formulation of R

Oswal's procedure[2] for deriving an orthogonal matrix from A consists in adding to it a matrix of corrections. The correction matrix is a function of the closures in the condition equations which the elements of A must satisfy as the elements of an orthogonal matrix. The closures for the rows and those for the columns, respectively, can be arranged as the elements of symmetric matrices:

$$V = AA^T - I$$

and

$$W = A^TA - I \quad .$$

Oswal's formulation, of which he does not give a derivation, can be derived as follows. If the closures are small, all elements of V and W are close to zero and increasingly higher powers of these matrices will be increasingly closer to the zero matrix. As a result, one can obtain an orthogonal matrix from A by multiplying it by a power series in V or W with a sufficient number of terms:

$$R = (I + aV + bV^2 + \ldots)A$$

or

$$R = A(I + aW + bW^2 + \ldots) \quad .$$

If one omits all terms of the second and higher degree in V and W and makes use of the property that $VR = RW$, this gives:

$$\begin{aligned} R^TR &= A^T(I + aV)(I + aV)A \\ &= A^TA(I + aW)(I + aW) \\ &- I + (1 + 2a)W \quad . \end{aligned}$$

Therefore, R will here be orthogonal if $a = -\frac{1}{2}$ and if also the contribution of the higher-degree terms is negligible. Thus,

$$R = (I - \tfrac{1}{2}V)A$$

or

$$R = A(I - \tfrac{1}{2}W) \quad .$$

Oswal employs the second one of these formulations.

It follows that Oswal's procedure is not, as is claimed, an exact one. However, where necessary, the orthogonalization can be improved either by the use of the formulation in an iterative procedure or by the inclusion of the second-degree term in the power series. In the latter instance, one has $a = -\frac{1}{2}$, $b = +\frac{3}{8}$.

Blais' formulation of R

Blais[1] has given a complete account of his derivation of the similarity transformation from the affine transformation. He prefers to call the transformation a similarity, as is done in some textbooks, especially in geometry. One cannot summarily reject such an abbreviation, which is in common use for such transformations as a rotation and a translation. However, he should then be consistent and speak also of an affinity and not of an affine transformation.

The derivation of an orthogonal matrix

from A is exact and is performed in two steps:

$$S = (I+A)^{-1} - ((I+A)^{-1})^T$$
$$R = (I-S)(I+S)^{-1} \quad .$$

The first step is the essential feature of the formulation. It is Blais' procedure for deriving a skew-symmetric matrix from any matrix A. The second step is the conventional formulation of an orthogonal matrix from a skew-symmetric one. If A is already orthogonal, it recovers this matrix.

Experiments

To compare the results obtainable with these three methods with those of the minimization of residuals by least squares, they have been incorporated in a Fortran program. Double-precision arithmetic (about 15 significant digits) has been used throughout the computations.

Schut's method has been programmed in two ways: first, as published earlier but using the centroid of the control points as center of rotation; second, introducing also the other two modifications described in the present paper. These two variants will be referred to as Schut (1968) and Schut (1972), respectively.

Two variants of Oswal's method have been programmed also: one in which the second-degree terms in the power series are included also and one in which the method is used iteratively. The inclusion of the second-degree terms did not improve the results appreciably. The iterative procedure made the results converge toward those of Blais, except in instances of extremely large residuals. Therefore, these two variants need not be reported on further.

The program has been checked with the help of transformations of two regular polyhedra: a tetrahedron and a cube, both with sides of 10,000 m and using the micrometer as unit of measurement. The vertices of these bodies have a geometrically ideal spatial distribution and, therefore, do not cause a problem for Blais' and Oswal's methods. The rotations used were 120° and 100°, respectively. Using exact coordinates, all methods leave zero residuals at the vertices. Introducing one coordinate error of 100 μm, all methods leave the same residuals at the vertices. The elements of the orthogonal matrix do not differ before the 11th decimal place.

The program has been used in experiments with the models of two strips. After independent analytical model triangulation, the models were transformed to the respective ground control systems and to the adjoining models.

The first strip is line 1 of the block of fictitious data prepared in 1971 by the U.S. Army Topographic Command for the Working Group on Analytic Block Adjustment of Commission III of the ISP. It contains 19 models. A wide-angle camera with 152 mm focal length is postulated, with a flying height of 10,000 m above average terrain. For the present experiment, nine measured points per photograph were used. These points are located near the points of a regular grid, but their heights range from 200 to 2,000 m. The photograph coordinates which have been used were perturbed by random and by systematic errors. Each model has been oriented using the six available terrain points and each model connection has been made on the common projection center and the three common terrain points. In all these transformations, all control points have been given equal weight.

The results obtained with the models of this strip are summarized in Tables 1 and 2. Table 1 gives root-mean-square values of the residuals at model scale for a base of 90 mm. Table 2 gives the differences between the orthogonal matrices derived by each of the three methods and those derived by the least-squares minimization of the residuals. These differences are expressed as the rotation differences.

Especially notable is the fact that the use of the space-time transformation, with the modifications introduced in the present article, produces transformed coordinates which are identical with those obtained by the least-squares minimization of residuals. The rotations computed by these two methods differ by less than $0.000,01^g$ (100 grads equals a right angle).

Oswal's and Blais' methods leave residuals after absolute orientation that are several times larger than those for the other two methods and they leave rotation errors that range up to 0.3^g. This is true even though the height variations are large and six control points were used in each orientation.

In the instance of the model connections, Blais' rotations are excessive and bear no resemblance to the actually required rotations. Oswal never even achieves orthogonalization to better than one decimal digit. These failures are caused by the fact that in each instance the projection center and the three terrain points lie close to a plane.

The second strip is a strip flown in 1966 at an average height of 450 m above terrain

TABLE 1. INTERNATIONAL TEST, LINE 1. RMS OF THE RESIDUALS AFTER MODEL TRANSFORMATION, IN MICROMETERS AT APPROXIMATELY PHOTOGRAPH SCALE

	Via space-time transformation		*Via affine transformation*	
	Schut 1968 using centroids	*Schut 1972**	*Oswal*	*Blais*
Absolute orientations				
smallest RMS	14 μm	13 μm	33 μm	33 μm
largest RMS	40	40	341	337
mean RMS	28	27	137	136
Model connections				
smallest RMS	12	11	Results not acceptable,	
largest RMS	101	97	see the rotations	
mean RMS	54	52		

* These results are identical with those of minimization of residuals by least squares.

TABLE 2. INTERNATIONAL TEST, LINE 1. DEVIATION OF ROTATION FROM THAT OF LEAST-SQUARES ADJUSTMENT, IN GRADES

	Via space-time transformation		*Via affine transformation*	
	Schut 1968 using centroids	*Schut 1972*	*Oswal*	*Blais*
Absolute orientations				
smallest dev.	0.0023	0.00000044	0.0281*	0.0281
largest dev.	0.0169	0.00000616	0.293†	0.293
mean dev.	0.0084	0.00000309	0.122	0.122
Model connections				
smallest dev.	0.0017	0.00000008	Orthogonalization	4.8
largest dev.	0.0112	0.00000872	not achieved	192.
mean dev.	0.0067	0.00000272	to one dec. digit	54.

* Orthogonalization achieved to 5 decimal digits only.
† Orthogonalization achieved to 3 decimal digits only.

over the Sudbury test area of the National Research Council of Canada. A wide-angle camera with a focal length of 152 mm was used and the comparator measurements were made shortly afterwards. The 13 models contain from 3 to 6 ground control points located in a random pattern and with heights ranging from 262 to 294 m. For relative orientation and model connection, four points were chosen in each triple overlap. Two of these points are located near the principal point and they have the largest possible spacing in the strip direction. This spacing is about 20 percent of the spacing of the two wing points. Together with the common projection center, these points provide the best possible pattern for the use of Oswal's and Blais' methods in the model connections.

The results are summarized in Tables 3 and 4. Again, the use of the space-time transformation with the introduced modifications gives identically the same result as the least-squares minimization of residuals and the rotations differ less than $0.000,01^g$.

Oswal's and Blais' methods give greatly inferior solutions of the absolute orientation. In four instances where only three control points are available and in one instance where the four available control points happen to lie very close to a plane, they fail to give a solution. Their model connections are acceptable, but are clearly inferior to those obtained by the use of the space-time transformation.

Oswal proposes to make model connections without making use of the common projection center. In view of the narrow overlap between models, and apart even from the above results, it should be obvious that this cannot produce connections of an acceptable accuracy.

CONCLUSION

The experiments show that the computation of the parameters of a similarity transformation based on the space-time formulation of special relativity, as described in the present article, gives results that are indistinguishable from those obtained where the parameters are computed by least-squares

TABLE 3. SUDBURY, 1:3,000. RMS OF THE RESIDUALS AFTER MODEL TRANSFORMATION

	Via space-time transformation		*Via affine transformation*	
	Schut 1968 using centroids	*Schut 1972°*	*Oswal*	*Blais*
Absolute orientations				
smallest RMS	7 μm	4 μm	37 μm	37 μm
largest RMS	36	15	418‡	419‡
mean RMS	15	11	—	—
Model connections				
smallest RMS	15	13	22	22
largest RMS	88	85	111	111
mean RMS	43	38	59	59

° These results are identical with those of minimization of residuals by least squares.
‡ Excluding four models with three and one with four control points, where no results or inacceptable results were obtained.

TABLE 4. SUDBURY, 1:3,000. DEVIATION OF ROTATION FROM THAT OF LEAST-SQUARES ADJUSTMENT, IN GRADS

	Via space-time transformation		*Via affine transformation*	
	Schut 1968 using centroids	*Schut 1972*	*Oswal*	*Blais*
Absolute orientations				
smallest dev.	0.0008	0.00000002	0.0470°	0.0469
largest dev.	0.0393	0.00000132	0.439‡	0.440‡
mean dev.	0.0118	0.00000066	—	—
Model connections				
smallest dev.	0.0045	0.00000043	0.0090†	0.0090
largest dev.	0.0265	0.00000903	0.101°	0.101
mean dev.	0.0121	0.00000349	0.0448	0.0448

° Orthogonalization achieved to 5 decimal digits only.
† Orthogonalization achieved to 6 decimal digits only.
‡ Excluding four models with three and one model with four control points, where no results or inacceptable results were obtained.

minimization of the residuals at the control points.

If all control points are complete, this computation has the advantage over the least-squares minimization that it does not require initial estimates of the parameters although requiring the same minimum number of control points.

The computation of the parameters from an affine transformation gives identical results only in the instance of an ideal spatial distribution of the control points. Also, it requires one more control point.

The use of this computation in topographic applications of photogrammetry is not advisable. As Table 1 shows, the absolute orientation of models gives residuals that are several times larger than those obtained by the least-squares minimization even in the instance of large height differences of the control points. As Table 3 shows, even in the use of a wide overlap between models and the best possible spacing of the points in this overlap, the residuals are considerably larger than those obtained by the least-squares minimization. In many practical applications the distribution of the control points is less advantageous to this method and either no solution or no satisfactory solution is obtained.

REFERENCES

1. Blais, J. A. R., "Three-Dimensional Similarity," *The Canadian Surveyor,* Vol. 26, No. 1, March 1972.
2. Oswal, H. L. and S. Balasubramanian, "An Exact Solution of Absolute Orientation," *Photogrammetric Engineering,* Vol. 34, No. 10, October 1968.
3. Schreier, O. and E. Sperner, *Modern Algebra and Matrix Theory,* Chelsea Publishing Company, New York.
4. Schut, G. H., "On Exact Linear Equations for the Computation of the Rotational Elements of Absolute Orientation," *Photogrammetria,* Vol. 17, No. 1, 1960/1961.
5. Schut, G. H., "Formation of Strips from Independent Models," *Photogrammetric Engineering,* Vol. 34, No. 7, July 1968.