A method of obtaining the relative positions of 4 points from 3 perspective projections

H. Christopher Longuet-Higgins
University of Sussex, Brighton BN1 9QG

Abstract

According to Ullman's Structure-from-Motion Theorem [U79], three orthogonal projections of four points in a rigid non-planar configuration uniquely determine their structure, and the relative orientations of the three views, up to a reflection in the image plane. It is here shown that a corresponding result holds for the more general "para-perspective" case, and leads to a rapidly convergent algorithm for the fully perspective case. Unless the four points are nearly coplanar, or the images closely similar, the output of this algorithm is not unduly sensitive to errors in the image coordinates.

1 The orthogonal case

We adopt one of the points P_0 as origin and denote the 3D coordinates of the others by (X_n, Y_n, Z_n), (X_n', Y_n', Z_n') and (X_n'', Y_n'', Z_n'') in the three projection frames, (X_n, Y_n), $(X_n' Y_n')$ and (X_n'', Y_n'') being the (relative) image coordinates and Z_n, Z_n' and Z_n'' the (relative) depth coordinates. Then there will exist rigid rotation matrices $U = [u_{ij}]$ and $V = [v_{ij}]$ such that

$$X_n' = u_{11}X_n + u_{12}Y_n + u_{13}Z_n, \quad X_n'' = v_{11}X_n + v_{12}Y_n + v_{13}Z_n, \qquad (1',1'')$$

$$Y_n' = u_{21}X_n + u_{22}Y_n + u_{23}Z_n, \quad Y_n'' = v_{21}X_n + v_{22}Y_n + v_{23}Z_n, \qquad (2',2'')$$

$$Z_n' = u_{31}X_n + u_{32}Y_n + u_{33}Z_n, \quad Z_n'' = v_{31}X_n + v_{32}Y_n + v_{33}Z_n. \qquad (3',3'')$$

The problem is to find U and V and the depth coordinates from the three sets of image coordinates. As each rotation involves 3 unknown parameters, and there are 9 depth coordinates, we have 18 equations for only 15 unknowns, and may expect to encounter 3 consistency conditions, useful for checking purposes.

Elimination of Z_n between (1') and (2') gives the three equations

$$X_n u_{32} - Y_n u_{31} + X_n' u_{23} - Y_n' u_{13} = 0, \qquad (n = 1,\ 2,\ 3), \qquad (4)$$

BMVC 1991 doi:10.5244/C.5.12

from which one can obtain the ratios of the "border elements" $u_{32} : u_{31} : u_{23} : u_{13}$. (The computation fails if either (i) the four points are coplanar, in which case equations (4) are no longer independent, or (ii) the Z and Z' axes coincide, in which case $u_{33} = \pm 1$ and all four border elements vanish.) For U to be a rigid rotation of the first frame into the second, its border elements must satisfy

$$u_{13}^2 + u_{23}^2 \ (= 1 - u_{33}^2) = u_{31}^2 + u_{32}^2.$$

This is one of the three consistency conditions mentioned above, and can be checked as soon as the ratios $u_{32} : u_{31} : u_{23} : u_{13}$ have been obtained from (4) .

Introducing the normalized quaternion

$$\mathbf{Q} = \mathbf{i}p + \mathbf{j}q + \mathbf{k}r + s, \quad p^2+q^2+r^2+s^2 = 1, \tag{5}$$

related to U by the equation

$$U(\mathbf{Q}) = \begin{matrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{matrix} = \begin{matrix} p^2-q^2-r^2+s^2, & 2(pq-rs), & 2(pr+qs) \\ 2(pq+rs), & -p^2+q^2-r^2+s^2, & 2(qr-ps) \\ 2(pr-qs), & 2(qr+ps), & -p^2-q^2+r^2+s^2 \end{matrix} \tag{6}$$

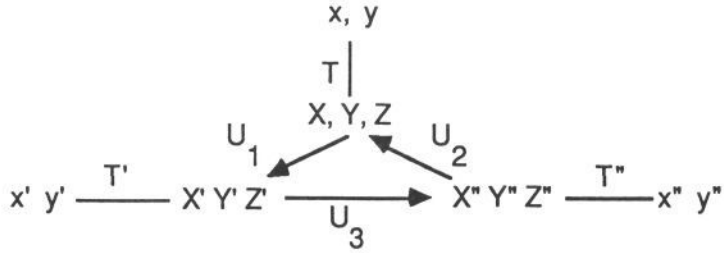
we see that the ratios $u_{32} : u_{31} : u_{23} : u_{13}$ determine the ratio of p to q and the ratio of r to s, but not the ratio of p to r. It follows that if **Q** is written in the parametric form

$$\mathbf{Q} = (\mathbf{i} \sin A + \mathbf{j} \cos A) \sin C + (\mathbf{k} \sin B + \cos B) \cos C, \tag{7}$$

then the two images (X_n, Y_n) and (X_n', Y_n') yield the values of the two parameters A and B, but not the "vergence" parameter C (equal to half of the angle between the Z and Z' axes). To compute C we need all three images, and the A and B parameters of the rotations connecting them, which we now denote by U_1 (= U), U_2 (= V^{-1}) and U_3 (= VU^{-1}), satisfying

$$U_1 U_2 = U_3^{-1}$$

(see figure at top of next page):



Armed with these parameters we substitute them in the parallel equation

$$Q_1Q_2 = Q_3^{-1}, \tag{8}$$

where

$$Q_1 = (\mathbf{i} \sin A_1 + \mathbf{j} \cos A_1) \sin C_1 + (\mathbf{k} \sin B_1 + \cos B_1) \cos C_1,$$

$$Q_2 = (\mathbf{i} \sin A_2 + \mathbf{j} \cos A_2) \sin C_2 + (\mathbf{k} \sin B_2 + \cos B_2) \cos C_2$$

and $$Q_3^{-1} = (-\,\mathbf{i} \sin A_3 - \mathbf{j} \cos A_3) \sin C_3 + (-\,\mathbf{k} \sin B_3 + \cos B_3) \cos C_3. \tag{9}$$

Using the rules of quaternion multiplication ($\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1$, $\mathbf{ij} = \mathbf{k} = -\mathbf{ji}$, etc.) we equate coefficients of **i**, **j**, **k** and unity on the two sides of (8), to obtain four equations (not given here), which we may call I, J, K and L . Elimination of $\sin C_3$ from I and J gives the ratio of $\tan C_1$ to $\tan C_2$, and elimination of $\cos C_3$ from K and L gives the product of $\tan C_1$ and $\tan C_2$. Eventually we obtain

$$\tan^2 C_1 = S_0S_1/S_2S_3, \quad \tan C_2/\tan C_1 = S_2/S_1 \quad \text{and} \quad \tan C_3/\tan C_1 = S_3/S_1, \tag{10}$$

where

$$S_0 = \sin (B_1 + B_2 + B_3), \qquad S_1 = \sin (B_1 - A_2 + A_3),$$

$$S_2 = \sin (B_2 - A_3 + A_1), \qquad S_3 = \sin (B_3 - A_1 + A_2). \tag{11}$$

These expressions for C_1, C_2 and C_3 in terms of the A's and B's enable us to determine $\mathbf{Q}_1$, $\mathbf{Q}_2$ and $\mathbf{Q}_3$, and from them the relative orientations U_1, U_2 and U_3. The relative depths Z_n then follow from (1') or (2'), though their absolute signs are subject to an overall "Necker" ambiguity associated with the arbitrary sign of $\tan C_1$ in equation (10).

2 The para-perspective case

Whereas in orthogonal projection the image coordinates are obtained by projecting 3D coordinates directly on to the (X, Y) plane, in perspective and paraperspective projection the equations for the image coordinates involve the distances T, T ' and T" of the three viewpoints from the reference point P_0. Without significant loss of generality we now assume that in each image P_0 lies on the optic axis of the camera; the plane projective image coordinates of the other 3 points are then

$$\begin{aligned} x_n &= X_n/(T + Z_n), & y_n &= Y_n/(T + Z_n), \\ x_n' &= X_n'/(T' + Z_n') & y_n' &= Y_n'/(T' + Z_n'), \\ x_n'' &= X_n''/(T'' + Z_n''), & y_n'' &= Y_n''/(T'' + Z_n''), \end{aligned} \tag{12}$$

where, as before, $X_n, \ldots, Z_n''$ are the 3D coordinates of P_n (relative to P_0) in the three frames. In the para-perspective or "small object" approximation one neglects the relative depths Z_n in these equations, and assumes that the image coordinates may be adequately approximated by the "reduced" 3D coordinates (note the reduced font size)

$$\begin{aligned} \mathrm{x}_n &= X_n/T, & \mathrm{Y}_n &= Y_n/T, & \mathrm{z}_n &= Z_n/T, \\ \mathrm{x}_n' &= X_n'/T', & \mathrm{Y}_n' &= Y_n'/T', & \mathrm{z}_n' &= Z_n'/T', \\ \mathrm{x}_n'' &= X_n''/T'', & \mathrm{Y}_n'' &= Y_n''/T'', & \mathrm{z}_n'' &= Z_n''/T''. \end{aligned} \tag{13}$$

This will be a good approximation so long as the relative depths of the points are small compared to tbeir distances from the three centres of projection. Substituting from (13) into (4) we obtain

$$\mathrm{x}_n T u_{32} - \mathrm{Y}_n T u_{31} + \mathrm{x}_n' T' u_{23} - \mathrm{Y}_n' T' u_{13} = 0, \quad (n = 1, 2, 3) \tag{14}$$

and these 3 equations yield the ratios $Tu_{32} : Tu_{31} : T'u_{23} : T'u_{13}$. *Assuming* the rigidity condition

$$u_{32}^2 + u_{31}^2 = u_{23}^2 + u_{13}^2 \tag{15}$$

makes it possible to calculate *both* $T : T'$ *and* $u_{32} : u_{31} : u_{23} : u_{13}$, from which the values of A and B follow. The gain in generality over the orthogonal case has been bought at the expense of a consistency check at this stage; but when a similar computation has been carried out for the other two rotations, U_2 and U_3, the ratios T' : T'' , T'' : T' and T : T' may be compared, to see whether their product is unity, as consistency demands. Thereafter, from the A and B parameters of U_1, U_2 and U_3 we can compute their C parameters, and hence the matrices themselves, as explained in the previous section. The final step is the calculation of the reduced depths. For each pair of images one obtains these from the relevant rotation U, using equations such as

$$T'x_n' = T(u_{11}x_n + u_{12}y_n) + u_{13}Z_n \tag{16}$$

$$T'y_n' = T(u_{21}x_n + u_{22}y_n) + u_{23}Z_n, \tag{17}$$

which are obtained directly from (1') and (2') by the para-perspective approximation. Multiplying (16) by u_{13}, (17) by u_{23} and adding the results we obtain eventually

$$z_n(1 - u_{33}^2) = (T'/T)(u_{13}x_n' + u_{23}y') + u_{33}(u_{31}x_n + u_{32}y_n). \tag{18}$$

3 The perspective case

The fact that (18) supplies values for the very quantities that are initially neglected in the para-perspective approximation raises the hope that one might be able, in favourable circumstances, to bootstrap one's way from the para-perspective to the fully perspective case. Having obtained provisional values of the reduced depths z_n, why not use them for recomputing the X_n and the Y_n, and feed the new values back into the original para-perspective computation?

With this idea in mind one rewrites (1') and (2') in terms of image coordinates, obtaining

$$(T' + Z_n')x_n' = (T + Z_n)(u_{11}x_n + u_{12}y_n) + u_{13}Z_n, \tag{19}$$

$$(T' + Z_n')y_n' = (T + Z_n)(u_{21}x_n + u_{22}y_n) + u_{23}Z_n. \tag{20}$$

Equations (19) and (20) immediately give the ratios of $(T' + Z_n')$, $(T + Z_n)$ and Z_n, and the reduced coordinates may then be recomputed from the relations

$$X_n = x_n(T + Z_n)/T, \quad Y_n = y_n(T + Z_n)/T \text{ and } z_n = Z_n/T. \tag{21}$$

With these new reduced coordinates — one set for each Necker alternative — one can return to equation (14) and recompute, first the A and B parameters and then the C parameters of the three rotation matrices. This time, however, the earlier choice of Necker alternative will affect, not only the absolute signs of the depths coordinates but their relative magnitudes as well, and also the magnitudes of the ratios T : T' etc. The product of these three ratios serves as a measure of the consistency of the chosen alternative with the three sets of image coordinates, and one will naturally prefer that alternative for which the product is closer to unity. Thereafter one may iterate cyclically through the various steps described until the process either converges or evidently fails to do so — because the viewpoints are too close to the object, the images too noisy or the views too similar.

4 Results

At the time of writing the only results available are those obtained by computer simulation. Each vector P_0P_n (n = 1, 2, 3) was assigned unit length and random direction, and the images were generated according to equations (12), with specified values of T, T' and T", and varying amounts of gaussian random noise. Typical runs of the relevant Pop-11 program are displayed on the next 2 pages. The function "twiddle" sets up a new configuration of 4 points, computes their coordinates in three randomly oriented frames, prints the triple product of the three unit vectors (a convenient measure of non-planarity) and the cosines of the angles between the three optic axes. The function "test", which takes the viewing distances T, T' and T" as parameters, prints the quaternion **Q** corresponding to U, computes the perspective images, contaminates them with noise of specified standard deviation "std" and prints, for "sgn" equal to +1 or –1, the computed value of (T/T')(T'/T")(T"/T) after each iterative cycle. It terminates as soon as (i) this value differs from unity by less than 0.00001 or (ii) 7 cycles have been completed, or (iii) the expression for $\tan^2 C_1$ is found to be negative, in which case the word "fail" appears, otherwise the current value of **Q**, as computed from the images.

Although in the orthogonal case the depth ambiguity is inescapable, in the perspective case the above method provides a way of resolving it, since different choices of "sgn" lead to different final solutions, only one of which satisfies the consistency condition. Informally speaking, if one adopts the wrong alternative the structure actually appears to deform as one views it from different angles—an effect which becomes more pronounced as the viewing distances are decreased.

```
: .twiddle;

-0.68415-0.235208-0.190722-0.286721

: ;;; No problems with this setup

: test(7,8,9);

-0.784  -0.055  -0.503   0.360

1.03764 0.983777 1.01396 0.992801 1.00379 0.997995 1.00106

-0.783  -0.055  -0.504   0.361

: ;;; Probably the correct Necker alternative; try the other:

: -sgn->sgn; test(7,8,9);

 -0.784  -0.055  -0.503   0.360

1.03764 1.11562 1.15003 1.18215 1.20755 1.23173 1.25887

  0.418   0.024  -0.788   0.450

: ;;; Obviously not as good. Come closer:

: -sgn->sgn; test(3,4,5);

-0.784  -0.055  -0.503   0.360

 1.09946 0.948895 1.1168 0.861048 1.17276 0.816809 fail

: ;;; Hardly surprising. Now for some noise

 0.001->std; test(7,8,9);

 -0.784  -0.055  -0.503   0.360
```

```
1.03169 0.977811 1.0086 0.987454 0.998203 0.992623 0.99552

-0.784  -0.055  -0.502   0.361

: ;;;Some typical image coordinates:

: x1.sh;

 -0.041   0.143  -0.050

: ;;; These have been disturbed in the 3rd decimal place,

: ;;; without serious effect on the solution; but

 ;;; the effects are more serious at long distances.

test(70,80,90);

 -0.784  -0.055  -0.503   0.360

1.22151 1.21287 1.21344 1.21341 1.21339 1.21338 1.21338

 -0.817   0.000  -0.417   0.398

: ;;; Another try:

: .twiddle;

-0.078179-0.158668 0.013091 0.334405

: ;;; Dangerously close to planar

: test(7,8,9);

 -0.543  -0.533  -0.496   0.417

fail
```

5 Discussion

It is evident from these sample results that the method is not guaranteed to converge on to the correct solution for any tetrahedron and any three viewpoints. But insofar as the orthogonal and paraperspective approximations are useful in the interpretation of image sequences it may also be useful to have a method of refining the crude estimates of structure and motion that result when the viewing distances are not very much larger than the separations between the points under inspection. In this respect the present work extends that of Tomasi and Kanade [TK91], who show how to decompose an essentially orthogonal image sequence into one matrix specifying the camera orientations and another encapsulating the structure of the object. In particular, the present procedure for determining U and V from three *orthogonal* projections supplies a simple and painless way of computing their 3x3 matrix A from three representative members of the sequence. In this connection it may be relevant to remark that the linear interdependence of 4 or more orthogonal projections of a rigid body carries over to the para-perspective approximation—a fact of considerable utility in the analysis of sequences of images at moderate viewing distances.

The main thrust of this work has been to make it possible to derive structure from motion without relying too heavily on the delicate "perspective effects" exploited in, for example [LH81]. To achieve this it is necessary, as Ullman [U79] and others have realized, to compare the images obtained from at least 3 sufficiently distinct viewpoints, and 4 is the minimum number of identifiable points to which the method can be applied. Fortunately, one consistency constraint survives the generalization from orthogonal to perspective projection, and this enables one to transcend Ullman's theorem and select the correct Necker alternative in a principled fashion.

Acknowledgment

Warmest thanks to Mike Brady for his invaluable encouragement and support.

References

[U79] Ullman, S. The Interpretation of Visual Motion. MIT Press, Cambridge Mass, 1979

[TK91] Tomasi, C. & Kanade, T. Shape and motion from image streams: a factorization method. 2. Point features in 3D motion. School of Computer Science, Carnegie-Mellon University, Pittsburgh, 1991

[LH81] Longuet-Higgins, H. C. The reconstruction of a scene from two projections. Nature 293: 133-135