On 3D Reconstruction from Two Perspective Views

Hans-Hellmut Nagel, Bernd Neumann

Fachbereich Informatik
Universitaet Hamburg
Schlueterstrasse 70
D-2000 Hamburg 13 / Germany

ABSTRACT

A concise derivation is given for a compact nonlinear equation which specifies possible 3D rotations of rigid objects, compatible with measurements of five object points in two views. The insight provided by direct geometrical interpretation of this equation = which contains ULLMAN's polar equation as a special case - may be exploited for attempts to categorize the set of possible solutions.

1 INTRODUCTION

Efforts towards an improved interpretation of image sequences from scenes with moving objects recently concentrated on approaches to derive a 3D description of rigid moving objects and their space trajectory. In the case of orthographic projection, the mathematical problems have been solved through the "structure-from-motion" theorem of ULLMAN [8]. In the case of perspective projection, various approaches have been used to derive equations for the unknown 3D point coordinates and motion parameters - see [1,2,4,5,6]. Usually minimization approaches are employed in order to obtain solutions. In special situations, some of the unknowns can be eliminated, yielding for example the "polar equation" in LB]. It can be shown that the image measurements of at least five points in at least two views are required in order to determine the remaining unknowns - see the references quoted above and [9]. For this basic situation, NAGEL [3] derived a compact equation for the unknown parameters specifying the rotation of the rigid object relative to the sensor between the first and second image frame. This equation can be written down in two lines rather than two pages required for a full specification of ULLMAN's "polar equation" which turns out to be a special case of it.

One of us (B.N.) found a very concise derivation of this equation and the ensueing discussion uncovered an immediate geometrical interpretation for it and - a fortiori - for ULLMAN's "polar equation". Based on these insights into the problem, it appears promising to investigate multiple interpretations which might satisfy the measurements in two views, especially degenerate ones which might upset minimization approaches.

II A COMPACT FORMULA FOR 3D RECONSTRUCTION

Let us assume that a certain number of perspective views of a rigid object can be obtained. Either object or observer or both may be in motion, so that in general each view will show a different aspect of the object. We shall further assume that a certain number of points rigidly fixed to the object can be traced through each view. Both, 3D-structure and motion of these points are unknown and unrestricted except of the rigidity constraint. In the following a vector equation will be derived relating the unknown rotation parameters to the observed object point coordinates.

The 3D-coordinates of the m-th object point Am (m = 1 ... M) are given with respect to an object coordinate system. For each frame time n = 1 ... N a translation vector Tn and a rotation matrix Dn relate the object coordinate system to a camera coordinate system whose origin coincides with the projection center of the camera. The image plane is parallel to the xz-plane of the camera system, at a distance f on the positive y-axis which represents the optical axis.

Let Cmn be the position vector of Am at time n in the camera system

$$Cmn = (Am + Tn)\ Dn$$

The projection of Cmn onto the image plane is given by a three-component image vector

$$Bmn = (f\ Cxmn/Cymn \qquad f\ Czmn/Cymn)$$

The x- and z-component of Bmn are known image coordinates. Carrying along the y-component Bymn = f permits a convenient vector notation of both image and spatial data, as will become evident soon. Similar to ROBERTS [7] we introduce an unknown scale factor $s_{mn} = C_{ymn}/f$ which lets us write

$$s_{mn} B_{mn} \quad (A_m + T_n)\ D_n \qquad (m = 1 \ldots M;\ n \quad N) \qquad (1)$$

Before combining several of these equations for the elimination of unknowns, two simplifications can be introduced. First, we are free to choose the object coordinate system to coincide with the camera system for $n = 1$. For convenience we set $T_1 = 0$ and $D_1 = 1$ (denoting the identity matrix).

Secondly, note that if s_{mn}, A_m, T_n and D_n constitute a solution for Eqs. 1, so do $q.s_{mn}$, $q.A_m$, $q.T_n$ and D_n, where q is an arbitrary factor. This, of course, is the well-known fact that the size of an object cannot be determined from any number of perspective views unless at least one absolute distance is known. While this prohibits exact 3D reconstruction, shape and apparent motion of a scaled variant can still be obtained. The scaling may be chosen by setting one of the factors s_{mn} to an arbitrary value, e.g. $s_{11} = 1$.

For a sequence of N frames at least $3 + 2/(2N - 3)$ points must be observed if the number of equations is to match or exceed the number of unknowns, see NAGEL 81. The following derivation will deal with two views of five points. Since due to the initial choice there is only one translation vector and one rotation matrix remaining (namely T_2 and D_2, resp.), the index will be dropped, and we have the following equations

$$s_{m1} B_{m1} = A_m \qquad (2)$$

$$s_{m2} B_{m2} = (A_m + T)\ D \qquad (3)$$

Substituting (2) into (3) we get

$$s_{m2} B_{m2} = (s_{m1} B_{m1} + T)\ D \qquad (4)$$

and for a particular point, say $m = 1$,

$$s_{12} B_{12} = (s_{11} B_{11} + T)\ D \qquad (5)$$

Subtracting (5) from (4) we get for $m > 1$

$$s_{m2} B_{m2} - s_{12} B_{12} = s_{m1} B_{m1} D - s_{11} B_{11} D \qquad (6)$$

To eliminate s_{m2} we take the vector product of both sides with B_{m2} and get

$$-s_{12}(B_{12} \times B_{m2}) = s_{m1}(B_{m1}D \times B_{m2}) - s_{11}(B_{11}D \times B_{m2}) \qquad (7)$$

Next, we eliminate s_{m1} by taking the scalar product with $D'B_{m1}{}^*$ (the apostroph denotes transpose).

$$-s_{12}(B_{12} \times B_{m2})D'B_{m1} \quad -s_{11}(B_{11}D \times B_{m2})D'B_{m1}'$$

A product $(A \times B)C^*$ can be interpreted as the volume of the parallelepiped formed by the 3 vectors A, B and C. Thus simultaneous rotation does not affect the product value and interchanging the order of the three vectors may only introduce a sign change. This allows us to write

$$s_{12}(B_{m1} \times B_{m2}D')DB_{12}' = s_{11}(B_{m1} \times B_{m2}D')B_{11}' \qquad (8)$$

and in particular for $m = 2$

$$s_{12}(B_{21} \times B_{22}D')DB_{12}' = s_{11}(B_{21} \times B_{22}D')B_{11}' \qquad (9)$$

Combining (8) and (9) we get

$$(B_{21} \times B_{22}D')B_{11}' \ (B_{m1} \times B_{m2}D')DB_{12}' = (B_{21} \times B_{22}D')DB_{12}' \ (B_{m1} \times B_{m2}D')B_{11}' \qquad (10)$$

For $m = 3,4,5$ this is a set of three equations for the three free parameters of the rotation matrix D. It contains the polar equation of ULLMAN [8] as a special case (see [3]). This rather compact equation will be the starting point for some interesting insights.

III GEOMETRICAL INTERPRETATION

Before entering a discussion concerning properties and applicability of Eq. 10 it may be helpful to gain some understanding of its meaning. First, it can be seen that the equation deals solely with vector <u>directions</u>. The magnitudes cancel out since each vector occurs in a product exactly once on each side. This is quite sensible since the directions of the projecting rays constitute the only information which can be exploited.

Note also that each image vector at time 2 is associated with the inverse rotation matrix, undoing the effect of rotation. Whatever difference remains between the directions of B_{m1} and $B_{m2}D'$ is caused by the translation vector T which must be coplanar with all such vector pairs. Two pairs give a solution for the direction of T in terms of the intersecting line of their planes for any choice of D. The planes of three pairs, however, will only intersect in a line if D is chosen properly. It will be shown now that this constraint is indeed expressed through Eq. 10 by rederiving it with geometrical arguments.

T must be coplanar with B_{m1} and $B_{m2}D'$ for all points m. Hence each vector product

$$(B_{m1} \times B_{m2}D^*) \qquad m = 1,\ 2,\ . \qquad (11)$$

defines a vector normal to a plane containing T. The planes for $m=1$ and $m=2$ intersect in a line whose direction is given by

$$(B_{11} \times B_{12}D') \times (B_{21} \times B_{22}D') \qquad (12)$$

This line must be oriented in the same direction as T. The plane defined by Bml and Bm2D' for any other object point must be compatible with this direction, i.e. its normal vector on T must be normal to the direction of T defined by the two other points. This can be expressed using the Cartesian product.

$$C(B11 \times B12D') \times (B21 \times B22D')] (Bm1 \times Bm2D')' = 0$$

The multiple vector product in brackets can be rewritten using (AxB)xC = B(CA)-A(CB). Rearranging terms gives Eq. 10. In summary, we have shown that this equation expresses a constraint on D arising from the condition that the translation vector must be coplanar with each vector pair Bml and Bm2D'.

IV DISCUSSION

The nonlinear system of Eqs. 10 for m = 3,4,5 can be transformed into a set of fourth order polynomials in three unknowns. As one possible solution approach iterative procedures may be employed. A linearized version of,' Eq. 10 has been proposed in [3] and has been used successfully to solve for the unknowns based on simulated image vectors.

Let us assume that a rotation matrix D satisfying Eq. 10 has been found. How can one obtain a 3D reconstruction? From the preceding section we know that a solution indicates: there exists a translation T compatible with the observations and this particular D. The translation vector can be determined by first solving Eq. 9 for s12 (using an arbitrary choice of s11 and the solution for D) and then solving Eq. 5 for T. Eq. 4 will give sml and sm2 for all other points. The 3D coordinates, finally, follow from Eq. 2.

Note that Eq. 9 leaves s12 and hence T undefined if the triple product on both sides is zero. This can happen in two cases, (I) when the translation vector T is coplanar with BIl and B21 and hence all vectors in Eq. 9 are coplanar or (ii) when B21 and B22D' are collinear. The first case can be remedied by choosing another pair of points i and j such that Bil, Bjl, and T are not coplanar. If all Bml are coplanar with T, none of them can be reconstructed from two views. The second case implies that the translation vector is zero or collinear with B21. Choosing another point will reveal whether T is zero since it cannot be simultaneously collinear with two image vectors in different directions. For zero translation (i.e. pure rotation about the origin) none of the scaling factors smn can be determined, hence no reconstruction iS possible. If T is nonzero and collinear with some Bml then only this point cannot be reconstructed. Hence tor all but certain exceptional situations a unique 3D reconstruction can be obtained for any D satisfying Eq. 10.

V CONCLUSION

The interpretation of image sequences recorded from real-world scenes with moving objects would be incomplete without a description of the 3D configuration of points on a rigid moving object and its space trajectory. To solve such a problem requires the image measurements from at least five object points in at least two views. Solutions could be used as start values for a minimization approach in the case where more than five points in more than two views are available and have to be explained as a single rigid object. Using appropriate choices for the available degrees of freedom, a compact equation for the unknowns remaining in this basic situation has been derived in [3]. Our current contribution has presented a much more direct derivation and an immediate geometrical interpretation of this equation, also some results concerning degenerate situations.

REFERENCES

[1] Gennery, D.B. Stereo-Camera Calibration, Proc. IUW November 1979, 101-107

[2] Lawton, D.T. Constraint-Based Inference from Image Motion, Proc. AAA1-80, 31-34

[3] Nagel, H.-H. On the Derivation of 3D Rigid Point Configurations from Image Sequences, Proc. PRIP-81

[4] Prazdny, K. Motion and Structure from Optical Flow, Proc. IJCAI-79, 702-704

[5] Prazdny, K. Egomotion and Relative Depth Map from Optical Flow, Biological Cybernetics 36 (1980) 87-102

[6] Roach, J.W. and Aggarwal, J.K. Determining the Movement of Objects from a Sequence of Images, IEEE Trans. PAM1-2 (1980) 554-562

[7] Roberts, L.G. Machine Perception of Three-Dimensional Solids, in: Optical and Electro-Optical Information Processing [Tippet et al., eds.) MIT Press, Cambridge/MA, 1965, 159-197

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

[9] Zvi Meiri, A. On Monoccular Perception of 3-D Moving Objects, IEEE Trans. PAMI-2 (1980) 582-583