[30] Double and Triple Ambiguities in the Interpretation of Two Views of a Scene[1]

H C Longuet-Higgins

Centre for Research in Perception and Cognition
University of Sussex, Brighton, BN1 9QG, UK

It is known (Longuet-Higgins 1984a and 1984b) that if two spherical images of a visually textured surface, from finitely separated viewpoints, allow more than one 3D interpretation, then the surface must be part of a quadric passing through the two viewpoints. It is here shown that this quadric is either a plane or a ruled surface of a type first considered by Maybank (1985) in a study of ambiguous optic flow fields; and that three is the maximum number of interpretations that the two images can sustain. An explanation is offered for the fact that two images of just 5 points generally permit two distinct 3D interpretations.

1 INTRODUCTION

A much discussed problem in computer vision (Hay 1966; Ullman 1979; Tsai and Huang 1981a, 1981b; Longuet-Higgins 1984a, 1984b; Maybank 1985; Subbarao 1986) is the degree of ambiguity of a pair of 2D projections of a scene; under what circumstances are they susceptible of more than one interpretation, and what is the maximum number of interpretations that can be entertained in exceptional circumstances? The problem arises when a camera is in motion through a rigid scene (Lee 1974; Nakayama and Loomis 1974; Koenderink and van Doorn 1976) and it is required to determine the notion of the camera, and the structure of the scene, from a pair of photographs taken one after the other (Longuet-Higgins 1981). If the time interval between them is very short, the problem reduces to that of determining the linear and angular velocity of the camera from an "optic flow field" (Longuet-Higgins and Prazdny 1980); a recent paper by Maybank (Maybank 1985) gives a definitive account of the ambiguities that can arise in that case. The present paper extends Maybank's results to the case in which the two camera positions are finitely separated.

The main existing results on the ambiguity of pairs of photographs taken from discretely separated viewpoints are as follows:

(a) If 8 or more points in the scene appear in both photographs then in general their 3D locations are unambiguously determined (Longuet-Higgins 1981).

(b) If all the points that appear in both photographs happen to lie in a plane, there will be two distinct interpretations (Hay 1966), or just one if the line joining the viewpoints is perpendicular to the plane.

(c) An algorithm that yields the unique interpretation of two views of 8 points in general position fails if and only if all the visible points of the scene lie on a quadric surface that passes through the two viewpoints (Longuet-Higgins 1984).

In this paper it will be shown that

(d) even when the scene is a visually textured quadric surface passing through the two viewpoints, the two images will generally possess a unique interpretation, but

(e) if this quadric is of a special type, first considered by Maybank(1985), then there may be two or even three distinct interpretations, but there cannot be more than three.

2 FORMULATION OF THE PROBLEM

The scene is supposed to consist of a set of visual texture elements disposed on a rigid surface. The two camera positions are denoted by O and O'. A typical element P is situated at vector position pQ relative to O, p being the distance OP and Q being a unit vector, which may be thought of as a point on the unit sphere centred at O (Hadani, Ishai and Gur 1980; Yen and Huang 1983). Relative to the second camera position the vector position of P is $p'Q'$, where Q' is a point on the unit sphere round O' and p' is the distance $O'P$. So if T and U^{-1} are the translation and rotation that carry the camera from O to O', then

(1) $p'Q' = U(pQ-T)$.

Equation (1) is to be regarded as a mapping between the space of the vectors Q, associated with the position O, and the space of the Q', associated with O'. In actual computations these vectors are represented by cartesian coordinates (x,y,z) or (x',y',z'), with squares adding up to 1. The translation T is taken to be a unit vector - a convention that fixes the otherwise indeterminate scale of distance - and the rotation U is represented as a proper orthogonal 3x3 matrix (one whose reciprocal U^{-1} equals its transpose $U^{\sim}$, and whose determinant equals unity).

In the coordinate system of the second viewpoint, $p'Q'$, pUQ and UT are vectors describing the three sides of the triangle $O'PO$. It follows at once that their triple product vanishes, and that

(2) $[Q', UQ, UT] = O$

- a relation known as "the epipolar constraint". It is the validity of (2) for all pairs of image points (Q,Q'), that

---

makes it possible, under favourable circumstances, to compute T and U and the structure of the scene from the two images alone.

There are, however, certain pairs of images that admit of two or more interpretations, in the sense that all the pairs (Q,Q') satisfy two distinct equations of type (1), namely

(3) $p_1'Q' = U_1(p_1Q-T_1)$ and

(4) $p_2'Q' = U_2(p_2Q-T_2)$.

In (3) and (4) the subscripts 1 and 2 refer, of course, to the two interpretations; in each interpretation p is a function of Q and p' is a function of Q':

(5) $p_1 = p_1(Q)$, $p_1' = p_1'(Q')$,
     $p_2 = p_2(Q)$, $p_2' = p_2'(Q')$.

An ambiguous pair of images therefore satisfies the identity

(6) $Q' = U_1(p_1Q-T_1)/p_1' = U_2(p_2Q-T_2)/p_2'$.

The second equality in (6) implies that, for every image point Q, $U_1(p_1Q-T_1)$ is a linear combination of U_2Q and U_2T_2; and it entitles us to infer that

(7) $[U_1(p_1Q-T_1), U_2Q, U_2T_2] = 0$.

Multiply the second term in (7) by p_1, and abbreviating the vector p_1Q as R, we arrive at the equation

(8) $[U_1(R-T_1), U_2R, U_2T_2] = 0$.

The triple product on the left hand side is clearly a second-order polynomial in (X, Y, Z), the components of R, and so (8) is the equation of a quadric passing through the points O and O', where R = 0 and R = T_1 respectively. But not every quadric passing through 0 and 0' can be represented in the form (8), since for given T_1 and U_1 this form has only 5 degrees of freedom (2 for the unit vector T_2 and 3 for the rotation matrix U_2). We deduce that although twofold ambiguity is quite likely to arise (and usually does - T S Huang, personal communication) if not more than 5 texture elements can be identified in both photographs, with 6 or more elements the two images will generally permit only one interpretation even when all the elements lie on a quadric passing through both 0 and 0'.

Equation (7) may be written as an explicit equation for $p_1(Q)$:

(9) $p_1(Q) = [U_1T_1, U_2Q, U_2T_2]/[U_1Q, U_2Q, U_2T_2]$.

Interchanging the subscripts 1 and 2 we deduce that on the other interpretation the equation for p(Q) is

(10) $p_2(Q) = [U_2T_2, U_1Q, U_1T_1]/[U_2Q, U_1Q, U_1T_1]$.

3 MULTIPLE AMBIGUITY

In his discussion of optic flow fields Maybank established (Maybank 1985) that 3 is the maximum number of distinct alternative interpretations of such a field, each being associated with distinct values of the camera's angular velocity and direction of motion. We shall show that the same is true of a pair of finitely separated projections.

If there are two interpretations of a pair of images, $p_1(Q)$ must satisfy (9); if there are 3, it must also satisfy

(11) $p_1(Q) = [U_1T_1, U_3Q, U_3T_3]/[U_1Q, U_3Q, U_3T_3]$.

It is the necessary equivalence of (9) and (11) that forms the basis of the following discussion.

The numerators of (9) and (11) are first order polynomials in the components of Q, and the denominators are polynomials of the second order. The most straightforward case is that in which each numerator divides its denominator algebraically, so that both (9) and (11) reduce to

(12) $p_1(Q) = 1/(N.Q)$.

This is the equation of a plane, the vector N being the inverse normal to the plane. The potential ambiguity of a pair of views of a plane has been fully discussed elsewhere (Longuet-Higgins 1984b), so we shall confine ourselves from now on to the case in which $p_1(Q)$ is not a plane, and the numerators in (9) and (11) do not divide their denominators. It follows at once that the numerators and the denominators in (9) and (11) are directly proportional - that there exists a constant c such that

(13) $[U_1Q, U_2Q, U_2T_2] = c[U_1Q, U_3Q, U_3T_3]$ and

(14) $[U_1T_1, U_2Q, U_2T_2] = c[U_1T_1, U_3Q, U_3T_3]$.

Though it is by no means obvious, the insertion of arbitrary values of U_1, U_2 and U_3 into (13), and subsequent comparison of the polynomial coefficients, determines the magnitude of c and the directions of T_2 and T_3. (The signs of the translation vectors cannot be determined until later, when the values of the distances p(Q) and p'(Q') are being computed; for each of the three interpretations of the sign of T must be such as to make all these distances positive, and if this is not possible the interpretation fails.) T_1 is then determined (again with unknown sign) by inserting the values of the other parameters into (14).

The detailed justification of these assertions will be given elsewhere; here we give the results of just one such computation, illustrating the fact that the three interpretations of a triply ambiguous pair of images may be uncomfortably close together.

First, the Q vectors of 5 points in the first image:

0.396	-0.172	0.902
0.180	0.438	0.881
0.171	0.371	0.913
-0.118	0.061	0.991
-0.272	0.164	0.948

Next, the Q' vectors of the corresponding points in the other image:

0.322	-0.200	0.925
0.172	0.411	0.895
0.162	0.342	0.926
-0.062	-0.011	0.998
-0.204	0.100	0.974

Finally, the vector T, the matrix U and the p values of the 5 points, in the 3 distinct interpretations:

Interpretation 1

-0.239	0.541	-0.807

0.999	0.019	-0.035
-0.015	0.996	0.092
0.036	-0.091	0.995

2.637	6.782	6.227	2.779	3.477

Interpretation 2

-0.846	0.435	0.310

0.995	0.056	-0.083
-0.056	0.998	0.001
0.082	0.004	0.997

62.751	21.478	19.194	6.095	5.405

Interpretation 3

-0.830	0.401	0.387

0.995	0.048	-0.088
-0.048	0.999	-0.002
0.087	0.006	0.996

52.076	18.157	16.592	5.813	5.066

4 GEOMETRICAL CONSIDERATIONS

In this section we briefly review what is known about triply ambiguous flow fields, and show that the surfaces from which they arise are of the same type as those that give rise to triply ambiguous image pairs. We conclude that the number of distinct interpretations of two views of a textured surface cannot exceed 3.

In his 1985 paper Maybank showed that ambiguous flow fields can only arise from planes or quadric surfaces of a special form, namely

(15) $M: [v', v, R] = (W'.R)(v'.R) - (W'.v')R^2$,

where (v, Ω) and (v', Ω') are alternative values of the camera's linear and angular velocity,

(16) $W' = \Omega' - \Omega$ and

(17) $R = pQ = (X, Y, Z)$.

The quadric M has several interesting properties: (a) it passes through the viewpoint R = 0; (b) its tangent at the viewpoint is the common plane of v and v' (since when R is small the triple product [v', v, R] is very small); (c) it contains the line $R = \lambda v'$ (such a value of R causing both sides of (15) to vanish); and (d) its quadratic part has a specially simple diagonal form. If the X and Z axes are taken as the internal and external bisectors of the angle 2ϕ between W' and v', the right hand side of (15) becomes

(18) $k[s^2 X^2 + (s^2 - c^2)Y^2 - c^2 Z^2]$

where k is the product of the lengths of W' and v', $c = \cos\phi$ and $s = \sin\phi$. Since the line $R = \lambda v'$ lies entirely in M, M must be a ruled quadric, the most general such surface being a hyperboloid of one sheet. A hyperbolic paraboloid is also a possibility (if $c^2 = s^2$), but in either case the middle coefficient in (18) is the sum of the other two.

We shall refer to the directions of W' and v' as the principal directions of the quadric M. A triply ambiguous flow field arises if in addition to (v, Ω) and (v', Ω') there exists a third pair of velocities (v'', Ω'') such that $W'' (= \Omega'' - \Omega)$ is parallel to v' and W'' is parallel to v''. Then the principal directions of M can, as it were, exchange roles, the one that was parallel to W' now being regarded as parallel to v'', and the one that was parallel to v' being seen as parallel to W''. It is, essentially, this duality that limits the number of interpretations; given the "correct" interpretation (v, Ω) of two views of M, the two other interpretations, (v', Ω') and (v'', Ω''), exhaust the possible ways of associating a linear velocity and an angular velocity difference with the principal directions of M.

At first sight equation (8), describing the type of surface that gives rise to ambiguous pairs of views, looks rather different from equation (18), for the optic flow case. But as we shall see in a moment, there is a close relation between the surfaces they represent.

We begin by writing (8) in the form

(19) $L: [U(R-T_1), R, T] = 0$,
    where $U = (U_2)^{-1} U_1$ and $T = T_2$.

Like Maybank's quadric M, the surface L is a ruled quadric containing the viewpoints R = O and R = T_1 and the straight line R = λT. We now show that the second-order terms of L are identical in form with those of M.

Writing R = (X, Y, Z), and using lower-case letters to denote the 3 components of T and the 9 components of U, we begin by expanding the second-order part of (19) in the form

(20) $[UR, R, T] = X^2(u_{31}t_2 - u_{21}t_3)$
      $+ Y^2(....) + Z^2(....)$
      $+ YZ(u_{22}t_1 - u_{12}t_2 + u_{13}t_3 - u_{33}t_1)$
      $+ ZX(....) + XY(....)$

where the dots indicate that the subscripts 1, 2 and 3 have been cyclically permuted. In order to proceed we need a parametric representation of the elements of U. A convenient one for the present purpose is in terms of 3+1 real numbers p, q, r and s whose squares add up to 1:

(21) $u_{11}=p^2-q^2-r^2+s^2$, $u_{12}=2(pq - rs)$, $u_{13}=2(rp + qs)$,
$u_{21}=2(pq + rs)$, $u_{22}=-p^2+q^2-r^2+s^2$, $u_{23}=2(qr-ps)$,
$u_{31}=2(rp -qs)$, $u_{32}=2(qr+ps)$,
$u_{33}=-p^2-q^2+r^2+s^2$.

(In point of fact $s = \cos(\psi/2)$, where ψ is the rotation angle of U and (p, q, r) are the direction cosines of the rotation axis, multiplied by $\sin(\psi/2)$.) Substituting from (21) we obtain the coefficient of X^2 in (17) as

(22) $2(pr - qs)t_2 - 2(pq + rs)t_3$

and that of YZ as

(23) $2(q^2-r^2)t_1 - 2(pq - rs)t_2 + 2(pr - qs)t_3$,

with analogous expressions for the other coefficients.

Defining three new vectors

(24) $u = (p, q, r)$, $v = u \times T$, $w = sT - v$,

we obtain, after some algebra,

(25) $[UR, R, T] = 2[(u.R)(w.R) - (u.w)R^2]$.

The right hand side of (25) is identical in form with that of (15), showing that L as defined by (19) is indeed a Maybank quadric.

To recapitulate: we showed in section 2 that in order to present an ambiguous pair of views from camera positions related by the relative orientation (T_1,U_1), a surface must be of the form (8) or equivalently (19). The second-order terms in its equation are of the same form as those of the Maybank quadric (15); the principal directions - those of u and w - are functions of U_1 and of the translation T_2 and the rotation U_2 associated with the alternative interpretation.

In the triply ambiguous case the same functions of U_1, T_3 and U_3 must also yield the principal directions of the same quadric; this is only possible if u' is parallel to w and w' is parallel to u, and leaves no room for any further distinct interpretation. Three is therefore the maximum number of distinct interpretations of two views of a visually textured surface.

5   DISCUSSION

What we have shown is that the existence of two alternative interpretations of a pair of views of a visually textured surface implies that the surface is either a plane or a quadric of the form (8). Given any "true" relative orientation (T_1, U_1) and any "spurious" one (T_2, U_2), one can construct a quadric of type (8) such that its two images will sustain either of the associated interpretations. This quadric passes through both viewpoints and contains the line $R = \lambda T_2$; it is, in fact, a surface of the type first considered by Maybank in connection with the interpretation of optic flow fields. The spurious interpretation will not necessarily satisfy the visibility conditions - that p(Q) and p'(Q') are both positive for all the image points. But the more nearly equal are the two relative orientations, the greater the likelihood that the alternative interpretation will survive the visibility test.

Three is the maximum number of distinct interpretations of a pair of views of a surface patch. Triply ambiguous view pairs may be constructed by assigning arbitrary values to the three associated rotation matrices; the corresponding translation vectors are then uniquely determined, as well as the three alternative surfaces on which the visible points may be deemed to lie.

Apart from their purely mathematical interest, these results have both a reassuring and a disturbing aspect for the designers of computer vision systems. Reassuring, in that they demonstrate the existence of an upper limit to the number of interpretations that two images will sustain if a sufficient number of visible texture elements (5 or more, in general) appears in both images; disturbing, in that they remind us of the untrustworthiness of vision algorithms based on the implicit assumption that there must be a single "best" interpretation of any given set of visual data.

Perhaps the most useful fact to emerge from the present analysis is that the most hazardous scenes for computational analysis are those in which all the visible texture elements lie in one smooth surface. The simplest scenes are often the most perceptually confusing!

REFERENCES

Hadani, I., Ishai, G. and Gur, M. (1980) Visual Stability and space perception in monocular vision: a mathematical model. *J. Opt. Soc. Amer.*, **70**, 60-65.

Hay, J.C. (1966) Optical motions and space perception: an extension of Gibson's analysis. *Psychological Review*, **73**, 550-565.

Koenderink, J.J. and van Doorn, A.J. (1976) Local structure of movement parallax of the plane. *J. Opt. Soc. Amer.*, **66**, 717-723.

Lee, D.N. (1974) Visual information during locomotion. *Perception: Essays in honour of James J. Gibson* (eds. R. B. MacLeod and H.L. Pick), 250-267, Cornell Univ. Press, Ithaca N.Y.

Longuet-Higgins, H.C. and Prazdny, K. (1980) The interpretation of a moving retinal image. *Proc. Roy. Soc. Lond. B 208*, 385-397.

Longuet-Higgins, H.C. (1981) A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**, 133-135.

Longuet-Higgins, H.C. (1984a) The reconstruction of a scene from two projections: configurations that defeat the 8-point algorithm. *IEEE: Proceedings of the first conference on artificial intelligence applications*, 395-397.

Longuet-Higgins, H.C. (1984b) The visual ambiguity of a moving plane. *Proc. Roy. Soc. Lond. B 223*, 165-175.

Longuet-Higgins, H.C. (1986) The reconstruction of a plane surface from two perspective projections. *Proc. Roy. Soc. Lond B 227*, 399-410.

Maybank, S.J. (1985) The angular velocity associated with the optical flow field arising from motion through a rigid environment. *Proc. Roy. Soc. Lond. A 401*, 317-326.

Nakayama, I. and Loomis, J.M. (1974) Optical velocity patterns, velocity sensitive neurons and space perception. *Perception*, 3, 63-80.

Subbarao, M. (1986) Interpretation of image motion fields: rigid curved surfaces in motion. *Research report CAR-TR-199*, Centre for Automation Research, Univ. of Maryland.

Tsai, R.Y. and Huang, T.S. (1981a) Estimating three-dimensional motion paramaters of a rigid planar patch. *Technical report R-922*, Coordinated Science Laboratory, Univ. of Illinois, Urbana.

Tsai, R.Y. and Huang, T.S. (1981b) Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *Technical report R-921*, Coordinated Science Laboratory, Univ. of Illinois, Urbana.

Yen, B.L. and Huang, T.S. (1983) Determining 3D motion and structure of a rigid body using the spherical projection. *Computer Vision Graphics and Image Processing.* 21, 21-32.