Proc. R. Soc. Lond. B **208**, 385–397 (1980)
Printed in Great Britain

The interpretation of a moving retinal image

BY H. C. LONGUET-HIGGINS, F.R.S.,† AND K. PRAZDNY‡

† *Centre for Research in Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, U.K.*
‡ *Department of Computer Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.*

(*Received* 1 *November* 1979)

It is shown that from a monocular view of a rigid, textured, curved surface it is possible, in principle, to determine the gradient of the surface at any point, and the motion of the eye relative to it, from the velocity field of the changing retinal image, and its first and second spatial derivatives. The relevant equations are redundant, thus providing a test of the rigidity assumption. They involve, among other observable quantities, the components of shear of the retinal velocity field, suggesting that the visual system may possess specialized channels for computing these components.

1. INTRODUCTION

When the eye is in motion relative to the visible environment, a moving pattern of light falls upon the retina, and the resulting 'optical flow field' supplies useful information not only about the motion but also about the three-dimensional structure of the scene (Helmholtz 1925; Gibson 1950, 1966, 1979; Gibson *et al.* 1955, 1957, 1959; Braunstein 1976). This information is inadequate, in general, for establishing the relative distances and velocities of all of the visible elements in the scene, if only because the line-of-sight velocity of a point source makes no difference to the retinal velocity of its image. But in practice the scene will usually consist of rigid objects of finite extent, and, if this condition is satisfied, the assumption of local rigidity can lead to a unique, and correct, three-dimensional interpretation of the moving retinal image (Ullman 1979).

There have been two somewhat different approaches to the interpretation of visual motion. One is based on an analogy with stereopsis, and the other appeals to the existence of receptors that respond to visual stimuli moving across the retina with specific velocities (Hubel & Wiesel 1968; Bridgeman 1972; Grusser & Grusser 1973; Sekuler & Levinson 1974, 1977). In the quasi-stereoscopic theory the visual problem is seen as that of collating the information from two or more discrete views of the scene. The first stage is to solve the 'correspondence problem' (Marr & Poggio 1976; Ullman 1979), that of establishing which elements in each image correspond to the same element in the scene; the second stage is to compute the structure from the finite disparity between the retinal positions of corresponding elements (Ullman 1979). The other approach (Gibson *et al.* 1955; Gordon 1965; Lee 1974; Koenderink & van Doorn 1976) takes as given the optical flow field itself, and attempts to infer the relative motion and the three-dimensional structure on the assumption that the scene is indeed locally rigid.

Whichever approach one decides to adopt, there is no great problem in calculating how the retinal image changes when the eye moves in a given manner relative to a scene of specified geometry. The real difficulties begin when one addresses the converse problem, that of proceeding from the optical flow field to conclusions about the motion and the structure. The difficulties are not severe if one assumes that the motion has no rotational component (Gibson *et al.* 1955; Lee 1974) or that its translational component is known (Koenderink & van Doorn 1976); but it is by no means obvious whether the visual system could, in principle, compute both the translational and the rotational motion of the eye relative to the scene, from the optical flow field alone. This is the problem that we consider in the present paper.

The plan of the paper is as follows. In §2 we examine the form of the optic flow field due to arbitrary motion relative to a rigid scene. In the most general case the field is found to be the vector sum of a translational component and a rotational component. The translational velocity at any point is directed towards or away from a unique 'vanishing point' determined by the relative translational motion. The rotational velocity field is fully determined by the angular velocity of the eye relative to the environment; it is entirely independent of the structure of the scene. Motion parallax cues (Helmholtz 1925), when they are available, are shown to provide a means of calculating both the translational and the rotational component of the relative motion; once this has been achieved, the structure of the scene can be fully determined from the translational component of the flow field.

In §3 we consider the case of a scene consisting of arbitrarily moving rigid objects with smooth, densely textured surfaces. The problem is to determine the translational and rotational motion of a given object, and the gradient of its surface at any point (Marr 1976), from the optic flow field due to the nearby texture elements. It is shown that all these unknowns may be computed from the field and its first and second spatial derivatives at the corresponding point on the retina. The first derivatives may be expressed in terms of the invariants discussed by Koenderink & van Doorn (1976), but the second derivatives are also needed for a full determination of the relative motion. They supply, incidentally, a check on the assumption of local rigidity. Plane surfaces are found to present special problems of visual interpretation.

Finally, in §4, we raise the possibility that the visual system may possess receptors that respond specifically to local deformations of the optic flow field, due to *relative* motion of neighbouring elements of the retinal image.

2. Motion through a static environment

To gain a qualitative insight into the form of the optic flow field it is helpful to think of the eye as a hemispherical pinhole camera, in arbitrary motion through a static environment. The motion at any instant may be resolved into two components: the translational velocity of the pinhole relative to the scene, and

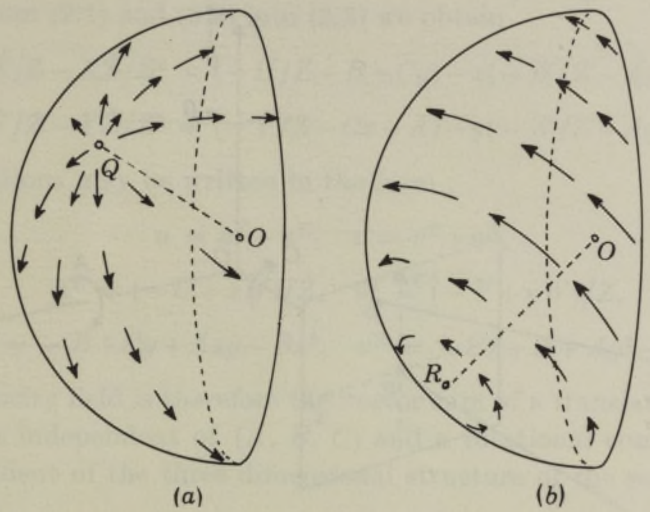


FIGURE 1. Typical translational ('polar') and rotational ('axial') components of the flow field on a hemispherical retina.

the angular velocity of the hemisphere about the pinhole, also measured relative to the scene. In purely translational motion (for which the angular velocity vanishes) the optic flow takes on a specially simple form: if the line of motion of the pinhole O intersects the hemisphere at the point Q, then every other image point on the hemisphere will move along the great circle that joins it to the point Q (Nakayama & Loomis 1974). So if Q is regarded as a 'pole' on the hemisphere, a purely translational field is one for which the image velocity is everywhere directed along 'lines of longitude', having a magnitude that depends on the detailed geometry of the scene (see figure 1*a*).

If the hemisphere is rotating as well as translating, then every image point will acquire an additional velocity component corresponding to a rigid rotation of the hemisphere about some radius OR. The most general flow field due to motion through a static environment is thus the vector sum of a 'polar' field due to the translation, and an 'axial' field due to the rotation; but there is no relation, in general, between the directions OQ and OR. The problem of 'interpreting' the flow field amounts, then, to resolving it into an axial field (completely determined by the three components of the angular velocity) and a polar field attributable to the translation. If such a resolution can be effected, the presumption of a rigid scene is confirmed, and the three-dimensional structure follows straightforwardly from the translational field component (Gibson *et al.* 1955; Lee 1974).

The analysis that follows is largely concerned with the task of resolving the optic flow field into its rotational and translational components. In describing the flow field due to motion through a rigid scene we have been idealizing the retina as a hemisphere, for which spherical polar coordinates might be thought appropriate; but the choice of retinal coordinates is entirely a matter of convenience,

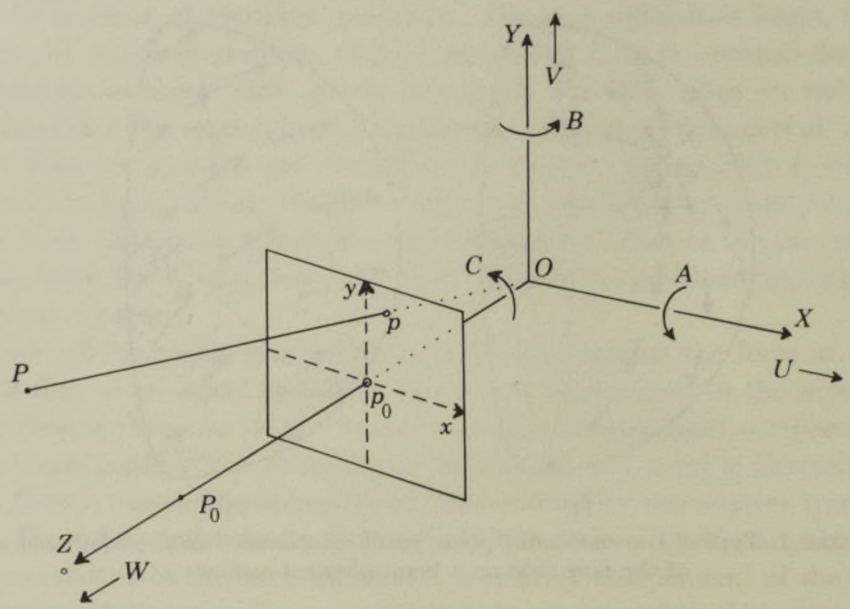


FIGURE 2. An external coordinate system $OXYZ$ moving with the eye, and the corresponding retinal coordinates (x, y). The distance OP_0 equals R.

and we prefer to work with plane projective coordinates (x, y), having the property that a great circle on the hemisphere corresponds to a straight line in the (x, y) plane. The reader who feels uncertain on this point may care to note that although the retina is not planar, neither is it a hemisphere with the lens at its centre; any retinal coordinate system is equally legitimate provided that it does not misrepresent the topology of the retinal image.

Consider a monocular observer moving through a static environment. Let O be the instantaneous position of the nodal point of the eye and let $OXYZ$ be an 'external' Cartesian coordinate system that is fixed with respect to the eye, OZ being the line of sight. Let (U, V, W) be the translational velocity of $OXYZ$ relative to the scene, and let (A, B, C) be its angular velocity. Then if (X, Y, Z) are the instantaneous coordinates of a texture element P in the scene, the velocity components of P in the moving frame will be (see figure 2):

$$\left.\begin{aligned} \dot{X} &= -U-BZ+CY, \\ \dot{Y} &= -V-CX+AZ, \\ \dot{Z} &= -W-AY+BX. \end{aligned}\right\} \qquad (2.1)$$

The retinal position of p, the image of P, may conveniently be represented by the 'internal' coordinates

$$(x, y) = (X/Z,\ Y/Z); \tag{2.2}$$

it will move across the retina with velocity

$$(u, v) = (\dot{x}, \dot{y}). \tag{2.3}$$

Substituting from (2.1) and (2.2) into (2.3) we obtain

$$u = \dot{X}/Z - X\dot{Z}/Z^2 = (-U/Z - B + Cy) - x(-W/Z - Ay + Bx), \tag{2.4}$$

$$v = \dot{Y}/Z - Y\dot{Z}/Z^2 = (-V/Z - Cx + A) - y(-W/Z - Ay + Bx), \tag{2.5}$$

and these equations may be written in the form

$$u = u^{\mathrm{T}} + u^{\mathrm{R}}, \quad v = v^{\mathrm{T}} + v^{\mathrm{R}}, \tag{2.6}$$

with
$$u^{\mathrm{T}} = (-U + xW)/Z, \quad v^{\mathrm{T}} = (-V + yW)/Z, \tag{2.7}$$

and
$$u^{\mathrm{R}} = -B + Cy + Axy - Bx^2, \quad v^{\mathrm{R}} = -Cx + A + Ay^2 - Bxy. \tag{2.8}$$

The retinal velocity field is therefore the vector sum of a translational component $(u^{\mathrm{T}},\ v^{\mathrm{T}})$ that is independent of (A, B, C) and a rotational component $(u^{\mathrm{R}},\ v^{\mathrm{R}})$ that is independent of the three-dimensional structure of the scene. Introducing the coordinates

$$x_0 = U/W, \quad y_0 = V/W, \tag{2.9}$$

we may write the translational component in the form

$$u^{\mathrm{T}} = (x - x_0)W/Z, \quad v^{\mathrm{T}} = (y - y_0)W/Z. \tag{2.10}$$

It follows that

$$v^{\mathrm{T}}/u^{\mathrm{T}} = (y - y_0)/(x - x_0), \tag{2.11}$$

and that the translational flow component is everywhere along straight lines that meet at the 'vanishing point' (x_0, y_0).

So, if the observer is able to resolve the retinal flow field into a rotational component of the form (2.8) and a translational component that is everywhere directed away from (or towards) some retinal point $(x_0,\ y_0)$, not only will this confirm his presumption that the scene is rigid, but also he will be able to compute his direction of motion,

$$U:V:W = x_0:y_0:1, \tag{2.12}$$

and even to determine the relative depths of all the texture elements in the scene:

$$Z/W = (x - x_0)/u^{\mathrm{T}} = (y - y_0)/v^{\mathrm{T}}. \tag{2.13}$$

But to effect this resolution he must first find values of $(A, B, C,)$ such that when the corresponding rotational field, given by (2.8), is subtracted from (u, v), the difference $(u^{\mathrm{T}},\ v^{\mathrm{T}})$ is indeed identifiable as a pure translational field from which the structure of the scene can be found. The question thus arises: how can the observer discover what value to assign to his relative angular velocity (A, B, C)?

One very useful source of information, when it is available, is motion parallax (Helmholtz 1925). If, for example, the observer is walking past a dusty window, then two distinct flow fields will be generated on his retina, one due to the dust particles and the other due to the texture elements behind the window. (This situation illustrates the fact that the optic flow field is not necessarily a single-valued function of retinal position; it is possible, and indeed common, for distinct texture elements to cast their images, momentarily, on the same retinal point.)

Suppose, then, that at the time of observation there are two texture elements, P_1 and P_2, lying in the same direction (x, y) but at different depths, Z_1 and Z_2. Then their images p_1 and p_2 will have the same rotational velocities $(u^{\mathrm{R}}, v^{\mathrm{R}})$ but will differ in their translational velocities $(u^{\mathrm{T}}, v^{\mathrm{T}})$. Hence the difference in the retinal velocities of p_1 and p_2 will be

$$u_1 - u_2 = (-U + xW)(1/Z_1 - 1/Z_2) \tag{2.14}$$

and

$$v_1 - v_2 = (-V + yW)(1/Z_1 - 1/Z_2), \tag{2.15}$$

from which it follows that

$$(v_1 - v_2)/(u_1 - u_2) = (-V + yW)/(-U + xW) = (y - y_0)/(x - x_0). \tag{2.16}$$

The relative velocity $(u_1 - u_2, v_1 - v_2)$ at (x, y) therefore points directly towards, or away from, the vanishing point (x_0, y_0), and this point can be located by using the motion parallax at a number of separate retinal positions. The concurrence of the relative velocity vectors at these positions supports the presumption that the scene is rigid.

Motion parallax thus enables the observer to locate the vanishing point (x_0, y_0) and hence to calculate his direction of motion from (2.12). But to calculate the relative Z coordinates of any two elements he also needs to know the angular velocity (A, B, C), and this he can compute as follows.

For any point on the line $(x = x_0)$, the value of u is given by

$$u(x_0, y) = -B + Cy + Ax_0 y - Bx_0^2. \tag{2.17}$$

It follows that a plot of $u(x_0, y)$ against y is a straight line of slope $(C + Ax_0)$ and intercept $-B(1 + x_0^2)$. Likewise, a plot of $v(x, y_0)$ against x is a straight line of slope $-(C + By_0)$ and intercept $A(1 + y_0^2)$:

$$v(x, y_0) = Cx + A + Ay_0^2 - Bxy_0. \tag{2.18}$$

Knowing x_0 and y_0, the observer can thus compute (A, B, C) without difficulty; having done so, he can then obtain the relative depths of the texture elements from (2.19),

$$Z/W = (x - x_0)/(u - u^{\mathrm{R}}) = (y - y_0)/(v - v^{\mathrm{R}}), \tag{2.19}$$

where u^{R} and v^{R} have been calculated from (2.8). To speak of the observer solving such equations is not, of course, to imply that the visual system performs such calculations exactly as a mathematician would: its modes of operation may well be more 'geometrical' than 'algebraic', and the same applies to the other computations envisaged in this paper. Our main point is that the equations demonstrate the feasibility of calculating both the motion and the structure from the optic flow field alone.

But, if motion parallax cues are not readily available, or, if parts of the scene are in relative motion, then these methods fail. We therefore turn, in the next section, to the problem of determining the motion and the structure of an object with a smooth, densely textured surface.

3. Motion relative to a visually textured surface

If the scene consists of a number of rigid objects in relative motion, then, to determine his motion relative to any one of them, the observer will have to rely on information from a limited part of the visual field. We therefore envisage a situation in which the object of interest has a smooth, densely textured surface S; the observer's task is then to determine his motion relative to S, and the gradient of the surface at any given point P_0 lying on it.

Since P_0 is not necessarily in the observer's line of sight, we now adopt a coordinate system, $OXYZ$, that is orientated in such a way that P_0 has the coordinates $(0, 0, R)$ at the time of observation. As before, $OXYZ$ is assumed to move with the observer's eye. The surface S may then be described by the equation

$$Z(X, Y) = R + \alpha X + \beta Y + O_2(X, Y),\dagger \tag{3.1}$$

where (α, β) is the gradient of S at P_0. If $(x, y) = (X/Z, Y/Z)$ are local retinal coordinates, then equations (2.4) and (2.5) hold in the new interpretation, so that

$$u = (-U + xW)/Z - B + Cy + Axy - Bx^2, \tag{3.2}$$

$$v = (-V + yW)/Z - Cx + A + Ay^2 - Bxy. \tag{3.3}$$

Introducing the dimensionless depth coordinate

$$z = (Z-R)/Z = \alpha x + \beta y + O_2(x, y) \tag{3.4}$$

and the depth-scaled velocities

$$u_0 = U/R, \quad v_0 = V/R, \quad w_0 = W/R, \tag{3.5}$$

we may write (3.2) and (3.3) in the forms

$$u = (-u_0 + xw_0)(1-z) - B + Cy + Axy - Bx^2, \tag{3.6}$$

$$v = (-v_0 + yw_0)(1-z) - Cx + A + Ay^2 - Bxy. \tag{3.7}$$

† $O_2(X, Y)$ means 'terms of the second order in X and Y'.

(The local coordinates of p_0, the image of P_0, are, of course, $(x, y) = (0, 0)$.) The question now arises: can the observer derive all the unknown parameters u_0, v_0, w_0, A, B, C, α and β from the velocity field (u, v) in the immediate neighbourhood of p_0?

Since S is smooth, by hypothesis, the derivatives of u and v with respect to x and y are well defined. At the point p_0, where $x = y = z = 0$, u and v take the values

$$u = -u_0 - B, \quad v = -v_0 + A; \tag{3.8}$$

their first derivatives are found to be

$$\left.\begin{aligned} u_x &= u_0\alpha + w_0, & u_y &= u_0\beta + C, \\ v_x &= v_0\alpha - C, & v_y &= v_0\beta + w_0, \end{aligned}\right\} \tag{3.9}$$

since

$$z_x = \alpha, \qquad z_y = \beta. \tag{3.10}$$

If u_0 and v_0 could only be found, (3.8) would immediately give the values of A and B, and (3.9) would supply the values of w_0, C, α and β; but, as the equations stand, the values of α and β are inseparable from those of u_0 and v_0, and, in particular, there is no way of finding the relative magnitudes of these two pairs of quantities. This is the familiar problem of the 'indeterminate depth scale', in a new guise.

There is, however, a way round the difficulty. The translational components of u and v correspond, as shown in the previous section, to flow along lines through the vanishing point

$$(x_0, y_0) = (u_0/w_0, v_0/w_0). \tag{3.11}$$

Writing (3.6) in the form

$$u = (x - x_0)\, w_0(1-z) - B + Cy + Axy - Bx^2, \tag{3.12}$$

we see that, if the (x, y) axes are reorientated in such a way that the y axis passes through (x_0, y_0), then (3.12) reduces to

$$u = xw_0(1-z) - B + Cy + Axy - Bx^2, \tag{3.13}$$

where all quantities now refer to the new coordinate system (see figure 3). On the y axis, where $x = 0$, u now assumes a particularly simple form, namely

$$u(0, y) = -B + Cy. \tag{3.14}$$

This equation asserts, in effect, that all those image points that lie on the (new) y axis at time t will still lie on a straight line at $t + \delta t$, namely the line

$$x = (-B + Cy)\,\delta t + \mathrm{O}_2(\delta t). \tag{3.15}$$

In general, furthermore, there will be only one straight line of image points, through the origin, that at the time of observation is not bending in the (x, y) plane. Equations for finding this line, when it is unique, are given in the appendix, where it is also shown that, if S is planar, every line through the origin has this property, so that the present analysis fails.

If the unbending line is indeed unique, then, once it has been located and adopted as the y axis, the problem is largely solved. For, by (3.11),

$$u_0 = x_0 w_0 = 0, \tag{3.16}$$

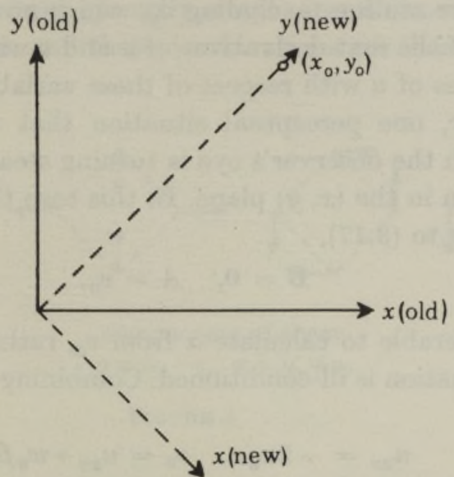


FIGURE 3

so that at the origin

$$u = -B, \quad v = -v_0 + A, \tag{3.17}$$

$$\left.\begin{aligned} u_x &= w_0, \quad u_y = C, \\ v_x &= v_0\alpha - C, \quad v_y = v_0\beta + w_0, \end{aligned}\right\} \tag{3.18}$$

and, by (3.13),

$$u_{xx} = -2w_0\alpha - 2B, \quad u_{xy} = -w_0\beta + A. \tag{3.19}$$

Expressions (3.17–19) constitute eight equations for the seven unknown quantities, A, B, C, v_0, w_0, α and β; their consistency is a necessary condition for S to be rigid. The solution is immediate. By (3.17) and (3.18),

$$B = -u, \quad w_0 = u_x, \quad C = u_y, \tag{3.20}$$

and substitution in the first of equations (3.19) gives

$$\alpha = (u - \tfrac{1}{2}u_{xx})/u_x. \tag{3.21}$$

Introducing the observable quantities (see figure 4)

$$\rho = u_x - v_y, \quad \sigma = u_y + v_x, \tag{3.22}$$

we infer from (3.18) that

$$v_0\alpha = \sigma, \quad v_0\beta = -\rho, \tag{3.23}$$

whence

$$v_0 = \sigma/\alpha, \quad \beta = -\rho/v_0. \tag{3.24}$$

Finally, then, by (3.17),

$$A = v + v_0 \tag{3.25}$$

and it only remains to check that u_{xy} satisfies the second of equations (3.19).

Equations (3.20)–(3.25) show that if and when a unique 'unbending line' of image points can be identified, then, if it is selected as the y axis, the seven parameters of relative motion (excluding u_0, which now vanishes) can be simply expressed in terms of the first derivatives of u and v with respect to x and y, and the second derivatives of u with respect of these variables.

There is, however, one perceptual situation that merits special discussion, namely that in which the observer's eye is turning steadily so as to maintain the point P_0 at the origin in the (x, y) plane. In this case the velocity (u, v) vanishes there, and, according to (3.17),

$$B = 0, \quad A = v_0. \tag{3.26}$$

It may then be preferable to calculate α from v_0, rather than from (3.21), since if w_0 is small this equation is ill-conditioned. Combining (3.19) and (3.26), we now obtain

$$u_{xx} = -2w_0\alpha, \quad v_0 = u_{xy} + w_0\beta, \tag{3.27}$$

and from (3.23) we deduce that

$$v_0 = u_{xy} + \rho u_{xx}/2\sigma. \tag{3.28}$$

The gradient (α, β) is then calculated from (3.23), and the final check of consistency is supplied by the first of equations (3.27), with w_0 set equal to u_x.

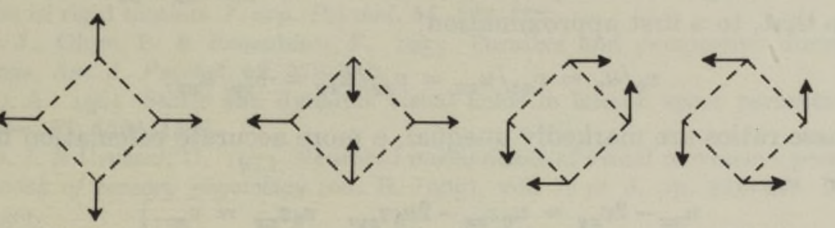
It appears, then, that the observer gains certain computational advantages from tracking with his eye any surface whose gradient and relative motion are of special interest to him; this is an intuitively reassuring result.

4. Discussion

What we have shown, in effect, is that an observer can *in principle* determine the structure of a rigid scene and his direction of motion relative to it from the *instantaneous* retinal velocity field. If the scene consists of separate objects in relative motion, then a separate computation must be carried out on each one, requiring access to the flow at some retinal point and its first and second spatial derivatives at that point. Each computation checks the rigidity assumption on which it is based.

Whether the visual system actually operates in this way is, of course, another matter. Ullman has argued cogently (Ullman 1979) for a 'polar-parallel' scheme, which requires a time long enough for the observer to obtain at least three sufficiently distinct views of each object to determine its structure. The present scheme requires, instead, that the neighbourhoods involved in local computations subtend an angle large enough for the second derivatives of u and v to be estimated with accuracy.

Leaving aside the relative merits and domains of application of the two schemes, there is one feature of equations (3.23) *et seq.* to which we would like to draw attention, namely the quantities ρ and σ defined in (3.22). As pointed out by Koenderink & van Doorn (1976), there are four independent combinations of the flow field derivatives that have specially simple transformation properties under rotation of the retinal axes, and these combinations are illustrated in figure 4.



dilatation $\mu = u_x + v_y$ | components of shear $\rho = u_x - v_y$, $\sigma = u_y + v_x$ | vorticity $\lambda = u_y - v_x$

FIGURE 4

The dilatation, μ, and the vorticity, λ, are separately invariant under rotation; the two components of shear, encountered in (3.22), are mixed together (but not with λ or μ) when the axes are rotated through an angle θ:

$$\rho \rightarrow \rho \cos 2\theta + \sigma \sin 2\theta, \tag{4.1}$$

$$\sigma \rightarrow -\rho \sin 2\theta + \sigma \cos 2\theta. \tag{4.2}$$

It would therefore not be surprising if the human visual system possessed channels tuned to these four basic types of 'relative' motion, as well as the 'absolute' motion channels implicated, for example, in the waterfall illusion. Evidence has quite recently been obtained (Regan *et al.* 1978*a*, *b*, 1979) for channels sensitive to dilatation; if there also exist shear- and vorticity-sensitive channels, it might be possible to demonstrate their existence psychophysically.

APPENDIX

The computation described in §3 hinges on the possibility of finding the line that joins the retinal origin to the vanishing point (x_0, y_0). This line has the peculiar property that image points lying on it at time t will still be in a straight line at time $t + \delta t$; and, if it is adopted as the y axis of a new coordinate frame, then the required kinematic parameters are simple functions of the retinal image velocity and its first and second derivatives in that frame.

In the original (x, y) system, by (3.11),

$$y_0/x_0 = v_0/u_0, \tag{A 1}$$

and so the problem reduces to determining this ratio. Now, according to (3.6) and (3.7),

$$\left.\begin{aligned} u_{xx} &= u_0 z_{xx} - 2w_0\alpha - 2B, & v_{xx} &= v_0 z_{xx}, \\ u_{xy} &= u_0 z_{xy} - w_0\beta + A, & v_{xy} &= v_0 z_{xy} - w_0\alpha - B, \\ u_{yy} &= u_0 z_{yy}, & v_{yy} &= v_0 z_{yy} - 2w_0\beta + 2A. \end{aligned}\right\} \tag{A 2}$$

If the surface is highly curved at P_0, with radii of curvature much less than R (the distance of P_0 from the eye), then the terms in z_{xx}, z_{xy}, z_{yy} will outweigh the others, so that, to a first approximation,

$$v_0/u_0 = v_{xx}/u_{xx} = v_{xy}/u_{xy} = v_{yy}/u_{yy}. \tag{A 3}$$

But, if these ratios are markedly unequal, a more accurate calculation is needed. By (A 2),

$$\left.\begin{aligned} u_{xx} - 2v_{xy} &= u_0 z_{xx} - 2v_0 z_{xy}, & v_0 z_{xx} &= v_{xx}, \\ v_{yy} - 2u_{xy} &= v_0 z_{yy} - 2u_0 z_{xy}, & u_0 z_{yy} &= u_{yy}. \end{aligned}\right\} \tag{A 4}$$

Elimination of z_{xx}, z_{xy}, z_{yy} between these equations leads to

$$u_0^3 v_{xx} + u_0^2 v_0 (2v_{xy} - u_{xx}) + u_0 v_0^2 (v_{yy} - 2u_{xy}) - v_0^3 u_{yy} = 0 \tag{A 5}$$

and this may be written in the alternative form

$$t = (v_{xx} + 2tv_{xy} + t^2 v_{yy})/(u_{xx} + 2tu_{xy} + t^2 u_{yy}), \tag{A 6}$$

where

$$t = v_0/u_0. \tag{A 7}$$

(A 6) is a cubic equation in the required ratio t, but the relevant root can be quickly found by iteration if a first approximation is available from (A 3).

Difficulties arise, however, if S is planar, because then z_{xx}, z_{xy} and z_{yy} all vanish, and so do all the coefficients in (A 5):

$$\left.\begin{aligned} u_{xx} - 2v_{xy} &= 0, & v_{xx} &= 0, \\ v_{yy} - 2u_{xy} &= 0, & u_{yy} &= 0. \end{aligned}\right\} \tag{A 8}$$

Planarity is thus easily detected, but to determine the kinematic parameters is surprisingly difficult. One finds, in fact, that, when the second derivatives of z are set equal to zero in (A 2), the solution of these equations involves extracting the roots of a cubic equation; but there is no obvious way of calculating the appropriate root. Ullman's polar–parallel scheme (Ullman 1979, p. 173) also encounters difficulties with planar objects.

References

Braunstein, M. L. 1976 *Depth perception through motion*. London: Academic Press.
Bridgeman, B. 1972 Visual receptive fields to absolute and relative motion during tracking. *Science, N.Y.* **187**, 1106–1108.
Gibson, E. J., Gibson, J. J., Smith, O. W. & Flock, H. 1959 Motion parallax as determinant of perceptual depth *J. exp. Psychol.* **58**, 40–51.
Gibson, J. J. 1950 *The perception of the visual world*. Boston: Houghton Mifflin.
Gibson, J. J. 1966 *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
Gibson, J. J. 1979 *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gibson, J. J. & Gibson, E. J. 1957 Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.* **54**, 129–138.
Gibson, J. J., Olum, P. & Rosenblatt, F. 1955 Parallax and perspective during aircraft landings. *Am. J. Psychol.* **68**, 372–385.
Gordon, D. A. 1965 Static and dynamic visual fields in human space perception. *J. opt. Soc. Am.* **55**, 1296–1303.
Grusser, O. J. & Grusser, U. 1973 Neuronal mechanisms of visual movement perception. In *Handbook of sensory physiology* (ed. R. Jung), vol. 7, pt 3, pp. 332–429. New York: Springer.
Helmholtz, H. von 1925 *Treatise on physiological optics* (ed. J. P. C. Southall). New York: Dover.
Hubel, D. H. & Wiesel, T. N. 1968 Receptive fields and functional architecture of monkey striate cortex. *J. Physiol., Lond.* **195**, 215–243.
Koenderink, J. J. & van Doorn, A. J. 1976 Local structure of movement parallax of the plane. *J. opt. Soc. Am.* **66**, 717–723.
Lee, D. N. 1974 Visual information during locomotion. In *Perception* (ed. Mcleod & Pick), pp. 250–267. Ithaca: Cornell University Press.
Marr, D. 1976 Early processing of visual information. *Phil. Trans. R. Soc. Lond.* B **275**, 483–534.
Marr, D. & Poggio, T. 1976 Cooperative computation of stereo disparity. *Science, N.Y.* **194**, 283–287.
Nakayama, K. & Loomis, J. M. 1974 Optical velocity patterns, velocity sensitive neurons and space perception. *Perception* **3**, 63–80.
Regan, D. & Beverley, K. I. 1978*a* Illusory motion in depth. *Vision Res.* **18**, 209–212.
Regan, D. & Beverley, K. I. 1978*b* Looming detectors in the human visual pathway. *Vision Res.* **18**, 415–421.
Regan, D., Beverley, K. I. & Cynader, M. 1979 The visual perception of depth. *Scient. Am.* **241**, 122–133.
Sekuler, R. & Levinson, E. 1974 Mechanisms of motion perception. *Psychologia* **17**, 38–49.
Sekuler, R. & Levinson, E. 1977 The perception of moving targets. *Scient. Am.* **236**, 60–73.
Ullman, S. 1979 *The interpretation of visual motion*. Cambridge, Massachusetts: M.I.T. Press.