

VISUAL MOTION AMBIGUITY

H. C. LONGUET-HIGGINS

Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, England

Abstract—This paper discusses a number of points that arise in the visual interpretation of optic flow fields. Particular attention is given to the case in which a flow field is visually ambiguous, and the paper concludes with a detailed description of the case of the moving plane.

Optic flow Structure from motion Visual ambiguity

In a “systems approach” to vision one attempts to understand how the machinery of vision enables an animal to form a useful model of its spatio-temporal environment. In this connection the phenomenon of optic flow—the motion of an optical image across the retina—reminds us that the world is not 3-dimensional but 4-dimensional, with time as the 4th dimension. In cinematography we need physical optics to help us understand how the optical image changes as the camera moves through the scene; but the theory of vision has an ulterior motive—to determine the structure of the scene from the motion of the retinal image. In this sense, vision is “inverse optics”—a thought which we owe largely to the computer vision people. In recent years, the theory of vision has benefited from an influx of ideas from computing science and artificial intelligence, the work of the late David Marr (Marr, 1982) having been particularly influential. His computational approach to vision invites us to think of the visual system as performing sophisticated computations on the optical input, in order to arrive at a representation of the visible world. According to this way of thinking, the central question about vision is: by what logical stages, exactly, does the human visual system construct a useful model of the world? In Marr’s view, the progress of vision research is to be evaluated by the light that it throws on this underlying problem.

In what follows I shall touch briefly on five matters relevant to the interpretation of optic flow fields, and will then summarize the main features of an important ambiguous flow field—that arising from a moving plane. The five matters are:

(1) The organic implementation of motion perception.

(2) The structure-from-motion problem in the presence of non-rigid deformations.

(3) The merits of the Rigidity Assumption, and possible relaxations thereof.

(4) The relative merits of partial versus complete, and qualitative versus quantitative, solutions of the structure-from-motion problem.

(5) The description of optic flow in terms of differential invariants.

My comments are as follows:

(1) Possibly the algorithms used by the visual system for deriving structure from optic flow are independent of the neuronal mechanisms by which the flow field is computed in the first place. But the concept of a retinal velocity field $\mathbf{v}(x, y)$ is itself a hazardous abstraction from the primary input $I(x, y, t)$, (I being the light intensity) as many authors have emphasized. The vector \mathbf{v} may be undefined in many regions and discontinuous or even many-valued in others; the “correspondence problem” and the “aperture problem”, which stand between the observation of I and the computation of the velocity \mathbf{v} , will not go away just because one chooses to ignore them. Nevertheless it is certainly a good idea to try to clear one’s mind about what the visual system might actually be computing before poking around inside it; it seems unlikely that the neuronal circuitry will somehow explain itself.

(2), (3) The Rigidity Assumption is undoubtedly the most useful of all the hypotheses available to the visual system in attempting to arrive at a $(3 + 1)$ -dimensional interpretation of the $(2 + 1)$ -dimensional retinal image. Actually there are quite severe constraints to be satisfied by a moving image if it is to have a rigid interpretation, and most non-rigid motions are

readily detectable as such in the retinal image itself. (An obvious example is the motion of a creature such as a cat.) The natural generalisation of the rigidity assumption is the assumption of piecewise rigidity or near rigidity. Ullman (Ullman, 1979) has proposed an algorithm for dealing with the image of a nearly rigid object in motion, based on what one may describe as a "rubberiness assumption", but personally I am a little doubtful about its psychological realism. Likewise, Koenderink's recent algorithm for deriving the structure of an articulated object composed of rigid parts is a mathematical *tour de force*, but I wonder how often circumstances arise in which the algorithm could usefully be applied. Indeed, the question of what hypotheses we adopt, and what we do when they become untenable, is one of the most tantalizing in the whole theory of vision.

(4) I suspect that the spatio-temporal interpretations we give to the optic flow field are much less complete than we often like to think. There is quite a bit of current work on the "interpolation" of visible surfaces between the few elements whose depths can be reliably estimated by motion parallax, stereopsis or other means; but it is at least likely that the visual system performs no such computation until it becomes involved in an action such as putting down a glass on a smooth white table. In the elaboration of a world model one has to stop somewhere; there may be much relevance in the dictum, "Out of sight, out of mind".

(5) To a mathematician the most perspicuous results in the "first-order" theory of optic flow fields are the expressions for the *div*, the *curl* and the *def* of the retinal velocity field of a smooth densely textured surface. These results were originally derived by Koenderink and van Doorn (1975), and usefully constrain the interpretation of a locally differentiable optic flow field. But in spite of having favoured such ideas a few years ago (Longuet-Higgins and Prazdny, 1980) I have come to doubt the computational realism of a velocity which is a *differentiable* function of retinal position (Regan and Beverley, 1978). It seems more likely that the *divergence* the *curl* and the *deformation* should be regarded not as much as first-order differential operators but as coefficients in a power series expansion of the velocity field. As for their visual significance, perhaps it should be emphasized that though these parameters are simple functions of the gradient of the surface and the motion of the camera, the converse is not true:

not only is the optic flow field incapable of revealing the scale of the scene, but its *div*, *curl* and *def* are compatible with infinitely many values of the surface gradient and the observer's velocity. The reason is that in Koenderink and van Doorn's equations the surface gradient F and the (reduced) transverse velocity A_t enter only as products with each other; so the observer might be moving rapidly past a low-relief surface or more slowly past a more steeply tilted one. This ambiguity, not merely of scale but of *shape*, persists even when the angular velocity R , and the (reduced) radial velocity A_r are reliably known. A few years ago (Longuet-Higgins and Prazdny, 1980) Prazdny and I showed that in principle this ambiguity could be resolved by taking account of the (spatial) second derivatives of the flow field; but I rather fear that such a computation would be altogether too ill-conditioned to be of any practical use.

Curiously enough, there seems to be an inverse relation between the "well-behavedness" of the image intensity (regarded as a function of retinal location and time) and the ease with which the visual system can extract depth information from it. The visually most perplexing surface for motion perception is the plane (of which more anon), and after that come curved surfaces such as the sphere. Much clearer impressions of depth are obtained from "rocky" surfaces, or when there is motion parallax involving the progressive occultation of one object by another. Most vivid of all—and giving the most accurate depth information (Rieger and Lawton, 1983)—are images in which a structured foreground and background are in relative visual motion—as when one walks through a wood, or past a dusty window. As various people (e.g. Blake, 1983) have pointed out, the visual image is often of most interest at those places where it is mathematically the most intractable.

In conclusion I will illustrate my thesis, that vision is the interpretation of images, by describing some recent results on motion relative to a textured plane—a situation in which the optic flow field may permit two quite distinct structural interpretations. An extended mathematical analysis of this situation has just been published (Longuet-Higgins, 1984); the main results are as follows:

(a) The two components of the flow field (u, v) are second-order polynomials in the reti-

nal coordinates (x, y) involving 8 independent coefficients altogether.

(b) The 8 parameters of the flow field depend, in turn, on (i) the observer's linear velocity (U, V, W) and angular velocity (A, B, C) in the camera coordinate system and (ii) the reciprocal vector (L, M, N) specifying the location of the plane.

(c) From the 8 field parameters one can construct a symmetric 3×3 matrix S , which can be diagonalized by a rigid rotation matrix T . This process involves solving a cubic equation; let its (real) roots be $f_1 < f_2 < f_3$.

(d) In the general case $f_1 < f_2 < f_3$, let θ be an angle satisfying

$$\cos 2\theta = (f_1 - 2f_2 + f_3)/(f_1 - f_3).$$

Then the unit vectors in the directions of (U, V, W) and (L, M, N) are $(\cos \theta, 0, \pm \sin \theta)$. T_1 (where the dot means matrix multiplication) but there is nothing in the values of the 8 field parameters to tell which vector is which.

(e) This ambiguity is resolved if *either* two of the roots f_1, f_2 and f_3 are equal *or* the field of view includes *both* points which lie "ahead" *and* points which lie "astern", where these adverbs refer to the direction in which the camera is moving, rather than the direction in which it is pointing. (If all three roots are equal, it may be inferred that either the camera's linear velocity is zero, or the plane is infinitely distant.)

The visual ambiguity of the moving plane is well illustrated by the flow field which arises when one is driving along a straight road looking into the distance. Exactly the same *instantaneous* flow field arises as one descends into a mine shaft keeping one's eyes fixed on a point at the top of the shaft. A human being would be most unlikely to mistake either situation for the other, at least for more than an instant; but a robot vision system which depended on optic flow would have to be carefully programmed to avoid this kind of error.

REFERENCES

- Blake A. (1983) *Parallel Computation in Low-level Vision*. Ph.D. thesis, Univ. of Edinburgh.
- Koenderink J. J. and van Doorn A. J. (1976) Local structure of movement parallax of the plane. *J. opt. Soc. Am.* **66**, 717-723.
- Longuet-Higgins H. C. and Prazdny K. (1980) The interpretation of a moving retinal image. *Proc. R. Soc. Lond. B* **208**, 385-397.
- Longuet-Higgins H. C. (1984) The visual ambiguity of a moving plane. *Proc. R. Soc. Lond. B* **223**, 165-175.
- Regan D. and Beverley K. I. (1978) Looming detectors in the human visual pathway. *Vision Res.* **18**, 415-421.
- Rieger J. H. and Lawton D. T. (1983) Sensor motion and relative depth from difference fields of optic flows. *Proc. 8th IJCAI*, Karlsruhe (F.R.G.).
- Marr D. (1982) *Vision*. Freeman, San Francisco, Calif.
- Ullman S. (1979) Relaxed and constrained optimization by local processes. *Comput. Graph. Image Processing* **10**, 115-125.