

MOTION AND STRUCTURE FROM OPTICAL FLOW

K. Prazdny
Computer Science Department
University of Essex
Wivenhoe Park
Colchester Essex
England

We show that a moving observer can (in principle) compute the relative depth map of the environment given only the instantaneous positional velocity field on his "retina". The depth map (and surface normals) are computed by recovering the egomotion parameters which in turn determine the ratio of depth associated with any two neighbouring "retinal" points. The surface normal map may be computed to a degree of accuracy determined by the resolution of the "retina", and by the accuracy of approximation obtainable in the solution of three non-linear equations of the third degree in three unknowns.

1. THE PROBLEM

The problem addressed here is computing the surface layout of the environment, and the parameters of observer's motion (egomotion) from the information on the image plane. The problem is similar (though not equivalent) to the one tackled by the workers in computer stereo vision [1].

Consider a monocular observer with his "eye" and "head" fixed, with planar "retina", and moving along a smooth curvilinear path in a stationary 3D environment. Each discriminable optical element on his "retina" will possess a velocity; the whole comprising a positional velocity field. Such an observer can be viewed, instantaneously, as simultaneously rotating and translating for (again instantaneously) any 3D motion can be viewed as a rotation about some centre on some osculating plane. The instantaneous positional velocity field (IPVF) is then expressible as being due to the vector sum of two angular velocities due to the translational and the rotational components of the movement [2], [3].

Given a IPVF our task is to recover the egomotion parameters (i.e., the translational and rotational components of the egomotion). We do this with respect to an external, static reference frame instantaneously coincident with our egocentric reference frame (centered at 0 in Fig. 1).

The actual velocity of any point Q on the image plane is given by

$$\underline{V}' = (V_1/S) (\bar{V}_1 \times \bar{Q}) + \underline{A} \times \underline{Q} \quad (1)$$

Here, V_1 is the magnitude of the translational component (the speed of egomotion). $\bar{V}_1$ is its direction vector. And S is the distance to the centre of polar projection of the environmental point P of which Q is the projection on the "retina" (see Fig. 1). $\underline{Q}$ is the position vector of the point Q in the egocentric reference frame centered at 0. $\bar{Q}$ is the unit vector in this direction. $\underline{A}$ is the rotational angular velocity vector the components of which are to be recovered in the first place (for more detail see [4]).

Refer to Fig. 1. As can be seen, the "retinal" velocity vector of point Q, $\underline{V}'$ must necessarily lie in the plane spanned by $\bar{Q}$ and $\underline{V}'$. Given the IPVF, $\underline{V}'$ may be derived for each point Q, since

$$\underline{V}' = \underline{V}'' - (\bar{Q} \cdot \underline{V}') \bar{Q} \quad (2)$$

(This follows from the equivalence of angular velocities generated by $\underline{V}'$ and $\underline{V}''$).

Given that we know $\underline{V}'$, we can rewrite (1) to obtain

$$-(V_1/S) \underline{Q} ((\bar{V}_1 \times \bar{Q}) \times \bar{Q}) = \underline{A} \times \underline{Q} - \underline{V}' \quad (3)$$

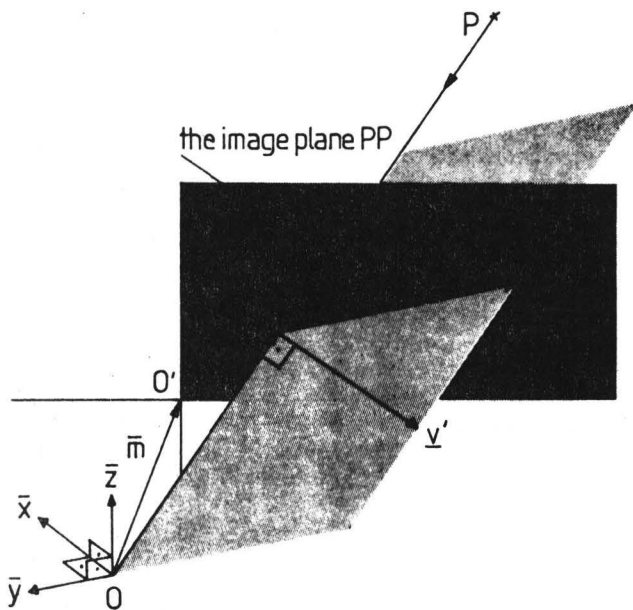


Fig. 1

An egocentric reference frame centered at the centre of polar projection, O , is defined, together with a planar projection surface PP . A "retinal" 2D rectangular coordinate frame centered at O is defined on PP .

V' and V'' lie on the same plane, the unit normal of which is $(\bar{V}' \times \bar{Q})$.

Refer to Fig. 2. The vector $(\bar{V}_1 \times \bar{Q}) \times \bar{Q}$ lies in the plane of $\bar{V}_1$. Given the fact that $\bar{V}_1$ must lie in all such planes corresponding to different points Q , its direction will be determined as the intersection of these planes. The direction of the line of intersection of two planes specified by their normals $\bar{N}_I, \bar{N}_J$ is $(\bar{N}_I \times \bar{N}_J)$. Referring back to (3) we see that

$$\bar{N}_I = (\bar{A} \times \bar{Q}_I - \bar{V}_I') \times \bar{Q}_I \quad (4)$$

lies in the direction of the vector normal to the plane P_{V_1} (see Fig. 2). From this, using simple vector algebra, we see that for any three distinct "retinal" points Q_I, Q_J , and Q_K the relation

$$[\bar{N}_I, \bar{N}_J, \bar{N}_K] = 0 \quad (5)$$

must hold. Here, the square brackets denote a scalar triple product of the three vectors. Equation (5), written out, is an equation of the third degree in 3 unknowns (the components of $\bar{A}$).

A set of 3 such equations (a minimum of 5 distinct "retinal" points is needed to define such a set) does not have an analytical solution: in general the solution can be obtained only using an iterative technique (a very fast and simple method reported in [5] is used in our computer implementation).

2. COMPUTER IMPLEMENTATION

The solution to a set of equations (5) can be computed locally, given some non-empty neighbourhood of a "retinal" point Q . In a 'real world' application one should not, however, use locally clustered points. To obtain accuracy in the presence of noise one should use points spread as widely as possible. Also, considering more points in the process of solving for $\bar{A}$ (perhaps by minimizing some suitable function of differences between the values of $\bar{A}$ computed at different regions locally) improves the accuracy with which $\bar{A}$ can be obtained.

A computer program incorporating the method has been written and experimented with using simulated input data (no distortions or noise). The program consists essentially of a local operator which computes $\bar{A}$, $\bar{V}_1$, and the surface normal corresponding to a given image plane point. All three values are computed locally from a neighbourhood of a given "retinal" point. The size of the neighbourhood is presently (due to some computational limitations) about 1 degree of arc. The surface normal produced by the operator is essentially that of the best fit planar scale at a given point. The application of the operator to the whole image results in a surface-normal map.

The results are encouraging. Using the single precision arithmetic the accuracy of the surface normal map lies within the error of (approximately) 1 degree of arc.

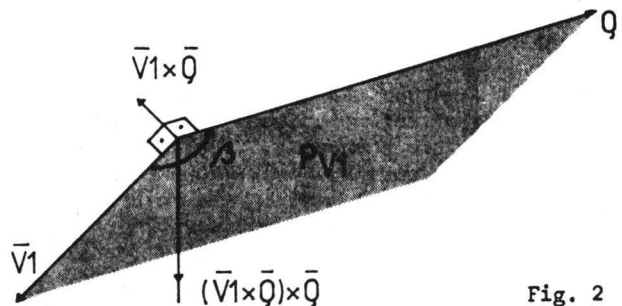


Fig. 2

$(\bar{V}_1 \times \bar{Q}) \times \bar{Q}$ and $\bar{V}_1$ lie in the same plane, P_{V_1} .

3. DISCUSSION

Using the preceeding analysis, it seems plausible that a relatively simple process may be used to recover the instantaneous egomotion parameters and surface normal map in a stationary world. The only assumptions made were the smoothness of the movement trajectory, rigidity of the objects, and the existence of the IPVF. This latter is taken for granted and constitutes the greatest weakness of this kind of approach. There is very little known about the behaviour of the correspondence computation-like methods suggested by Ullman [6] using real images, and these or similar mechanisms are, at present, the only way to generate the IPVF.

On the other hand, the above approach applies equally well to the case of a rigid 3D world of moving objects. To see this note the reciprocity between the motion of an object and the motion of the observer; for any objection motion there is an observer motion which, assuming the object to be stationary, generates the same IPVF in the region of the object. (This is the main reason that smoothness of path was the ONLY restriction put on the observer's motion.) Using a simple technique suggested by Nakayama and Loomis [3], we may compute the depth contours to perform a figure-ground separation. Applying the analysis above to each separated image region leads then directly to the surface normal map corresponding to the region. It is not possible, however, to recover the object motion since absolute distances (of which the optical flow is a function) are not recoverable from IPVF's.

4. CONCLUSION

A method is outlined which is capable (at least in principle) of delivering a surface normal map from an instantaneous positional velocity field (IPVF) generated by a curvilinearly moving observer. While the method is not as elegant and simple as the one developed in Clocksin [7] for the case of a rectilinearly moving observer, it is considerably more general.

It is possible that the analysis outlined above for the case of moving objects may offer some insights into the phenomena of position constancy ("How is it possible that when we move, and thus every optical element on the retina moves, we still perceive a stationary world?"). In that case, after applying the method locally, an "egomotion map" may be constructed which maps each point on the "retina" to this "egomotion space" (spanned by the orthogonal components of A and V). Some evaluation mechanism may be

operating in this space, choosing, for example the point into which the 'majority' of image plane points maps as the stationary reference frame. All point mapping into this point (or perhaps some small neighbourhood of it) would then be regarded as projections of the stationary environment.

ACKNOWLEDGEMENTS

I'm in debt to Drs M. Brady and P. Hayes for many helpful suggestions, and to my wife Dasha who did the layout.

REFERENCES

- [1] Gennery, D.B. "A stereo vision system for an autonomous vehicle", Fifth Int. Joint Conf. Artif. Intel., MIT, August, 1977.
- [2] Wheatherburn, C.E. "Elementary vector analysis", Bell and Sons, London, 1965.
- [3] Nakayama, K. and Loomis, J.M. "Optical velocity patterns, velocity sensitive neurons, and space perception", Perception, 3, (1974).
- [4] Prazdny, K. "Egomotion and relative depth map from optical flow", (submitted for publication), 1978.
- [5] Broyden, C.G. "Algorithm 214: solution of non-linear simultaneous equations", Computer Journal 12 (1969), pp.406-408.
- [6] Ullman, S. "The interpretation of visual motion", Ph.D. thesis, MIT, 1977.
- [7] Clocksin, W. "Perception of surface slant and edge labels from optical flow", DAI Working Paper, No. 33, University of Edinburgh, 1978.

