OPTIC FLOW FIELD STRUCTURE AND PROCESSING IMAGE MOTION

Daryl T. Lawton

Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

The structural properties of optic flow fields are used to develop constraints on the motion of image features over time. For several restricted cases of motion, these constraints greatly simplify the determination of inter-frame motion and the inference of environmental information. These structural properties are briefly reviewed. Procedures are presented for computing, and making environmental inferences from, optic flow produced when camera motion is known, translational, or restricted to a plane. Some applications are discussed.

I. INTRODUCTION

Optic flow is the field of velocity vectors generated on an imaging surface by the moving projections of environmental points. The analysis of flow fields provides useful environmental and movement control information——————without "top-down", semantic knowledge and often by simple computations [1]. There are, however, significant problems with computing optic flow from real world image sequences, notably how techniques for environmental inference can be applied to computed flow fields which are noisy and imprecise and how to seperate particular image transformations, such as light source and occlusion effects, from image motion which corresponds to optic flow.

In this paper we consider restricted cases of motion for which the processing of optic flow, from real world image sequences, is robust. These cases occurr frequently in important situations. The structural properties of flow fields are reviewed and used to develop constraints on the motion of image features overtime. For several restricted cases of motion, these constraints greatly simplify the determination of optic flow and the inference of environmental information.

Finally, in processing image sequences, a flow field is generally approximated by the displacements of features between successive images. While there are significant differences between a flow field and a displacement field, in this communication they are referred to interchangeably.

// OPTIC FLOW FIELD STRUCTURE

A. Translational/Rotational Components

The motion of features in successive images, IMAGE(i) and IMAGE(i+1), produced by a camera moving relative to a stationary environment can be described by 5 parameters [2,3]: two (T1(i),T2(i)) for the unit vector describing the direction of translation with respect to the camera coordinate system; two (R1(i),R?(i)) for the unit vector describing the axis of rotation with respect to the camera coordinate system; and one (R3(i)) for the extent of rotation about the axis of rotation.

For a planar retina, the translational flowpaths are straight lines which intersect at a point called the Focus of Expansion/Contraction or FOE for short. The positioned of the FOE is determined by the intersection of the retinal surface with the translational axis. The direction of motion along a flowpath, either towards or away from the FOE, reflects whether the corresponding environmental point is becoming more or less distant. The environmental depth *of* a point is a simple function of its position and displacement along a translational flowpath [4]

$$Z = \frac{D * z'}{d'}$$

where D is the distance of an image point from the FOE at i; d' is the displacement of the image point from i to i+1; z' is the environmental displacement of the point relative to the observer along the z axis from i to i+1; and Z is the environmental depth at time i.

Rotational flowpaths, for planar retinas, are conic sections. The important fact about rotational flow is that it is not a function of environmental depth. That is, the parameters R1(i) ,R2(i) ,R3(i) *are* sufficient to determine the notion of a point, due to camera rotation, from IMAGE(i) to IMAGE(i+1).

•This research was supported by NIH Grant No. R01 NS14971-02 COM and NSF Grant MCS79-18209.

B. Recovery of Camera Motion

These properties enable the recovery of the camera motion parameters from a displacement field. Given such a field, one technique [5,6] involves searching through the bounded space of rotational parameters (R1(i) ,R2(i) ,R3(D) to determine image displacements, due to rotation, which, when subtracted from the established displacements, yield translational components which nearly intersect at a point corresponding to a FOE.

The inference of these parameters is simplified when environmental motion is constrained. Amoung such constraints are restrictions on the positions of the translational and rotational axes, their relative orientation and motion overtime. Such constraints restrict the motion of image points, simplifying the determination of image motion. Some basic cases are presented below.

III. CASES OF CAMERA MOTION

A. Known Camera Motion

Processing known camera motion (when all 5 parameters and the extent of the translational displacement are given) involves applying the rotation specified by R1(i), R2(i). R3(i) to IMAGE(i) and to determine the position of the translational axis between the rotated IMAGE(i) and Image(i+1). The inference of environmental depth then involves finding the 1 dimensional displacements along the translational flowpaths for image points (fig. 1).

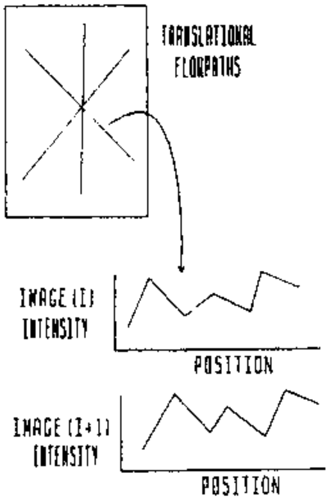

FIGURE 1

These displacements can be determined in several ways. Since image motion is constrained to a known line, temporal-spatial gradient analysis[7,8] (based upon approximating the product dl(s)/dt • ds/dl(s), where l(s) is the intensity at position s along a line) will provide solutions, provided certain requirements concerning image smoothness and the extent of motion hold. Another general technique is determining the optimal match for an image subarea from the rotated IMAGE(i) along a flowpath with respect to IMAGE(i+1). Such a window need only be a ID array of interpolated intensity values along a portion of a flowpath. Figure 2 shows a flow field produced by camera motion with respect to some industrial parts using such a feature. Edge-based descriptors, such as zero-crossings, region boundaries, and intensity iso-contours can also be matched easily.

FIGURE 2

Given a computed depth map and known camera motion, the positions and displacements of image points in later images can be predicted. Differences between predicted and the determined positions can be used to refine the initially computed depth map or supply evidence for some other type of image transformation or independently moving object.

B. Translational Motion

Processing translational motion requires determining a FOE. The techniques for known motion are then applicable. Some techniques for this are 1) searching through T1(i), T2(i) for a FOE and corresponding flowpaths along which the strength and number of matches of extracted features are maximized; 2) Determining the FOE from computed image displacements. For a perfect displacement field, this is given by finding a common intersection. Given an errorful displacement field, approximations to such a point must be found, using, for example, least squares error. Additionally, if the direction of translation stays fixed over several images, then using the straightness of the flow paths as a constraint can filter out bad matches; 3) Using a measure of the consistency of the environmental inferences with a prior surface models to guide FOE search[9].

In figure 3 is shown a determined FOE from a portion of a messy flow field. A least squares intersection is determined for strong matches and used to initialize a more refined search. It is difficult to infer the FOE when it is positioned far from the actual image and computed matches tend to be parallel.

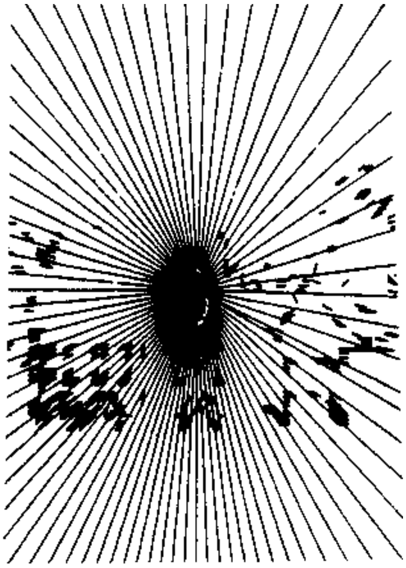

FIGURE 3

C. Planar Motion

When motion is constrained to a known plane P, the axis of rotation (R1(i) ,R2(i)) is determined and the position of the FOE is constrained to a line L, determined by the intersection of the image surface with a plane, parallel to P_f and containing the focal point. Processing involves finding, for a computed displacement field, a value for the extent of rotation R3(i) which yields translational components whose intersections with L form a tight cluster corresponding to a FOE.

This is done by searching through the bounded set of values for R3(i) (the extent of rotation between successive Images is generally limited). The scatter of the intersections of the translational components along L can be used as an error measure. If several displacement vectors are used, the intersections of the translational components can be histogramed along L and a value for R3(i) selected which produces a strong, sharp peak.

IV. APPLICATIONS

A. Robot Control

An industrial robot generally exists in an environment where known or restricted motions occur. Potential uses of motion processing involve image sequences produced from a stationary camera positioned with respect to a moving conveyor belt or one directly attached to a robot arm. I have recently begun the analysis of image sequences produced in this latter case. There is much useful information for collision determination, controlling grasping, and object recognition. There are also considerable technological problems, notably the necessary speeds of processing and the load and bulk of attached cameras.

The load and bulk problems may be resolvable through the use of optic fiber transduction elements. In figure 4, several such elements are shown attached to an idealized cartesian arm. This particular arrangment shows some of the appealing properties of such arms and the use of multiple optic sensors: an encompassing view of the workspace and a direct decomposition of image motion into translational and rotational components by the placement of the sensors.

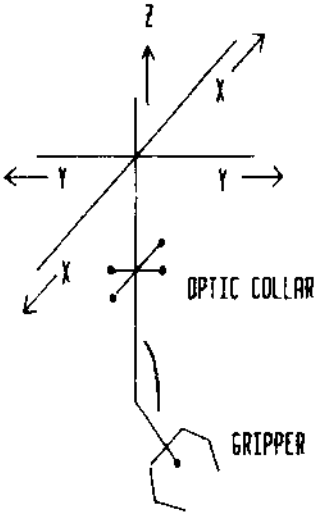

FIGURE 4

B. Motion Knowledge Source for VISIONS

These techniques are being applied towards the development of a Motion Knowledge Source for the UMASS VISIONS system [10,11]. The restrictions of constrained camera motion, with respect to stationary outdoor house scenes are significantly weaker than the semantic constraints currently used to control image interpetation by the VISIONS system.

ACKNOWLEDGEMENTS

I would very much like to acknowledge a wide range of support from the UMASS VISIONS group, and in particular, Ed Riseman, Al Hanson and Ralf Kohler. And thanks especially to Dr. Nelson Corby and his associates from the Intelligent Machine Systems group at the General Electric Corporate Research and Development Center in Schenectady, New York for the use of their CART arm, cameras, skill, and patience.

REFERENCES

[1] Lee.D.N., "The Optic Flow Field: The Foundation of Vision",Phil.Tran3.R.Soc.Lond.B 290,169-179.1980.

[2] Nakayama.K. and Loomis,J.M., "Optical Velocity Patterns, Velocity Sensitive Neurons and Space Perception: A Hypothesis", Perception 3, 63-80.

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

[4] Lee.D.N. ."Visual Information during Locomotion", in perception: Essays in honor of James J. Gibson (ed.R.B. MacLeod and H.L. Pick Jr.),pp. 250-267. Ithaca: Cornell University Press,1974.

[5] Lawton.D.T., Processing Dynamic Images and the Control of Robot Behaviour,Ph.D. Thesis in preparation, Univ. of Massachusetts, Amherst ,MA.

[6] Prazdny.K., "Determing the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinearly Moving Observer", Proc. IUW(Darpa) April, 1981.

[7] Thompson, W.B. and Barnard, S.T.. "Low-Estimation and Interpretation of Visual Motion", draft paper, 1981.

[8] Fennema, C.L. and Thompson, "Velocity Determination in Scenes Containing Several Moving Objects," Computer Graphics and Image Processing, vol. 9. PP. 301-315, April 1979-

[9] Williams, T.D., "Depth for Camera Motion in a Real World Scene," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-2, pp7~511-516, November 1980.

[10] Parma,C.C., Hanson.A.R., and Riseman,E.M., Experiments in Schema-Driven Interpretation of a Natural Scene. COINS Technical Report 80-10, University of Massachusetts, Amherst. April, 1980.

[11] Hanson.A.R. and Riseman, E.M. (eds.), Computer Vision Systems, Academic Presss, 1978.