CONSTRAINT-BASED INFERENCE FROM IMAGE MOTION

Daryl T. Lawton
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

We deal with the inference of environmental information (position and velocity) from a sequence of images formed during relative motion of an observer and the environment. A simple method is used to transform relations between environmental points into equations expressed in terms of constants determined from the images and unknown depth values. This is used to develop equations for environmental inference from several cases of rigid body motion, some having direct solutions. Also considered are the problems of non-unique solutions and the necessity of decomposing the inferred motion into natural components.

Inference from optic flow is based upon the analysis of the relative motions of points in images formed over time. Here we deal with environmental inferences from optic flow for several cases of rigid body motion and consider extentions to linked systems of rigid bodies. Since locality of processing is very important, we attempt to determine the smallest number of points necessary to infer environmental structure for different types of motion.

I INTRODUCTION

The processing of motion information from a sequence of images is of fundamental importance. It allows the inference of environmental information at a low level, using local, parallel computations across successive images. Our concern is with processing a particular type of image motion, termed optic flow, to yield environmental information. Optic flow [1] is the set of velocity vectors formed on an imaging surface by the moving projections of environmental points. It is important to note that there are several types of image transformations, caused by environmental motion, which are not optic flow. For example, image lightness changes due to motion relative to light sources, the motion of features produced by surface occlusion, moving shadows, and a host of transduction effects. The occurrance of these different types of image transformations requires explicit recognition so the appropriate inference technique can be applied for each.

This work was supported by NIH Grant No. R01 NS14971-02 COM and ONR Grant No. N00014-75-C-0459.

II CAMERA MODEL AND METHOD

The camera model is based upon a 3-D Cartesian coordinate system whose orgin is the focal point (refer to figure 1 throughout this section). The image plane (or retina) is positioned in the positive direction along, and perpendicular to, the Z-axis. The retinal coordinate axes are A and B. They are aligned with, and parallel to, the X and Y axes respectively. For simplicity and without loss of generality, the focal length is set to 1.

A point indexed by the number i in the environment at time m is denoted by Pmi. The time index will generally correspond to a frame number from a sequence of images. The projection of an environmental point Pmi onto the retina is determined by the intersection of the retinal surface with the line containing the focal point and Pmi. The position of this intersection in the 3-D coordinate system is represented by the position vector Imi. In this paper, any subscripted I, A, or B, is a constant determined directly from an image. The significant relations concerning Pmi and Imi are

1) $$\mathbf{P}_{mi} = \left(x_{mi}, y_{mi}, z_{mi}\right)$$

2) $$\mathbf{I}_{mi} = \left(a_{mi}, b_{mi}, 1\right)$$

3) $$\mathbf{I}_{mi} = \left(\frac{x_{mi}}{z_{mi}}, \frac{y_{mi}}{z_{mi}}, 1\right)$$

4) $$\mathbf{P}_{mi} = z_{mi}\mathbf{I}_{mi}$$

In the method used here, Equation 4 is used to transform expressed relations between environmental points into a set of equations in terms of image position vectors and unknown Z values. Solving these equations yields a set of Z values which provide a consistent interpretation for the positions, over time, of the corresponding set of environmental points under the assumed relations.

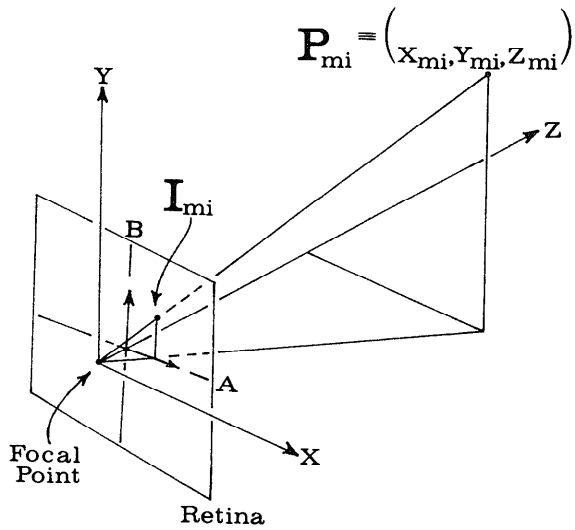

Fig. 1

III INFERENCE FROM RIGID BODY MOTION

A. Arbitrary Motion of Rigid Bodies

The constraint equations developed for this case reflect the preservation of distances between pairs of points on a rigid body during motion. For two points i and j on a rigid body at times m and n, the preservation of distance yields

$$5) \quad \left|P_{mi}-P_{mj}\right| = \left|P_{ni}-P_{nj}\right|$$

which expands into the image-based equation

$$6) \quad \begin{aligned} & Z_{mi}^2\left(I_{mi}\cdot I_{mi}\right)+Z_{mj}^2\left(I_{mj}\cdot I_{mj}\right) \\ & -2Z_{mi}Z_{mj}\left(I_{mi}\cdot I_{mj}\right) \\ & -Z_{ni}^2\left(I_{ni}\cdot I_{ni}\right)-Z_{nj}^2\left(I_{nj}\cdot I_{nj}\right) \\ & +2Z_{ni}Z_{nj}\left(I_{ni}\cdot I_{nj}\right) \\ & = 0 \end{aligned}$$

To determine a solution, we find the minimum number of points and frames for which the number of independent constraints (in the form of equation 6) generated equals or exceeds the number of unknown Z values. It is then necessary to solve the resulting set of simultaneous equations. Note that each such constraint is a second degree polynomial in 4 unknowns.

We begin with the number of unknown Z values. For N (N>2) points in K (K>1) frames there are (NK)-1 unknown Z values. The minus 1 term reflects the degree of freedom due to the loss of absolute scale information. Thus, one of the Z-values can be set to an arbitrary value.

The number of rigidity constraints generated by a set of N (N>2) points in K (K>1) frames is the product of 3*(N-2) and (K-1). The first term is the minimum number of unique distances which must be specified between pairs of points, in a body of N points, to assure its rigidity. Thus 4 points require 6 pairwise distances (all that are possible). For configurations of more than 4 points, its is necessary to specify the distance of each additional point to only 3 other points to assure rigidity. The second term is the number of interframe intervals. Each distance specified must be maintained over each interframe interval.

The number of constraints is greater or equal to the number of unknowns when

$$7) \quad NK-1 \leq 3(K-1)(N-2)$$

$$8) \quad 0 \leq 2NK-6K-3N+7$$

Thus minimal solutions (But not necessarily unique! see below) can be found when (N=5,K=2,number of constraint equations=9) or (N=4,K=3,number of constraint equations=12), in agreement with [2].

The rigidity equations can be simplified by adding restrictions on allowable motions of environmental points. In the following sections we investigate two such restrictions.

B. Motion Parallel to the XZ Plane

Here the Y component of an environmental point is assumed to remain constant over time. Otherwise its motion is unrestricted. This corresponds to an observer moving along an arbitrary path in a plane, maintaining his retina at an orientation perpendicular to the plane, with the motion of objects also so restricted. For point i at times m and n this is expressed as

$$9) \quad Y_{mi} = Z_{mi}B_{mi} = Z_{ni}B_{ni} = Y_{ni}$$

$$10) \quad Z_{ni} = Z_{mi}\frac{B_{mi}}{B_{ni}}$$

This allows a substitution, for points i and j, which simplifies the rigidity constraint to

$$11) \quad Z_{mi}^2\left[\left(I_{mi}\cdot I_{mi}\right)-\left(\frac{B_{mi}}{B_{ni}}\right)^2\left(I_{ni}\cdot I_{ni}\right)\right] + Z_{mj}^2\left[\left(I_{mj}\cdot I_{mj}\right)-\left(\frac{B_{mj}}{B_{nj}}\right)^2\left(I_{nj}\cdot I_{nj}\right)\right] + Z_{mi}Z_{mj}\left[2\left(\left(\frac{B_{mi}}{B_{ni}}\right)\left(\frac{B_{mj}}{B_{nj}}\right)\left(I_{ni}\cdot I_{nj}\right)-\left(I_{mi}\cdot I_{nj}\right)\right)\right] = 0$$

where the bracketed expressions are constants determinable from an image. This case has a direct solution using 2 points in 2 frames. To see this, consider points 1 and 2 at times 1 and 2. This yields a system of 4 unknowns: Z11,Z12,Z21,Z22. The substitution allowed by equation 10 reduces it to a system of 2 unknowns, Z11 and Z12. Z11 can then be set to an arbitrary value, reflecting scale independence. Z12 is then determined from a constraint of the form of equation 11 relating Z12 and the evaluated variable Z11. This is a quadratic equation of Z12.

C. Translations

The constraint expressing the translation of points i and j on a rigid body at times m and n is

$$12) \quad P_{mi}-P_{mj}=P_{ni}-P_{nj}$$

$$13) \quad P_{mi}=P_{mj}+P_{ni}-P_{nj}$$

where the operation is vector subtraction. This reflects the preservation of length and orientation under translation. Setting Zmi to a constant value C, to reflect scale independence, in equation 13 yields 3 simultaneous linear equations in 3 unknowns

$$14) \quad \begin{aligned} CA_{mi} &= Z_{mj}A_{mj}+Z_{ni}A_{ni}-Z_{nj}A_{nj} \\ CB_{mi} &= Z_{mj}B_{mj}+Z_{ni}B_{ni}-Z_{nj}B_{nj} \\ C &= Z_{mj}+Z_{ni}-Z_{nj} \end{aligned}$$

Thus environmental inference from translation requires 2 points in 2 frames. A potential implication of this case is for interpreting arbitrary, and not necessarily rigid body, environmental motion. If the resolution of detail and the rate of image formation relative to environmental motion are both very high, then, in general, the motion of nearby points in images can be locally approximated as the result of translational motion in the environment.

D. Solving the Constraints

The rigidity constraints are easily differentiable and can be solved using conventional optimization methods (taking care to avoid the solution where all the Z-values equal zero). There are, however, in the case of arbitrary rigid body motion, generally many solutions. Here we consider ways of dealing with this.

One way utilizes feedforward. It is crucial to note that the rigidity equations needn't be solved anew at each point in time. If the environmental structure has been determined at time t and an image is then formed at time t+1, half the unknowns in the system of constraint equations disappear. This greatly simplifies finding a solution. Additionally, the solution process can be further simplified by extrapolation of infered motion, if enough frames have been processed. But how can the positions of the environmental points be determined initially? 1). Prior knowledge of the environment could supply the initial estimates of the relative positions. 2). There may be a small number (perhaps less than 50) of generic patterns of image motion (which may be termed flow fields), each associated with a particular class of environmental motion. For example, translational motion is characterized by straight motion paths which radiate from or converge to a single point on the retina. Other flow fields we have analyzed also have such distinguishing characteristics. These characteristics would be used to recognize particular types of image motion, associated with particular types of environmental motion, to initialize and constrain the more detailed solution process based upon solving the constraints for arbitrary rigid body motion. 3). The observer could constrain his own motion for one sampling period to a case of motion for which environmental structure can be unambiguously determined. For example, by stabilizing the retina of a moving observer with respect to rotations relative to a stationary environment, all image motion could be interpreted as the result of translation.

Another possibility is to use more than the minimum required number of points in the inference process to supply additional constraints.

IV APPROACHES TO OTHER CASES OF MOTION

A. Sub-Minimal Rigid Configurations

A sub-minimal configuration is one consisting of 1,2 or 3 points over an arbitrary number of frames or 4 points in 2 frames. Human subjects can get an impression of 3-D rigid motion from displays of such configurations [4], even though there are not sufficient generated constraints, by the above analysis, for a solution. How is this possible?

Other assumptions must be used for the inference. A potential one reflects an assumption of smoothness in the 3-D motion. This can be had by minimizing an approximation of the acceleration of a given point. For a point i at times t, t+1, t+2 this can be expressed as

$$15) \qquad |P_{ti} - P_{t+1,i}| - |P_{t+1,i} - P_{t+2,i}|$$

Perhaps a sum of such expressions, formed using the substitution of equation 4, should be minimized for a set of points over several time periods along with the satisfaction of expressible rigidity constraints.

B. Johansson Human Dot Figures

Experiments initiated by Johansson have shown the ability of subjects to infer the structure and changing spatial disposition of humans performing various tasks with only joint positions displayed over time [5]. Here we consider how such inference could be performed.

First, it is necessary to determine which points are rigidly connected in the environment. Work by Raschid [6] as shown that this is possible on the basis of image properties only, using relative rates of motion between images. That is, without any inference of environmental structure.

Given the determination of rigid linkages, it is necessary to find the relative spatial disposition of the limbs. An approach is to infer environmental position, for each limb, using the rigidity constraints and optimizing the smooth motion measure discussed above. If the figure is recognized as being human, several object specific constraints can also be used. These would involve such things as allowable angles of articulation between limbs, their relative lengths, and body symmetries.

ACKNOWLEDGEMENTS

I greatly appreciate Ed Riseman, Al Hanson, and Michael Arbib for their support and the intellectual environment they have created. I would also like to acknowledge the last minute (micro-nano-second!) heroics of Tom Vaughan, Earl Billingsley, Maria de LaVega, Steve Epstein, and Gwyn Mitchell.

REFERENCES

[1] Gibson, J.J. The Perception of the Visual World. Boston: Houghton Mifflin and Co. 1950.

[2] Ullman, S. The Interpretation of Visual Motion. Cambridge, Massachusetts: The MIT Press. 1979.

[3] Rogers, D.F. and Adams, J.A. Mathematical Elements for Computer Graphics. McGraw-Hill Book Company. 1976.

[4] Johansson, G., and Jansson, G. "Perceived Rotary Motion from Changes in a Straight Line." Perception and Psychophysics, 1968, Vol. 4 (3).

[5] Johansson, G, "Visual Perception of Biological Motion and a Model for its Analysis." Perception and Psychophysics 14:2 (1973) 201-211.

[6] Rashid, R.F. "Towards a System for the Interpretation of Moving Light Displays", Technical Report 53, Department of Computer Science, University of Rochester, Rochester, New York 146727, May 1979.