Proceedings of the Ninth International Joint Conference on Artificial Intelligence
18-23 August 1985
Los Angeles CA

UNIQUE RECOVERY OF MOTION AND OPTIC FLOW VIA LIE ALGEBRAS

*A. Peter Blicher**
Computer Science Department, Stanford University, Stanford, CA 94305
and *Mathematics Department, University of California, Berkeley*
and
*Stephen M. Omohundro***
Physics Department, University of California, Berkeley

ABSTRACT

We use some ideas from the theory of Lie groups and Lie algebras to study the problem of recovering rigid motion from a time-varying picture. We are able to avoid the problem of finding corresponding points by considering only what can be determined from picture point values and their time derivatives. We do not assume that we can track individual points in the image, nor that we are given any of their velocities (i.e., the optic flow). Among our results are:

The 6 point df/dt theorem, showing that generically*** the values of df/dt at 6 points of the monochrome image f are necessary and sufficient to specify the motion of a given object.

The 2-color theorem for optic flow, which states that the optic flow vector is uniquely specified at a generic point of the image if there are 2 or more color dimensions.

Also, we get the color version of the 6 point theorem, the **2 colors, 3 points corollary**, which reduces the number of points required to 3, if there are at least 2 color dimensions.

I INTRODUCTION

For the past several years, many researchers have been investigating problems of moving objects and observers (see e.g., [Tsai and Huang 1984], [Prazdny 1983], [Nagel 1983], [Horn and Schunck 1980], [Bruss and Horn 1983], [Ullman 1979]). A conventional paradigm is to consider 2 subproblems: finding the optic flow in the image, then computing 3-dimensional motion. Finding the optic flow in monochrome images by point tracking is, however, degenerate except for special points, just as for the point matching problem [Blicher 1983, Blicher 1984]. E.g. at a single point, the image function and its time derivative tell us nothing about motion perpendicular to the gradient of the image function.

We consider a rigid object undergoing an arbitrary motion in space. Our data is a time-varying image, i.e. a map $f : \mathbf{I} \times M^2 \to \mathbf{R}^n$, where $\mathbf{I}$ is a time interval, M^2 is some 2-dimensional manifold, specifically the image plane, and n is the number of independent color dimensions; $n = 1$ for monochrome pictures. We concern ourselves here with the problem of finding the motion of the object, particularly, how much data is necessary and sufficient. Rather than make assumptions about first finding point correspondences or optic flow, we consider the full situation of a map from the rigid motion group to the time-varying image (but only for the interior of a single object), and we develop the differential theory, based on the data of the picture and its time derivative.

We regret that space limitations preclude defining mathematical terms. A fuller presentation, as well as more extensive references, can be found in [Blicher 1984].

II THE MATHEMATICAL STRUCTURE

The situation is that of Fig. ($*'$), just as in [Blicher 1983], except now the nature of the transformation g will be paramount.

We are interested in rigid motions in $\mathbf{R}^3$, so $g \in E(3)$, the Euclidean (rigid motion) group of $\mathbf{R}^3$. The time evolution of the motion is then given by $\gamma : \mathbf{R} \to E(3)$, i.e., as a path in the transformation group. In fact, γ defines a 1-parameter family of transformations. Since we are interested only in small changes from the current state, we take $\gamma(0) = I$, the identity in $E(3)$ (we could have done this anyway by using the group structure to translate back to the identity). For every t, γ gives a rigid motion of $\mathbf{R}^3$, since we are identifying $E(3)$ with the rigid motions of $\mathbf{R}^3$, i.e. $\gamma(t) : \mathbf{R}^3 \to \mathbf{R}^3$. Each point of $\mathbf{R}^3$ is carried along with this motion, and describes a path in $\mathbf{R}^3$ (defined by $\gamma_p(t) = (\gamma(t))(p), \quad p \in \mathbf{R}^3$). In particular, every point of our surface of interest, embedded in $\mathbf{R}^3$, has such a path. Now apply the imaging projection, and restrict attention only to the visible surface of the embedded object. By composition, this leads to a path through each point that gets hit in the image. (defined by $\tilde{\gamma}_q(t) = \pi((\gamma(t))(p)), \quad q \in M^2$). Now consider only a single time, $t = 0$. The structure we have presented thus far is summarized in Fig. (flow).

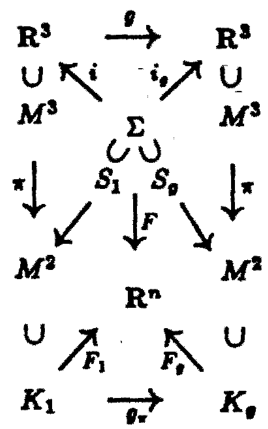

Fig. ($*'$)

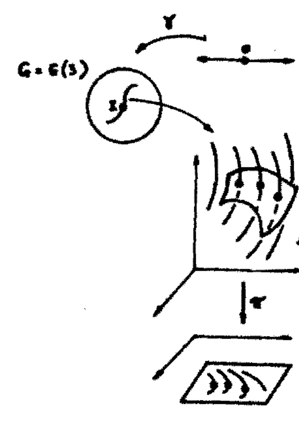

Fig. (flow)

Each such path in the picture has a velocity vector, and each point in the image has a path, so there is a vector field defined on the image. This is usually called the *optic flow*, but it is more consistent with mathematical terminology to call its integral, i.e. the paths in the image, the optic *flow*. We will reserve the term *optic flow* for this integral, i.e. the map $\varphi_t : U \to \mathbf{R}^2$ which specifies the paths of corresponding points in the picture with initial points in the region U, while using *optic velocity field* or *optic vector field* for its instantaneous velocities, the vectors $d\varphi_t/dt$. Similarly, the paths in $\mathbf{R}^3$ define a vector field on $\mathbf{R}^3$, and the path γ in $E(3)$ defines a tangent vector at the identity in $E(3)$.

The available data, however, is not the optic flow or vector field, but the time-varying picture function f_t which is just the projection of the intrinsic surface function F, which we assume is carried along with the motion, i.e. we neglect changes in f due purely to photometric effects, such as specular reflection. Since we are considering only the differential theory, we regard our data as telling us only the instantaneous value f_0, and all the time derivatives at $t = 0$. This is the same as knowing the Taylor series for f_t. We will only use the 1st

*This work was supported in part by DARPA contracts N00039-82-C-0250 and N00039-84-C-0211.

**Present address: Thinking Machines, Cambridge, Mass.

***A *generic* property is one which is true for a *typical* element of a space, i.e for a very dense subset of the space. For this paper, we take this to mean an open dense subset. See [Blicher 1984] for a discussion of genericity.

derivative. At a point p of the image, call the optic flow vector v. Then in a frame with velocity v at p in the image, f_t does not appear to change; the optic flow specifies the motion of corresponding points. Thus if we leave the frame fixed, we see that

$$\frac{d}{dt} f_t(p) = -D_v(f_t)(p) = -v \cdot \nabla f_t(p), \qquad (*)$$

where D_v means differentiation by the vector v, equivalent to $v \cdot \nabla$. (This is well-known in the context of optic flow; see e.g. [Horn and Schunck 1980], [Ballard and Brown 1982].) Equation (*) shows how it is that we only have partial information about v: we only know 1 component. We can immediately see, also, that if f had multiple dimensions, i.e. if there were more than 1 color dimension, we would have information about multiple components, and v would be uniquely determined for generic f. This is the differential version of the 2-color theorem we have proved earlier [Blicher 1983, Blicher 1984]. Finding optic flow, like matching, is much easier with color. We formalize this in

Theorem. (2-color theorem for optic flow) For a generic time-varying image function $f_t : M^2 \to \mathbf{R}^n$, the optic flow vector is uniquely specified at a generic point of the image if $n \geq 2$, i.e. for 2 or more color dimensions.

When we fix $t = 0$, each side of equation (*) is just a number, so for each p we have a map $D_\cdot(f)(p) : v \mapsto$ *a real number*. We have thus defined a string of linear mappings (*v.f.* stands for *vector field*, *v.b.* for *vector bundle*):

tangent vector on $E(3) \mapsto$ *v.b. section on object* $\mapsto$

v.f. on image $\mapsto$ *vector at* $p \mapsto$ *real number*

(We must consider *sections* of a vector *bundle* on the object rather than vector *fields* (sections of the tangent bundle) because the vectors we are interested in are tangent vectors to paths in $\mathbf{R}^3$ going through points of the object. Since the paths generally do not lie in the object, their tangent vectors needn't be in the tangent space of the object, but rather are merely tangent vectors in $\mathbf{R}^3$.)

The Lie algebra $\mathfrak{g}$ of a Lie group G is a vector space which can be identified with the tangent space of G at the identity. $E(3)$ is a Lie group, and therefore associated with it is the Lie algebra $\mathfrak{e}(3)$; and since $E(3)$ is a 6-dimensional manifold, $\mathfrak{e}(3)$ is a 6-dimensional vector space. The tangent vector $\gamma'(0)$, which is the instantaneous motion, can therefore be thought of as an element of the Lie algebra $\mathfrak{e}(3)$.

We can do this for every path γ, hence for every element of $\mathfrak{e}(3)$, giving us a homomorphism from the Lie algebra $\mathfrak{e}(3)$ to sections of the vector bundle on the object, and likewise again to a Lie algebra of vector fields on the image of the object in the image plane. The composition of these is a Lie algebra homomorphism. The sequence of linear maps can therefore be written

Lie algebra $\mathfrak{e}(3) \to$ *v.b. sections on object* $\to$

v.f.'s on image $\to$ *vectors at* $p \to$ *real numbers*

This defines a map $\mathfrak{e}(3) \to \mathbf{R}$, i.e. an element of $\mathfrak{e}^*(3)$, the dual of $\mathfrak{e}(3)$.

Now we have enough machinery to attack some questions. The first question is whether there is enough information in df/dt to uniquely specify the instantaneous motion, for generic f. The instantaneous motion is an element of $\mathfrak{e}(3)$. As we just saw, for each point p of the image, the geometry defines an element of $\mathfrak{e}^*(3)$. The question then becomes whether we can span all of $\mathfrak{e}^*(3)$ by ranging over all points of the image, for knowing the value of applying a dual basis in $\mathfrak{e}^*(3)$ uniquely specifies the original vector in $\mathfrak{e}(3)$. $\mathfrak{e}^*(3)$ is 6-dimensional, so if this is possible, it is possible for 6 points corresponding to a dual basis. This doesn't say anything yet about finding the shape or position of the object; we only want to know whether we can recover the motion for fixed shape and position.

Theorem (6 point df/dt theorem). Let

$$f : \mathbf{I} \times U \to \mathbf{R}$$
$$(t, p) \mapsto f(t, p)$$

be a time-varying picture for some time interval $\mathbf{I}$ around 0, and some neighborhood U in the image plane of regular values of the imaging projection of some 2-dimensional object (i.e. a 2-manifold) embedded in $\mathbf{R}^3$. If f comes from the projection of a generic intrinsic function on an object undergoing rigid motion in $\mathbf{R}^3$, then the values of $\partial f/\partial t(0, p)$ at 6 generic points $p \in U$ are necessary and sufficient to uniquely specify the instantaneous motion of the object.

Proof. We are in effect measuring the optic velocity field with our image function; this is what equation (*) says. To be able to tell the difference between different elements of $\mathfrak{e}(3)$, i.e. different motions, the mapping from $\mathfrak{e}(3)$ to velocity fields on the picture must be 1-1. Since the mapping is a vector space homomorphism, this is the same as saying it has no (nontrivial) kernel. The homomorphism $\mathfrak{e}(3) \to$ *v.b. sections on object* has no kernel, because any kernel would leave the entire object fixed, but a rigid motion of $\mathbf{R}^3$ can leave at most a line fixed. So $\mathfrak{e}(3)$ is mapped 1-1 to sections of bundles on the object. Now we must show that the kernel of the homomorphism *v.b. sections on object* $\to$ *v.f.'s on image* doesn't contain anything that comes from the previous map from $\mathfrak{e}(3)$. The kernel of the current map is just the sections whose vectors lie along the rays of projection to the picture. For orthogonal projection, vertical translation would of course be in this kernel, but we are assuming a projective projection, i.e. that the rays all meet at a point; for a planar retina this is the usual perspective projection.

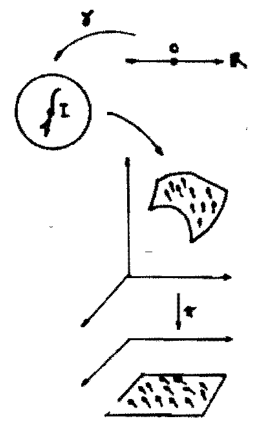

Fig. (vector fields)

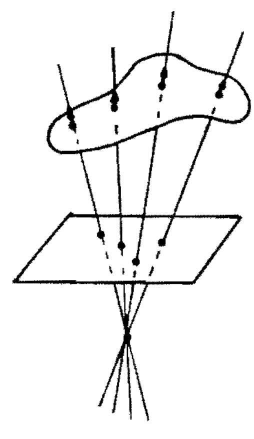

Fig. (kernel-rays)

We have to show that any such motion, where points move only along rays, cannot come from a rigid motion. This is easy to see; take 3 points a, b, c on the object not all on the same line in $\mathbf{R}^3$. Since a rigid motion of $\mathbf{R}^3$ can only leave a single line axis (or nothing) fixed, at least 1 of the points must move, say a. If a moves down (toward the image plane), b must move up, to keep their distance constant (rigid motion). Since b is moving up, c must move down. But then a and c are both moving down and therefore narrowing their distance, showing that the motion cannot be a rigid motion, i.e. the kernel of *v.b. sections on object* $\to$ *v.f.'s on image* is not in the image of $\mathfrak{e}(3) \to$ *v.b. sections on object* (except for 0, of course). So we know that the composition $\mathfrak{e}(3) \to$ *v.f.'s on image* has no kernel, i.e. is 1-1. This means that every rigid motion gives a unique optic velocity field, and the vector space of such fields is 6-dimensional.

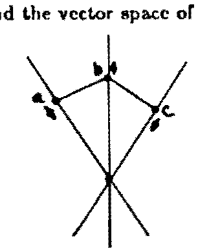

Fig. (3 points)

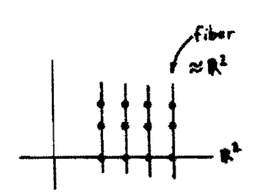

Fig. (3 fibers)

Actually, we showed more than that. We showed that a generic set of 3 points cannot stay fixed in the image—we didn't even have to consider the whole vector field. The set of vectors at 3 such points in the image make up a 6-dimensional vector space, so what we showed is that the map $\epsilon(3) \rightarrow$ *vectors at 3 given points in image* has no kernel, i.e. is 1-1.

That means that to specify a motion, i.e. an element of $\epsilon(3)$, we only have to figure out the optic velocity vectors at 3 points. A generic function, via equation (*), tells us 1 component of each of the vectors (by genericity, the gradient is nonzero at all 3 points). If we had 2 generic functions, then we could recover both components of each of the 3 vectors by using equation (*) for both functions (generically, the gradients will be linearly independent, i.e. in different directions at the 3 points). Parenthetically, we have just proved

Corollary (2 colors, 3 points). For generic f taking values in 2 or more color dimensions, the values of $\partial f/\partial t(0,p)$ at 3 noncollinear points $p \in U$ are necessary and sufficient to uniquely specify the instantaneous motion of the object.

Now we must show that 1 component at each of 6 points is as good as 2 components at each of 3 points. We saw earlier that df defines an element of $\epsilon^*(3)$. Thus the geometry defines a map $T^*\mathbf{R}^2 \rightarrow \epsilon^*(3)$. What we saw earlier is

Lemma (3 fiber lemma). If we choose 3 generic points in $\mathbf{R}^2$, and 2 linearly independent covectors in each fiber over those points, the 6 resulting points of $T^*\mathbf{R}^2$ are mapped to a spanning set in $\epsilon^*(3)$.

What we will now show is that we can choose *any* 6 generic points in $T^*\mathbf{R}^2$, i.e. 6 generic points in the image, and 6 generic values of df at those points (i.e. a generic f). This is pretty easy by making use of the 3 fiber lemma. The lemma still applies for any neighborhood of $\mathbf{R}^2$, i.e. we can choose the 3 points arbitrarily close together. This gives us

Lemma (local spanning). Every neighborhood of every point in $T^*\mathbf{R}^2$ contains 6 points which are mapped to a spanning set in $\epsilon^*(3)$.

Proof. Choose a point and neighborhood in $T^*\mathbf{R}^2$. It projects to a neighborhood of $\mathbf{R}^2$, in which we can choose 3 generic points. We can then choose 6 points in $T^*\mathbf{R}^2$, 2 to a fiber, by the 3 fiber lemma. QED (local spanning).

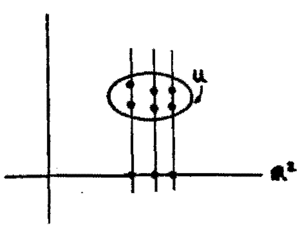

Fig. (local spanning)

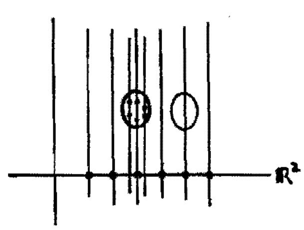

Fig. (6 points)

Now we can see what happens when we choose 6 points in the image. df gives us 6 points in $T^*\mathbf{R}^2$. We can perturb these points to guarantee that $df \neq 0$. Now since every neighborhood of each point maps to a spanning set of $\epsilon^*(3)$ (local spanning lemma), we can always perturb the nth point so that it is mapped to something outside the span of the first $n-1$ points (at least through $n = 6$, anyway). This gives a perturbation of the 6 points which maps to a spanning set. Since spanning sets are open, these points will still span under sufficiently small perturbation. (In general, one might need a perturbation of both the location of the points and of f to guarantee a spanning set. The degenerate situation occurs when the optic velocity vector is in the direction of constant f.) QED∎

III AFTERWORD

By virtue of the local spanning lemma and the 3 fiber lemma, our results are local, i.e. they hold in an arbitrarily small neighborhood—generically every neighborhood has 6 points yielding sufficient data. This is significant because it implies that an estimate of the motion can be obtained from any neighborhood. In practice, of course, using a very small neighborhood would lead to a very bad estimate. One would rather use many points over a large region to obtain a least squares estimate. But the localness means that estimates can be made over a range of scales, and that a procedure for segmentation based on local estimates is well-founded.

ACKNOWLEDGMENTS

We are grateful to Tom Binford for his help and encouragement throughout the period of this work.

REFERENCES

[Ballard and Brown 1982]
Ballard, Dana Harry and Christopher M. Brown, **Computer Vision**, Prentice-Hall, Englewood Cliffs, 1982. [TA1632.B34, ISBN 0-13-165316-4].

[Blicher 1983]
Blicher, A. Peter, "The Stereo Matching Problem from the Topological Viewpoint," *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 1983, 1046-1049.

[Blicher 1984]
Blicher, A. Peter, "Edge Detection and Geometric Methods in Computer Vision," Ph.D. thesis, Mathematics Department, University of California, Berkeley, December 1984. Also Stanford University Computer Science Department technical report STAN-CS-85-1041 and AI memo AIM-352, February 1985.

[Bruss and Horn 1983]
Bruss, Anna R. and Berthold K.P. Horn, "Passive Navigation," *Computer Vision, Graphics, and Image Processing*, 21, 1983, 3–20.

[Horn and Schunck 1980]
Horn, Berthold K.P. and Brian G. Schunck, "Determining Optical Flow," MIT AI Memo 572, MIT Artificial Intelligence Laboratory, April 1980; also *Artificial Intelligence*, 17, 1981, 185-203.

[Nagel 1983]
Nagel, Hans-Hellmut, "Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, 1983, 85–117.

[Prazdny 1983]
Prazdny, K., "On the Information in Optical Flows," *Computer Vision, Graphics, and Image Processing*, vol. 22, 1983, 239–259.

[Tsai and Huang 1984]
Tsai, Roger Y. and Thomas S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 1, 1984, 13–27.

[Ullman 1979]
Ullman, Shimon, **The Interpretation of Visual Motion**, MIT Press, Cambridge, 1979. [BF241.U43 ISBN 0-262-21007-X].