Rigid Body Motion from Depth and Optical Flow

Dana H. Ballard and O.A. Kimball
Computer Science Department
University of Rochester
Rochester, NY 14627

TR70

November 1981

Abstract

Motion is an important cue that facilitates the perception of rigid bodies. This perception can be viewed as finding the correct values for the nine parameters necessary to describe rigid body motion. These parameters are computable in parallel from depth and optical flow information. When coupled to the flow computations, the rigid body computations can resolve difficult singularities in the flow calculations.

The preparation of this paper was supported in part by the National Institutes of Health under Grant No. 2R01 HL21253-04.

Table of Contents

1. Introduction

Low level vision may be viewed as an active process of building explicit parsimonious descriptions of physical parameters from image intensities [Marr, 1978]. For human beings, there is much compelling evidence that the computation is done in parallel and that it involves "indexing into" or activating existing structures. Thus the challenge of modeling the initial steps in human vision, cast in this framework, is to specify parallel algorithms which activate explicit descriptions of the world.

Motion, which at one time was considered a by-product of segmenting static sequences of images by using color and form cues, has come to be regarded as a separate perceptual channel. Much recent neurophysiological and psychological work supports this viewpoint [Ullman, 1980; Gibson, 1979; Lishman, 1981]. Thus there is hope that motion can be studied *ad inizio* without involving color and form. We take this view, defering studies of the couplings between these different sensory cues.

Of all the common motions, rigid body motions are pervasive. Also, in many cases where the motion is non-rigid, for example in human movements, an articulated rigid body will suffice. Rigid body motion is completely specified at any instant by nine parameters: three to describe the body's local coordinate origin with respect to a reference coordinate frame; three to describe the translational velocity of the local origin; and three to describe the rotational velocity. In general, these parameters will change with time, especially if the body is subjected to external forces. These parameters together with knowledge of the body's surface allow the velocity of points on the body to be determined. This is important, since it provides a constraint whereby known motions can affect the computation of velocities derived from intensity data. However, it also serves as a constraint for the inverse problem, which is: given velocity data for points on the rigid body, compute the nine parameters describing the motion. In general, not all the velocity data will be from a single body, so the computational techniques for finding the parameters must be able to segment just those velocities which are relevent to the rigid body description. The main thrust of this paper is to show how these parameters can be computed.

The above discussion assumes that three-dimensional velocity information is available from the time-varying image. Other parametric approaches to representing rigid body motion have used the two-dimensional optic flow image [Prazdny, 1981; O'Rourke, 1981]. In particular situations, such as translatory motion, interesting parameters can be computed from the optic flow [Lee and Reddish, 1981]. However, general rigid body motion can not be easily extracted from optic flow without depth information. The three-dimensional velocities that correspond to optic flow are not immediately available, but can be computed from depth (disparity) and optical flow information. We term these velocities *3-d flow*, and show how they can be computed.

Current research has shown the feasibility of computing both depth [Marr and Poggio, 1976; Barnard and Thompson, 1979] and optic flow [Ullman, 1979; Horn and Schunck, 1980] by parallel-local relaxation algorithms. These algorithms work well when boundary conditions are specified a priori but otherwise are underdetermined. The rigid body motion parameters are one way of providing the boundary conditions that make the problem well-specified. It might seem that this situation is paradoxical; that the optic flow is a prerequisite for computing motion parameters, and that the motion parameters are a prerequisite for computing optic flow. However, a technique originally developed for shape-from-shading [Ballard and Kimball, 1981] allows these two computations to be coupled in a way that both should converge to correct values.

Three-dimensional flow can be regarded as a kind of *intrinsic image* [Barrow and Tenenbaum, 1978] in that (a) the parameters represent physical phenomena (psychophysical phenomena are also allowed), (b) the parameters may only be defined for a part of image space, and (c) the parameters can be computed by parallel-local relaxation techniques. The rigid body motion parameters are global parameters that describe the appropriate portion of the 3-d flow. They can be detected by exploiting a general Hough-transform technique for relating intrinsic images and parameters [Ballard, 1981]. In brief: each velocity vector "votes" for its set of possible parameters given the rigid body constraint; the parameter values receiving the most votes are selected to describe the motion.

2. Mathematics

To describe the computations, we first develop the general relations introduced by the perspective geometry, then we briefly review the Newtonian equations for rigid body motion. Next we describe the Horn/Schunck method for computing flow [1980]. As a final prerequisite we review the Hough technique which is a general parallel method for solving distributed constraints.

The Perspective Transform

The perspective transform describes how points in three-dimensional space are projected onto the retina. Referring to Figure 1, an argument using similar triangles shows that the retina position $x' = (x',y')$ is related to a given three-dimensional point $x = (x,y,z)$ by

$$x' = xf/(f-z) \tag{2.1a}$$

$$y' = yf/(f-z) \tag{2.1b}$$

where f is the optical focal length.

Figure 1.

Differentiating these equations with respect to time results in a relationship between optic flow and what we call 3-d flow. The term optic flow describes the field of instantaneous retinal velocities of a set of retinal points $\{x',y'\}$. Optic flow will be represented by a vector (u,v) representing the velocity components in the x' and y' directions, respectively. Three-dimensional flow or the velocities of points in 3-d space will be described by a vector $v(x) = (v_x(x),v_y(x),v_z(x))$. Following the conventions used in dynamics, we will often drop the explicit spatial reference in v(x) using just v to mean the velocity associated with a point x. With these conventions, the result of differentiating Eqs. 2.1 is

$$u = v_x f/(f-z) + xfv_z/(f-z)^2$$

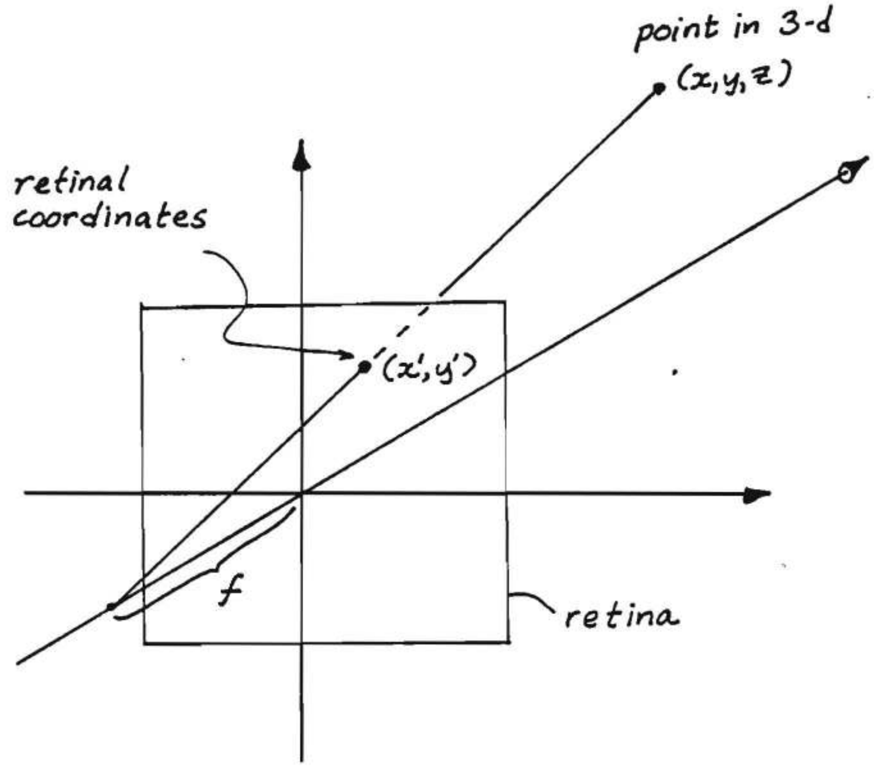


Figure 1.

$$v = v_y f/(f-z) + y f v_z/(f-z)^2$$

These can be simplified using Eqs. (2.1) to:

$$(f-z)u = v_x f + x' v_z \quad (2.2a)$$

$$(f-z)v = v_y f + y' v_z \quad (2.2b)$$

The significance of these equations is that by knowing the retinal information (f,x',y',u,v) and the depth information (z,v_z), the other two components of the three-dimensional velocity vector, v_x and v_y, can be recovered.

Rigid Body Motion

The description of the motion of a rigid body is relative to a reference body. For most of our purposes the reference body will be the viewer, and the viewer coordinate system will be the one shown in Figure 2 that is naturally related to retinotopic coordinates.

Figure 2.

The equations are based on a theorem that the most general displacement of a rigid body can be produced by superimposing, on the translation of the body, its rotation about an axis through a body point B that, because of the translation, has already reached its final position. This point serves as the origin of a coordinate system fixed in the body itself. Every point in the body can be described by a vector ρ described with respect to this frame (see Figure 2). Thus the vector x can be described by the sum

$$x = x_b + \rho \quad (2.3)$$

Thus the velocity $\mathbf{v}$ of the point x can be described as

$$\mathbf{v} = \mathbf{v}_b + d\rho/dt \quad (2.4)$$

which, for a rigid body ($|\rho|$ = constant) can be expressed as

$$\mathbf{v} = \mathbf{v}_b + \Omega \times \rho. \quad (2.5)$$

where Ω is the rotation vector of the body. The vector Ω has the direction of the rotation axis and the mafgnitude of the rotation. Note that from (2.3), $\rho = \mathbf{x} - \mathbf{x}_b$ so that Equation (2.5) expresses the 3-d flow vector at x as a function of nine rigid body parameters: $\mathbf{v}_b$, Ω, and $\mathbf{x}_b$. The acceleration of the body can be obtained by differentiating again

$$a = a_b + (d\Omega/dt) \times \rho + \Omega \times (d\rho/dt) \quad (2.6)$$

which, again using the rigid body constraint of $|\rho|$ = constant, can be expressed as

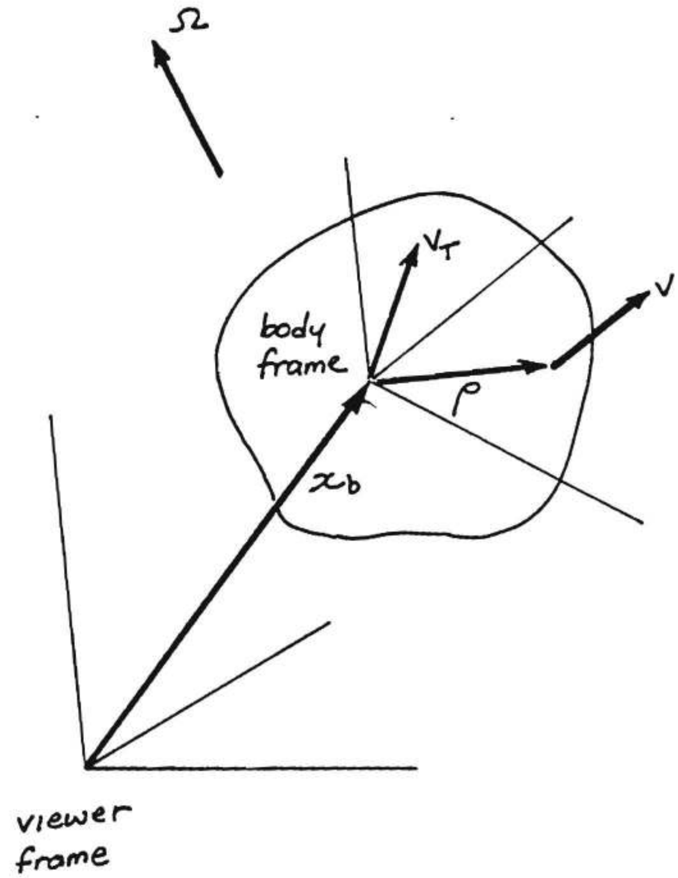


Figure 2.

$$a = a_b + (d\Omega/dt) \times \rho + \Omega \times (\Omega \times \rho) \tag{2.7}$$

or

$$a = a_b + (d\Omega/dt) \times \rho + (\Omega\cdot\rho)\Omega - |\Omega|^2\rho \tag{2.8}$$

Computing Flow

One of the important features of optical flow is that it can be calculated simply, using local information [Horn and Schunck, 1980]. They model the motion image by a continuous variation of image intensity as a function of position and time, then expand the intensity function f(x,y,t) in a Taylor series.

$$f(x+dx,y+dy,t+dt) = f(x,y,t) + f_x\Delta x + f_y\Delta y + f_t\Delta t + \text{higher-order terms} \tag{2.9}$$

where f_x, f_y, and f_t are the appropriate partial derivatives and higher-order terms are ignored. The crucial observation which is to be exploited is the following: if indeed the image at some time $t+\Delta t$ is the result of the original image at time t being moved translationally by Δx and Δy, then in fact

$$f(x+\Delta x, y+\Delta y, t+\Delta t) = f(x,y,t) \tag{2.10}$$

Consequently, from (2.9) and (2.10),

$$-f_t = f_x(\Delta x/\Delta t) + f_y(\Delta y/\Delta t) \tag{2.11}$$

The approximation (2.10) will not hold at boundaries between different objects (or at boundaries arising from a self-occluding object), but may be expected to hold for large patches of the image if the variations Δx, Δy, and Δt are appropriately small. Now the partial derivatives are all measurable quantities, and $\Delta x/\Delta t$ and $\Delta y/\Delta t$ are estimates of the optic flow (u,v). Thus

$$-f_t = \nabla f\cdot v \tag{2.12}$$

where ∇f is the spatial gradient of the image and $\mathbf{v} = (u,v)$ the velocity.

Consider a fixed camera with an otherwise static scene moving past it. Equation (2.12) implies that the *time* rate of change in intensity of a point in the image is (to first order) explained as the *spatial* rate of change in the intensity of the scene multiplied by the *velocity* that points of the scene move past the camera. Furthermore, the velocity (u,v) must lie on a line perpendicular to the vector (f_x,f_y) where f_x and f_y are the partial derivatives with respect to x and y respectively (Fig. 3). In fact, the magnitude component of the velocity in the direction (f_x,f_y) from (2.12) is

$$-f_t/\sqrt{(f_x^2 + f_y^2)}.$$

Equation (2.12) constrains the velocity but does not determine it uniquely. The component in the direction (f_x,f_y) is determined, and the component perpendicular to this direction unknown. To make the problem well-posed, one is motivated to seek a solution that satisfies (2.12) as closely as possible and also is locally smooth. One can measure local smoothness by Laplacians of the two velocity components, i.e., $\nabla^2 u$ and $\nabla^2 v$.

Figure 3.

Horn and Schunck seek a minimum of the flow error expressed as

$$E^2(x,y) = (\text{Error in } (2.12))^2 + \lambda^2(\text{Deviation in Smoothness})$$

or

$$E^2(x,y) = (f_x u + f_y v + f_t)^2 + \lambda^2((\nabla^2 u)^2 + (\nabla^2 v)^2) \tag{2.13}$$

where λ is an appropriately chosen constant. Differentiating this equation with respect to u and v provides equations for the change in error with respect to u and v which must be zero for a minimum. Writing $\nabla^2 u$ as $u-u_{ave}$ and $\nabla^2 v$ as $v-v_{ave}$, these equations are

$$(\lambda^2 + f_x^2)u + f_x f_y v = \lambda^2 u_{ave} - f_x f_t \tag{2.14a}$$

$$f_x f_y u + (\lambda^2 + f_y^2)v = \lambda^2 v_{ave} - f_y f_t \tag{2.14b}$$

which may be solved for u and v yielding

$$u = u_{ave} - f_x P/D \tag{2.15a}$$

$$v = v_{ave} - f_y P/D \tag{2.15b}$$

where

$$P = f_x u_{ave} + f_y v_{ave} + f_t$$

$$D = \lambda^2 + f_x^2 + f_y^2$$

To turn this into an iterative equation for solving u(x,y) and v(x,y) use the following Gauss-Seidel method:

t = 0. Initialize all u(x,y,0), v(x,y,0)

for t = 1 until maxframes do

$$u(x,y,t) = u_{ave}(x,y,t-1) - f_x P/D \tag{2.16}$$

$$v(x,y,t) = v_{ave}(x,y,t-1) - f_y P/D \tag{2.17}$$

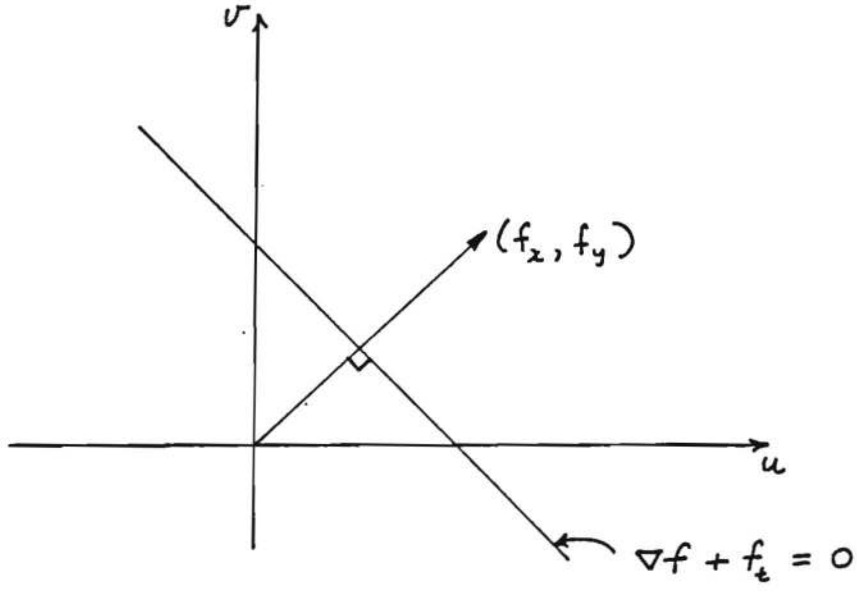


Figure 3.

Horn and Schunck implemented (2.16) and (2.17) and showed that for the case of *static* optical flow, the algorithm converges. Static flow means that although the image data is time-varying, the flow field is not: u and v are only functions of retinal position. One example is a rotating sphere. In the case of dynamic flow (time-varying flow), such as an object translating across a stationary background or a rotating non-sphere, there will be boundaries between the different physical objects. In general the flow equation, which is based on local smoothness, will not hold at these boundaries. If these boundaries could be detected, then the flow algorithm could be limited to smooth regions. Horn and Schunck have shown some methods for detecting these boundaries based on local heuristics [Horn and Schunck, 1981]. We show how the rigid body constraint could be used to find these boundaries.

As a footnote to this technique, we note that the function f need not be *intensity* but could be any parameter which is varying temporally and spatially. Hue is another obvious choice. Derivatives of different functions could also be used in a cooperative manner to improve their estimates.

Hough Methodology

A way of describing the relationship between parts of the retinotopic 3-d flow image and associated non-retinotopic rigid body motion parameters is an interpretation of the *Hough transform* which we have termed **constraint transforms** [Ballard, 1981]. Constraint transforms relate intrinsic images to global parameter values. If an intrinsic image parameter is a vector unit $(\mathbf{x},\mathbf{a}(\mathbf{x}))$ in an intrinsic image space A and an element of feature space is a vector $\mathbf{b}$ in a feature space B then there is usually a *physical constraint* that relates $\mathbf{a}(\mathbf{x})$ and $\mathbf{b}$, i.e., some relation $f(\mathbf{a},\mathbf{b})$ such that $f(\mathbf{a},\mathbf{b}) = 0$.

The space A represents all possible intrinsic image values. A particular intrinsic image is described by a set of values $\{\mathbf{a}_k\}$ where $\mathbf{a}_k = \mathbf{a}(\mathbf{x}_k)$. Now the set $\{\mathbf{a}_k\}$ is only consistent with certain elements in the space B, owing to the constraint imposed by the relation f. This restriction can be exploited in the following manner.

For each $\mathbf{a}_k$ compute the set

$$B_k = \{\mathbf{b} \mid \mathbf{a}_k \text{ and } f(\mathbf{a}_k,\mathbf{b}) \leq \delta_b\}$$

where the constant δ_b is related to the quantization in the space B. Define $H(\mathbf{b})$ as the number of times the value $\mathbf{b}$ occurs in $U_k B_k$ (the union of all sets B_k). $H(\mathbf{b})$ is the constraint transform from the space $\mathbf{a}$ to the space b and is the number of points in intrinsic image space which are consistent with the parameter value $\mathbf{b}$. $H(\mathbf{b})$ makes the most sense when the values both $(\mathbf{a}(\mathbf{x}),\mathbf{x})$ and $\mathbf{b}$ are discrete. H is also best normalized by defining $C(\mathbf{b}) := H(\mathbf{b})/\Sigma_b H(\mathbf{b})$. In that case, the value $C(\mathbf{b})$ can stand for the confidence that the segment with feature value $\mathbf{b}$ is present in the image. $H(\mathbf{b})$ can be thought of as a histogram and $C(\mathbf{b})$ a normalized histogram. Maxima in $H(\mathbf{b})$ correspond to likely global parameter values b.

To avoid describing the above computations in detail each time we need to use them, we use a shorthand notation for constraint transforms. Each transform can be described as the triple <a,b,f> where the necessary computations are implicit. Note that the order of a and b is important in the notation; in general, <a,b,f> is not equivalent to <b,a,f>. For many vision applications one can think of the triple as representing a transformation from a less abstract, retinotopic space to a more abstract, non-retinotopic space. Thus the triple generally represents

<parameters of retinotopic space, parameters of non-retinotopic space, the relation between elements in the two spaces>

but many variants are possible. As we shall see, (a) some of the parameters in **a** may not be retinotopic, (b) different triples may have overlapping subspaces in **a** or **b**, and (c) the relationship f may be expressed as a set of relations.

3. Rigid Body Motion: Special Cases

To introduce the Hough transform, we first consider a case where interesting information can be derived directly from the optic flow (without requiring 3-d flow). If an observer is translating in a stationary environment, then the optic flow will be radially outward from a single point called the focus-of-expansion (FOE).

The solution parallels Kender's Hough transform for detecting vanishing points in an image from oriented line segments [Kender, 1978]. Such line segments which are part of a given vanishing point form a radial field which emanates from the point. Different vanishing points have different sets of associated radial line segments (Fig. 4a). The same geometrical situation occurs with respect to optical flow due to pure translation.

Figure 4a

This example requires two levels of abstraction. The first transforms colinear flow elements into points (representing radial flow lines). Radial sets of flow vectors correspond to circles through the origin in flow-space. Thus the second transformation is between circles in flow-space to points in focus-of-expansion space.

The first level is easy if a (r,θ) line space is used where

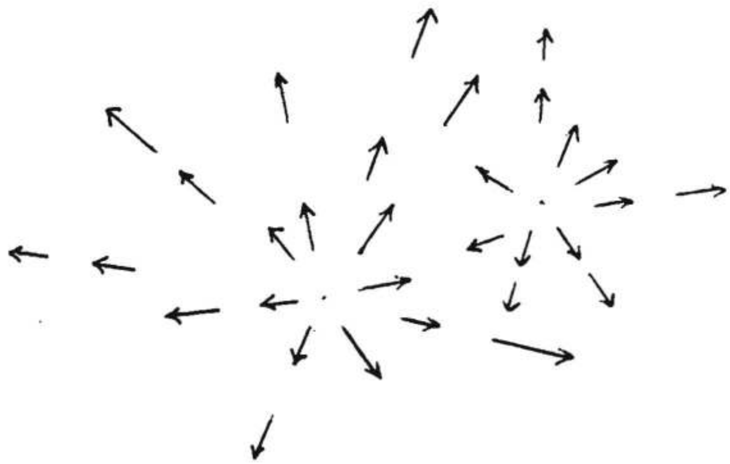
$$r = x \cos\theta + y \sin\theta \text{ and } \tan\theta = u/v.$$

Since a flow vector has direction θ (Fig. 4b), each such element maps onto precisely one point in (r,θ) space: $(x \cos\theta + y \sin\theta, \theta)$. Thus the constraint transform, in the notation of Section 2, is:

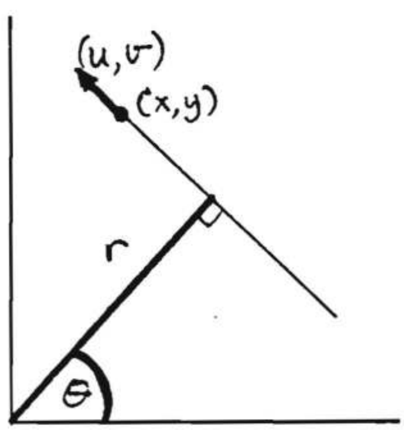
$$\langle (x,y,u,v),\ (r,\theta),\ (\theta = \tan^{-1}(u/v);\ r = x \cos\theta + y \sin\theta)\rangle.$$

Figure 4b

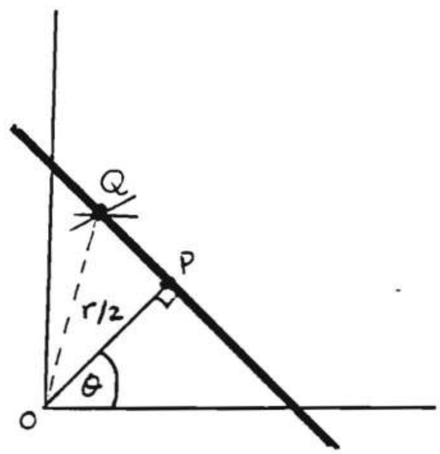
Now maxima in $C(r,\theta)$ correspond to lines in the image. Also, radial flow will form a circle of local maxima in (r,θ)-space. To see this note that the triangle OPQ in Figure 4b is always a right triangle, and therefore OQ must be the diameter of a



a.



first level: flow vectors to line parameters



second level: line parameters to vanishing points

b.

Figure 4

circle. Note that this circle is constrained to go through the origin so that its diameter must be on the line

$$r/2 = a \cos\theta + b \sin\theta$$

where (2a,2b) is the location of the focus of expansion. Thus the second transform is

$$\langle(r,\theta),\ (a,b),\ (r/2 = a \cos\theta + b \sin\theta)\rangle.$$

An interesting relationship exists between the flow vector from translatory motion and the FOE. Where d is the distance from the FOE, the time that the flow point will take to be adjacent to the observer is simply $d/||v||$. Consider the case of approaching a plane that is perpendicular to the direction of travel. In this case every flow vector will predict the same time to adjacency. Thus time to adjacency could be reliably detected by the Hough transform

$$\langle(FOEx,\ FOEy,\ x,\ y,\ u,\ v),\ (t),\ t = d/||v||\rangle$$

Lee and Reddish [1981] argue that time-to-adjacency is the control parameter used by diving gannetts. These birds must fold their wings just before hitting the water after starting from different heights.

4. General Rigid Body Motion

The algorithm for rigid body motion will be developed in four parts. First, we show how to calculate three-dimensional velocities (3-d flow) from optical flow and depth. Second, we show formally how to decompose high-dimensional Hough transforms into sequential low-dimensional Hough transforms. Third, we show how to use the equations of motion to derive the rigid body motion parameters using a Hough transform decomposition. Fourth, we describe how knowing the parameters can determine boundary conditions for the optic flow.

Calculating 3-D Flow

A pivotal first problem in finding the motion parameters is to compute 3-d flow, the actual three-dimensional velocities whose projections on the retina are optic flow. Three-dimensional flow consists of not only the three-dimensional velocity at points corresponding to retinotopic coordinates but also the three-dimensional position corresponding to those coordinates.

Knowing the optic flow constrains the three-dimensional velocities but leaves one degree of freedom. The depth provides the necessary additional information from which the three-dimensional velocities can be uniquely determined, but care must be taken in its use. Following the general methodology of [Horn and Schunck, 1980] and [Ikeuchi, 1979], one might try to impose a smoothness condition on the three-dimensional flow in order to compute it by a parallel iterative technique but this formulation turns out to be underconstrained; many smooth three-dimensional flows can produce the same optic flow. A feasible way of obtaining 3-d flow is to extract v_z by differentiating the depth map. Some care must be exercised in computing this derivative because of the correspondence problem. Owing to 3-d movement, the point $z(x',y',t+\Delta t)$ does not correspond to the point $z(x',y',t)$. The solution is to use the fact that the optic flow determines the correspondence in the depth map. (This idea was suggested and shown to work by A. Tevanian.) Thus the three-dimensional z-component, v_z, is given by

$$v_z \approx [z(x'+u\Delta t, y'+v\Delta t, t+\Delta t) - z(x',y',t)]/\Delta t \qquad (4.1)$$

The difference shows the mistake in differentiating the depth map at a fixed retinal point; in fact, v_z is given by a directional derivative in the (u,v,1) direction. Once v_z is known, v_x and v_y can be computed from (2.2a and b).

Hough Decompositions

A general feature of constraint transforms is that if the algorithms are completely parallel, the space required is exponential in the number of parameters. This can lead to immense space requirements. For example, consider an eight-parameter space of 100 discrete values for each parameter. The total number of parameter nodes required to represent the space is 100^8! Fortunately this problem can generally be alleviated by *detecting groups of parameters sequentially* [Ballard and Sabbah, 1981]. The advantage of this extremely powerful decomposition technique is that the dimensionality of the computation at each stage is much less than the single computation involving all of the parameters simultaneously.

In terms of the notation, the feature space B can be partitioned into two subspaces (B_1, B_2). Then the corresponding computation, denoted by $\langle a,b,f\rangle$, can be decomposed into two successive computations. First, compute $\langle a_1,b_1,f_1\rangle$ which has a set of maxima b^*, followed by $\langle (a,b_1^*),b_2,f_2\rangle$. Naturally it follows that $H_1(b_1) = \Sigma b_2 H(b)$ and $H_2 = H(b_1^*,b_2)$.

Adapting the Equations of Motion

Rather than solve for nine parameters for the object, one can find six by assuming that at a reference time $t=0$ the origins of the reference coordinate frame and that of the rigid body coincide. For times greater than zero, the origin position can be determined by knowing the motion parameters at the previous time step, i.e.,

$$x_b(t+\Delta t) = x_b(t) + \dot{v}_b\Delta t \qquad (4.2)$$

The drawback of this assumption is that the origin will not, in general, be in the best place for the most intuitive description of the motion. In fact there are many equivalent descriptions of the motion which arise from different body coordinate frame locations. The most intuitive are those on the axis of rotation. (We are working on iterative methods for moving the origin to one of these points.) For the remaining six parameters we use the decomposition technique. First the direction of rotation is determined. Next, the magnitude of the rotation and the three components of v_b are determined.

To determine direction of motion, we first show that the direction of rotation ω, defined as $\Omega/|\Omega|$, is perpendicular to the acceleration a of any point in the body provided the body is not subjected to large external forces. Recall that

$$a = a_b + d\Omega/dt \times \rho + (\Omega\cdot\rho)\Omega - \Omega^2\rho \qquad (4.3)$$

If a_b and $d\Omega/dt \approx 0$, (3.3) reduces to

$$a = (\Omega\cdot\rho)\Omega - \Omega^2\rho$$

Thus $a\cdot\omega = ((\Omega\cdot\rho)\Omega - \Omega^2\rho)\cdot\Omega/|\Omega|$ which is identically equal to zero. Thus **a** is perpendicular to ω. Now consider $a(t+\Delta t)$ and $a(t+2\Delta t)$ denoted by a_1 and a_2. Since ω is perpendicular to both these vectors,

$$\omega = a_1 \times a_2 / (|a_1||a_2|) \qquad (4.4)$$

Once 3-d flow is determined, two accelerations can be determined from three successive measurements of $v(x,t)$. Let $v_1 = v(x,t)$, $v_2 = v(x+v_1\Delta t, t+\Delta t)$. Then

$$a_1 = (v_2 - v_1)/\Delta t$$

In an analogous fashion, a second acceleration measurement (of the same point) at the next discrete time instant can be calculated as

$$a_2 = (v_3 - v_2)/\Delta t.$$

Each pair of successive accelerations derived from the 3-d flow field allows the determination of a value for ω. If these measurements all come from the same rigid body, then the values for ω should all be the same with the accuracy of the measurement. Thus ω can be detected by the constraint transform

$$\langle (a_1,a_2), \omega, (4.4)\rangle. \qquad (4.5)$$

(Recall the notation for Hough transforms introduced in Section 2.)

Once ω has been determined, the remaining parameters can be found via a second constraint transform. In the equation

$$v = v_T + |\Omega|\,\omega \times \rho \qquad (4.6)$$

only v_T and $|\Omega|$ are unknown. These can be determined by three constraint transforms

$$\langle (\rho,\omega), (|\Omega| v_{Tk}), k^{th} \text{ component of } (4.6)\rangle \qquad (4.7)$$

where $k = x,y,z$.

Recursive Filtering

If the set of rigid body parameters were known, they could be used to filter the optic flow. Recall from the description of optic flow that algorithms for computing it have difficulty deciding on flow boundaries. The key point is that the flow computation depends on smoothing the flow, but the smoothing should not be carried out across boundaries that differentiate different physical objects.

The rigid body parameters allow the determination of flow boundaries. In other words, it is possible to separate the portion of the 3-d flow corresponding to a particular set of parameters from the rest. Given these parameters, one can compute the set of 3-d velocities $\{v(x)\}$ such that each member of $v(x)$ is consistent with the parameter values, i.e., if $v(x)$ satisfies (4.6) then it is a member of this set. These $v(x)$ can be marked in parallel. Consequently the $v(x)$ which (a) do not satisfy Eq. (4.6) for the particular set of parameters and (b) are neighbors of $v(x)$ that do, can also be determined. This is also a parallel operation. After this operation, each $v(x)$ has only neighbors which are part of the same rigid body. Since the 3-d flow projects onto the optic flow, the corresponding portions of the optic flow are also determined. The optic flow calculations could be carried out as before but restricting the smoothing to optic flow points with the same rigid body parameters. This technique, which we term *recursive filtering*, allows model information in the form of rigid body motion parameters to be used in a "top-down" manner. In general form, this is also a way of focusing attention [Feldman and Ballard, 1982].

Formally we can let $\{link(x',y',k)\}$ denote a set of k links to neighbors of a retinotopic point $x'y'$. Averaging for the flow computation is carried out using neighbors defined by these links. If $v(x)$ corresponding to the point x',y' satisfies (4.6) but that defined by link k does not, then the link k is deleted from the set. Figure 5 shows an example.

Figure 5.

Combined Computations

Knowing the rigid body motion parameters can thus help the computation of flow, but the obvious question is: how can the parameters be determined without finding the flow in the first place? The answer is that both the rigid body motion calculations and the flow calculations can proceed *in parallel*. The partial results from each help the other calculation in a way that both calculations converge. The reason for this is that the flow calculations are local and so many flow points interior to the boundary will be close to the appropriate values. At the outset of the calculations the under-constrained situation at the boundaries will be too far away to affect these interior points. Thus the interior flow points allow the approximation of the motion parameters.

This estimate ~~is~~ should be sufficient to initiate the backward masking and arrest the flow smoothing across physical boundaries. Thus the rigid body motion constraint is sufficiently powerful that, when it is applicable, it makes the combined calculation of flow and motion parameters well defined.

In the real world the rigid body constraint is very common since most objects are rigid to a first approximation or articulated with rigid parts.

5. Implementation

In principle, the algorithms described in the previous sections can be solved in parallel using a network formalism [Feldman and Ballard, 1982]. The Hough formalism described earlier is part of this more general formalism and specifies the

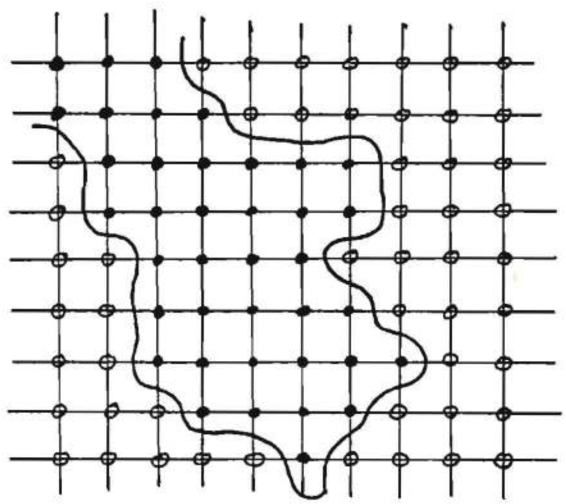


Figure 5 One method for finding boundaries in optic flow. Dark points have flow measurements which have the same rigid body parameters.

connections between elements in the networks. Parallel-iterative relaxation describes the *dynamic behavior* of this network. The intent of our implementation was to explore the feasibility of the model by rigorously simulating the rigid body motion parameter portion but less rigorously simulating the 3-d flow portion.

The rigid body parameters are implemented as an array, one location for each *value* of depth, optical flow, 3-d velocity, and rigid body motion parameters. This is true for the rigid body motion parameter space which is divided into three subspaces of units of x_b, v_b, and Ω. The 3-d velocity spaces are represented as *variable records*, i.e., one record per each two-dimensional position.

Arrays of depth and optical flow information determine appropriate 3-d velocities as determined by (2.2). Three-dimensional velocities, together with motion parameters, should determine optical flow. In our simulation, this part is currently not implemented. Instead, we begin with 3-d flow vectors which we generate.

Experiment 1

In Experiment 1, 3-d flow vectors were generated from a rigid body model whose parameters were determined *a priori* (Fig. 6). These were used as input data. In a more complete experiment these vectors would be determined from spatially registered depth and optical flow fields.

Figure 6.

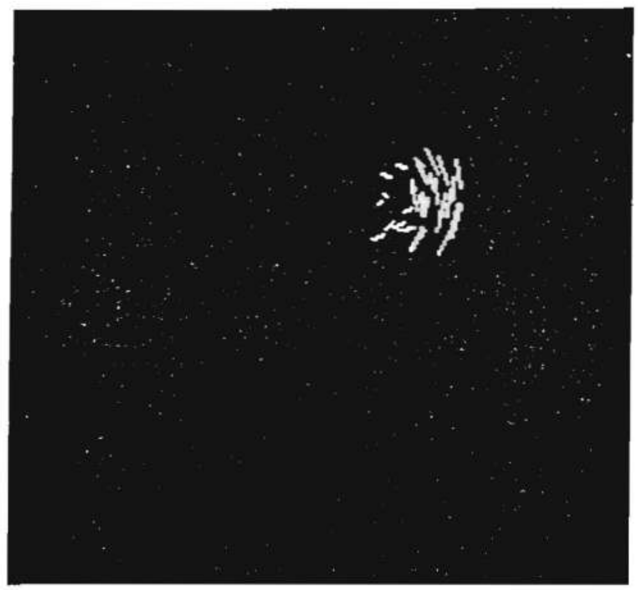
Figure 7 shows the results of detecting ω, the unit vector in the direction of the rotation vector. Since ω is a unit vector, only two components need be determined, so the figure shows only $H(\omega_y, \omega_z)$. The single maximum shows that different pairs of acceleration vectors give rise to a common rotation vector, as expected. Figure 7a shows the results of applying the Hough transform technique (3.5) to the first three frames of flow information. Figure 7b shows the results of nine consecutive time frames including those of Figure 7a. In this example, the parameter space is rescaled to be more finely distributed around the maximum. Nice ways of doing this are described in [O'Rourke, 1981; Sloan, 1981].

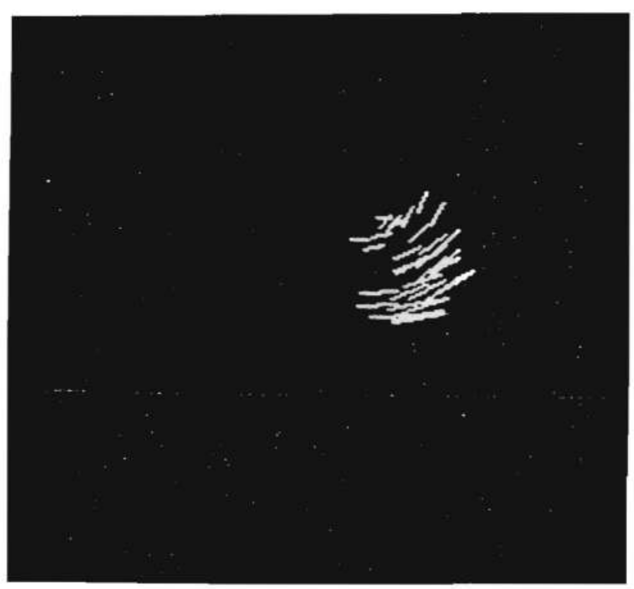
Figure 7.

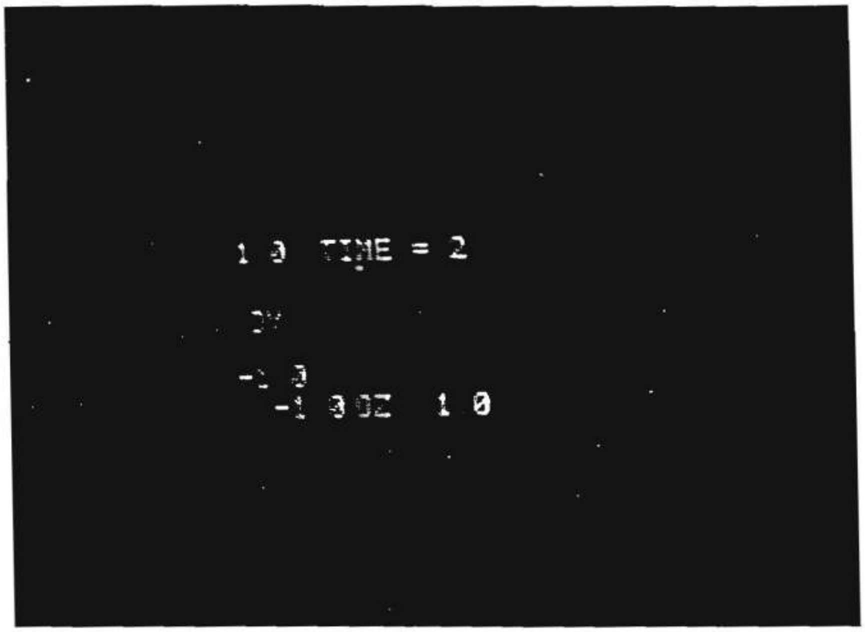
Figure 8 shows the results of detecting $|\Omega|$ and v_T. In Figure 8a the Hough transform (3.7) has been applied to the third frame using ω as a known constant vector. Also the decomposition technique is clearly shown. The four-parameter space is represented as three two-parameter spaces: $(|\Omega|, v_{TX})$, $(|\Omega|, v_{TY})$, and $(|\Omega|, v_{TZ})$. Figure 8b shows the corresponding results for nine consecutive frames.

Figure 8.

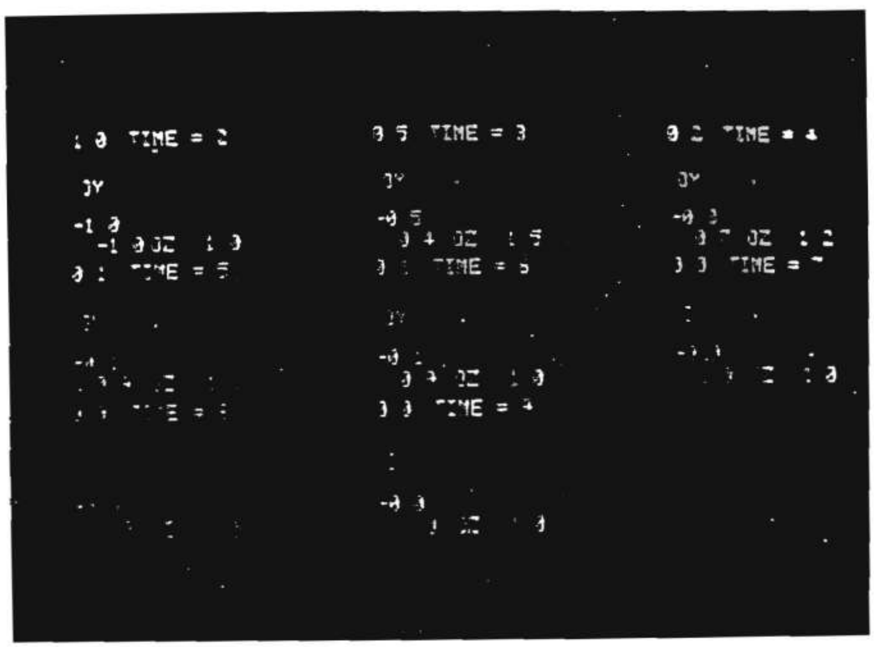


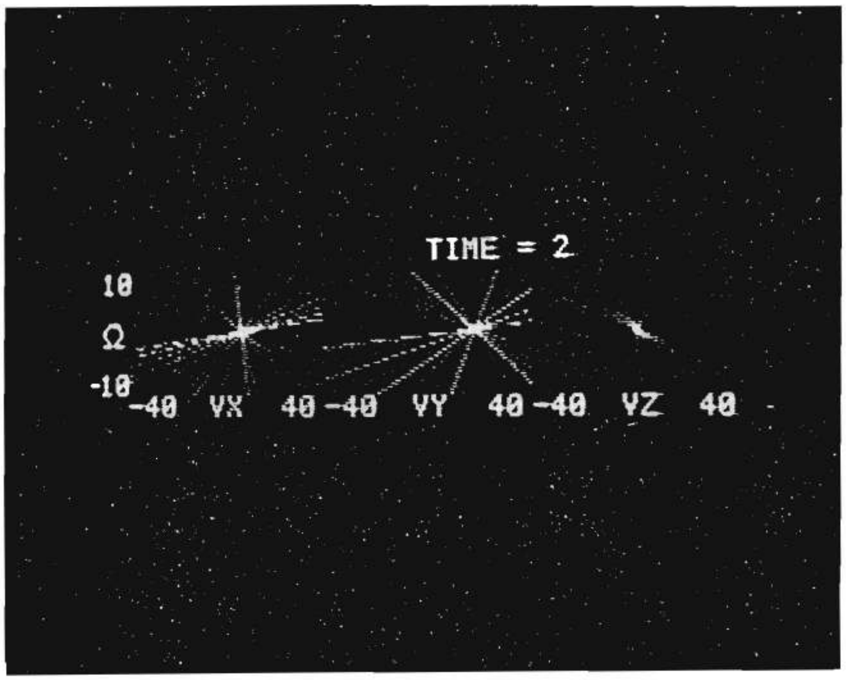




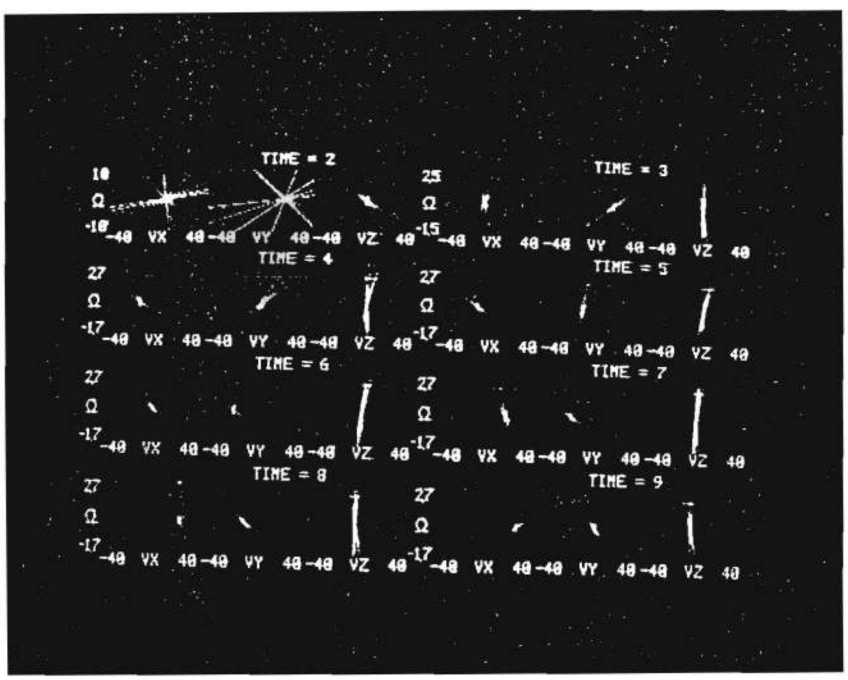
a



b


TIME = 2
10
Ω
-10
-40 VX 40 -40 VY 40 -40 VZ 40

a


TIME = 2
10
Ω
-10 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 3
25
Ω
-15 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 4
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 5
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 6
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 7
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 8
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40
TIME = 9
27
Ω
-17 -40 VX 40 -40 VY 40 -40 VZ 40

c

Experiment 2

In Experiment 2, 3-d noise vectors were added to the flow so that their number approximately equaled that of the rigid body flow vectors. The results of this experiment show that the techniques are extremely noise insensitive. Figures 9-11 are analogous to Figures 6-8, and show the results of detecting the rigid body parameters in the presence of spurious 3-d flow information. Figure 10 has been deliberately overexposed to show that singleton rotation direction parameters are produced by the noise vectors. Figure 11 also shows spurious parameter values produced.

Figures 9-11.

In this experiment, we tested the feasibility of recursive filtering. Using the motion parameters derived from the transform, we were able to select out the vectors that contributed to those parameters. Each vector in the 3-d flow field was tested to see if it satisfied (3.6); if so, it was deemed to be part of the rigid body. Naturally a noise vector could be selected in this manner, but the probability of this happening is very low. Figure 12 shows 3-d flow frames with the rigid body motion vectors displayed with increased illumination.

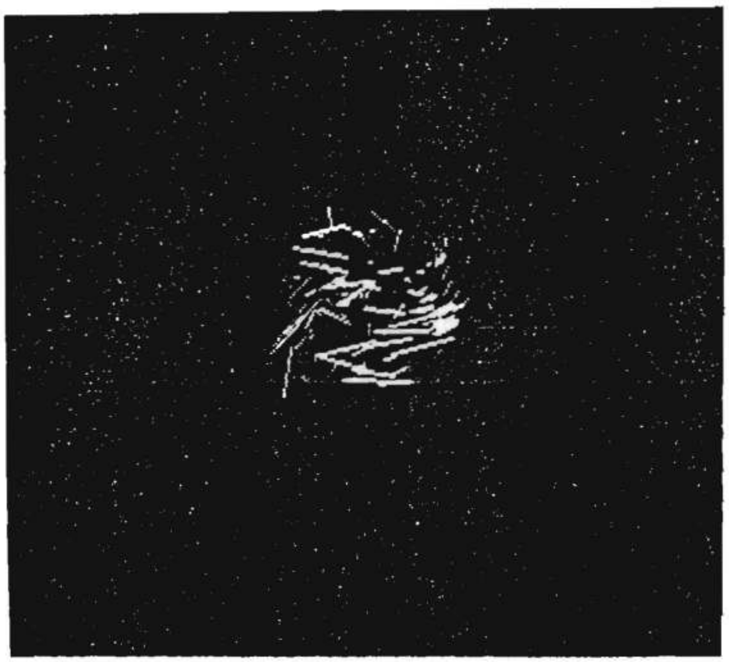
Figure 12.

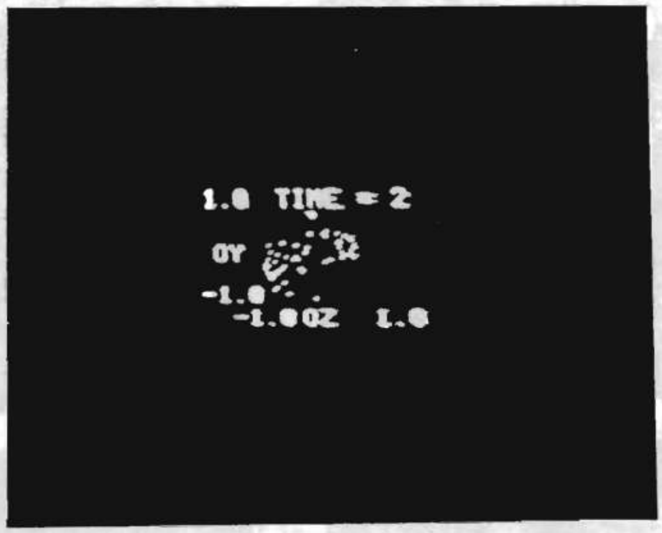
6. Discussion

Throughout the presentation we assumed very simple environments, such as one moving object against a stationary background. How can multiple objects be handled? The principal problem is introduced by the decomposition of parameters. If there are two moving objects, which maxima in ω space corresponds with which maxima in the $(|\Omega|, v_k)$ spaces? Fortunately there is an elegant solution to this problem which takes advantage of the fact that the right pairing of parameters is coupled since they are related to the same 3-d flow vectors. Thus once a maximum in ω-space has been determined, the velocity space can be filtered by backward masking. In other words, all v(x) which cannot have the parameter ω have their confidences set to zero, i.e., are temporarily removed from contention. Thus the transform for $(|\Omega|, v_b)$ is taken using a conditioned set of v(x). This should have the effect of removing the other maxima in the $(|\Omega|, v_b)$ spaces. Once an appropriate set of parameters have been associated in this manner, another set can be associated.

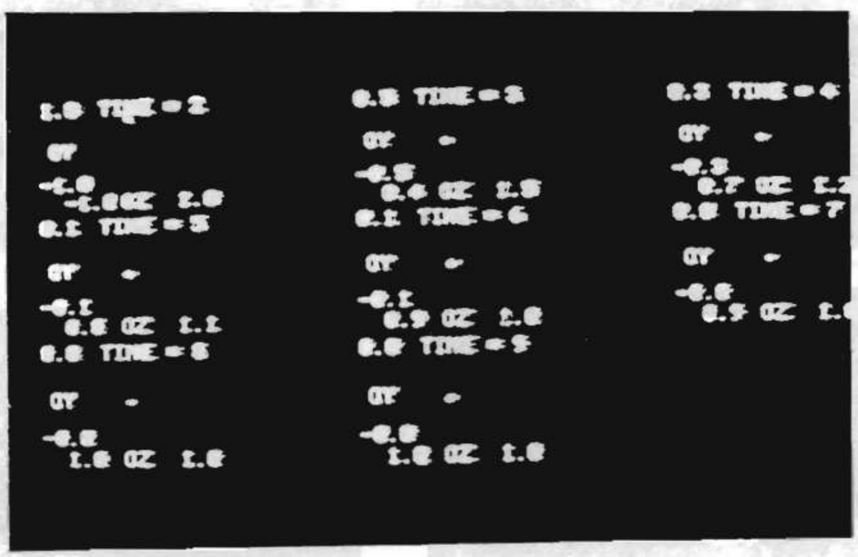
This exposition might imply that these computations are sequential and to some extent they must be. However, (a) the computations for a particular $\omega - |\Omega|, v$ association may be overlapped, and (b) the computations for sets of $\omega - |\Omega|, v$ associations may be interleaved. These details are pursued in [Ballard, 1981].

The multi-body problem also introduces the prospect of computing the motion of one body relative to another. All the necessary information is available, the principal question is can the computations be done in parallel using the network formalism? The answer is in the affirmative and the technical details will appear in [Hrechanyk, 1982].

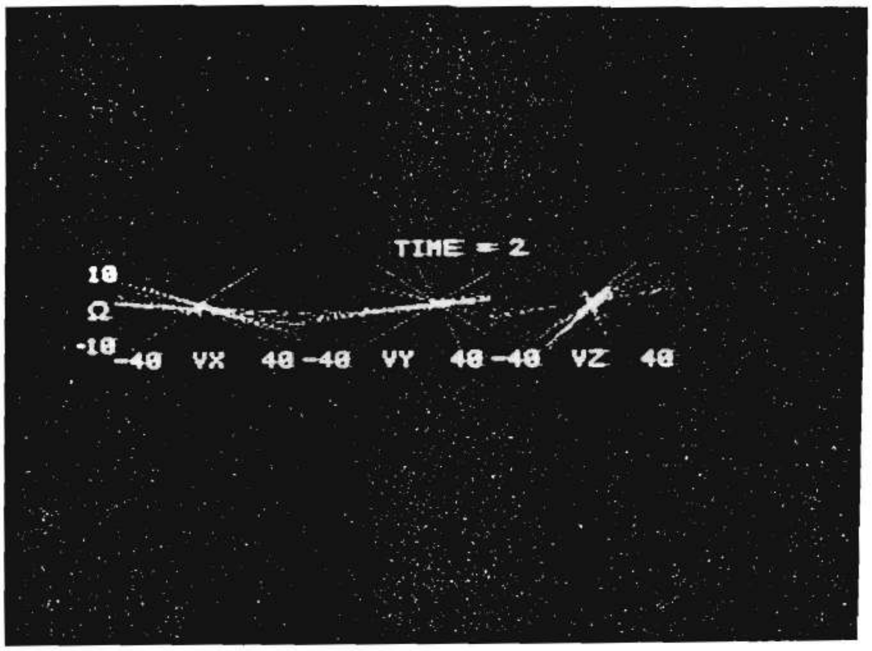



1.0 TIME = 2
OY
-1.0
-1.0 OZ 1.0

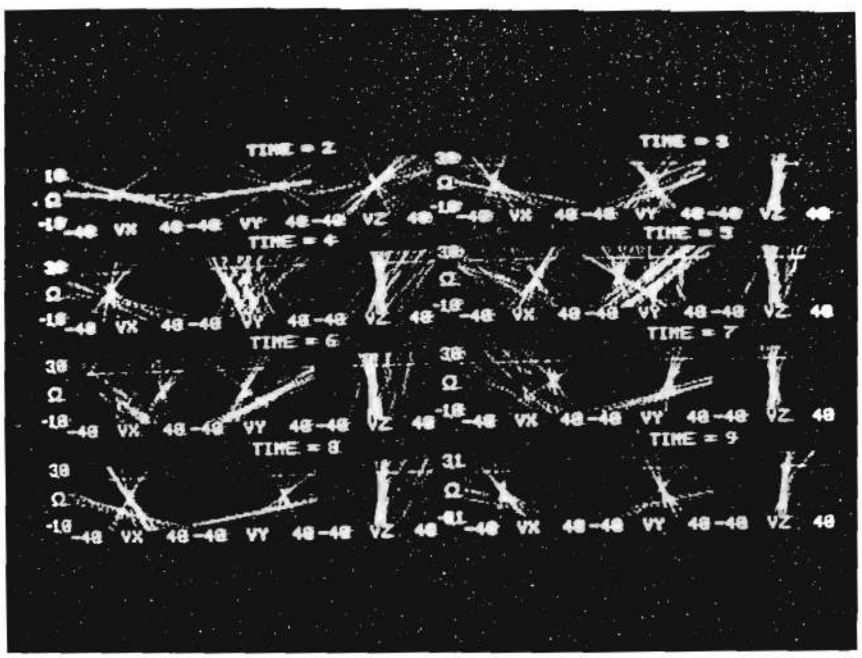
a


TIME = 2
TIME = 3
TIME = 4
TIME = 5
TIME = 6
TIME = 7
TIME = 8
TIME = 9

b


TIME = 2
10
Ω
-10
-40 VX 40 -40 VY 40 -40 VZ 40

a


TIME = 2
TIME = 3
TIME = 4
TIME = 5
TIME = 6
TIME = 7
TIME = 8
TIME = 9
Ω
VX
VY
VZ

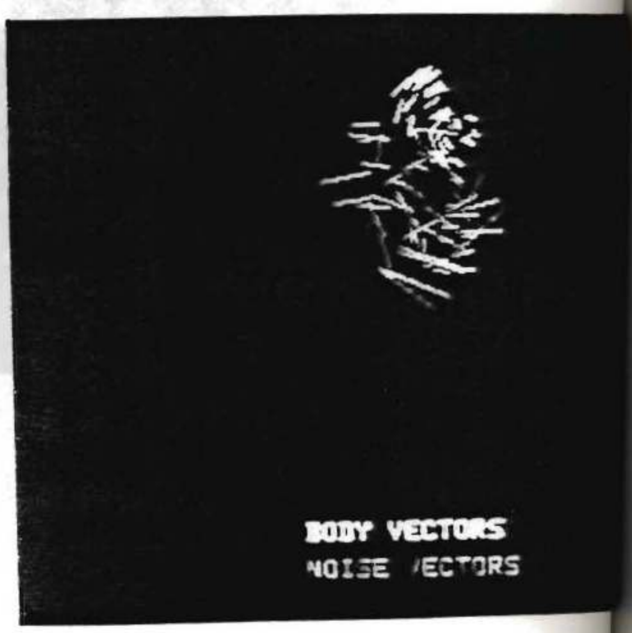
b


BODY VECTORS

BODY VECTORS

BODY VECTORS

BODY VECTORS
NOISE VECTORS

BODY VECTORS

BODY VECTORS
NOISE VECTORS

The methodology presented here has several psychological implications. First, we required three time frames to reliably compute direction of rotation. Experiments with human perception of dot patterns [Lappin et al., 1980] have shown that the perception of a rigid body motion is extremely noise sensitive when only two frames are presented. Second, if these constraints are used by human perception, then the prediction is that humans will have difficulty with accelerating objects where the magnitudes of a_T and $d\Omega/dt$ are significant compared to the rest of the accelerations. Alternately, the methodology could be extended to incorporate these parameters. Third, the decomposition technique used makes it difficult to deal with bodies that have different parameters. These difficulties have also been observed in humans by [Lappin and Kottas, 1981].

In summary, we have developed a method for recognizing rigid body motion from depth and optic flow. This method also facilitates the computation of the optic flow itself by providing important boundary constraints on the flow. The method is completely parallel and could serve as the basis for an abstract model of neural computation of the same information. The method also has psychological implications which could be tested.

References

Ballard, D.H., "Parameter networks: Towards a theory of low-level vision," *Proc.*, 7th IJCAI, Vancouver, B.C., August 1981.

Ballard, D.H. and O.A. Kimball, "Shape and light source direction from shading," TR, Computer Science Dept., U. Rochester, to appear, 1981.

Ballard, D.H. and D. Sabbah, "On shapes," *Proc.*, 7th IJCAI, Vancouver, B.C., August 1981.

Barrow, H.G. and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images," Technical Note 157, AI Center, SRI Int'l., April 1978.

Barnard, S.T. and W.B. Thompson, "Disparity analysis of images," TR 79-1, Computer Science Dept., U. Minnesota, January 1979.

Feldman, J.A. and D.H. Ballard, "Connectionist models and their properties," to appear in *Cognitive Science*, 1982.

Gibson, J.J. *The ecological approach to visual perception.* Houghton-Mifflin, 1979.

Horn, B.K.P. and B.G. Schunck, "Determining optical flow," AI Memo 572, AI Lab, MIT, April 1980.

Horn, B.K.P. and B.G. Schunck, *Proc.*, Pattern Recognition and Image Processing Conf., Dallas, TX, August 1981.

Hrechanyk, L., forthcoming TR, Computer Science Dept., U. Rochester, 1982.

Ikeuchi, K., "Numerical shape from shading and occluding contours in a single view," Technical Memo AIM 566, AI Lab, MIT, 1979.

Kender, J.R., "Shape from texture: A brief overview and a new aggregation transform," *Proc.*, DARPA IU Workshop, 79-84, Pittsburgh, PA, November 1978.

Lappin, J.S. and B.L. Kottas, "The perceptual coherence of moving visual patterns," to appear, *Acta Psychologica*, 1981.

Lappin, J.S., J.F. Doner, and B.L. Kottas, "Minimal conditions for the visual detection of structure and motion in three dimensions," *Science 209*, 717-719, 8 August 1980.

Lee, D.N. and P.E. Reddish, "Plummeting gannets: A paradigm of ecological optics," *Nature 293*, 293-294, 24 September 1981.

Lishman, J.R., "Vision and the optic flow field," *Nature 293*, 263-264, 24 September 1981.

Marr, D., "Representing visual information," in A.R. Hanson and E.M. Riseman (Eds). *Computer Vision Systems.* NY: Academic Press, 1978.

Marr, D.C. and T. Poggio, "Cooperative computation of stereo disparity," *Science 194*, 283-287, 1976.

O'Rourke, J., "Dynamically quantized spaces for focusing the Hough transform," *Proc.*, 7th IJCAI, 737-739, Vancouver, B.C., August 1981.

Prazdny, K., "A simple method for recovering relative depth map in the case of a translating sensor," *Proc.*, 7th IJCAI, 698-699, Vancouver, B.C., August 1981.

Sloan, K.R., Jr., "Dynamically Quantized Pyramids," *Proc.*, 7th IJCAI, Vancouver, B.C., August 1981.

Ullman, S., *Behav. Brain Sci. 3*, 373, 1980.

Ullman, S. *The interpretation of visual motion.* Cambridge, MA: MIT Press, 1979.