MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo 835 1 April 1985

BOUNDARIES OF VISUAL MOTION

John M. Rubin & W.A. Richards

Abstract. A representation of visual motion convenient for recognition should make prominent the qualitative differences among simple motions. We argue that the first stage in such a motion representation is to make explicit boundaries that we define as starts, stops, and force discontinuities. When one of these boundaries occurs in a motion, human observers have the subjective impression that some fleeting, significant event has occurred. We go farther and hypothesize that one of these subjective motion boundaries is seen if and only if one of our defined boundaries occurs. We enumerate all possible motion boundaries and provide evidence that they are psychologically real.

Acknowledgment. This report describes work done at the Department of Psychology and the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology under NSF grant IST-8312240. We appreciate the helpful comments of A. Bobick, E. Grimson, D. Hoffman, E. Saund, S. Ullman, and A. Yuille.

1 Introduction

The human visual system is adept at recognizing different kinds of motion. We easily identify moving things as walking, waltzing, bouncing, or slithering, to name just a few types of motion. How do we do it? The visual system must somehow represent motions and match these representations to descriptions in memory. In this paper, we establish criteria for a motion representation and begin its construction.

We consider only motion of points or blobs, that is, motion without shape information. There are three reasons for this approach (see Rubin, 1985). First, it makes sense to start with the simplest case. As shapes are allowed to increase in complexity, the number of degrees of freedom of motion skyrockets. Whereas a blob in space can be specified by three coordinates, the articulated characters in Lucasfilm's computer-generated cartoon, *The Adventures of André & Wally B.*, require over five hundred parameters of motion. Second, even for an elaborate shape, it seems useful to understand as much as possible about its movement as a whole. The motion of the center of mass, say, will often be informative, independent of the wrigglings of movable parts. The third reason is that the motion of parts will eventually be related to some object-centered frame (Wallach, 1959; Marr & Vaina, 1980). Representing the motion of this frame is an important first step.

2 Representing Motion

What aspects of motion must a visual system represent if its goal is to recognize simple types of motion? A point moving in three dimensions is completely described by its position over time, $\vec{p}(t)$. This representation has three shortcomings. First, all that it makes explicit is position, and *where* a motion occurs has nothing to do with *what* kind of motion it is. Second, $\vec{p}(t)$ is *unstable* in the sense of Marr & Nishihara (1978): all detectable variations in motion are represented independent of their importance to motion recognition. By contrast, a stable representation will have some explicit component that remains invariant over unimportant changes. Finally $\vec{p}(t)$ depends on the choice of units for measuring space and time.

The representation we develop below overcomes these three objections.

We seek a motion representation that captures the blatant qualitative differences among such motions as bouncing, planetary orbits, and bat flight, yet is insensitive to minutia such as the particular value of the viscous drag coefficient of air. In pursuit of such a representation, we must examine in some detail what is meant by a *kind* of motion.

Consider the bouncing motion of a tossed ball. What defines bouncing is not a particular trajectory, but rather the sequence of free-fall, impact, free-fall, impact, and so on (see Forbus, 1981). Intuitively, the trajectory is divided into natural parts or *eras*. Each period of free-fall is an era, as indicated in Figure 1a. Separating two consecutive eras is a brief application of force—a bounce—which seems to be a natural motion boundary. Some motions, like planetary orbits, lack such boundaries.

As the foundation of our motion representation, we will define motion boundaries in the following section. (Motion eras will be described in Rubin (1985); boundaries and eras together will constitute a complete motion representation.) We will then show that motion boundaries can in principle be detected in images from almost any[1] viewpoint.

3 Elementary Motion Boundaries

Viewers of a bouncing ball perceive fleeting, significant events—subjective boundaries—at the bounces (Figure 1a). Why aren't these boundaries perceived at the apices, as in Figure 1b? Why are subjective boundaries seen at all? An explanation lies in our motivations for the motion boundary definitions below.

3.1 Starts and Stops

Starts and stops are obvious candidates for motion boundaries, and we define them as such. They are illustrated in Figure 2. If they were not made explicit in the representation of motion, we would be unable to demarcate

[1] "Almost always" or "almost any" means *with probability one* if the item in question is chosen randomly.

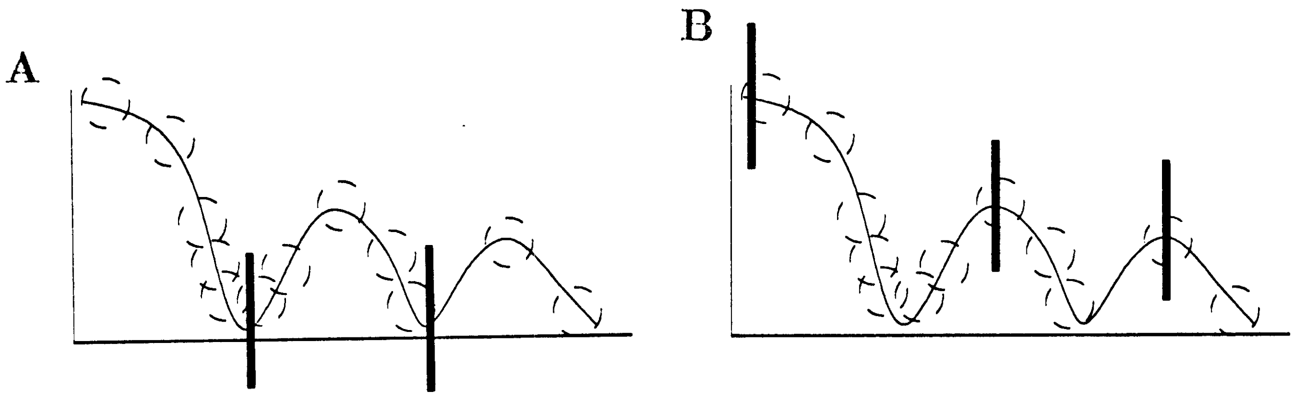

Figure 1: A bouncing ball. Circles on the trajectory show the ball's position at fixed time intervals. **a)** Parts of the trajectory as predicted by the theory here, and as seen by most observers. **b)** A possible division of the bouncing ball trajectory into parts.

a period of activity from a period of rest. Starts and stops must be defined with respect to a reference frame since their definition will require a well-defined zero of velocity. There are several choices for reference frames: the viewer, the ground, or moving objects in the scene. Given but a single moving object, the only choice is the viewer's frame. However, for the case of articulated shapes, the motion of parts will often be most conveniently referred to the motion of the whole, perhaps hierarchically (Marr & Nishihara, 1978; Marr & Vaina, 1980).

3.2 Dynamic Boundaries

We define a second type of motion boundary that is independent of starts and stops: discontinuities of force. A motion boundary based on force, unlike starts and stops, will be independent of (inertial) reference frame. We choose force discontinuities to supplement starts and stops as motion boundaries because discontinuities are robust force events; they can be detected even in non-inertial frames[2] We will henceforth call starts and stops

[2] Force discontinuities can be detected in any smoothly accelerated frame. A non-example is a reference frame tied to a Brownian particle (Lavenda, 1985). We take "reference

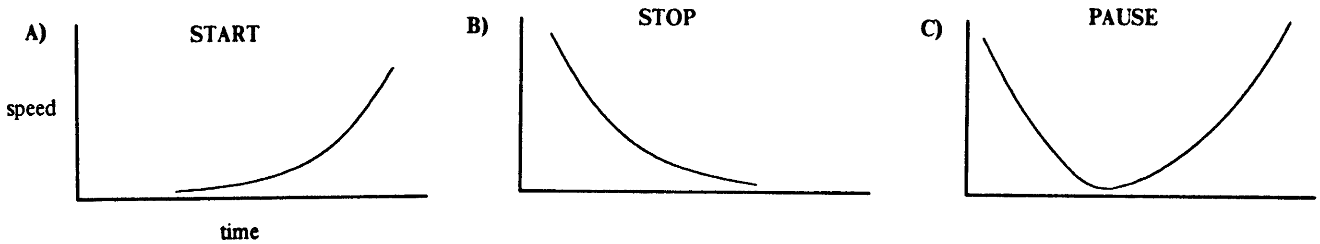

Figure 2: Two elementary reference frame boundaries shown in one dimension, and their conjunction. **a)** Start. **b)** Stop. **c)** Pause (see section 6.1).

"reference frame boundaries" and force discontinuities "dynamic boundaries."

There are two advantages to adding force discontinuities as motion boundaries. First, the choice captures the intuition that dynamic processes are continuous, and that an abrupt change in force is likely to indicate that one process has been succeeded by another. Second, given a rigid body, if one point undergoes a dynamic boundary, then almost all points on the body must simultaneously undergo dynamic boundaries (see Appendix I). This means that a visual system can monitor an indiscriminately chosen point on a rigid body and still detect dynamic boundaries.

Steps and Impulses

We claim that there are two fundamental motion boundaries that are independent of the reference frame[3]. These are step discontinuities and impulses of force. Detecting discontinuities involves issues of scale which are discussed in Appendix II.

For a function f that is continuous at t_0 (Thomas, 1972), we have

$$f(t_0) = \lim_{t \to t_0+} f(t) = \lim_{t \to t_0-} f(t) \qquad (1)$$

frame" to mean an inertial frame, or any frame smoothly accelerated with respect to an inertial frame.

[3] While the arguments in this section are given for single-valued functions of time, they generalize straightforwardly to vector-valued functions of time. That is, vector-valued functions have the same two elementary types of discontinuity.

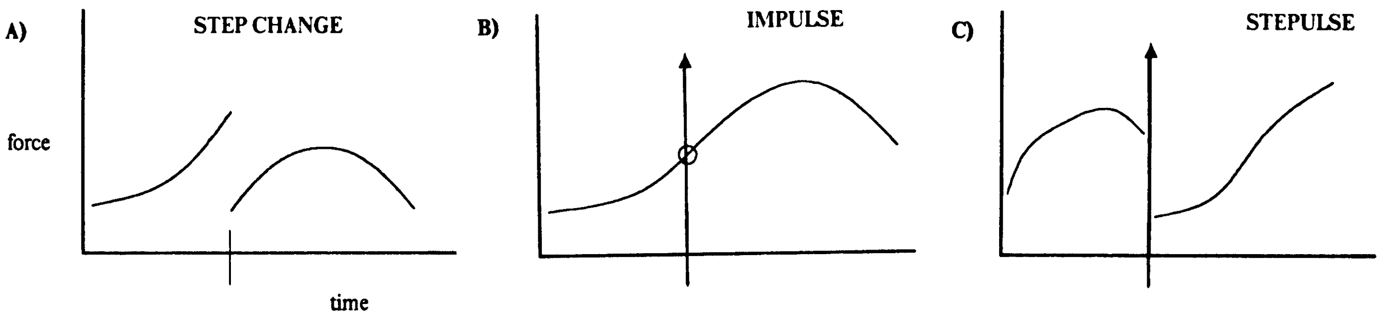

Figure 3: Two elementary types of discontinuity of a single-valued function f of one variable t, and their conjunction. **a)** A step-change; f can take any finite value at t_0. **b)** An impulse; f at t_0 is equivalent to an impulse function, and the limits at t_0 are equal. The impulse is depicted by a vertical line capped by an arrow to indicate nonfinite value. **c)** A "stepulse" (see section 6.1): f at t_0 is equivalent to an impulse, and the limits of f at t_0 are unequal.

What are the ways a function can violate (1) and therefore be discontinuous? Taking f to be force as a function of time, we will assume the right and left limits exist for motion at a biological scale[4]. Furthermore, we will assume that f is defined at t_0, either taking on some finite value or the value of an impulse function[5].

Given the assumptions above, there are only two ways for f to be discontinuous at t_0. One possibility is that the left and right limits of f at t_0 are unequal. In this case we say f is step-discontinuous at t_0. This is the first elementary force discontinuity; it is shown in Figure 3a.

Consider next the case that the left and right limits of f at t_0 are equal. Then there are two ways for f to be discontinuous at t_0: $f(t_0)$ can have a finite value not equal to the value of the limits, or $f(t_0)$ can have the value of an impulse. We claim the former sort of discontinuity is *in principle undetectable*. To see this, note that the choice of a finite value of $f(t_0)$ cannot affect velocity or position, since more generally, changing the value of a function at isolated points does not affect its integral (Bracewell, 1965).

[4] Brownian motion is an example where such limits can fail to exist.

[5] An impulse function has value zero except at a single point, yet its integral from $-\infty$ to ∞ is finite and nonzero (Bracewell, 1965). Thus the value of an impulse is not finite.

It is the remaining subcase in which $f(t_0)$ has the value of an impulse that is of interest to us. An impulse is the second elementary force discontinuity; it is shown in Figure 3b.

4 Three-Dimensional Kinematics

So far we have argued that starts, stops, and force discontinuities should be explicit boundaries in the representation of motion. To find these boundary conditions in the image, one must understand their three-dimensional kinematics. In this section, we will express boundaries as conditions in three-dimensional kinematics. Two points must be made with regard to this task. First, force, per se, is invisible; we must infer it from the acceleration it causes. Second, it is inconvenient to seek an impulse discontinuity in a (force) function. It is easier to find a step discontinuity in the integral of that function (Bracewell, 1965). Both force discontinuities will therefore be expressed in terms of kinematic step discontinuities.

4.1 Starts and Stops

Starts and stops will now be precisely defined. Let a reference frame—the "scene frame"—be chosen. We have the intuition that an object stops at a certain time if it moves for a period prior to that time and is stationary for a period after that time. More formally, let $s(t)$ be speed as a function of time in the scene frame. Define a stop at t_0 when $\exists\epsilon > 0$ such that $\forall t \in (t_0 - \epsilon, t_0)$, $s(t) \neq 0$, and $\forall t \in [t_0, t_0 + \epsilon)$, $s(t) = 0$. The definition of a start is analogous. (In practice, measurement of speed will be subject to the spatiotemporal resolution limits of a visual system; see Appendix II.)

4.2 Dynamic Boundaries

Consider any kind of force discontinuity. Begin again with Newton's Second Law: $\vec{F}(t) = m\vec{a}(t)$. Note immediately if $\vec{F}(t)$ is discontinuous[6] at some t_0,

[6] A vector-valued function of one variable is continuous at a point if its components are continuous at that point (Seeley, 1970). Therefore, a vector-valued function has a discontinuity at a point if one or more of its components has a discontinuity at that

then $(m\vec{a}(t))$ is also discontinuous at t_0. But mass, at a biological scale, does not fluctuate much at all, let alone discontinuously[7] Thus given any type of dynamic discontinuity, and an *assumption* of constant mass, $\vec{a}(t)$ must have a discontinuity at t_0. In particular, step discontinuities of force bring about step discontinuities in the acceleration vector.

We turn next to the case of impulses. Since an impulse of force will change the velocity of an object in an instant, an object moving in three dimensions that is subject to an impulse will undergo a step discontinuity of its velocity vector. A simple approach to the result is to note that the integral of acceleration is velocity, and the integral of an impulse is a step function. Hence, an impulse in force yields a step discontinuity in velocity.

5 Image Motion

In this section we examine how the three-dimensional conditions of the previous section project to the image. It will be shown below that from almost any viewpoint, the three-dimensional boundaries will also be two-dimensional boundaries in the image. Furthermore, it will be shown that a boundary in the image *always* indicates a three-dimensional boundary; that is, there are no "false targets". Finally, we show that *to find all motion boundaries, a visual system must detect exactly four features of image motion: starts, stops, and step discontinuities of velocity and acceleration.*

5.1 From Three Dimensions to Two

First we must show that almost any two-dimensional image of a three-dimensional boundary contains a boundary. Starts and stops in three-dimensions are also starts and stops in the image if the viewer is at rest with respect to the scene frame. Likewise, we show in Appendix III that step

point.

[7] We have assumed throughout that the mass of the blob is constant. If mass is allowed to vary, do new motion boundaries obtain? Continuous variation of mass, as exemplified by a rocket, most of whose mass is fuel, cannot cause an acceleration discontinuity. A discontinuity of mass—a break, explosion, or agglomeration—will cause an acceleration discontinuity.

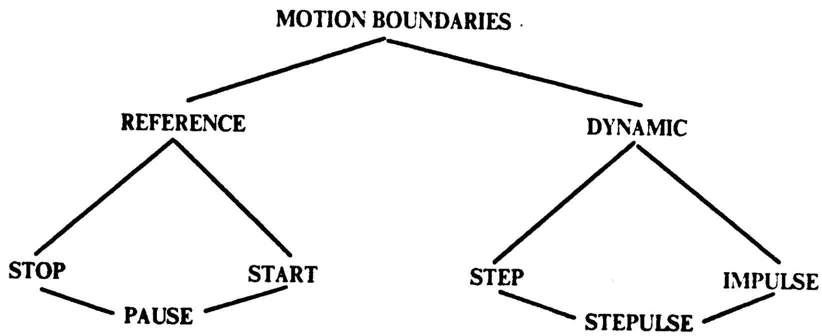

Figure 4: The relations among the six types of motion boundaries.

discontinuities in the three-dimensional velocity and acceleration vectors will almost always appear as the step discontinuities in the corresponding two-dimensional image vectors.

5.2 False Targets

Does the appearance of a boundary in the image imply a boundary in three dimensions? In Appendix III, we show this is always the case for dynamic boundaries. It remains to examine image speed zeroes. In orthographic projection, a false target occurs when $\vec{v}_o = \vec{0}$ but $\vec{v} \neq \vec{0}$. There is thus a false target whenever $\dot{x} = \dot{y} = 0 \neq \dot{z}$. The probability of this exact occurrence is zero. Hence false targets will be rare[8]. (The argument for perspective projection is similar.)

6 Compound Motion Boundaries

In this section, we consider the possibility of co-occurrence of two or more of the quartet of elementary motion boundaries thus far defined. We show that the two reference frame boundaries can co-occur, as well as the two dynamic boundaries. Furthermore, reference frame and dynamic boundaries are independent. When such combinations are considered, a total of

[8] It is also important to ask whether the three-dimensional kinematic conditions of Section 4 could have arisen from circumstances other than the dynamic boundaries we wish to detect. The answer is no because Newton's Second Law can be considered a *definition* of force. That is, a step discontinuity of velocity is *equivalent* an impulse of force, and a step discontinuity of an object's acceleration vector is *equivalent* to a step-change in the force on that object.

fifteen mathematically distinct motion boundaries emerge. Some of these are physically odd. We must emphasize that this set of boundaries does not constitute a complete motion representation; eras of motion—periods between successive boundaries—must also be described (Rubin, 1985). Also, we make no claim that all fifteen compound boundaries are psychologically distinct.

6.1 Conjunctions Within a Boundary Type

6.1.1 Conjunction of Start and Stop: Pause

There is an event that is the limiting case of both starts and stops as defined in section 4.1. We call this occurrence a *pause*; it is illustrated in Figure 2c. A pause is more formally described as a speed zero such that there are open intervals that contain it but no other speed zeroes. Pauses occur naturally in simple systems: consider the motion of an inchworm, or a pendulum at the ends of its swing. Note that the terms stop, start, and pause are mutually exclusive.

6.1.2 Conjunction of Step and Impulse: "Stepulse"

A step and an impulse can also co-occur. We call this event a "stepulse;" it is illustrated in Figure 3c. A stepulse occurs when a tetherball is struck so hard that the tether breaks. Henceforth, the terms step, impulse, and stepulse will be taken as mutually exclusive. The elementary boundaries and their conjunctive progeny are shown in Figure 4.

6.2 Conjunctions Across Boundary Types

Starts and stops can coincide with dynamic boundaries. Consider a beanbag striking the ground, an event that involves the coincidence of a stop and a stepulse. More generally, at a given moment, velocity (and hence reference frame boundaries) and acceleration (and hence dynamic boundaries) are independent. To enumerate the combinations, note that at a motion boundary, there are four possible dynamic circumstances: step, impulse, stepulse, and continuity (no dynamic boundary). Similarly, there are four

REFERENCE BOUNDARIES

DYNAMIC BOUNDARIES	NONE	START	STOP	PAUSE
NONE	*motion era; see Rubin (1985)*	**car begins to move as light turns green**	**deceleration of a shuffleboard puck**	**extrema of pendular motion**
STEP	**gas pedal floored while car is moving**	**object is released and begins to fall**	**plane lands**	parachute opens at apex of vertical flight
IMPULSE	**bounce of ball w/ horizontal velocity**	stationary hockey puck slapped on frictionless ice	**falling feather hits ground**	bounce of vertically dropped ball
STEPULSE	swinging tetherball struck, breaking tether	**golf ball is putted**	**beanbag falls to ground**	plate dropped from air into water

Table 1: Compound motion boundaries: the fifteen possible motion boundaries, shown as combinations of dynamic and reference frame boundaries. Examples of all boundaries are given. The common conjunctions are shown in boldface; peculiar combinations are in small font. When there is neither a reference nor a dynamic boundary, there is a period of nonzero speed with continuously changing force: this is a motion era (Rubin, 1985).

options for reference frame boundaries: start, stop, pause, and no speed zero. There are thus 15 distinct motion boundaries; from 4^2 we subtract 1 for the case in which there is neither a dynamic nor a reference frame boundary. These compound motion boundaries are shown in Table 2. All 15 compound boundaries make mathematical sense, but only some of them are physically common.

To illustrate the notion of compound motion boundaries we examine in Figure 5 the four types of start distinguished in Table 1. (The following argument applies to stops as well.) Let $s(t)$ characterize speed as a function of time, and let $\dot{s}(t)$ be its derivative. Then the four dynamic possibilities for starts are simply described by considering s and $\dot{s}$ are step-discontinuous at the start time t_0. If both functions are continuous, there is no co-occurring

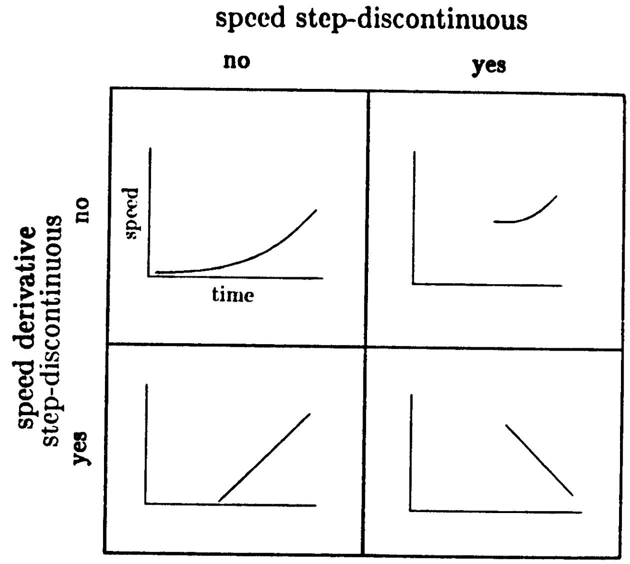

Figure 5: Four types of start. Time is plotted on the abscissas; speed s or its derivative $\dot{s}$ are plotted on the ordinates. Start occurs at time t_0. The four possible starts are determined by the continuity of s and $\dot{s}$ at t_0.

dynamic boundary. If s is continuous, but not $\dot{s}$, then there is a co-occurring step discontinuity of force. If $\dot{s}$ has no step discontinuity, but s does, there must be an impulse. Finally, if both s and $\dot{s}$ are step-discontinuous, a stepulse has occurred.

7 The Trace of Motion

Above we showed how velocity and acceleration step discontinuities, starts, and stops in the image imply motion boundaries. We have yet to explore *position* in the image. That is, can information in the static trace of motion in the image be useful in determining what motion boundaries have occurred? We show in this section that the answer is yes, *given a continuous trace in the image that has been created continuously*[9] *in time*. The results are summarized in Table 2.

7.1 Terminal Points

First, it is clear that given a generic viewpoint, there is a terminal point in the image if and only if there is a terminal point in the three-dimensional trace. The moving blob must have had zero speed at the terminal point, otherwise the blob would have moved $\vec{v}\Delta t$ in the next small time interval

[9] The correct traversal of the trace must be specified at a point of self-intersection.

2D TRACE FEATURE	**DIAGRAM**	**INFERRED MOTION BOUNDARIES**
terminal point		speed zero
cusp		speed zero
corner		impulse of force or speed zero
step-change of curvature		step discontinuity of force or speed zero

Table 2: Illustration of features of image traces that allow inferences about motion boundaries. The boundary interpretations here are *necessary*; *possible* boundaries—such as an impulse coinciding with the speed zero at a cusp—are not listed.

$\triangle t$, making the point nonterminal[10].

7.2 Cusps and Corners

First we will define cusps and corners. Next we will show that cusps and corners in a three-dimensional curve almost always project orthographically to cusps and corners, respectively, in the image. Finally, we will prove that cusps and corners necessarily entail motion boundaries.

[10] Another possibility is that the blob exploded. We assume a visual system would be able to detect such a dramatic event and distinguish it from a speed zero.

7.2.1 Definitions

Let $\vec{T}(s)$ be the unit tangent vector to the three-dimensional curve, where s is arclength. At a certain distance s_0 along the curve, let $\vec{T}^{+}(s_0)$ and $\vec{T}^{-}(s_0)$ be the left- and righthand tangent vectors. When the tangent is continuous at s_0, $\vec{T}^{+}(s_0) = \vec{T}^{-}(s_0)$. By definition, there is a cusp at s_0 if the two tangents are anti-parallel: $\vec{T}^{+}(s_0) = -\vec{T}^{-}(s_0)$. Finally, there is a corner at s_0 if the two tangents are neither parallel nor anti-parallel. A concise way of expressing the conditions above is with the dot product $d = \vec{T}^{+}(s_0) \cdot \vec{T}^{-}(s_0)$. At a particular place on a curve, if $d = 1$, the tangent is continuous; if $d = -1$, there is a cusp; and if $d \in (-1, 1)$, there is a corner.

7.2.2 Detection

In Appendix III we show that, almost always, there is a tangent discontinuity in the image iff there is a tangent discontinuity in the three-dimensional curve. We must show more specifically that image cusps (corners) are reliably related to three-dimensional cusps (corners). This is clear: any reasonable projection to the image will map a pair of antiparallel vectors (at a point in $\Re^3$) to a pair of antiparallel image vectors. Furthermore, two arbitrary $\Re^3$ tangent vectors (as in a corner) will almost never map to antiparallel image vectors.

7.2.3 Motion Boundaries

We interpret cusps and corners in the image by noting that the velocity vector of a curve is always parallel to the tangent vector to the trace. Consider a corner. We claim that a corner implies either a speed zero (a reference frame boundary) or a step-discontinuity of velocity (force impulse). Suppose the object has finite speed at the corner. Then speed must change instantaneously at the corner, else the trace would have been extended along the direction of velocity; that is, there would have been no corner. Therefore, there is a speed zero or a force impulse at a corner.

At a cusp, velocity exactly reverses direction. Thus cusps entail speed zeroes. (A cusp could also involve a force impulse, as a corner, but the impulse must be exactly antiparallel to the direction of motion.)

7.3 Curvature Discontinuities

We show in Appendix III that (almost always) image velocity and acceleration are continuous iff three-dimensional velocity and acceleration are continuous. If $\vec{p}(t)$ is a plane or space curve, then whenever the curve has nonzero speed, curvature is given by (Flanders *et al.*, 1970, p. 489):

$$k(t) = \frac{[|\dot{p}(t)|^2|\ddot{p}(t)|^2 - (\dot{p}(t) \cdot \ddot{p}(t))^2]^{\frac{1}{2}}}{|\dot{p}(t)|^3} \tag{2}$$

Curvature of plane and space curves is thus a continuous function of velocity and acceleration. By Appendix III, velocity and acceleration in the image are continuous iff they are continuous in the space curve. But then curvature in the image is continuous iff space curvature is continuous. Contrapositively, there is (almost always) a curvature discontinuity in the image of the trace iff there is a curvature discontinuity in the trace.

Furthermore, by inspection of (2), we can see that almost all step discontinuities of force (acceleration) bring about discontinuities in curvature. Note that force impulses cause corners which are infinities of curvature. Therefore, almost always, there is a step-change of curvature (as opposed to a corner, which is an isolated infinity of curvature) iff there is step-change of force.

8 Psychophysical Evidence

While watching motion, human observers sometimes have subjective impressions of fleeting, significant events (see again the bouncing ball of Figure 1a). We will call perceived motion boundaries "subjective" and those that we have defined mathematically "theoretical." We hypothesize that our four elementary theoretical motion boundaries describe a competence of a human observer (Chomsky, 1965; see also Yilmaz, 1962 and Marr, 1982); the human visual system in principle represents these elementary events. (We do not claim the human visual system distinguishes all fifteen compound motion boundaries.) That is, there is a subjective motion boundary

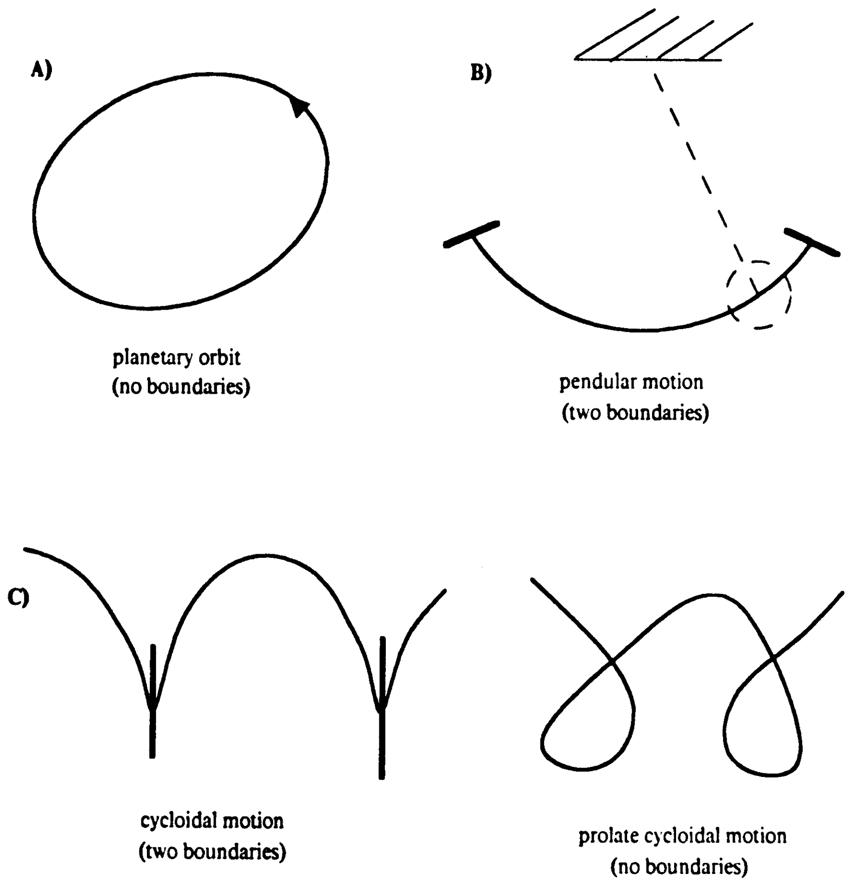

Figure 6: Motions and motion boundaries (marked by heavy slashes) a) Planetary orbit—no boundaries. b) Pendulum. The two terminal points are the only boundaries. c) Cycloidal motion: the cusps are motion boundaries. d) The motion of a prolate cycloid: there are no boundaries.

if and only if there is a theoretical motion boundary. For example, the planetary orbit shown in Figure 6a has no theoretical motion boundaries, and none are seen.

We present evidence for our claim below. While it seems to be the case that there are subjective boundaries if and only if there are theoretical boundaries, it is not clear that the visual system distinguishes the fifteen compound boundaries of Table 1. We begin by showing that the human visual system is sensitive to aspects of acceleration, a capacity necessary for the detection of step-changes of force.

8.1 Sensitivity to Acceleration

There is solid evidence to suggest human sensitivity to aspects of motion acceleration. A display can be generated of a bouncing ball. A second display can be made that creates a trace identical to that of the first display, but traverses it at constant speed. The two displays, differing in acceleration but not average speed, appear strikingly distinct (Rubin, 1985). The constant-speed ball seems to whip around the apices of its path.

Additional evidence that the human visual system is capable of representing aspects of acceleration comes from work in a motion extrapolation paradigm. Rosenbaum (1975) showed subjects an object that moved horizontally at constant acceleration and disappeared behind a barrier. Subjects indicated when they thought the object—now no longer in view—reached a marked location on the barrier. Results showed that subjects extrapolated motion at constant acceleration, as opposed to, say, the average velocity of the object while it was visible. In an extension of Rosenbaum's work, Jagacinski *et al.* (1983) found that extrapolated trajectories of constant-acceleration were modeled by a period of constant acceleration followed by a period of constant velocity. Thus it is clear the human visual system is sensitive to *some* aspects of acceleration; it remains to be shown that step-changes and impulse are among these aspects.

8.2 Reference Frame Boundaries

Pauses, as in the pendular motion of Figure 6b and the cycloidal motion of Figure 6c, are subjective boundaries. If a constant horizontal velocity is added to a cycloid, prolate and curtate cycloids obtain. These new trajectories have no pauses, and—when the trajectories are sufficiently distinct from a cycloid—no motion boundaries are seen (Rubin, 1985). See Figure 6d.

The perception of starts has been examined by Runeson (1974, 1977). He presented a variety of starts differing in how speed changed as a function of time[11]. Runeson argued that human perception distinguishes two sorts of starts. Undramatic starts are seen when speed increases smoothly

[11] All of Runeson's displays were of linear motion.

from zero to some asymptotic velocity. Otherwise, dramatic or eventful starts are seen. Perception of undramatic starts involves simply a sensation of velocity and its inception. Eventful starts are perceptually more complex: observers report something *happening* at the beginning of motion distinct from their sensation of velocity. Runeson's results are consistent with our claim that starts cause subjective motion boundaries. However, our scheme distinguishes four types of starts to Runeson's two. Whereas all of our start types are motion boundaries, they are not all perceptually distinct: stepulse and impulse starts are dramatic; smooth and step starts are uneventful. (The perception of stops is analogous; stepulse and impulse stops are violent, whereas smooth and step stops are peaceful.)

8.3 Dynamic Boundaries

Force impulses, as exemplified by bounces, are seen as motion boundaries. To examine the appearance a step-change of force, we created a display of a ball moving with constant acceleration in one direction. When the direction of acceleration was suddenly switched, the change was conspicuous to observers as a motion event. Furthermore, force step-changes are perceptually distinct from a force impulses; the former are legato, the latter, stacatto.

9 Discussion

9.1 Properties of Motion Boundaries

To be useful for recognizing different kinds of motion, a motion representation should be *psychologically relevant, mathematically convenient*, and *physically apropos.* Psychological relevance means that the primitives of the representation should be computable by the human visual system, and the scheme should divide the class of trajectories into roughly the same equivalence classes as human observers. Our treatment of motion boundaries satisfies this criterion. Motion boundaries are the *transient* aspect of a complete motion representation, the *ongoing* aspect of which is eras (Rubin, 1985), descriptions of the periods between successive boundaries. The human visual system is sensitive to a few motion *events*, and preliminarily, it seems that our theoretical boundaries are at worst a refinement of the psychological classification of motion transients[12].

A mathematically convenient representation is one that has useful invariant properties. Our motion boundaries, based on local properties of $\vec{p}(t)$, the description of three-dimensional position as a function of time, exhibit three useful invariances. The motion boundaries do not depend on the units for measuring space; they are invariant over spatial scaling. Also, if a given motion is repeated twice as fast, the boundaries maintain their relative positions: this is speed scaling. Finally, the motion boundaries are *transparent*. A local feature of a three-dimensional curve is said to be "transparent" when there is an associated feature in the image of the curve—call it the "shadow" of the three-dimensional feature—such that whenever the shadow is found in the image, the feature is guaranteed present in the world, and whenever the feature is present in the world, images of the curve from almost all viewpoints contain the shadow. (See Appendix IV.) Stated more simply, a transparent feature is one that can be found without error from almost any viewpoint.

The result is even stronger than just stated: not only do the motion boundaries have useful invariant properties; they are the *only* reasonable local properties of motions having those invariances. Consider the class of local properties of curves that are invariant over spatial scaling. This class consists of zeroes and impulses of a curve and its derivatives. This is the same class as that of local curve properties invariant over speed (or force) scaling. It is also the same as the class of transparent properties. *Each of these three types of invariance thus independently specifies the same class of local properties.* (See Appendix IV.) Furthermore, our motion boundaries are the lowest-order members of this class[13].

---

[12] Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of sets that partition a universe $\mathcal{U}$ into equivalence classes. Then $\mathcal{X}$ is said to be a *refinement* of $\mathcal{Y}$ if the members of a member of $\mathcal{X}$ are members of exactly one member of $\mathcal{Y}$. A refinement of a classificatory system thus makes all the original distinctions, and then some. Crucial to the notion is the fact that refinements do not carve up the universe in an independent way.

[13] An example of a higher-order term that is not a motion boundary is an impulse in the eleventh derivative of the curve.

There is no a priori reason that mathematically convenient boundaries should be physically meaningful. In fact, the higher-order members of the class of invariants described in Appendix IV (impulses in the eleventh derivative of position, say) are probably without physical importance. But, as we have argued, the lower-order members—starts, stops, and dynamic discontinuities—signify meaningful force events.

9.2 Relation to Previous Work

Some of our suggestions for the representation of motion have been made by others. Our work is distinguished from previous motion studies by the following combination of features. First, the scope of our representation is large, encompassing any piecewise-continuous motion of any shape that can be construed as a point or blob. Second, the definition of motion boundaries is founded in the physics of the macroscopic world; our representation makes force explicit. Third, we show rigorously that our theoretical motion boundaries are in one-to-one correspondence with certain kinematic image conditions, and, more importantly, are the lowest order members of the class of reliably detectable local properties of motion. Finally, our scheme is "bottom-up;" no knowledge of the shape or motion of particular objects is necessary.

Gibson (1979, p. 101) advanced the idea that motion can be divided into natural parts, writing ". . . the flow of ecological events consists of natural units" that are arranged hierarchically so that "[w]hat we take to be a unitary episode is therefore a matter of choice . . . ." In contrast, we suggest that motion boundaries are rigidly defined. (We do not rule out the possibility of description at two or more scales; see Appendix II.)

Runeson (1977) focused on how material properties of objects (relative mass, elasticity, and so on) could be inferred from kinematics. While Runeson would be interested, say, in inferring the elasticity of a bouncing ball, we are primarily interested in recognizing bouncing motion. Though his goal differs from ours of motion recognition, his distinction between *events* and *processes* is similar to our division between continuous motion eras and transient motion boundaries. Events for Runeson are abrupt, evanescent occurrences that signify energy transfer; processes are enduring kinematic goings-on. Runeson's events seem related to the dynamic subset of motion boundaries in this paper.[14] Runeson argued that a perceptual system must give priority to dynamic events over processes because the former are more informative about causal relations. By contrast, we give equal weight to dynamic and reference frame boundaries in our motion representation.

Forbus (1981) undertook to describe complex motions by what he called an "Action Sequence." An Action Sequence is a concatenation of Acts, each of which is a period (or a moment) of a single type of motion that can be described by a particular equation. Bouncing motion is, for example, represented as FLY UP, FLY DOWN, COLLISION, FLY UP, FLY DOWN, and so on.

Some important differences between Forbus's work and ours must be noted. Forbus was interested in *reasoning* about motion from diagrams or word problems; we are interested in *perception*. Forbus used a restricted two-dimensional domain in which gravity is the only force. Furthermore, his program only reasoned about Acts that are on a menu, namely, Collide and Fly. The scope of our theory is greater; any sort of (piecewise-continuous) force in three dimensions is acceptable, and more importantly, our scheme does not rely on a menu of force equations; novel forces can be represented.

Two motion representations have been proposed to describe the motion of complex, articulated shapes. Laban (1975) developed an elaborate notation for transcribing choreography. "Labanotation," as it is called, is necessarily specific to the human form. Marr & Vaina (1980) offered a means of representing the motion of objects that admit 3D model descriptions. These two schemes thus have narrower scope than ours since they apply only to certain classes of shapes. Moreover, their descriptions relate the motion of parts to superordinate parts or the whole, paying less attention to the overall motion of the whole. (Marr & Vaina suggested without motivation that the representation of the motion of the entire 3D model mark speed zeroes and step discontinuities of velocity.)

Badler (1975) took a top-down approach to the description of motion based on a sequence of static drawings. That is, his scheme required that objects be recognized so that knowledge about them can be used in the mo-

[14]Events are not precisely defined in Runeson's work. He seems to have velocity but not acceleration step discontinuities in mind.

tion description. By contrast, we suggest a visual system extract the most informative possible description of motion independent of object recognition. Badler did, however, suggest that the representation of motion make explicit an assortment of conditions that included starts, stops, and trajectory discontinuities.

10 Summary

We have proposed that blob motion be cut into eras at certain natural boundaries. The eras can then be given qualitative descriptions (Rubin, 1985). This approach to representation for recognition is reminiscent of Hoffman & Richards' (1982,1984) work on static planar contours. They cut contours at a boundary condition and then describe the resulting natural parts qualitatively. A significant difference, however, is that the motion boundaries of this paper are in themselves meaningful; Hoffman & Richards' contour boundary condition serves only to separate parts.

We began by noting criteria for a motion representation suitable for recognition: *stability* over niggling variations in trajectory, and *invariance* over space and time scaling. We defined two types of motion boundaries that satisfy the two criteria above. Dynamic boundaries (force discontinuities) are a good foundation for a motion representation in that their appearance in the image is reliably related to events in the three-dimensional world, regardless of the (smooth) motion of the observer. Reference frame boundaries—starts, stops, and pauses—depend on the viewer's frame. We gave evidence that our theoretical motion boundaries underlie subjective boundaries.

Appendix I: Rigid Bodies and Impulses

Claim. If a rigid body is subjected to a force impulse, then almost all its points will move as if they have been subjected to a force discontinuity.

Proof sketch. Any three-dimensional motion of a rigid body is equivalent at each instant to a unique *twist* (Coxeter, 1961). A twist is defined by an axis l, an angular velocity ω about l, and a translational velocity v

along l. An impulse will cause a discontinuity in at least one of l, ω, and v.

Case I: The impulse changes translation v instantaneously (and possibly ω as well). Then all points are affected, since translations have no invariant points. *Case II:* The impulse affects the rotation of the object (but not its translation). All points on the rigid body that lie on l are invariant, and will not show the impulse. But for a two- or three-dimensional rigid body, such invariant points constitute a measure-zero set. Hence there is zero probability of choosing an invariant point on the body at random. *Case III:* The impulse changes the twist axis l instantaneously into a new axis l'. Then the only invariant points on the rigid body will be those that lie on the intersection of l and l'. By Case II, there is zero probability of selecting such a point.

Appendix II: Scale Problems

Here we describe (but do not solve) some problems of scale that affect the detection of motion boundaries. These problems are analogous to scale issues that arise when trying to represent static figures.

Resolution

Any visual system that detects starts will have to face the following problem of scale. Let speed be given by $s(t) = |\sin \frac{1}{t}|$ for, say $t \in (0, 1]$. Note there are an infinite number of values of $t \in (0, 1]$ that satisfy the definition of start. The number of starts actually perceived will be finite and depend on the spatiotemporal resolution of the system. Analogously, if one inspects (the static) graph of the function $s(t)$, one will see only a finite number of points where the graph of the function touches the x-axis, depending on the spatial resolution of the human visual system.

Discontinuity

The detection of dynamic discontinuities is a scale problem that is not solved simply by knowing the spatiotemporal resolution of a visual system. Consider the analogous static problem of deciding when a continuous curve has a corner. That is, when does the tangent change rapidly enough (with arclength) to be considered discontinuous? Clearly no fixed "$\triangle$ tangent" will suffice; rather, the critical value seems to depend on how rapidly the tangent is changing in a neighborhood. We expect similar considerations to apply to the detection of velocity and acceleration step discontinuities.

Description of Independent Scales

Consider how the human visual system might represent the shape of a tire. It is reasonable to suspect the description has at least two distinct scales (see Mandelbrot, 1977). At the larger scale, the overall rounded torus is described; at the smaller scale, the terrain of the tread is represented. Certain motions will also be best described at two separate scales. Consider the motion of a reaching hand: it moves through space along a smooth arc. Looking closer, one might notice the hand is trembling. These two sorts of motion are independent—independent in spatial and temporal scale and independent in cause. Note that the representation of the tremble might have motion boundaries (the pauses where the oscillation reverses direction), whereas there might be no boundaries in the larger scale description. The punchline is that, for complex motions, motion boundaries must be sought at a particular scale, and that descriptions at two or more scales might be necessary.

Appendix III: Images of Discontinuities

Below we show that images of continuous curves are continuous. It is intended that $\vec{p}$ be interpreted either as the position (section 5), the tangent (section 7.2), or the acceleration vector (section 7.3) of a three-dimensional curve as a function of time, and I is any projection function that maps space to an image plane such that the pre-image of every image point is a one-manifold in space (a generalized "line of sight"). It is clear that orthographic and perspective projection are reasonable in this sense.

Claim. Let $\vec{p}(t) : \Re^1 \mapsto \Re^3$ and $I : \Re^3 \mapsto \Re^2$ be continuous functions such that $I^{-1}(x,y)$ is a one-manifold in $\Re^3$ (i.e., rank(Jacobian(I))=2).

Then $\vec{p}(t)$ is continuous $\Longrightarrow I(\vec{p}(t))$ is continuous, and $I(\vec{p}(t))$ is continuous $\Longrightarrow$ (almost always) that $\vec{p}(t)$ is continuous.

Proof. Since the composition of continuous functions is continuous (Seeley, 1970), we have immediately that the continuity of $\vec{p}$ implies the continuity of $I(\vec{p})$. Next, suppose that at some t_0, $I(\vec{p}(t_0))$ is continuous, but that contrary to the claim we are to prove, $\vec{p}$ is discontinuous at t_0. We consider only step discontinuities. Let $\vec{p}^+(t_0) = \lim_{t \to t_0+} \vec{p}(t)$, and $\vec{p}^-(t_0) = \lim_{t \to t_0-} \vec{p}(t)$. A step discontinuity at t_0 implies $\vec{p}^+(t_0) \neq \vec{p}^-(t_0)$. By continuity of $I(\vec{p})$, we know $I(\vec{p}^+(t_0)) = I(\vec{p}^-(t_0))$. But I assigns the same $\Re^2$ value only to points in $\Re^3$ lying on a particular one-manifold. There is zero probability that the two points $\vec{p}^+(t_0)$ and $\vec{p}^-(t_0)$ lie on one of those special one-manifolds.

Corollary. Let $\vec{p} : \Re^1 \mapsto \Re^3$ be a vector-valued function, and let I be a reasonable and continuous imaging function as before. Then, almost always, $I(\vec{p})$ is discontinuous at t_0 iff $\vec{p}$ is discontinuous at t_0.

Proof. (Contrapositive of claim above.)

Appendix IV: A Unique Class of Boundaries

In this appendix we investigate *reliably detectable* local properties of space curves parameterized by time. Reliably detectable events are those that can be found with a 100% hit rate and no false targets. More specifically, a reliably detectable curve event is one that is associated with a particular image feature, such that almost all images of the curve possess that feature, and whenever that feature appears in an image, it is certain the space curve event has occurred. We are particularly interested in reliably detectable events that are invariant over speed and space scaling.

We will show that $\vec{0}$ and impulses of the derivatives of a space curve are the only reliably detectable events. The definitions and claims that follow serve more to make precise the special properties of $\vec{0}$ and impulses than to derive surprising mathematical results.

We begin by describing a moving point.

Definition. A *position curve in* $\Re^n$ (PC_n) is a continuous function $p : \Re \mapsto \Re^n$, $n \geq 2$.

We would like to generalize the notion of function in order that all position curves be differentiable infinitely many times (see Lighthill [1958] for a more elaborate generalization of functions). To this end we add a value "$*$" to the range of real vector-valued functions, where $*$ represents infinity, as in an impulse. The following definition assures that the values $*$ are sparse.

Definition. A *piecewise real curve* (PRC_n) is a function $h : \Re \mapsto \Re^n \cup \{*\}$, $n \geq 2$, such that $\forall r^* \in \Re$ such that $h(r^*) = *$, $\exists \epsilon > 0$ such that $\forall r \in (r^* - \epsilon, r^*) \cup (r^*, r^* + \epsilon)$, $h(r) \neq *$.

It is clear that a PC is a PRC. We next define the "generalized derivative" of a PRC so that impulses are treated correctly (see Bracewell, 1965). That is, the generalized derivative at a step or stepulse is an impulse; the generalized derivative at an impulse is $\vec{0}$; and the derivative at an ordinary point matches the normal definition of derivative.

Definition. $h^{(i)}$, the i^{th} *generalized derivative of a* PRC_n h is given recursively by

$$\forall \hat{r} \in \Re,\ h^{(i)}(\hat{r}) = \begin{cases} * & \text{if } \lim_{r \to \hat{r}+} h^{(i-1)}(r) \neq \lim_{r \to \hat{r}-} h^{(i-1)}(r) \\ \vec{0} & \text{if } h^{(i-1)}(\hat{r}) = * \\ \lim_{\epsilon \to 0} \frac{h^{(i-1)}(\hat{r}) - h^{(i-1)}(\hat{r}+\epsilon)}{\epsilon} & \text{otherwise} \end{cases}$$

where $h^{(0)}$ denotes h. The following is an immediate consequence of the definition above:

Claim. The generalized derivative of a PRC is a PRC; hence a PRC is infinitely generalized-differentiable.

We now discuss orthographic projections $I : \Re^3 \mapsto \Re^2$. We must state precisely how the projection is related to the coordinate frame of the position curve. We want to rule out the troublesome possibility of I varying erratically with r, as would be the case of a reference frame tied to a jackhammer. Any I can be exactly specified by six numbers: three rotations α, β, and γ, and the location of the image origin (a, b, c). We will call I an ***aristotelian view*** or *projection* of a PC_3 when $0 = \frac{d\alpha}{dr} = \frac{d\beta}{dr} = \frac{d\gamma}{dr} = \frac{da}{dr} = \frac{db}{dr} = \frac{dc}{dr}$. Simply put, an aristotelian view is stationary with respect to the world reference frame.

Next we prove a claim about reasonable images of position curves.

Claim. Given an aristotelian projection $I : \Re^3 \mapsto \Re^2$ and a PC_3, p, then $I(p) : \Re \mapsto \Re^2$ is a PC_2. Hence, $I^{(j)}(p)$, the j^{th} generalized derivative of $I(p)$, exists for all j.

Proof. First note that $I(p)$ is the composition of continuous real-valued functions and is therefore continuous and real-valued itself. It is, therefore, a PC_2. But PCs are infinitely generalized-differentiable, so the claim is proved.

We now want to formalize the idea of a reliably detectable event. Below, the local event of the curve to be detected is some particular value p_0 (the only *strictly* local property of a function is its value); its associated image feature is i_0.

Definition. A *j-transparent value of a PC_3, p, under aristotelian view* is any $p_0 \in \Re^3 \cup \{*\}$ such that $\exists i_0 \in \Re^2 \cup \{*\}$ such that for all aristotelian projections $I : \Re^3 \mapsto \Re^2$ and for all r, $p^{(j)}(r) = p_0 \iff I^{(j)}(p(r)) = i_0$.

Definition. A *transparent value of a PC_3 under aristotelian view* is a j-transparent value from some j.

We are now ready for the first major claim.

Claim. The only transparent values of a three-dimensional position curve under aristotelian view are $\{*, \vec{0}\}$.

Proof. Consider the second generalized derivative of the position curve. We claim $*$ is transparent; this follows from Appendix III, where we show that images of discontinuities are almost always discontinuities, and from the fact that $*$ is the derivative of a step-discontinuity. We claim that $\vec{0}$ is transparent due to the restriction to aristotelian view. The second derivative of the change of viewpoint is zero, hence $\vec{0}$ on the curve maps to $\vec{0}$ in the image, and almost always, $\vec{0}$ in the image arises from $\vec{0}$ in the space curve. It remains to show that other non-zero, non-impulse values of curves cannot be transparent. This follows because the orientation of a aristotelian image with respect to three-space is arbitrary (though fixed). The only vector invariant over rotation is $\vec{0}$, but transparency quantifies over all aristotelian views.

We next investigate another sort of invariance.

Definition. A *scaling of a PRC_3 h* is a function

$$g(r) = \begin{cases} \alpha h(r) & \text{if } h(r) \in \Re^3 \\ * & \text{if } h(r) = * \end{cases}$$

where α is some constant.

For a given position curve, suppose all velocities are doubled. The resulting motion is a scaling of the original motion. So is the motion that obtains when all forces (accelerations) are multiplied by some constant.

Definition. A *scalable value of a* PRC_3 h is any $p_0 \in \Re^n \cup \{*\}$ such that $\forall r \in \Re$ and for all scalings g of h, $h(r) = p_0 \iff g(r) = p_0$

We state the following without proof.

Claim. A PRC_n can have at most two scalable values, $\{*, \vec{0}\}$.

We have shown that transparency (under aristotelian view) and scalability independently select special values $*$ and $\vec{0}$. Note that a scaling of the position curve is equivalent to a change of spatial units. Scalings of the first and higher generalized derivatives of the position curve are speed and force scalings, and so on.

References

Badler, N.I., "Temporal scene analysis: conceptual descriptions of object movements." University of Toronto Department of Computer Science Technical Report # 80, 1975.

Bracewell, R.N., *The Fourier Transform and its Applications.* McGraw-Hill Book Company, New York, 1965.

Chomsky, N. *Aspects of the Theory of Syntax.* M.I.T. Press, Cambridge, Massachusetts, 1965.

Coxeter, H.S.M., *Introduction to Geometry.* John Wiley & Sons, New York, 1961.

Flanders, H., Korfhage, R.R. & Price, J.J., *Calculus.* Academic Press, New York, 1970.

Forbus, K.D., "A Study of Qualitative and Geometric Knowledge in Reasoning about Motion." M.I.T. Artificial Intelligence Laboratory Technical Report 615, 1981.

Gibson, J.J., *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, 1979.

Hoffman, D.D. & Richards, W.A., "Representing Smooth Plane Curves for Recognition: Implications for Figure-Ground Reversal." Proceedings of the National Conference on Artificial Intelligence, 5–8, 1982.

Hoffman, D.D. & Richards, W.A., "Parts of Recognition." *Cognition* **18**, 65–96, 1984.

Jagacinski, R.J., Johnson, W.W. & Miller, R.A., "Quantifying the Cognitive Trajectories Of Extrapolated Movements." *J. Exp. Psych: Human Perception & Performance* **9**, 43–57, 1983.

Laban, R., *Principles of Dance and Movement Notation.* Plays, Inc., Boston, 1975.

Lavenda, B.H., "Brownian motion." *Sci. Am.* **252**(2), 70–85, 1985.

Lighthill, M.J. *Introduction to Fourier Analysis and Generalized Functions.* Cambridge University Press, 1958.

Mandelbrot, B.B., *Fractals: Form, Chance, and Dimension.* W.H. Freeman & Co., San Francisco, 1977.

Marr, D., *Vision.* W.H. Freeman & Co., San Francisco, 1982.

Marr, D. & Nishihara, H.K., "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes." *Proc. Roy. Soc. Lond. B* **200**, 269–294, 1978.

Marr, D. & Vaina, L., "Representation and Recognition of the Movements of Shapes." M.I.T. Artificial Intelligence Laboratory Memo # 597, 1980.

Rosenbaum, D.A., "Perception and Extrapolation of Velocity and Acceleration." *J. Exp. Psych: Human Perception & Performance* **1**, 395–304, 1975.

Rubin, J.M., "Categorizing Visual Motion." M.I.T. Ph.D. Thesis (in preparation), 1985.

Runeson, S., "Constant velocity—Not Perceived as Such." *Psych. Res.* 37, 3–23, 1974.

Runeson, S., "On Visual Perception of Dynamic Events." Ph. D. dissertation, Uppsala, 1977.

Seeley, R.T., *Calculus of Several Variables.* Scott, Foresman and Co., Glenview, Ill. , 1970.

Thomas Jr., G.B., *Calculus and Analytic Geometry, Part I.* Alternate Edition. Addison-Wesley Publishing Co. Inc. , Reading, Massachusetts, 1972.

Wallach, H., "The Perception of Motion." *Sci. Am.*, **201** #1, 56–60, 1959.

Yilmaz, H., "Color Vision and a New Approach to General Perception," in *Biological Prototypes and Synthetic Systems, Volume I*, 126–141, E.E. Bernard & M.R. Kare, eds., Plenum Press, New York, 1962.