

Shapes, Shocks, and Deformations I: The Components of Two-Dimensional Shape and the Reaction-Diffusion Space

BENJAMIN B. KIMIA

Laboratory for Engineering Man-Machine Systems, Brown University, Providence, Rhode Island 02912

ALLEN R. TANNENBAUM

Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455

STEVEN W. ZUCKER

Center for Intelligent Machines, Department of Electrical Engineering, McGill University, Montreal, Quebec, Canada H3A 2A7

Abstract. We undertake to develop a general theory of two-dimensional shape by elucidating several principles which any such theory should meet. The principles are organized around two basic intuitions: first, if a boundary were changed only slightly, then, in general, its shape would change only slightly. This leads us to propose an operational theory of shape based on incremental contour deformations. The second intuition is that not all contours are shapes, but rather only those that can enclose “physical” material. A theory of contour deformation is derived from these principles, based on abstract conservation principles and Hamilton-Jacobi theory. These principles are based on the work of Sethian (1985a, c), the Osher-Sethian (1988), level set formulation the classical shock theory of Lax (1971; 1973), as well as curve evolution theory for a curve evolving as a function of the curvature and the relation to geometric smoothing of Gage-Hamilton-Grayson (1986; 1989). The result is a characterization of the computational elements of shape: deformations, parts, bends, and seeds, which show where to place the components of a shape. The theory unifies many of the diverse aspects of shapes, and leads to a space of shapes (the reaction/diffusion space), which places shapes within a neighborhood of “similar” ones. Such similarity relationships underlie descriptions suitable for recognition.

1 Introduction

Our visual world contains a vast arrangement of objects, yet we are amazingly robust in recognizing them. This includes objects projected from novel viewpoints, or partially occluded objects. We are even able to describe totally unfamiliar objects, or to recognize unexpected ones out of context. Examples include the shapes displayed in Fig. 1; even though they are hand-drawn, they are immediately recognizable. We seek a theory of shape sufficiently robust to support recognition tasks such as this.

While there is a sense in which the meaning of shape is effortlessly and intuitively understood, a formal definition has been elusive: there is currently no generally accepted definition of shape in either computational vision or psychology. Part of the difficulty is that different representations may be appropriate for different tasks: what is required for navigation is not necessarily required for recognition, and what is required for recognition of generic object types may be different from what is required for recognizing particular instances of

a given type. That is, recognizing the generic “person” or “car” in Fig. 1 is different than recognizing an image of your car from among several car images.

Particular object recognition from among a pre-specified class of objects is amenable to a variety of template- and iconic-matching techniques (Fischler and Elschlager 1973; Ferrie, Levine and Zucker 1982; Solina and Bajcsy 1990). These are based on metric properties, and thus are inappropriate for the generic problem. A theory of generic object recognition must be robust to variations within scenes, e.g. due to the hand-drawn character of the images in Fig. 1, or (more realistically) to small changes in viewpoint, to the changing appearance of objects due to local motion and emergent occlusions, as well as to variations within objects, e.g. due to flexibility, growth, and inflation. We seek to develop a theory of generic object recognition that meets these needs.

Many early visual processes contribute to shape: edges, texture, color, shading, and so on. We observe from the examples in Fig. 1, however, that the generic shape recognition problem is intuitively well defined

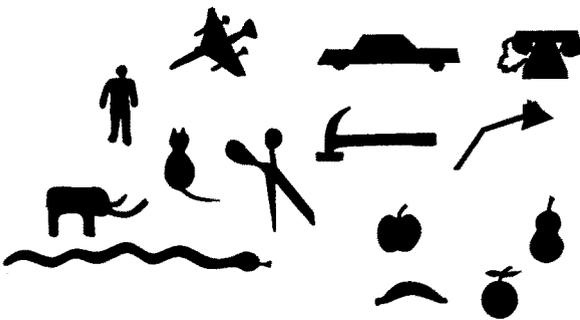


Fig. 1. Despite the poor quality of these hand-drawn shapes, we are able to recognize the underlying objects effortlessly. What aspect of the geometry of these figures should be computed to allow robust recognition?

for two-dimensional outlines of objects, and it is on this problem that we focus. In particular, we derive a formal framework for our theory from a mathematical model of deformations, and, in so doing, consider the notion of curve evolution via partial differential equations. The mathematical framework underlying this is based on the work of Sethian (1985a–c; 1989), Sethian and Osher (1988), and the fundamental Osher-Sethian level set algorithm (Osher and Sethian 1988) which has already proven to be of enormous use in image processing and interface motion. We should also add that while the mathematics is different, the spirit of our work was prefigured in many ways by Koenderink (1986) in his important paper on “dynamic shape.” But we do not only want to emphasize the mathematics in such approaches: in an attempt to capture the intuitions underlying shape, we postulate a series of natural principles to which any such theory should be subject, and show how our theory is consistent with them.

The paper is organized as follows. The next section motivates the need for a novel geometry for shape, and we introduce a basis for this which lies in the differential evolution of deformations. Although the mathematics of deformation are well studied, we here introduce them to the field of computer vision. We focus on a special class of uniform deformations, and show that they can be represented as the combination of two basis deformations: a constant deformation and a curvature deformation. To handle shapes with discontinuities in their outlines, we next abstract the mathematical framework considerably, by showing that the deformations are equivalent to a hyperbolic conservation law with viscosity as discussed extensively in (Sethian 1985b, c). This is significant because such nonlinear conservation laws lead to the formation of shocks and to a notion of entropy (Lax 1973; Sethian

1985a). We are now finally able to close the loop back to shape, and to present the most novel results in the paper: namely, how different classes of shocks correspond to the computational components of generic shapes. Intuitively this connection between shocks and generic shapes is pleasing, because, just as shocks are the singular events in the continuum of geometric variation, generic shapes are the singular ones (the categories, in the sense of Rosch (1976)) in the continuum of physical objects. The remaining task is to define spaces of shapes, and to show how the notion of scale (or significance) arises within them. This leads to what we call the reaction/diffusion space, which induces a topology over “similar” shapes, and to a hierarchical description of a shape, with parts, protrusions, and bends specified. The result is a mathematical framework for shape that unifies earlier approaches in a unique way, using the aforementioned mathematical techniques for their analysis.

2 The Multidimensional Nature of Shape

Shape is multifaceted, in that it involves a range of “dimensions,” or aspects, and much of the earlier research on shape has been to elucidate one or another of them. This is certainly reasonable if the goal is selecting a particular shape from among members of a family, because the differences can be quantitatively expressed; an example is the “bent paper clip” shapes studied by (Bulthoff and Edelman 1992), in which the differences are captured uniquely by sequences of angles. However, such differences are irrelevant for other tasks, such as classifying bananas, for which boundary curvature might be more relevant. It follows that a theory of generic shape must span these many individual dimensions, several of which are reviewed below. We later show how our approach accomplishes this.

2.1 Bounding Contour vs Region

Should a shape be represented by its bounding contour, or by its interior region (Ballard and Brown 1982)? Physically the representations are different, in that boundaries arise when one substance interpenetrates another, e.g., oil flowing into water, while interior regions arise via growth processes. Conceptually the two approaches are equivalent in the sense that the interior is accessible via the boundary and vice versa. As such, most approaches concentrate on either representing one or the other (see Table 1). Nevertheless,

Table 1. Most methods concentrate either on the boundary or on the interior of a shape. However to fully capture shape a simultaneous representation of both boundary and regional properties is needed.

Boundary	Region
Strip tree (Ballard 1981)	Quad tree (Samet 1980)
Chain code (Freeman 1974)	SAT (Blum 1973)
Polygonal approximation (Ramer 1972)	Generalized ribbons (Brooks 1981)
Deformable snakes (Kass et al. 1988)	MDL (Leclerc 1989; Pentland 1989)
Fourier descriptors (Zahn and Roskies 1972; Granlund 1972; Persoon and Fu 1977; Wallace and Wintz 1980)	Fourier descriptors (Gardenier et al. 1986)
Codon (Richards and Hoffman 1985; Leyton 1988)	Superquadrics (Pentland 1987)

representations make certain information explicit while implicitly encoding the remainder. For example, when a shape is represented by the boundary, the orientation information is explicit while closeness of points along the “necks” and symmetry are implicit. In contrast, in a region-based representation, the orientation of the boundary tangent is implicit while the closeness of points through the region and object symmetry is explicit. A different trade-off may arise when considering computational complexity: boundary representations are one-dimensional and therefore inexpensive to process, while two dimensional regional representations are more expensive. We submit that a simultaneous representation of the boundary and the interior is needed for a full understanding of shape. While the distinction is in part semantic, since one could always blur it by basing region computations on a global boundary function (e.g. by deriving an implicit characteristic function), there is an important technical aspect to the distinction, as will emerge subsequently.

2.2 Local vs Global

One method of shape classification is based on shape features, e.g., area, eccentricity, centroid, compactness, shape moments, and others (Ballard and Brown 1982; Rosenfeld and Kak 1982). These shape features capture the shape by a few numbers and, as such, they are *global* approximations of the shape. In other words, information about the shape from *all* portions of it combine to form a global description. More powerful representations of shape can be global too. The Fourier representations (Zahn and Roskies 1972; Granlund 1972; Persoon and Fu 1977; Wallace and Wintz 1980) are global, in that each Fourier descriptor is dependent on all portions of the shape. Bolles et al. introduced *focus features*, in which global relationships of local features are represented (Bolles and Cain 1982). Hough

transform techniques gather votes for certain features and, as such, can also be classified as global.

The major problem with a global representation of shape occurs in the presence of occlusions. When an object is partially occluded, all global descriptors change drastically. As such, while they may work in a particular instance, they are not suitable for generic object recognition. Furthermore, a notion of approximation in the global domain does not correspond to that in the shape domain. For example, the shapes corresponding to a set of Fourier descriptors with or without some higher-order terms do not resemble each other closely. Local features in isolation, on the other hand, do not give a global sense of shape and are sensitive to noise. The challenge is to capture the general shape of the object without losing its partial representations. For generic object recognition, the representation must degrade gracefully as portions of the object are occluded.

2.3 Primitives vs Transformations

One way to characterize shape is to somehow define its constituent components. These components may be defined *a priori* as primitives that can either model the boundary, the region, or the three dimensional volume enclosed by the object; see Table 2. An alternative point of view is that shape is best characterized as a sequence of transformations from simpler shapes (Leyton 1988; 1989; 1987; Koenderink and van Doorn 1986; Koenderink 1990). How can one consolidate these seemingly conflicting approaches? We will show that our principles lead to a framework whose computational elements allow both compositions of arbitrary parts and deformations of shapes. Our notion of “parts” is novel, and, we believe, more in agreement with our intuition. Similarly, deformations allow for general representations of form, e.g., biological form, where no natural *a priori* primitives exist.

Table 2. A review of shape primitives.

Boundary-based primitives	Region-based primitives	Volume-based primitives
Codons (Richards and Hoffman 1985)	Generalized ribbons (Brooks 1981)	Generalized cylinders (Binford 1981; Marr and Nishihara 1978; Ulupinar and Nevatia 1990)
Arcs of circles (Dudek 1990)		Superquadrics (Pentland 1987; Solina and Bajcsy 1990)
Primitive Curvature Changes (Asada and Brady 1983)	MLD parts (Pentland 1989)	Geons (Biederman 1987)
Polygons (Ramer 1972)	SAT (Blum 1973)	Polyhedra (Waltz 1975)

2.4 Scale and a Hierarchy of Significance

The remaining elusive property of generic shapes is their composite organization. On one side is the fact that shapes consist of arrangements of “sub-ordinate” shapes, and so on—just consider a body. The torso is the most significant component, in that it dominates the shape relative to the arms, which in turn dominate the hands which in turn dominate the fingers. Thus there is a hierarchy of significance around shape (Marr and Nishihara 1978). However, such hierarchies are often confused with *scale spaces*, or hierarchies of image operators of different sizes (Witkin 1983), because the most significant components of a shape often have large image support. However, these notions become confused when the image smoothing properties of operator scale spaces are considered with respect to noise. It is key for shape that the significance hierarchy be tied to the object, and hence only indirectly to the image of the object. We conclude this paper with one such scale space, the *reaction/diffusion space*.

In summary, previous approaches have highlighted certain aspects of shape. While each of these choices works well in some situations, it may fail in others. Since the challenge is to cover all situations, we seek to unify these approaches, and to simultaneously represent the many facets of shape (Kimia et al. 1989; 1990).

3 The Need for a Novel Geometry of Shape

Objects come in all forms. As they deform and grow incrementally, their shape does not change drastically. For example, our perception of a tree is not drastically altered each day as it grows, or as a flock of birds rests on it. That the primary perception is one of a generic object with modifications is so intuitive to us

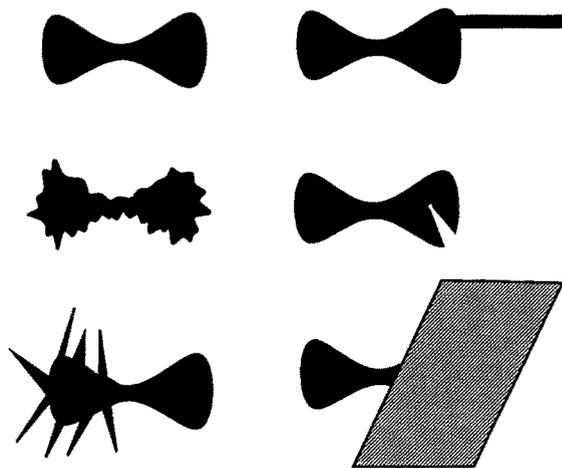


Fig. 2. Shapes are categorized into equivalence classes despite their differences. This is partly due to our ability to abstract the shape of an object in the presence of occlusions, protrusions, chips, noise, and various degradations. We seek a theory capable of supporting such generic competence.

that to mention it might appear redundant. Yet, it is essential that our computational representation of shape behave similarly, so that, e.g., an industrial tool that is slightly bent or chipped will be described as “a tool that was bent or chipped.” Analogously, when objects deform with motion, growth, erosion, etc., our perception is only slightly modified, as Fig. 2 illustrates. There seems to be great stability with regard to such changes.

3.1 Standard Geometries and Shape

Unfortunately, standard geometries of traditional and modern mathematics do not satisfactorily address these aspects of shape for the purposes of object recognition. Topology is so general that bounding contours of non-fractal physical objects (planar, closed, and simple)

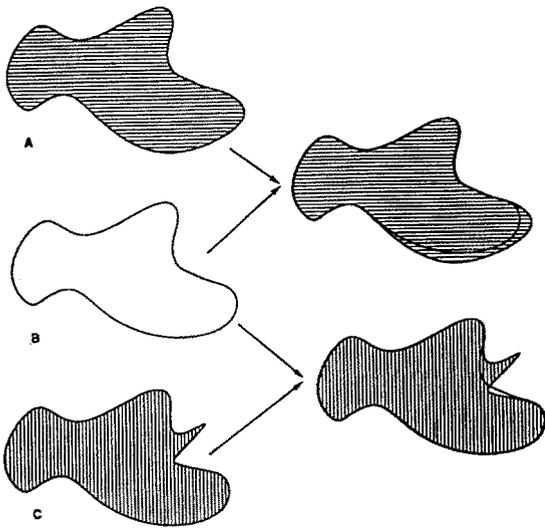


Fig. 3. Standard geometries do not properly capture the natural topological and metric issues associated with shape, as becomes apparent when different shapes are compared to determine which are more similar. For example, does figure (B) or figure (C) look more similar to (A)? Our intuitions suggest that, for general, unfamiliar objects such as these, (A) and (B) are most similar, because (C) has a visually salient “spike” protruding from it. But the Euclidean metric places less emphasis on this perceptually significant protrusion than on the slight metric differences in the boundary of the main lobe. This is illustrated by the overlapped outlines on the right side of the figure, which shows that the area difference between (A) and (B) is larger than that between (A) and (C). Thus the Euclidean metric would give the counter-intuitive result that (A) is more different from shape (B) than it is from shape (C)! We seek to develop a geometry in which the descriptions of (A) and (B) will be most similar, and significantly different from (C).

are equivalent. On the other hand, “Congruence geometries, such as Euclidean, affine, and projective geometries require an exact match, or some distance or area tolerance from it” (Blum 1973). Mumford questioned the success of a theory of shape description for recognition and categorization tasks without having first defined what is meant by a “nearby” shape (Mumford 1987). In other words, what is needed is to define a space for shape, and then to impose a topology on it. It is clear, however, that the Euclidean metric is not natural for shape, in the sense that certain “close” objects are perceived as different, and certain “distant” objects are visually indistinguishable (as in Fig. 3). A number of other metrics, e.g. the Hausdorff metric, have been considered. Koenderink and van Doorn (1986) point out that useful notions of “partial order, similarity, and relatedness” have no equivalent in the usual geometrical shape theories. Indeed, without these notions, the task of object recognition seems impossible.

These ideas point to the need for a language that makes the morphogenesis of shape explicit.

3.2 Salient Singularities

A second problem concerns the treatment of singularities. Singularities have often been reduced to limits of highly bent structures. Attneave argued that the most salient portions of a shape are corners and high curvature points (Attneave 1954). However, singularities do occur in nature, and they play a different role than their smoothed versions (Link and Zucker 1987). Given our predisposition to highly developed geometrical descriptions of curves and surfaces, it is not surprising that smooth curves and surfaces have been the common tools used in describing the visual world. Nevertheless, singularities must have an explicit place in a theory of shape, as well as in some other areas of vision.

3.3 A Novel Framework for Shape

What kind of geometry, then, do we require for shape? And, how does one define a metric and/or a topology for shape? What are our constraints and guidelines?

Note that objects come in all forms and sizes resulting in a dense varied collection of possible shapes. The task of generic object recognition is to ascribe a category to any given shape based on resemblance. Since the space of possible shapes is so varied, a language for representing it must be powerful enough to capture all the salient features. Moreover, shapes are not static in the world, since they are the projections of moving, deforming objects going into changing occlusion relationships with other objects. Since these transformations of shape are not discrete in nature, one can imagine a continuous space of shapes in which any shape is a point, and its transformation is a trajectory. Our task, then, is to discover how one might structure such a space through the relationship of a shape to its immediate neighbors.

We claim that the key to discovering such a structure is based on *deformations*. Observe that slight deformations of the boundary of a shape induce only a slight change in the original shape. We will show how to characterize local deformations in the next section. The result is that they may be qualitatively described as a combination of two *basis deformations*. It is precisely the combinations of these deformations that span the *reaction-diffusion space* for a shape, i.e., a “manifold” sampling the structure of the shape space.

A second key point concerns contours and shape. It is often assumed that a representation for a closed

contour is sufficient as a representation of shape. However, three dimensional objects are bounded by surfaces enclosing “material.” This translates into a two dimensional image consisting of a contour enclosing an interior. Hence, not all contours are eligible to be projected boundaries of objects; see next section. This inseparable combination of boundary and region is a significant aspect of shape. As such, when deforming shapes we must not only assure that the boundary remains a “nice curve,” but also that it encloses a “nice shape.”

Finally, we observe that in such spaces of continuously varying shapes, certain singular, or categorical, events will occur. Mathematically these singularities are called *shocks*, and we show how they define the generic components of shapes. Thus we are now in a position to explain the title of this paper: we study shapes by a process of curve evolution, during which singular “shock” events occur. These shocks within the space of deformed shapes define the categorical components of generic shapes.

3.4 Preview of Results

We close the introductory portion of the paper with a preview of our results, to provide a concrete focus for the ensuing theoretical discussion, and to show how they may be applied for object recognition. First, we illustrate the notion of deformation and how it leads to robust descriptions of parts. Figure 4 contains four images of pears, presented by Richards et al. (1986), and which were intended as gross modifications of an object category (pear). The original shapes are across the top, and each column contains samples from a continuous sequence in which the bounding contour has evolved according to our deformation rules. The samples were chosen to illustrate how the deformation process eliminates the noise (first row) to reveal the fundamental part structure for the pear (second row). This structure is a pair of lobes, with the most significant one on the bottom. The relevant shocks in this case signal the part structure, and correspond to the orientation discontinuities that develop on the evolving contour in ordered pairs. Note how the lobe structure, and the dominant lobe (bottom row) are comparable for each of these different pair images. The continuous space of shapes which supports such descriptions is called the *Reaction-Diffusion Space*.

A second example illustrates the notion of hierarchy in more detail. An image of a doll was chosen to show how the different “parts” emerge according to our natural intuitions about significance, Fig. 5. Note how

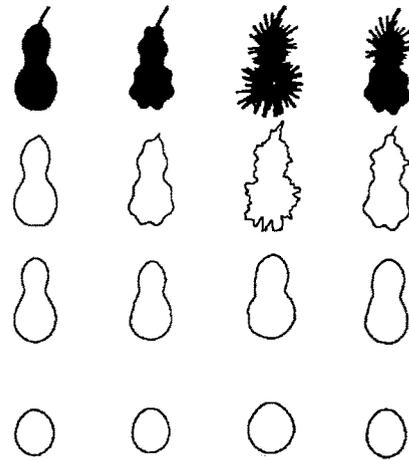


Fig. 4. An illustration of how our deformation approach to shape leads to natural descriptions despite large quantities of noise and texture. Four pears were proposed by Hoffman and Richards as gross modifications of a single object category (pear). The original shapes are across the top, in black. Each column contains samples from a continuous sequence in which the bounding contour has evolved according to our deformation rules. The samples were chosen to illustrate how the deformation process eliminates the noise (first row) to reveal the fundamental part structure for the pear (second row). This structure is a pair of lobes, with the most significant one on the bottom. The part structure is signaled by the shocks (discontinuities) that develop on the contour in opposing pairs. Note how the lobe structure, and the dominant lobe (bottom row) are comparable for each of these different pair images, even though the noise and texture were so prominent. Details on how the shocks form are developed in the paper; the images were obtained by normalizing scale by running the reaction term backwards from the shocks. Thus in our framework the description for each of these pears is a variant of “a large bottom, a small middle and a very small top.”

hands and feet are less significant than limbs, which are in turn less significant than the torso. This example also illustrates that several different types of shocks arise within our system, with first-order shocks signaling deformations, second-order shocks signaling part connections, third-order shocks signaling bends, and forth-order shocks signaling part centers. Note that occlusion will not affect decomposition into parts, a desirable feature for recognition.

4 Shape from Deformations

We now begin the development of our framework for shape. Since any recognition strategy requires a notion of similarity between shapes, or of a “neighborhood” around each shape, we begin with the proposition that the shape of an object should be intimately interconnected to “nearby” shapes. To illustrate, consider the

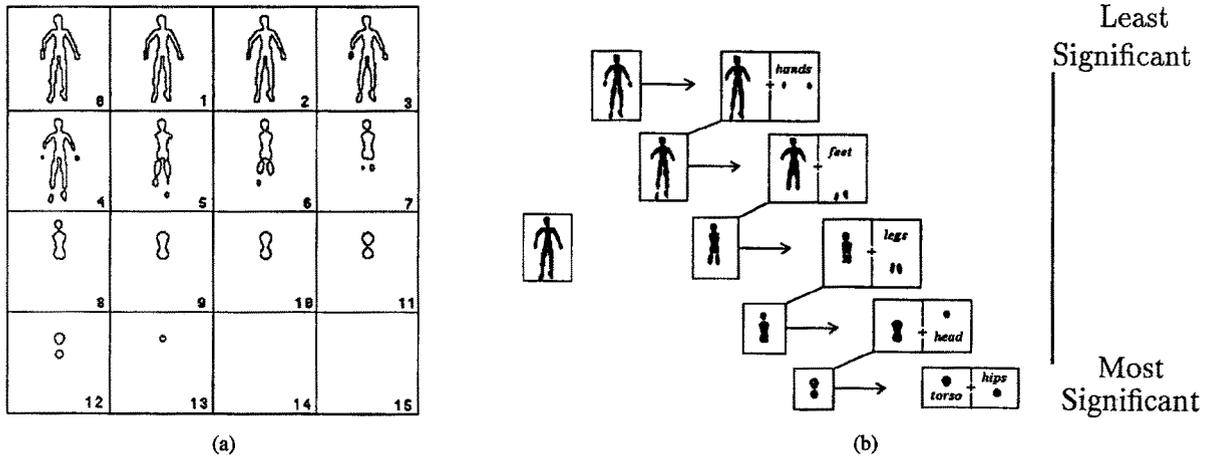


Fig. 5. a) The evolution of shocks leads to parts, protrusions, and bends. This figure shows the development of an image of a doll (National Research Council of Canada Laser Range Image Library CNRC9077 Cat No 422; 128 × 128). The contour shown in box *N* corresponds to increasing boundary evolution (time) steps. Observe that the “feet” partition from the “legs” (via second-order shocks) between frames 3 and 4, and the “hands” from the “arms” between frames 2 and 3. Following these second-order shocks, first-order shocks develop as the “arms” are “absorbed” into the chest. Running this process in the other direction would illustrate how the arms “protrude” from the chest. b) The hierarchical decomposition of a doll into parts. Selected frames were organized into a hierarchy according to the principle that the significance of a part is directly proportional to its survival duration.

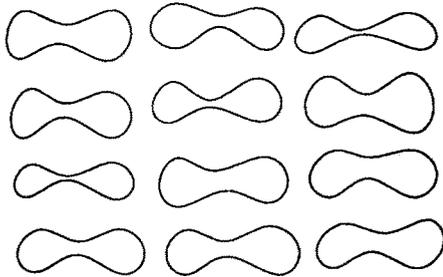


Fig. 6. These shapes seem to belong to the same group of objects. This concept of a neighborhood of “nearby” shapes is key to recognition.

shapes in Fig. 6; these are readily seen as similar, and as variations within the category of “peanuts.” We capture this kind of variation via deformations as characterized in differential geometry. Our approach is to apply arbitrary deformations to shapes and, through incremental change, to observe the emerging organization of the space of shapes. In the first subsection, we will discuss a model which captures a number of key shape operations and which is spanned by two simple deformations: constant deformations and curvature deformations. In the next subsection, a distinction is made between the evolution of *contours* and the evolution of *shapes*. The entire development is guided by several principles that we take to be fundamental and self-evident. More formally, they are proposed to ensure that evolving curves remain “valid” shapes,

i.e., possible projections of three dimensional objects onto a two dimensional image. Then, we show that it is in the interaction between constant and curvature deformations that the contrasting and complementary properties of shape are captured.

4.1 Shape from Deformations of Contours

Much of early vision is organized around inferring boundaries (Zucker et al. 1989). We therefore ask: How does the percept of a shape change as its boundary is modified slightly? We begin with the assertion that:

PRINCIPLE 1. *Slight changes in the boundary of an object cause only slight changes to its shape.*

Note that there are two different domains involved in this principle. The change in the boundary belongs in the domain of planar geometry of curves as measured by, say, the Hausdorff metric, or the domain of differential geometry as measured by the Euclidean metric. The change in shape, on the other hand, belongs in the space of shapes as measured by a similarity metric.

Thus we consider a shape represented by the curve $C_0(s) = (x_0(s), y_0(s))$ undergoing a deformation, where s is the parameter along the curve (not necessarily the arclength), x_0 and y_0 are the Cartesian coordinates and the subscript 0 denotes the initial curve prior to deformation. Now, let each point of

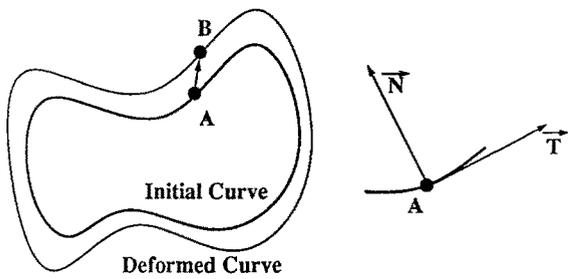


Fig. 7. The points on the initial curve A move to B to generate a new curve. The direction and magnitude of this motion is arbitrary in order to capture general deformations. However, with mild restrictions appropriate to shape, one can classify this deformation to be a sum of *constant deformation* and *curvature deformation* along the normal.

this curve move by some arbitrary amount in some arbitrary direction; see Fig. 7. This evolution is then described as

$$\begin{cases} \frac{\partial C}{\partial t} = \alpha(s, t)\vec{T} + \beta(s, t)\vec{N} \\ C(s, 0) = C_0(s), \end{cases} \quad (1)$$

where \vec{T} is the tangent, \vec{N} is the outward normal, s is again the parametrization, t is the time duration (magnitude) of the deformation, and α, β are arbitrary functions. This, by a reassignment (i.e. reparametrization) of points, can be reduced to (Kimia 1990),

$$\begin{cases} \frac{\partial C}{\partial t} = \beta(s, t)\vec{N} \\ C(s, 0) = C_0(s), \end{cases} \quad (2)$$

where β is again arbitrary, but not necessarily the same as that of the previous equation. Now, we concentrate on intrinsic deformations¹ that depend only on the *local geometry* of the curve at that point, namely those dependent on the curvature (do Carmo 1976),

$$\begin{cases} \frac{\partial C}{\partial t} = \beta(\kappa(s, t))\vec{N} \\ C(s, 0) = C_0(s), \end{cases} \quad (3)$$

where κ is the curvature.

Since we have an evolution equation for shapes, we now address the time at which deformations are applied. Recall that, since our deformations are intended to bring out the relationships among shapes, it is reasonable to require that the process relating shape S_1 to shape S_2 , is independent of *when* it is applied to S_1 . For example, the way an ellipse relates to a circle in the space of deformations should not be dependent on the time of the deformation, but rather on the amount and

nature of the deformation itself. Hence our second principle concerns time-invariance and we propose that:

PRINCIPLE 2. *The class of contour deformations necessary to articulate shape consists of those deformations that do not depend on the time the deformation is applied.*

Then from (3) we get,

$$\begin{cases} \frac{\partial C}{\partial t} = \beta(\kappa(s))\vec{N} \\ C(s, 0) = C_0(s). \end{cases} \quad (4)$$

There are a number of interesting possibilities to take for the function β . In this paper, we will choose a simple “first-order” model which captures both morphological operations as well as smoothing for our study of shape. Explicitly, we consider the following deformation:

$$\begin{cases} \frac{\partial C}{\partial t} = (\beta_0 - \beta_1\kappa)\vec{N} \\ C(s, 0) = C_0(s). \end{cases} \quad (5)$$

The remaining terms in a Taylor expansion of an analytic β involving odd higher orders of κ qualitatively resemble κ for the purposes of shape (Kimia 1990), or are not stable at high curvature points. We should note that this case has been extensively studied in (Osher and Sethian 1988; Sethian 1985a, b; Sethian and Osher 1988; Sethian 1985c; 1989).

The above equation contains two terms. The first term describes a deformation that is a constant motion along the normal, or *constant deformation*. The second term, describes a deformation that is proportional to the curvature along the normal, or *curvature deformation*. Such deformations will be fundamental to our framework, and will provide the basis for forming a topology over shape.

4.2 Shape Deformation vs Contour Deformation

Our next principles relate to the observation that not all contours are valid shapes. Recall that, informally, shape derives from the projection of three-dimensional objects, or volumes of material, onto two-dimensions. What kinds of contours, then, can represent shapes, and what kinds cannot? A “figure eight,” for example, cannot possibly represent the occluding boundary of an object: Even though a twisted paper clip may project to a figure eight, any clip composed of physical material projects to a region. The basic constraint is thus

that for contours to represent shapes they must be (the projection of) boundaries which could enclose “material”. This notion also seems to hold psychophysically (Elder and Zucker 1993).

The next principles articulate different aspects of this constraint that are necessary to restrict curve evolution to shape contours. First, we ask: what if, during the process of deformation, two remote points of the boundary touch each other? This would occur in the process of pinching a ball of clay. At the point when the two extremal points of the object come together, the object falls apart into two pieces.

PRINCIPLE 3. *If, during the process of deformation, distinct points of the boundary touch at a single point, then the evolved shape splits into two subshapes.*

It follows, of course, that once a shape has split it cannot be joined together again by continuing the process of deformation:

PRINCIPLE 4. *During the process of deformation the boundary of the shape must not cross over itself.*

This principle had an earlier expression in the grass-fire transformation of Blum (1973), who observed that grass could not burn twice.

What other properties exclude contours as projections of objects? Again, as volumes containing material, objects must project to closed contours. In other words, open curves cannot contain material. Therefore,

PRINCIPLE 5. *The boundary of a shape must remain closed during the process of deformation.*

How do singularities of the contour, such as corners and cusps, affect it as a valid candidate for shape? Since objects often have sharp edges, bends, etc., these project to corners and cusps in the contour. In fact, as was argued previously, these are among the salient points of a shape and deserve an explicit representation. However, there cannot be infinitely many such singularities, or for that matter extrema in curvature, because physical objects are composed of materials with a finite grain size and are observed by devices with finite resolution limits. This implies a finite total undulation in the two dimensional shape, and such total variation may be measured by total absolute curvature as defined by

$$\bar{\kappa}(t) := \int_0^{2\pi} |\kappa(s, t)| g(s, t) ds,$$

where $g(s, t)$ is the length metric along the curve:

$$g(s, t) := \left| \frac{\partial C}{\partial s} \right| = [x_s^2 + y_s^2]^{1/2}.$$

Note that this definition allows for the representation of curves with tangent discontinuities, e.g., a square, for the infinite curvature can be countered by infinitesimal speed (Kimia 1990). Therefore,

PRINCIPLE 6. *During the process of deformation the boundary of the shape must remain of finite total absolute curvature.*

Notice that closed curves evolving by Eq. (5) must remain closed (as long as the classical solution² exists). Moreover, from the maximum principle for parabolic equations, one can show (see e.g., (Angenent 1988)), that

THEOREM 1. *Simple closed curves evolving by Eq. (5) remain simple and closed (as long as the classical solution exists).*

Also see (Kimia et al. 1994) for further detail.

4.3 Shape Deformation and Preservation of Similarity

This next principle relates the deformation process to the change in similarity.

PRINCIPLE 7. *The deformation of shape is required to preserve similarity.*

In particular, given two shapes which differ by a rigid Euclidean motion (rotations and translations), they must deform in such a way that the resulting shapes at any fixed later time t differ by the same motion. This is an invariance principle with respect to the Euclidean group. In recent work, this principle has been extended to other groups of interest, e.g., affine group (Alvarez et al. 1992; Sapiro and Tannenbaum 1994; 1993). More generally, the intuition is that, if two shapes are very similar and are then deformed according to the same evolutionary process, their similarity shall be retained. This is a simple stability criterion indicating that our deformations are designed to bring out connections through convergence, not divergence. Another interpretation relates this principle to Principle 1: A shape which obtained from a slight change of another, should quickly converge to it in the process of deformation.

5 Constant Deformation versus Curvature Deformation

We have obtained an important class of deformations of shape that can be viewed as the combination of two basis deformations: constant deformation and curvature deformation. These deformations have drastically different properties. While constant deformation will often lead to singularities in shape, curvature deformation will smooth shapes (Gage and Hamilton 1986; Grayson 1989). While constant deformation operates primarily through the region, the curvature deformation needs boundary information. Sethian (1985c) has shown that the curvature evolution equation under the above motion corresponds to a reaction diffusion equation, where constant deformation is the reaction part and curvature motion is the diffusion part. Thus in the differential equation representation, constant deformation will be referred to as *reaction* or the *hyperbolic* part of the equation, while the curvature deformation will be referred to as *diffusion* or the parabolic part of the equation. We will see below that curvature deformation leads to a quasi-linear equation in certain coordinates. A key distinction between the hyperbolic and parabolic components of the deformation is that constant deformation is local, while curvature deformation is instantaneously global. We now show that our framework formally unifies aspects of two of the common techniques in computer vision: Smoothing and Mathematical Morphology. Also, constant and curvature deformations can be viewed as opposing forces, the interaction of which captures certain physical analogies for shape.

5.1 Curvature Deformation and Nonlinear Diffusion

Curvature deformation of a shape is a (non-linear) analogue to (linear) Gaussian smoothing of the boundary coordinates (Gage and Hamilton 1986; Grayson 1989):

THEOREM 2. *Consider the family of curves $C(s, t) = (x(s, t), y(s, t))$ satisfying*

$$\begin{cases} \frac{\partial C}{\partial t} = -\kappa(s, t)\vec{N} \\ C(s, 0) = C_0(s), \end{cases} \quad (6)$$

where $C_0(s) = (x_0(s), y_0(s))$ is the initial curve, s is some arbitrary parameter along the curve, t is time, κ is curvature, and \vec{N} is the normal. Then the coordinates

satisfy the diffusion equation

$$\begin{cases} \frac{\partial x}{\partial t} = \frac{\partial^2 x}{\partial \bar{s}^2} & x(\bar{s}, 0) = x_0(\bar{s}) \\ \frac{\partial y}{\partial t} = \frac{\partial^2 y}{\partial \bar{s}^2} & y(\bar{s}, 0) = y_0(\bar{s}), \end{cases} \quad (7)$$

where \bar{s} is the arclength parameter along the curve.

PROOF. The proof is very simple, and is based on the classical Frenet formulas. For the convenience of the reader, we give the details. (All the facts from elementary differential geometry we use may be found in (do Carmo 1976). As above, we denote by \vec{T} the tangent, and by \vec{N} the outward normal.

$$\vec{T} = C_{\bar{s}}.$$

Then $\vec{T} \cdot \vec{T} = 1$. Thus differentiating the latter expression with respect to \bar{s} , we get that

$$\vec{T} \cdot C_{\bar{s}\bar{s}} = 0.$$

We thus have that $C_{\bar{s}\bar{s}}$ is orthogonal to \vec{T} , and it is easy to check that it points in the inward normal direction (the curve is parametrized so that the interior is on the left in the direction of increasing \bar{s}). The length of $C_{\bar{s}\bar{s}}$ by definition is the curvature κ . Since \vec{N} denotes the outward normal, we have that

$$C_{\bar{s}\bar{s}} = -\kappa\vec{N},$$

as required.

This equation has been called the *geometric heat equation*. It is a parabolic diffusion equation. In fact, by a remarkable result due to Grayson (1989), any embedded curve shrinks to a circular point under the geometric heat equation. Thus, it smooths shapes. The equation is nonlinear since the arclength \bar{s} is a function of time.

In classical computer vision and image processing, one uses convolution with the Gaussian for smoothing (Rosenfeld and Kak 1982; Witkin 1983). Since the Gaussian is the kernel for the diffusion (heat) equation (Widder 1975), this is equivalent to running the shape through the linear heat equation. Thus the geometric heat equation may be regarded as a nonlinear Gaussian smoothing process. In fact, it gives an anisotropic smoothing in the sense of Perona and Malik (1990), and thus has a number of advantages in image processing, e.g., in edge detection (Alvarez et al. 1992).

For shape, a scale space was proposed by (Mokhtarian and Mackworth 1986) in which shapes

are increasingly smoothed and whose inflection points are tracked for matching. It follows from the above theorem that this scale-space is a special case of the one we shall derive using the full reaction-diffusion system.

Finally, we should note that the geometric heat equation is Euclidean invariant. Recently, it has been shown how to write heat type-flows invariant with respect to a given Lie group acting on the plane (Sapiro and Tannenbaum 1994). From the point of view of vision, an important flow is that which is affine invariant. It turns out that, in this case, the *affine invariant geometric heat equation* is given by

$$C_t = -\kappa^{1/3} \vec{N}.$$

See (Sapiro and Tannenbaum 1994; 1993; Alvarez et al. 1992) for the details.

5.2 Constant Deformation and Mathematical Morphology

Constant motion acts in a complementary fashion to curvature motion. While curvature motion introduces the singularity-removing linear Gaussian Smoothing, constant motion captures nonlinear smoothing as in mathematical morphology (Serra 1982; 1988; Vincent 1991); it introduces singularities. The development of the curve evolution paradigm has enabled mathematical morphology to be viewed as the solution of an evolutionary partial differential equation, in effect giving a continuous implementation of classical discrete morphology (Brockett and Maragos 1992; Sapiro et al. 1992; Aehart et al. 1993; Alvarez et al. 1992). The approach of these works is based on the following elementary observation:

THEOREM 3. *Consider the family of curves $C(s, t) = (x(s, t), y(s, t))$ satisfying*

$$\begin{cases} \frac{\partial C}{\partial t} = \beta_0 \vec{N} \\ C(s, 0) = C_0(s), \end{cases} \quad (8)$$

where $C_0(s) = (x_0(s), y_0(s))$ is the initial curve, s is some arbitrary parameter along the curve, t is time, β_0 is the coefficient determining speed and inward/outward direction, and \vec{N} is the normal. Then the interior of the shape evolves by the mathematical morphology operations of erosion/dilation with a ball as the structuring element. The size of the ball is β_0 .

6 Hyperbolic Conservation Laws

What is most intriguing about our basis of local deformations is that an arbitrary combination of these deformations satisfies a *conservation law with viscosity* as was pointed out and analyzed in (Sethian 1985a–c). (For the classical theory of hyperbolic conservation laws and the associated theory of shocks, see (Lax 1971; 1973).) To formulate this connection, however, the mathematical model of conservation of matter, energy, etc. is first reviewed. This material is classical; see (Smoller 1993) and the references therein. However, for the convenience of the reader, we include it to motivate our use of these concepts in the theory of shape. Following this, the notions of shock and entropy and their role in these models are explained. We will then be able to examine its relevance to shape.

6.1 Conservation in Nature

Hyperbolic conservation laws appear frequently in physical sciences. Examples include conservation of matter, energy, electric charge, and heat, among others. To illustrate, consider the conservation of matter: “the net amount of matter that flows into a volume is exactly the amount of increase of matter within that volume.” In other words, “matter is neither created nor destroyed.” To derive an equation expressing the conservation of a quantity u , such as heat, consider the volume G with boundary ∂G . The total quantity of u in the volume is $\int_G u dv$, where dv is the volume element, and the total quantity passing through the boundary is $\int_{\partial G} u dS$; where dS is the surface element. Then, conservation holds if

$$\frac{\partial}{\partial t} \iiint_G u dv = - \iint_{\partial G} f(u) \cdot \vec{n} dS, \quad (9)$$

where f is the flux. Using the Divergence Theorem, the right hand side is simply $\iiint_G \nabla \cdot f(u) dv$, so that

$$\iiint_G \left[\frac{\partial u}{\partial t} + \nabla \cdot f(u) \right] dv = 0. \quad (10)$$

Since this holds for any volume G ,

$$\frac{\partial u}{\partial t} + \nabla \cdot f(u) = 0, \quad (11)$$

which is the differential equation representing the above integral equation. Such an equation represents a hyperbolic conservation law.

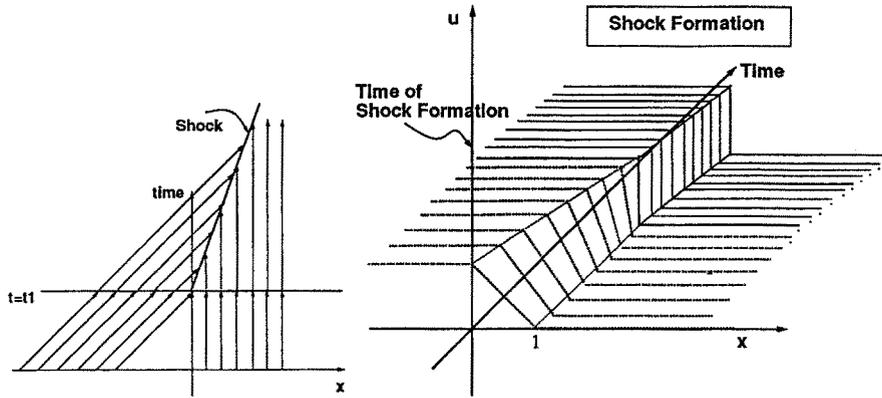


Fig. 8. This figures illustrates how characteristics clash and shocks form. Note that after the shock forms, it travels as a shock. Therefore, singularities are explicitly represented in the context of generalized functions.

6.2 *The Formation of Shocks and the Role of Entropy*

The hyperbolic conservation laws (including those we consider in vision) are nonlinear equations and, as such, will frequently lead to singularities, even when the initial data are smooth. (It is these singularities, of course, that will be important for expressing the connection between curve deformation and categories of shapes.) Mathematically this is important because the usual spaces of differentiable, or even continuous functions, no longer contain the physical solutions beyond the development of a singularity. Thus we need to consider spaces of *generalized functions* (Smoller 1993), but this introduces the additional problem that there are too many solutions in this space satisfying the above partial differential equation with the initial condition. Since a single physically-realized solution obviously exists, we must face the question of how to determine it. The answer lies in the notion of *entropy* (Lax 1971; Oleinik 1957), which in the case of gas dynamics reduces to “entropy of the particles must increase as they cross a *shock* front.” We will formulate an entropy condition for shape in Section 7.

To illustrate the notions of shock and entropy, let us consider the *characteristics* of Eq. (11),

$$\frac{dx}{dt} = \frac{df}{du}. \tag{12}$$

To recall, characteristics are trajectories in the (x, t) , or space and time domain, over which u satisfying Eq. (11) remains constant. Consider then, a well-studied example of a hyperbolic conservation law, the Burgers’ equation (Hopf 1950), where $f(u) = \frac{1}{2}u^2$, leading to

$dx/dt = u$. Given initial condition

$$u_0(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ 1 - x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } 1 \leq x, \end{cases} \tag{13}$$

all points on the negative x -axis will move to the right with speed 1, all points with $x \geq 1$ stay put, while all points with $0 \leq x < 1$ will move to the right with intermediate speeds. It is clear from Fig. 8 that for $t < 1$ the function $u(x, t)$ remains single valued. However, for $t \geq 1$, the characteristics clash, and there exists the potential for the formation of a *shock*. At this point, the two characteristics enforce two different values for u , which is clearly not possible. The dilemma of which of the two values is physically meaningful is solved by enforcing conservation at the shock, leading to the so-called *jump condition* (or the Rankine-Hugoniot condition in gas dynamics.)

$$s(u_r - u_l) = f_r - f_l, \tag{14}$$

where l and r denote left and right, respectively, and s is the speed with which the shock moves. For the case of Burgers’ flux, the shock will move with the average speed of the two incoming characteristics.

A second problem arises when we consider diverging characteristics. Consider the initial condition

$$u_0(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } 0 < x \end{cases} \tag{15}$$

where there will be points whose value can not be determined. The gap may be filled using the jump condition, however we find that many solutions exist in conjunction. The entropy condition rules out the

above example as a possible discontinuous solution by imposing that characteristics always flow into the discontinuity (Lax 1971; Oleinik 1957; Smoller 1993). A discontinuity satisfying both the jump condition and the entropy condition is called a *shock*. The important theorem is provided by Lax, who shows that a generalized solution of (11), which has only shocks as discontinuities, exists and is unique (Lax 1973; 1957).

7 Conservation and Shape: The Role of Shocks and Entropy

The relevance of conservation laws to shape is subtle. First, there is an intuitive connection in which our deformations leave certain aspects of the shape conserved—e.g., contour orientation. The second relevant connection is technical: the conservation laws allow our deformation models to continue beyond the formation of singularities (Sethian 1985a–c). We proceed by first deriving conservation laws for shape. We then show that in the process of deformation corners, or orientation discontinuities, can and often do form. These singularities, as we will show, are among the *shocks* of shape. It is in this context that the role of the conservation law and entropy become clear.

We note that the conservation concepts that we are introducing here are different from the classical notions of invariance (affine, projective, Euclidean) that are commonly used in computer vision (Blake and Zisserman 1987). *Invariance* refers to properties which are preserved under rigid transformation groups while *conservation*, as used in our sense, is dynamic. That is, conservation refers to properties that are preserved under shape deformations. Having clarified this distinction, we further note that all our conservation laws are Euclidean invariant.

7.1 Conservation of Orientation

To begin we study how the orientation of an infinitesimal piece of a curve changes as it evolves according to Eq. (5). It is straightforward to show that deformation of orientation is governed by (Gage and Hamilton 1986; Grayson 1989):

$$\begin{cases} \frac{\partial \theta}{\partial t} = \frac{-1}{g} \frac{\partial [\beta(\kappa)]}{\partial s} \\ \frac{\partial \theta}{\partial s} = \kappa g, \end{cases} \quad (16)$$

where θ is the orientation of the curve in some Cartesian coordinate frame, i.e., the angle the curve's tangent makes with the x axis, i.e., $\angle(T, \vec{x})$, and where $g(s, t)$ denotes length along the curve

$$g(s, t) = \left| \frac{\partial C}{\partial t} \right| = [x_s^2 + y_s^2]^{1/2}. \quad (17)$$

Therefore, when $\beta_1 = 0$, i.e., when there is pure constant motion, we have

$$\theta_t = 0 \quad (18)$$

In words, the orientation of a curve at a point does not change as the curve evolves; orientation is *conserved* when a local piece of the curve is viewed locally and intrinsically.

It is perhaps more interesting to examine how an infinitesimal piece of curve changes its orientation when viewed externally. To motivate, recall our discussion of Subsection 6.1, in which matter is conserved in the sense that the amount of matter flowing into a small piece of pipe is precisely equal to that which flows out plus that which stays in. Similarly then, consider a small piece of the external x -axis coordinate, the interval $(x, x + \Delta x)$. This infinitesimal interval can be regarded as a small section of a “pipe” through which “orientation flows.” We show below that orientation is not annihilated or created by this flow for curves evolving according to Eq. (5) with $\beta_1 = 0$, i.e., orientation is conserved. When $\beta_1 \neq 0$, viscosity is introduced into the system (Sethian 1985a, b; Sethian and Osher 1988; Sethian 1985c; 1989):

THEOREM 4. *Orientation of a curve deformed by constant deformation satisfies*

$$\frac{\partial \theta}{\partial t} + \mathcal{H}_\theta(\theta)_x = 0, \quad (19)$$

where $\mathcal{H}_\theta(\theta) = \cos(\theta)$ is the flux of orientation flow, $-\pi/2 < \theta \leq \pi/2$; clearly a hyperbolic conservation law for orientation θ .

Intuitively, the “pipe” through which orientation “flows” is then each coordinate frame's horizontal axis. The conservation law asserts that in this process orientation does not annihilate or regenerate. Rather, it flows from one section to another, governed by a flux $\mathcal{H}_\theta(\theta) = \cos(\theta)$. Adding curvature motion, on the other hand, adds viscosity to the system:

THEOREM 5. *Orientation of a curve deformed by a combination of constant motion and curvature motion satisfies*

$$\theta_t + \beta_0[\mathcal{H}_\theta(\theta)]_x = \beta_1 \cos^2(\theta)\theta_{xx}, \quad (20)$$

where $\mathcal{H}_\theta(\theta) = -\cos(\theta)$, namely a viscous hyperbolic conservation law for orientation θ .

PROOF. We sketch a proof; see (Sethian 1985c; Osher and Sethian 1988).

Treat x as the independent variable and let $\hat{\theta}(x, t) = \theta(s, t)$, where x is the first coordinate in $\mathcal{C}(s, t) = (x(s, t), y(s, t))$. Then, differentiating both sides with respect to t ,

$$\hat{\theta}_x \cdot x_t + \hat{\theta}_t = \theta_t. \quad (21)$$

Note from Eq. (4) that

$$(x_t, y_t) = \beta(\kappa)\vec{N}. \quad (22)$$

Since $\vec{N} = (\sin(\theta), -\cos(\theta))$, we have $x_t = \beta(\kappa)\sin(\theta)$. Therefore, substituting in Eq. (21), we obtain

$$\hat{\theta}_x \cdot \beta(\kappa)\sin(\hat{\theta}) + \hat{\theta}_t = \theta_t. \quad (23)$$

Now, specializing to curves evolving by Eq. (5), namely where $\beta(\kappa) = \beta_0 - \beta_1\kappa$,

$$\hat{\theta}_x \cdot (\beta_0 - \beta_1\kappa)\sin(\hat{\theta}) + \hat{\theta}_t = \beta_1 \frac{\kappa_s}{g}. \quad (24)$$

where θ_t was computed using Eq. (16). Rearranging,

$$\hat{\theta}_t + (\beta_0 - \beta_1\kappa)[- \cos(\hat{\theta})]_x = \beta_1 \frac{\kappa_s}{g}, \quad (25)$$

or,

$$\hat{\theta}_t + \beta_0[- \cos(\hat{\theta})]_x = \beta_1\kappa \cdot [- \cos(\hat{\theta})]_x + \beta_1 \frac{\kappa_s}{g}, \quad (26)$$

Now, we need to replace all the above terms with terms involving $\hat{\theta}$. First, note that $\frac{x_s}{g} = \cos(\hat{\theta})$ and

$$\begin{aligned} \hat{\theta}_x &= \frac{\hat{\theta}_s}{x_s} \\ &= \frac{\kappa g}{x_s} \\ &= \frac{\kappa g}{g \cos(\hat{\theta})} \\ &= \frac{\kappa}{\cos(\hat{\theta})}, \end{aligned} \quad (27)$$

so that $\kappa = \cos(\hat{\theta})\hat{\theta}_x = [\sin(\hat{\theta})]_x$. Similarly,

$$\begin{aligned} \kappa_s &= [[\sin(\hat{\theta})]_x]_s \\ &= [\sin(\hat{\theta})]_{xx} \cdot x_s. \end{aligned} \quad (28)$$

Hence,

$$\begin{aligned} \frac{\kappa_s}{g} &= [\sin(\hat{\theta})]_{xx} \cdot \frac{x_s}{g} \\ &= [\sin(\hat{\theta})]_{xx} \cdot \cos(\hat{\theta}). \end{aligned} \quad (29)$$

As such Eq. (26) simplifies as

$$\begin{aligned} \hat{\theta}_t + \beta_0[- \cos(\hat{\theta})]_x &= \beta_1\kappa \cdot [- \cos(\hat{\theta})]_x + \beta_1 \frac{\kappa_s}{g} \\ &= \beta_1 \cos(\hat{\theta})\hat{\theta}_x \cdot \sin(\hat{\theta})\hat{\theta}_x + \beta_1 [\sin(\hat{\theta})]_{xx} \cdot \cos(\hat{\theta}) \\ &= \beta_1 \sin(\hat{\theta}) \cos(\hat{\theta})\hat{\theta}_x^2 + \beta_1 [\cos(\hat{\theta})\hat{\theta}_x]_x \cdot \cos(\hat{\theta}) \\ &= \beta_1 \sin(\hat{\theta}) \cos(\hat{\theta})\hat{\theta}_x^2 \\ &\quad + \beta_1 [- \sin(\hat{\theta})\hat{\theta}_x^2 + \cos(\hat{\theta})\hat{\theta}_{xx}] \cdot \cos(\hat{\theta}) \\ &= \beta_1 \sin(\hat{\theta}) \cos(\hat{\theta})\hat{\theta}_x^2 + -\beta_1 \sin(\hat{\theta})\hat{\theta}_x^2 \cos(\hat{\theta}) \\ &\quad + \beta_1 \cos^2(\hat{\theta})\hat{\theta}_{xx} \\ &= \beta_1 \cos^2(\hat{\theta})\hat{\theta}_{xx}. \end{aligned} \quad (30)$$

This completes the proof of both Theorems.

Note that, when $\beta_1 = 0$, this gives the conservation law. When $\beta_1 \neq 0$, viscosity in the form of second order terms is added to the system.

Viscosity, or ‘‘diffusion,’’ changes the character of the deformation. Whereas before, deformations conserved the local orientation identity of each piece, with diffusion the local orientation of each piece is ‘‘blended’’ with that of neighbors. Informally, one view of the constant deformation and curvature deformation trade-off is that of area versus length or region versus boundary.

Other quantities are conserved as well. For example, one can also show that a conservation law holds for the curvature-metric, or the product of curvature and metric as shown in (Kimia 1990). Specifically if κ denotes curvature and g the Euclidean metric, then $(\kappa g)_t = 0$.

7.2 The Formation of Shocks

In addition to their intuitive appeal, the conservation laws have a very significant role: they show how the original process of deformation, as defined by Eq. (4), a local differential model, can be extended to handle singularities using an ‘‘integral’’ form of the law and the notion of ‘‘weak’’ solutions.

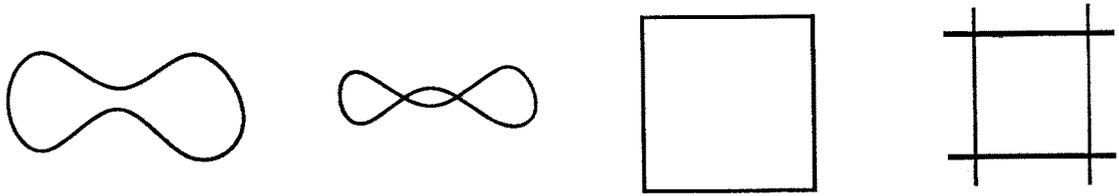


Fig. 9. Curve evolution is not shape evolution. Note that in the process of deformation of the curve, local portions of the boundary may cross over each other, like the corner of the square. Similarly, remote portions of the boundary may cross over each other. Since shapes are curves that are filled with material as required by principles 4 and 5, the local curve deformation does not always lead to shape deformation. To resolve this dilemma, the interior must be represented explicitly.

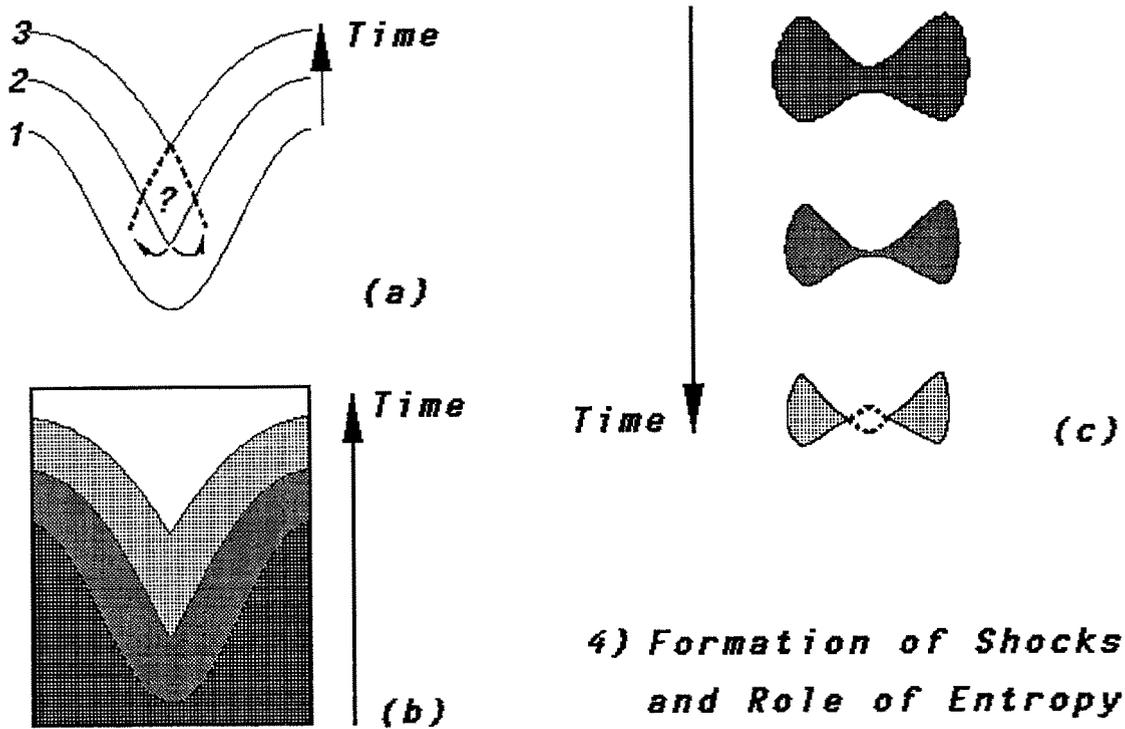


Fig. 10. Nonlinear processes can transform initially smooth functions to functions with singularities. (a) shows a curve with a negative curvature extremum which, when evolved by constant motion along the normal, leads to a singularity. This evolution can be based entirely on boundary information until the singularity arises. However, at this point the entropy condition is required to further control evolution, so that the curve does not cross over itself and the swallowtail configuration can be properly handled (b). The entropy condition is region-based, and controls how interior information interacts with the boundary. It plays another key role in controlling topological evolution, by globally managing the splitting of a single boundary into two closed boundaries (c). In both cases the entropy condition dictates that the solution does not include the “dashed” portions of the contour—these annihilate into the shock.

To motivate the problem consider Fig. 9 which depicts situations where curves representing shapes of objects now evolve to curves that cannot possibly represent the occlusion boundaries of real objects. These situations violate Principles (4) and (5), even though they follow the local differential model. To illustrate that this situation is not exclusive to shapes with initial singularities, consider Fig. 10, taken from Sethian (1985c), where initially smooth shapes develop

singularities. Since the differential model can be utilized at all points except singularities, the question arises of how to continue the deformation once shocks form. This situation is analogous to the example of Burgers equation in Section 6, where the smooth initial condition (13) deforms according to the differential Eq. (11), up to the point where a singularity forms in the solution. From this point onward, however, the differential equation in its classical form is no longer valid

at such points; yet the principle of conservation as represented by its integral form (9) remains in force. The integral form is the manifestation of the conservation principle for the domain of *weak solutions*. While in the case of the Burgers equation the intergal form of the equation could be derived based on the original physical principles of conservation, what is the “integral form” of Eq. (5)?

To summarize, when the boundary is smooth, the local differential model is in force for the process of deformation. However, beyond the time of formation of singularities, the law is no longer valid and an alternative must be sought.

As a first attempt toward a solution to this problem, one might simply decompose the problem into the independent local deformation of pieces of contour between some arbitrary points, say high curvature points or singularities. While this may be done arbitrarily for smooth shapes, for singular shapes this unfortunately generates non-valid shapes as Fig. 9 illustrates. How then can the process of deformation continue beyond this point?

We believe that a natural solution lies in Principle 7 and the introduction of a notion of *entropy* for shape. Such an entropy condition has been introduced by Sethian (1985a); Sethian and Osher (1988) for propagating fronts. (For general discussions of entropy and the connection to hyperbolic conservation laws, see Lax (1957; 1971; 1973) and Smoller (1993).) Since the conservation law for orientation, which we derived from the local differential model, is valid even beyond the point when singularities form, we can postulate the principle of conservation as the fundamental principle underlying the deformation of shape. This is in agreement with the situation of the Burgers example: as Fig. 11 illustrates, the shape evolution can continue beyond the formation of “corners.” Note that the “shock” remains sharp past its formation and that the extra dashed lines resulting from independent evolution of each curve segment on either side of the discontinuity are not present. Informally, the entropy condition, which abstractly forces characteristics to lean into a shock, translates into the condition of removing the dashed portions of the curves for shape. These dashed segments are portions of the curve which cross over each other. Similar effects occurred in the Burgers example, when particles traveling at unit velocity had to “reach an agreement” with stationary ones. This agreement was the formation of a shock moving with a compromise velocity depending on the form of the flux (in the case of Burgers’ flux, this speed is exactly

the average of the two speeds). Clearly, certain particles annihilate into a shock. In our case, the dashed lines would be present if each portion could evolve independently. However, given the collision a similar “agreement” must be reached. Now the crossed-over portions—the dashed lines—annihilate into a shock. Since the dashed segments are points over which the boundary has crossed already, the following definition of *entropy* for shape is appropriate:

DEFINITION 1 (Entropy Condition). In the process of inward deformation, once a point is dislodged from a shape, it remains disjoint from it forever. Similarly, in the process of outward deformation, once a point becomes part of a shape, it remains part of it forever.

This entropy condition is a reformulation of the one given by Sethian (1985a, c) for the propagation of flame fronts. Namely, the propagating flame front satisfies the entropy condition if once a particle burns, it remains burnt. It is also reminiscent of the “grass fire” algorithms introduced to vision by Blum.

The entropy condition is one of three related ways of picking out a physically meaningful solution to a hyperbolic conservation law, and each of them provides a different perspective on understanding the solution; they are reviewed in the Appendix. To summarize briefly, the entropy condition above relates to shock waves in dynamics, and to the burning interface around an island of fuel. More abstractly, it is well-known mathematically that the type of hyperbolic conservation laws we are considering have a Hamilton-Jacobi formulation (see (Smoller 1983) and our discussion below). In fact, it is this formulation which is the key for the numerical implementation of curve evolution equations which was introduced in the fundamental work of Osher-Sethian (1988), and which will be explicated below. We now note the second way of viewing solutions, which was introduced by Crandall and Lions (1983): Informally, the viscosity solution of a conservation law is the one obtained in the limit as viscosity is reduced to zero. The beautiful connection established by Barles (1985) is that the entropy solution of the conservation law given above is precisely the viscosity solution of the corresponding Hamilton-Jacobi equation. Thus, from a practical point of view, it is precisely the viscosity solutions of the corresponding Hamilton-Jacobi equations which are implemented.

We finally note that there is a third equivalent approach to the “weak” solutions of conservation laws

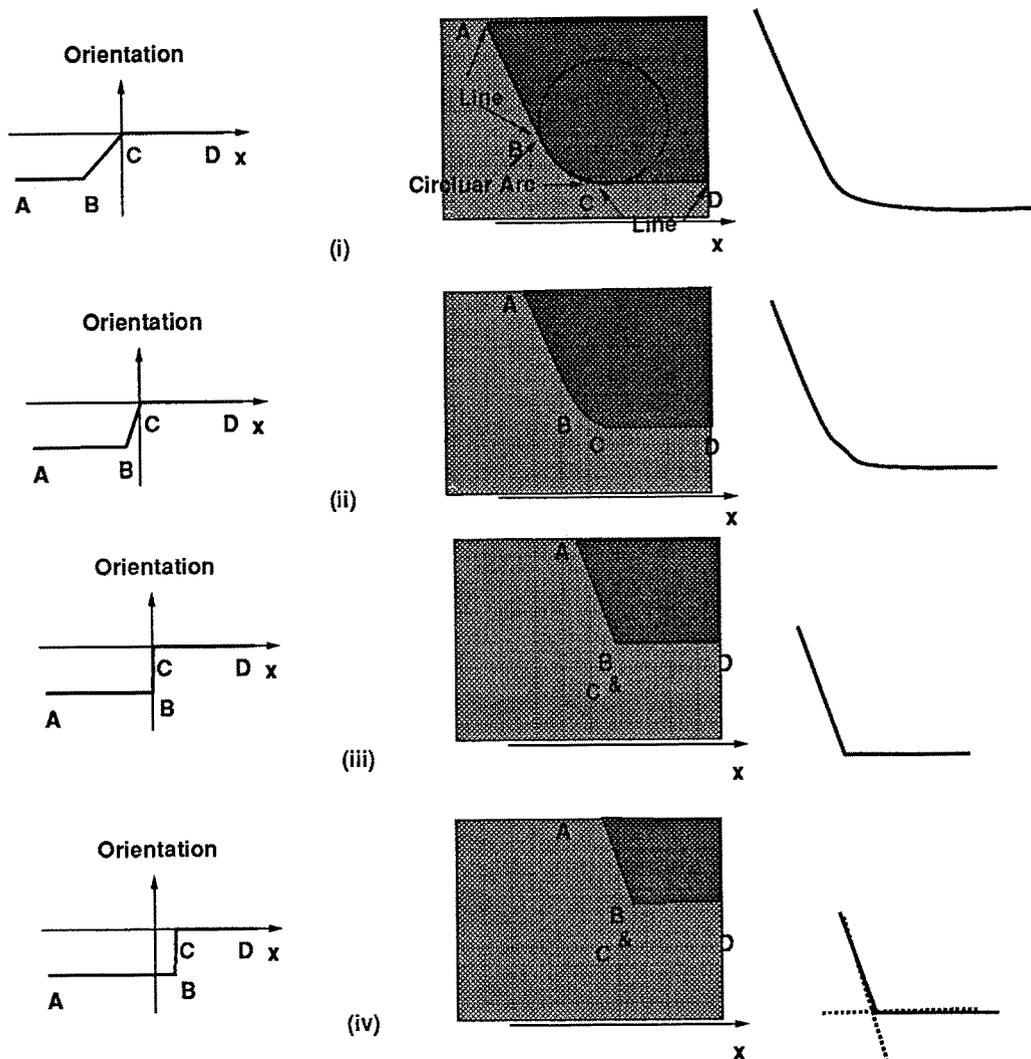


Fig. 11. The shock as depicted for the Burgers equation in Fig. 8 is now applied to shape. Recall that orientation satisfies a conservation law which produces a similar shock (left column). On the other hand, the deformation in the shape domain is constant motion along the normal. Note that with time, the circular arc will dissolve into one point, leaving the orientation of the boundary discontinuous. The continuation of the deformation keeps the shock intact.

and the Hamilton-Jacobi equation; namely, the theory of nonsmooth analysis from optimal control (Clarke 1989). Here certain *generalized gradients* can be defined to allow the “differentiation” of certain singular functions. The value function in optimal control theory is also the solution of a Hamilton-Jacobi equation, which in this case inherits its interpretation from dynamic programming. Historically, viscosity solution theory and nonsmooth analysis developed from the fact that the value function usually has singularities as well. Barles’ connection, noted above, again connects them. (See also (Crandall et al. 1992) for an extensive list of

references on this methodology). Each of these ideas is illustrated in the Appendix.

7.3 Physical Analogies

The study of shape is not only interesting when viewed as a visual entity, but also when viewed as an interface separating one material from another. Under such a view, the theory of shape description and evolution is relevant to many physical problems, for example, crystal growth (Langer 1980), flame propagation (Sivashinsky 1977; Sethian 1985c), the oil-water

boundary problem, and the deterioration of the shapes of stones (Firey 1974). Typically, in a class of models for these physical phenomena, a reactive term is in conflict with a diffusive term. To illustrate, consider the case of crystal growth in which the growth pattern of a solidification front is determined by the interaction of two forces: the driving force of the instability due to heat diffusion and the restabilizing force due to surface tension (Smith 1981). Common to these models is a reactive force which corresponds to our constant motion and a diffusive force depending on surface curvature which corresponds to our curvature motion. For an application of these reaction-diffusion models see (Smoller 1993).

There is yet another connection between the entropy condition and the Huygens' principle and eikonal equations that is relevant. Namely: the entropy solutions are exactly those constructed by the Huygens' principle, say as a solution in geometric optics as illustrated by Barles (1985); for a specific example, see Appendix.

7.4 *Embedding Curve Evolution in a Higher-Dimensional Space*

While the conservation law formulation resolves the first of the two problems depicted in Fig. 9, i.e. the local collision of the boundary and the consequent formation of singularities, this is not the only problem that invalidates the curve evolution as shape evolution. As the peanut shape of Fig. 9 evolves in time, remote portions of the boundary collide and pass over each other. This collision does not manifest itself in either the local curve deformation model of Eq. (4) or the more general conservation model of Eq. (20). What additional constraints are needed to restrain this behavior so that the curve evolution can claim to model shape evolution? The missing ingredient is a notion of "interior." To recall our discussion of Section 2.1, a comprehensive understanding of shape involves *both* the notion of its boundary *and* its interior. Up to this point, however, our formulation has been focused mostly on the evolution of the boundary of the shape. The current problem as illustrated above not only confirms the distinction but also points to the solution: an explicit role for the region bounded by the boundary. This region represents the material that "glues" the various portions of the boundary together, in distinction to a thin wire. Our approach, then, is to allow for this extra "dimension" of information by considering an evolution in a higher-dimensional space, e.g., the evolution of a two-dimensional surface in a

three-dimensional space constrained to embed the original problem. We should add that this approach is also essential to the computer implementations of the curve evolution equations. The algorithms based on this embedding concept have been derived in the elegant work of Osher-Sethian (1988); Sethian and Osher (1988).

To motivate these ideas, let us consider the field of fluid dynamics and the two formulations capturing the regions bounded by the flow of fluids, a problem not unlike ours. In the *Lagrangian formulation*, equations of motion are based on the flow of particles, whereas in the *Eulerian formulation* the physical quantities are constrained as a function of their position. One may view the first framework as local and boundary-based, and the latter as global and region-based. To accommodate the regional and global attributes, points distant along the boundary but close through the region may have to be connected. The solution lies in explicitly representing the regional information as a surface or a "tent" built around the boundary. The boundary, then, is the zero level set of the surface, which represents the interior. Consider, then, the shape as the zero level set of some function $z = \phi(x, y, t)$, where ϕ represents some physical quantity e.g., density, intensity, depth, etc., which indicates where the region of interest, i.e., shape, is located. The simplest scheme is to consider all points for which $\phi(x, y, t) \geq 0$ as belonging to the region. In shape representation, Koenderink has utilized the characteristic function of some region as an indicator of the shape (Koenderink 1984; Koenderink and van Doorn 1986). Similar representations have been proposed for the propagation of flame fronts (Sethian 1985b; Barles 1985).

What principles govern the evolution of the surface $z = \phi(x, y, t)$, and what equations model it? First, note that the surface is initially only constrained by its zero level set. As such, one degree of freedom exists in choosing the initial surface. Second, observe that the evolution of this initial surface is underdetermined as the only constraint is on the evolution of its zero level set. Namely, the zero level set of the evolving surface ought to be precisely the curve evolution defined by Eqs. (4) and more specifically (5), or their integral form of Eq. (20). (Practically, in the computer implementations, the surface is constructed over the evolving curve via the distance function.)

We now reproduce, in the rest of this section, the level set formulation introduced in Osher-Sethian (1988). To mathematically capture the constraint that "the surface evolution be consistent with the original

curve evolution," consider the surface defined by

$$z = \phi(x, y, t). \quad (31)$$

This constraint requires that the zero level set of the surface, namely,

$$\phi(x, y, t) = 0 \quad (32)$$

be exactly the trace of the curve $\mathcal{C}(s, t) = (x(s, t), y(s, t))$ defined by Eq. (4). Note that by differentiating Eq. (31) we have

$$\phi_x x_s + \phi_y y_s = 0, \quad (33)$$

or equivalently,

$$(\phi_x, \phi_y) \cdot (x_s, y_s) = 0. \quad (34)$$

Since the vector (x_s, y_s) is along the tangent \vec{T} , the vector (ϕ_x, ϕ_y) is along the normal \vec{N} . Therefore, up to a change of sign,

$$\vec{N} = \frac{1}{(\phi_x^2 + \phi_y^2)^{1/2}} \cdot (\phi_x, \phi_y). \quad (35)$$

Next, differentiating Eq. (31) with respect to t , we have

$$\phi_x x_t + \phi_y y_t + \phi_t = 0, \quad (36)$$

or equivalently,

$$(\phi_x, \phi_y) \cdot (x_t, y_t) + \phi_t = 0. \quad (37)$$

Since $\mathcal{C}_t = (x_t, y_t)$, for deformations defined by Eq. (4), we have

$$(\phi_x, \phi_y) \cdot \beta(\kappa(x, y))\vec{N} + \phi_t = 0, \quad (38)$$

or equivalently,

$$(\phi_x, \phi_y) \cdot \beta(\kappa(x, y)) \frac{1}{(\phi_x^2 + \phi_y^2)^{1/2}} \cdot (\phi_x, \phi_y) + \phi_t = 0. \quad (39)$$

Thus, we see that

$$\phi_t + \beta(\kappa(x, y))(\phi_x^2 + \phi_y^2)^{1/2} = 0. \quad (40)$$

Finally, since $|\nabla\phi| = (\phi_x^2 + \phi_y^2)^{1/2}$,

$$\phi_t + \beta(\kappa)|\nabla\phi| = 0. \quad (41)$$

The expression of κ for points on the zero level set of surface, ϕ , is easily seen to be

$$\kappa(x, y) = -\frac{(\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2)}{(\phi_x^2 + \phi_y^2)^{3/2}}, \quad (42)$$

for all points (x, y) which satisfy $\phi(x, y) = 0$.

Specializing to the case of Eq. (5) where $\beta(\kappa) = \beta_0 - \beta_1\kappa$, and using Eq. (42), we have

$$\phi_t + \left\{ \beta_0 - \beta_1 \left[-\frac{(\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2)}{(\phi_x^2 + \phi_y^2)^{3/2}} \right] \right\} \cdot (\phi_x^2 + \phi_y^2)^{1/2} = 0, \quad (43)$$

or,

$$\begin{aligned} \phi_t + \beta_0(\phi_x^2 + \phi_y^2)^{1/2} \\ = \beta_1 \left[-\frac{(\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2)}{(\phi_x^2 + \phi_y^2)^{3/2}} \right]. \end{aligned} \quad (44)$$

Finally,

$$\phi_t + \beta_0|\nabla\phi| = \beta_1 \left[-\frac{(\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2)}{(\phi_x^2 + \phi_y^2)^{3/2}} \right], \quad (45)$$

for all points (x, y) which satisfy $\phi(x, y, t) = 0$, i.e., are on the zero level set. This is precisely what our earlier constraint implies for the evolution of the surface, i.e., it determines the evolution of the points on the zero level set.

The above constraint restricts the evolution of points on the zero level set only. How should other points evolve? For simplicity, we opt to move each level set according to the same process. In other words, all points (x, y) should evolve according to Eq. (45). This constraint, together with an initial definition for the surface, defines the surface evolution such that it is consistent with the curve evolution problem.

Recapping, consider a surface $z = \phi(x, y, t)$ evolving according to Eq. (45), with the initial condition $z = \phi(x, y, 0) = \phi_0(x, y)$. If the explicit representation of $\phi_0(x, y) = 0$ is denoted by $\mathcal{C}_0(s) = (x_0(s), y_0(s))$, then the (x, y, t) satisfying $\phi(x, y, t) = 0$ also satisfy evolution Eq. (5).

7.5 Hamilton-Jacobi Equations and Conservation Laws

In the last section, we showed following Osher-Sethian (1988); Sethian (1985b, c) that in order to embed the curve evolution problem in a higher dimension, i.e., as the evolution of the zero level set of the surface, the surface is required to evolve according to Eq. (45) at all the points. This equation may be rewritten as

$$\begin{aligned} \phi_t(x, y, t) + \beta_0\mathcal{H}_0(\phi_x, \phi_y, t) \\ = \beta_1\mathcal{H}_1(\phi_{xx}, \phi_{yy}, \phi_x, \phi_y, t), \end{aligned} \quad (46)$$

a first-order Hamilton-Jacobi equation with a parabolic right hand side, where in our case,

$$\mathcal{H}_0(\phi_x, \phi_y, t) = (\phi_x^2 + \phi_y^2)^{1/2}, \quad (47)$$

and,

$$\begin{aligned} \mathcal{H}_1(\phi_{xx}, \phi_{yy}, \phi_x, \phi_y, t) \\ = - \frac{(\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2)}{(\phi_x^2 + \phi_y^2)^{3/2}}. \end{aligned} \quad (48)$$

A Hamilton-Jacobi equation is a first-order partial differential equation of the form

$$\mathcal{J}_t + \mathcal{H}(t, x, \mathcal{J}_x) = 0, \quad (49)$$

where \mathcal{H} in analytical dynamics is known as the Hamiltonian, or the total energy of the system (Garabedian 1964).

There is an interesting connection between first-order Hamilton-Jacobi equations and hyperbolic conservation laws. Recall the conservation law of Eq. (11) in the one-dimensional case:

$$u_t + f(u)_x = 0. \quad (50)$$

Informally, as in conservative systems of the analytical mechanics analogy, external forces can be derived from a potential function giving rise to a Hamiltonian, and the connection between conservation laws and the first-order Hamilton-Jacobi equation. Formally, using the transformation

$$u = \mathcal{J}_x, \quad (51)$$

C^2 solutions of Eqs. (49) and (50) are the same (Garabedian 1964; Hopf 1950; Lax 1957).

REMARK. In fact, using a formula of Lax-Oleinik (1985; 1981), we get the viscosity solution for the corresponding Hamilton-Jacobi equation of a conservation law of the form (50). Namely, the Hamilton-Jacobi equation is

$$U_t + f(U_x) = 0,$$

where $u = U_x$, and the viscosity solution is given by

$$U(t, x) = \min_y \left\{ U_0(y) + tf^* \left(\frac{x-y}{t} \right) \right\},$$

for initial function U_0 , where

$$f^*(w) := \max\{uw - f(u)\},$$

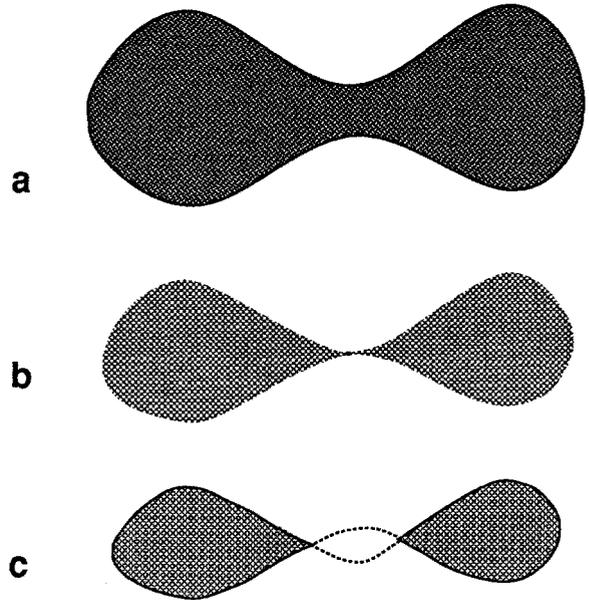


Fig. 12. This figure depicts the case when two points of a shape (A) that are distant along its boundary come together during an arbitrary deformation (B). How should the deformation proceed beyond this point? A pointwise deformation along the normal would produce the dashed-lines, which clearly violate Principle (4) since they do not correspond to an actual object.

is the conjugate function (we assume that f is convex here). The entropy solution may then be immediately derived from this. (This is the basis of Barles proof (1985) of the equivalence of entropy and viscosity; see Appendix.)

Now, we are in a position to consider the significance of the Hamilton-Jacobi formulation as applied to the case of Fig. 12. To restate, the problem is that in the local boundary-based model of evolution, remote portions of the shape as measured along the boundary may be close and in fact pass through each other as shown in Fig. 12. By introducing a surface to represent the interior, this information becomes explicit. Invoking the entropy condition as presented in Definition 1, when portions of the shape's boundary collide and pass over each other, the shape is segmented into two disjoint subshapes, each evolving separately, according to Principle 4. In this way, topologically connected shapes like that in Fig. 12.A, and those with two disjoint components, like that in Fig. 10.C, become neighbors in a deformation process and therefore are similar.

In summary, solutions of Eq. (46) satisfying the entropy condition 1 are the proper "physical" solutions in the viscosity sense.

Although this is not the paper to discuss numerical implementations, we do observe that Sethian (1990) proved that simple, Lagrangian, difference approximations require an impractically small time step in order to achieve stability for Hamilton-Jacobi equations. The algorithm proposed by Osher-Sethian (1988); Sethian and Osher (1988), which is based on Hamilton-Jacobi theory (and therefore optimal control), has provided reliable numerical solutions to these problems.

8 The Reaction-Diffusion Space

In the previous section we showed how shocks formed in the course of evolution of shapes. In this section, we classify these shocks and show how they lead to our proposal for the elements of shape: parts, protrusions, and bends. These shocks occur for various combinations of constant deformation and curvature deformation, or reaction and diffusion. The space generated by these combinations and by time is referred to as the *reaction-diffusion space*. It is in the context of this space that shocks will be related to shape.

Before providing a formal definition of the reaction-diffusion space, a kind of physical analogy might help to focus intuitions about what is being developed. Imagine a particular shape, say a peanut, made out of material such as wax. Among the uniform (with respect to the boundary) operations that are possible on the peanut, two are especially important. First, material can be added to, or withdrawn from, the peanut, to simulate growth. Second, the temperature can be varied to simulate melting, or deformation. These two operations correspond, of course, to the reaction and diffusion terms in our boundary evolution equations. Now, to build the reaction-diffusion space, imagine a collection of peanut-shapes in a row. We shall uniformly add (or withdraw) material to (from) to each of the peanuts, but of increasing temperatures along the row. Observe that the conservation law dictates that these changes (addition/withdrawal of material and melting) take place as uniformly as possible; i.e., conserving the shape as much as possible. On the colder end of the scale, “folds” or shocks will form, while, on the warmer, more viscous end of the scale, the wax will tend to flow together and the shape will be smoother. Pure melting, of course, will result in a round puddle of wax. Thus, the space of shapes generated by the peanut-shape is spanned by the ratio of reaction to diffusion (the “temperature” scale), and time (the “temporal” scale indicating the duration of evolution,

or the quantity of material added or withdrawn). This is our reaction-diffusion space for the peanut. Reaction-diffusion spaces for more complex shapes require a temporal ordering of events, which describes when a seed for a part is born (for material being added) or annihilated (for material being withdrawn). This temporal ordering, of course, is the significance hierarchy.

The different aspects of shape are well represented in the reaction-diffusion space. For example, pure reaction is a region process, in that “area” determines what is significant for survival over time. Pure diffusion, on the other hand, is a boundary process, and “length” is the critical feature. In addition, pure reaction is a process that is local with regard to the time of deformation, while pure diffusion is global.

8.1 The Reaction-Diffusion Space

In order to define the reaction-diffusion space, we need to quote the following result from (Kimia et al. 1992):

THEOREM 6. *Consider a C^2 embedded curve C_0 evolving through a function of curvature as in (3). (We regard our curves as mappings from the unit circle S^1 to the plane \mathbf{R}^2 .) Let $C_t(s) = C(s, t)$ be a classical solution for $t \in [0, t']$. Then,*

$$\lim_{t \rightarrow t'} C_t = C^*, \quad (52)$$

in the Hausdorff metric where curve C^ has finite total curvature. Moreover, C^* is locally the graph of a Hölder continuous function with exponent $1/2$.*

With this result, we can define now the *reaction-diffusion space*:

DEFINITION 2. Let S denote the space of all images of embedded C^2 curves regarded as mappings from $S^1 \rightarrow \mathbf{R}^2$. We will call these shapes. The representation of a shape, S , in all possible time and all possible ratios β_0/β_1 ($\beta_1 < 0$) is called the **REACTION-DIFFUSION SPACE** for that shape.

In other words, the reaction-diffusion space for a shape S is the set of all shapes S' generated by

$$\mathcal{RD}_{(\alpha,t)} S \mapsto S', \quad (\alpha \in (-\infty, \infty), t \in [0, \infty)). \quad (53)$$

where \mathcal{RD} is the deformation with α as the ratio of constant motion to curvature motion magnitudes and t

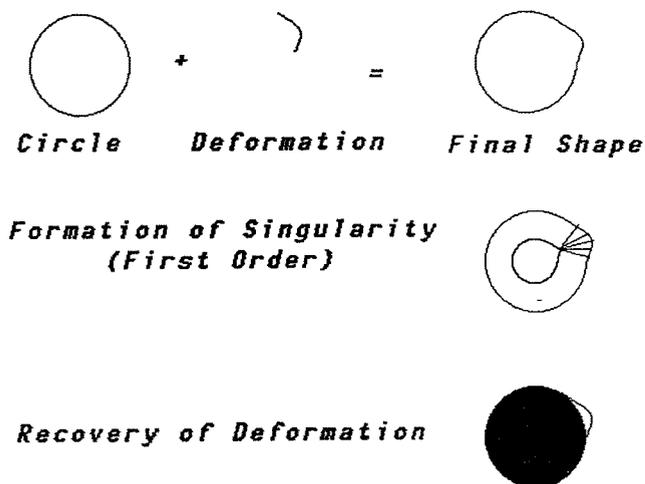


Fig. 13. The shape on the right is perceived as a circle with a deformation. While a number of other interpretations are possible, this interpretation seems to be favored naturally. How can this deformation be recovered?

is time. From the theorem stated above, S' is the space of all curves which are images of mappings $S^1 \rightarrow \mathbf{R}^2$ which are locally the graphs of Hölder continuous functions³ of exponent $1/2$. This space therefore spans all combinations of reaction and diffusion, and time. Note that $\beta_1 > 0$ represents the heat equation running backward in time and $\beta_1 = 0$ represents the case of no diffusion; a nonrealistic situation.

Notice that the reaction-diffusion space leads to a topological notion of *path-connectedness* in the space of shapes. Namely, two shapes lie in the same connected component if one can be deformed to the other in the reaction-diffusion space.

In the reaction-diffusion space, the formation of shocks is key to its representation. The shock formation process is governed differently as the ratio of reaction to diffusion is altered. Also, the time of formation of a shock is related to its significance. This two-dimensional space affords a much richer analysis than the representation of a single curve in isolation, Figs. 18, 19, 20 and 21. It is in the context of this space that we analyze shocks, as follows (Kimia 1990).

8.2 First-Order Shocks

Consider the shape in Fig. 13 which is formed by pushing a portion of a circle outwards. It would not be uncommon to describe this shape as a “circle with a protrusion.” While other descriptions, e.g. a half circle glued on a half deformed rectangle, are possible,

nevertheless our perception is clear (unless we have been primed to another category previously). Now, let us consider the effect of a constant motion type of deformation on this shape. Recall that in Fig. (13), constant motion produces a single isolated orientation discontinuity from a negative curvature minimum. Adhering closely to the terminology of classical conservation laws, then, let us preserve the term shock and define:

DEFINITION 3. A **FIRST-ORDER SHOCK** is a discontinuity in orientation of the boundary of a shape.

THEOREM 7. *In the process of evolution by constant motion, each local curvature extremum leads to a first order shock, provided that only this local portion of the curve evolves.*

PROOF. For constant motion of evolution Eq. (5), $\beta_0(\kappa) = 1$, so that the evolution of the metric, or length along the curve g , and curvature κ is governed by (Grayson 1989; Osher and Sethian 1988)

$$\begin{aligned} \frac{\partial g}{\partial t} &= \kappa g, \\ \frac{\partial \kappa}{\partial t} &= -\kappa^2. \end{aligned} \tag{54}$$

These equations can be solved explicitly by noting that

$$\kappa(s, t) = \frac{\kappa(s, 0)}{1 + \kappa(s, 0)t}, \tag{55}$$

and then substitute to solve for the metric equation

$$\begin{aligned} \frac{\partial g}{\partial t} &= \kappa(s, t)g(s, t) \\ \frac{g_t}{g} &= \frac{\kappa(s, 0)}{1 + \kappa(s, 0)t} \\ \frac{\partial \ln(g)}{\partial t} &= \frac{\kappa(s, 0)}{1 + \kappa(s, 0)t} \\ \frac{\partial \ln(g)}{\partial t} &= \frac{\partial}{\partial t} [\ln(1 + \kappa(s, 0)t)] \end{aligned}$$

$$\begin{aligned} \ln(g(s, t) - \ln(g(s, 0)) &= \ln(1 + \kappa(s, 0)t) - \ln(1) \\ \ln(g(s, t) &= \ln(g(s, 0)) \\ &+ \ln(1 + \kappa(s, 0)t) \\ g(s, t) &= g(s, 0)(1 + \kappa(s, 0)t) \end{aligned} \tag{56}$$

Hence, the metric changes linearly in time with a curvature dependent coefficient. As such, for points of negative curvature, as $t_0 \rightarrow \frac{-1}{\kappa(s, 0)}$, the metric will tend to zero, while curvature will tend to negative infinity; this is exactly when the first order shock forms.

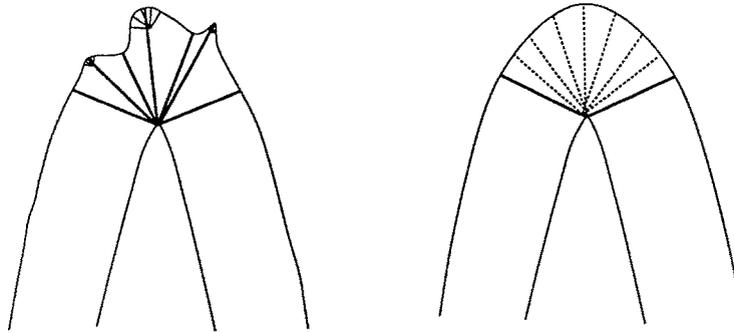


Fig. 14. RIGHT: Illustration of how curvature concentrates to form first-order shocks. Once a shock is formed, the information contained in the section of the curve which mapped to the shock in the forward direction (between dashed lines) is now lost forever. Note that the removal of information is local, and nonlinear. An attempt to reconstruct the original signal by running the reaction term backwards in time would yield a circular arc for the section that mapped to the shock. LEFT: This is an illustration of the significance hierarchy imposed by shocks: less significant shocks corresponding to protrusions get absorbed into more significant shocks describing a more global portion of the shape of the object. Reconstruction (by running reaction backwards in time) again gives the same circular arc as in (RIGHT), even though the section mapping to the shock is now more complicated.

Clearly, the smaller the negative curvature at a point, the smaller this time, t_0 . Then, negative curvature minima will first reach first-order shocks, provided remote portions of the boundary do not interfere.

In summary then, first order shocks are associated with protrusions (indentations) in absence of other shocks. They arise because curvature accumulates most rapidly at extrema. Note that several smaller protrusions may merge to form one at a larger “scale” as in Fig. 14.

8.3 Second-Order Shocks

A second kind of shock forms, not due to curvature build-up as in the first type, but due to a collision of boundaries. Consider the shape in Fig. 12. As the shape (A) evolves in time due to a *constant deformation*, portions of the boundary collide and give rise to two cusps (B). These cusps are discontinuities, not in tangent, but in curvature. We call these *second-order shocks*. Note the change of connectivity at the instant it forms. Beyond this instant, portions of the boundaries cross each other (the dashed lines). The role of entropy in this case is to remove portions of the boundary that have reached a previously visited point (C). Formally,

DEFINITION 4. When in the process of deformation two distinct non-neighboring boundary points join and not all the other neighboring boundary points have collapsed together, a **SECOND-ORDER SHOCK** is formed.

The second-order shocks define *parts* of a shape. This notion of *parts* is different than that proposed in (Hoffman and Richards 1985), where parts were defined by negative minima of curvature. Our parts are more intuitive e.g. consider the examples in Figs. 15 and 16. These ideas have been extended in (Siddiqi and Kimia 1994; Siddiqi et al. 1992) to include neck-based and limb-based parts.

We should note that shocks of the first and second orders are generic. “Generic” means in the sense that such singularities cannot be removed by a small perturbation. “Small perturbation” is in the sense of the theory of singularities of smooth mappings. For details, see Appendix C.

8.4 Third-Order Shocks

A third type of shock is generated when distinct boundary points are brought together as in second-order shocks, but unlike the second-order shock, the neighboring boundary points on each side have also joined with other distant boundary points. Formally,

DEFINITION 5. When in the process of deformation two distinct non-neighboring boundary points join in a manner such that neighboring boundaries of each point also collapse together, a **THIRD-ORDER SHOCK** is formed.

As defined above, third-order shocks cannot possibly change the topological connectivity of the shape. Rather, they indicate a symmetric axis, as in the case of

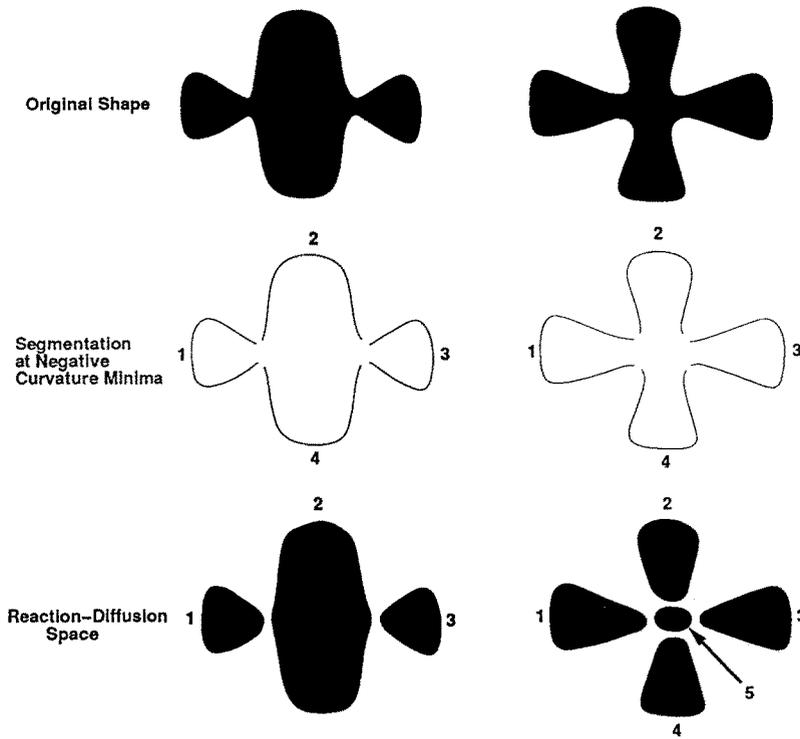


Fig. 15. Partitioning of a two-dimensional shape requires not only boundary, but also region information. We show two shapes (top row) and the un-natural part structure implied by (Hoffman and Richards 1985). Our theory leads to much more natural descriptions (bottom row).

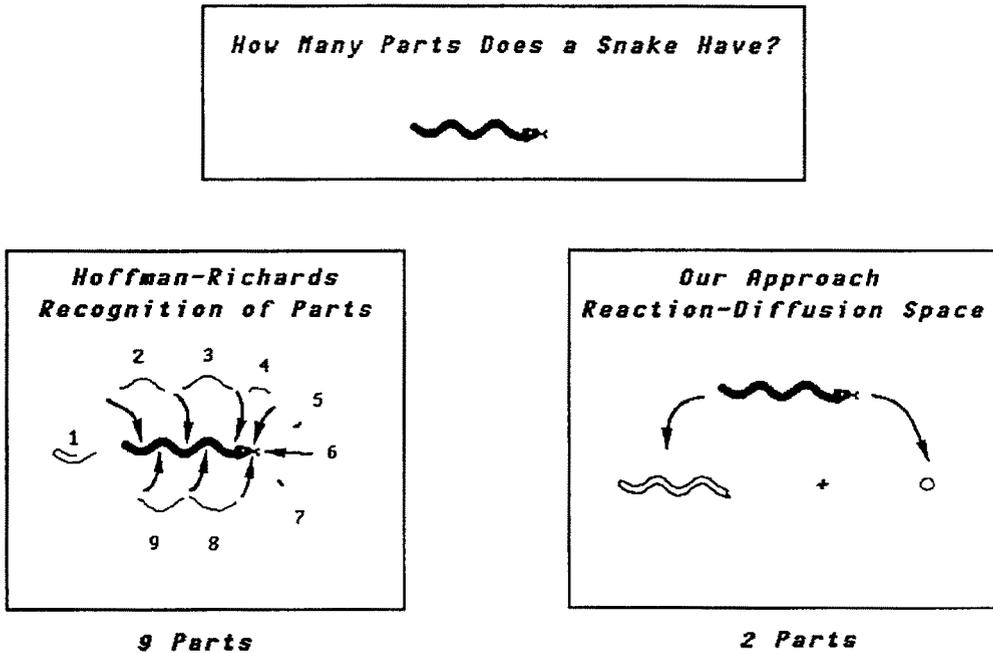


Fig. 16. Basing a partitioning theory on only the boundary information, may confuse “bent” bars with parts as the SNAKE example illustrates.

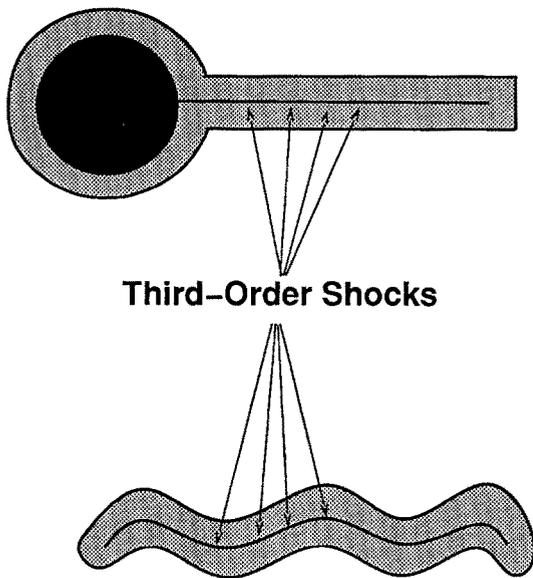


Fig. 17. The snake shape forms third-order shocks when distant points of the boundary come together not in isolation, but rather in conjunction with neighbors. Third-order shocks indicate the “bending” of an object. The interpretation of the snake therefore is as a “bent stick”.

an ellipse. However, this axis is not composed of first-order shocks where portions of the boundary collapse into a single point. Rather, this axis is the result of a region collapsing into points, Fig. 17. Therefore, the locus of these points indicates a *bending* of an extended region, rather than a protrusion of the boundary.

8.5 Fourth-Order Shocks

In the process of inward evolution of a shape, regions shrink and form shocks. In time, remaining regions finally shrink to a point and disappear due to the entropy condition. All parts of a shape must eventually annihilate to a point, since the shape may be entirely embedded inside some circle of radius R which will, in $\frac{R}{\beta_0}$ units of time, disappear. These are the fourth-order shocks and are the *seeds* for shape.

DEFINITION 6. When in the process of deformation a closed boundary collapses to a single point, a **FOURTH-ORDER SHOCK** is formed.

8.6 Examples

We now illustrate the reaction-diffusion space and the formation of shocks within it. To recall, the *reaction-*

diffusion space is the collection of all deformations of a shape for all combinations of reaction, β_0 , diffusion, β_1 , and time t . Since of these three variables only two are independent, we keep time and the ratio of reaction to diffusion, β_0/β_1 as the two variables which span the space. This choice is motivated by the fact that there is always some amount of diffusion present in any numerical implementation, so that the vertical lines at infinity on the x -axis are associated with pure reaction and the y -axis is associated with pure diffusion. Furthermore, the case of pure diffusion is a natural seam between inward and outward reaction; see Fig. 19.

To interpret our reaction-diffusion convention, then, each vertical line at x is a deformation of the original shape with the ratio $\beta_0/\beta_1 = x$ (the absolute values of β_0 and β_1 are not relevant in that they are absorbed in time t). Note that for $\beta_1 = 0.0$, or pure reaction, some diffusion manifests itself in the numerical implementation, so that shapes along this line diffuse minimally. The vertical dimension of the line represents time, or the amount of deformation. This vertical axis is conveniently depicted on a logarithmic scale and the numerals indicate the time frame of the computation from which the image was taken. However, note that there is no inherent “vertical topology” in this space and that the space is intended to generate a topology in both dimensions; this is a visual choice for representing this space.

There is an interesting connection between the reaction-diffusion and the symmetric axis transform (SAT): the locus of points at which shocks form and move under “pure” reaction defines axes that correspond to the loci of the symmetric axis transform. Since SAT is susceptible to noise, various smoothing algorithms have been proposed (Dill et al. 1987; Pizer et al. 1987). In our framework, however, diffusion is naturally part of the deformations. As such, the description of axes resulting from our shock provides a “coloration” of the symmetric axis into meaningful portions, and it also imposes a measure of significance on each; these ideas are developed in (Siddiqi et al. 1993).

Another interesting connection is that the reaction-diffusion space, under the pure reaction axes, embeds the mathematical morphology operations of erosion and dilation with a ball structuring element. The smoothing of these operations can now be viewed as annihilation of information through shocks. This information-theoretic view of mathematical morphology relates to the shape not only as a set, but also as a boundary. Furthermore, the reaction-diffusion space

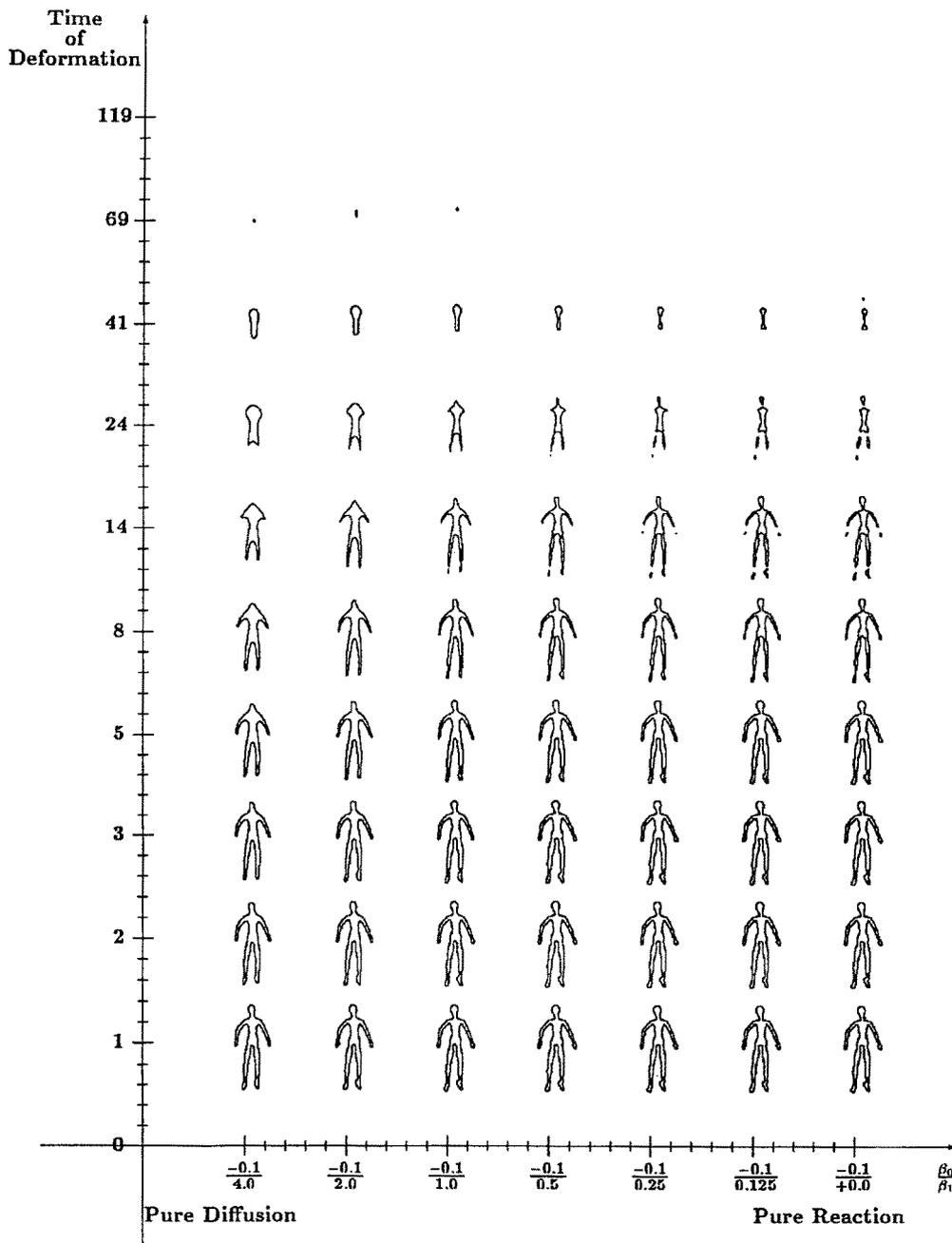


Fig. 18. This figure illustrates the right half of the reaction-diffusion space for the DOLL image. The DOLL image was taken from a range image collection of the National Research Council of Canada's Laser Range Image Library CNRC9077 Cat No 422. The image was thresholded and stored as a 128×128 image. The numbers on the x -axis are indicative of the two values β_0, β_1 in relation to time. Note that even in such low resolution the algorithm is robust and the trade-offs between reaction and diffusion are made clear: *Diffusion* "melts" the boundary of the shape by propagating and amalgamating boundary information to finally converge on a circular point (Grayson 1989); *Reaction*, on the other hand, is a rigid process which breaks the shape into pieces.

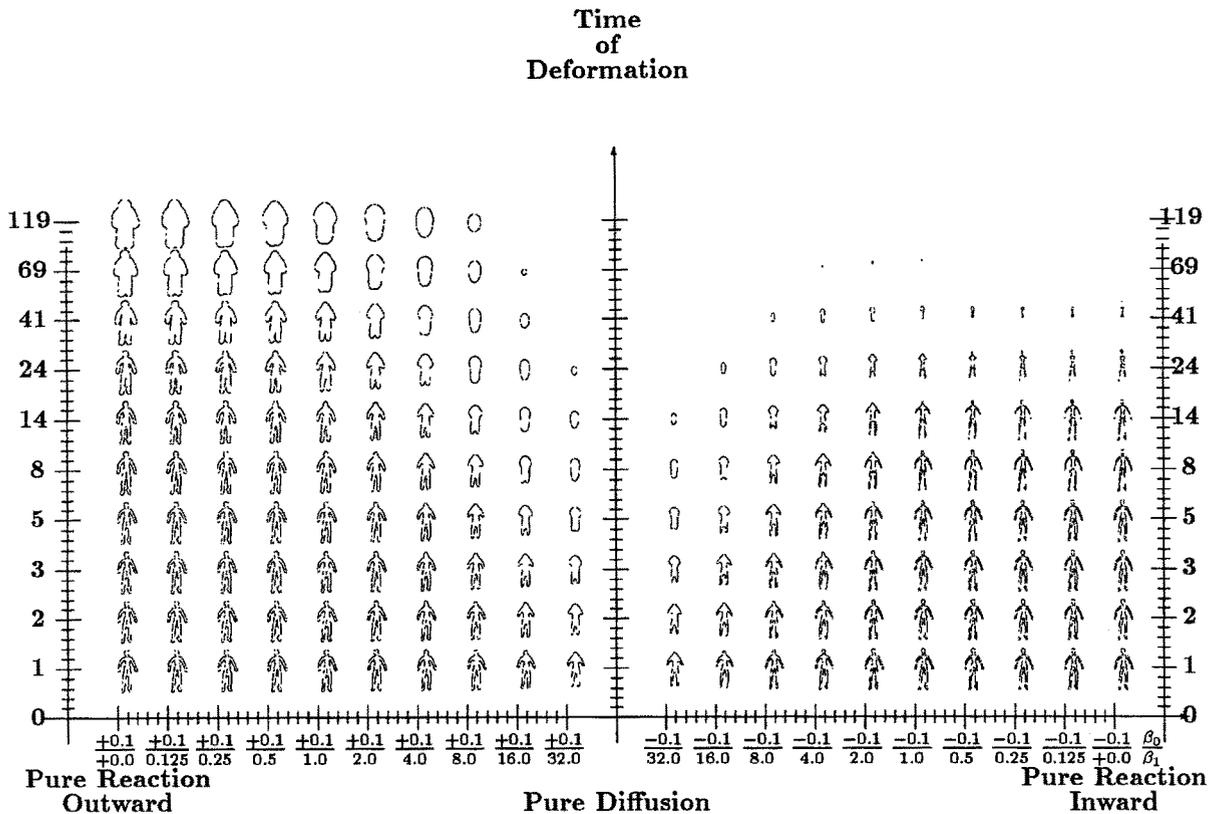


Fig. 19. This figure is an example of a reaction diffusion involving both inward and outward deformation. Note the formation of shocks with outward reaction. Also, observe that outward reaction may be thought of an inward reaction when the role of figure and ground is reversed.

can now be viewed as extending the mathematical morphology framework by adding a notion of significance through a diffusion process (Arehart et al. 1993; Sapiro et al. 1992). Finally, our algorithms are robust even for low resolution images (128×128) and may be viewed as a new method for numerically implementing mathematical morphology for a ball structuring element.

Shape representation is perhaps most important to object recognition. Any object matching method employs a similarity metric, whether it is explicit or implicit in the algorithm. As we have seen, the formation of shocks in the reaction-diffusion space and their classification yields a complete representation of the shape. These shocks as discrete events represent the shape not only statically, but also dynamically in relation in to its “nearby” shapes (compare (Koenderink and van Doorn 1986)).⁴ Since these deformations simplify shapes in time, the longer it takes two shapes to become similar under these deformations, the more dissimilar they are. Therefore, the

degree of similarity of the shock-based representation of shapes in the reaction-diffusion space is indicative of their degree of similarity for object recognition. Also, the discrete property of the shocks captures the generic categorical classification of shapes. These ideas will be described in more detail in a future paper.

An important concern in object recognition is occlusion. A representation of shape that is robust with occlusion is necessarily organized such that half the shape gives half the representation, leading to the need for a stable partitioning of shapes. Figure 5 illustrates how second-order shocks of the reaction-diffusion space in conjunction with the entropy scale space reconstruction give a hierarchy of parts for the doll. See (Siddiqi and Kimia 1994; Siddiqi et al. 1992) for first steps toward partitioning shapes.

Finally, the images are chosen to illustrate the reaction-diffusion space for a number of shapes and features: The CAT is a hand-drawn image (Gombrich

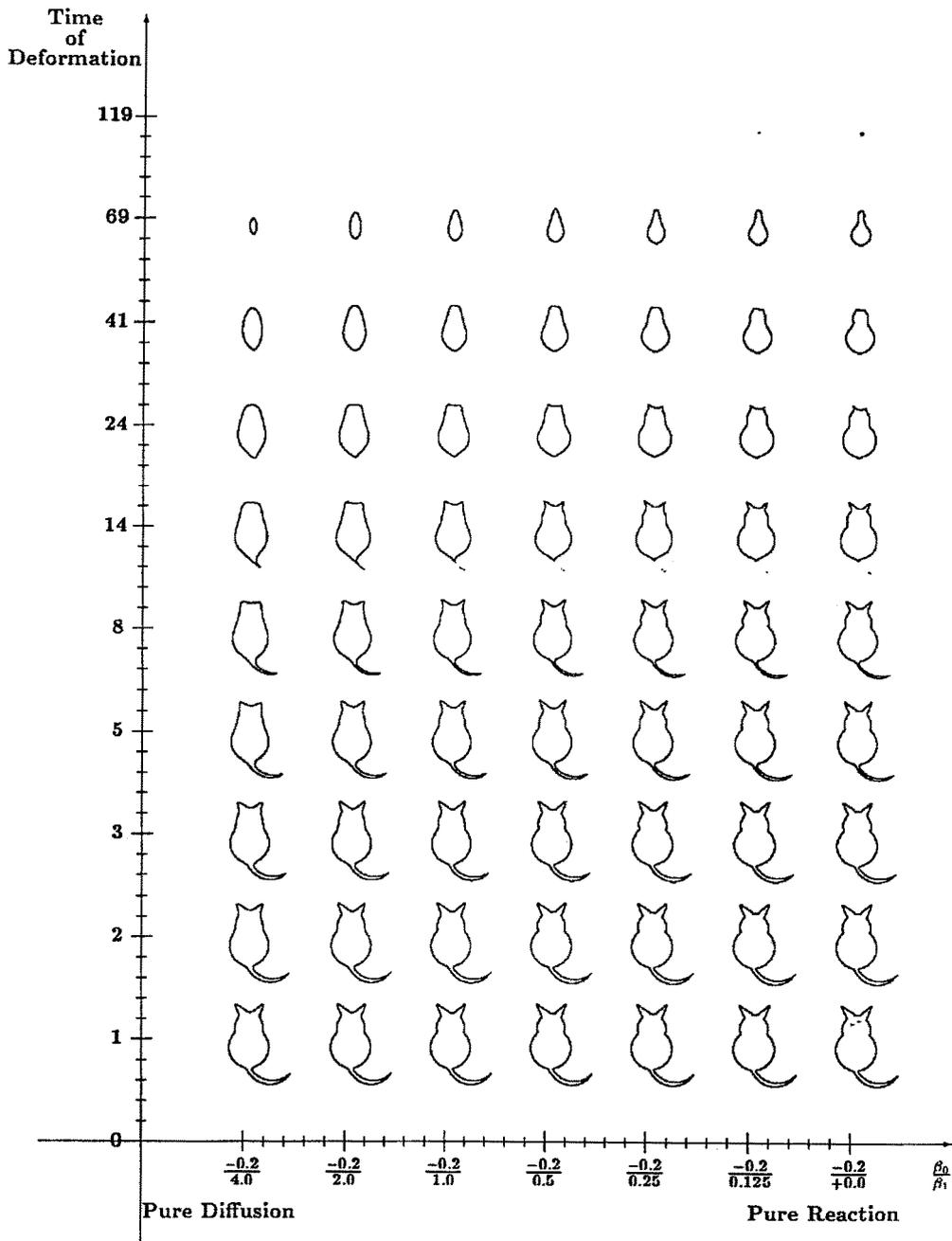


Fig. 20. The right half of the reaction-diffusion space of the CAT image. Note the tail as the shape undergoes the various deformations: *reaction* breaks the tail off (iterations 8 to 14 show this clearly), while *diffusion* melts the tail into the rest of the shape, resulting in a pointed deformation that persists over many iterations.

1956) consisting of simple geometric shapes; the DOLL is the binary version of the range image of a doll which is an example of an object with a hierarchy of body parts; the don Quixote image (NO141) is a

scan of a Picasso drawing; the F16 image is another man-made object representable as a collection of simple geometrical parts; and finally the HAND image is representative of biological shapes.

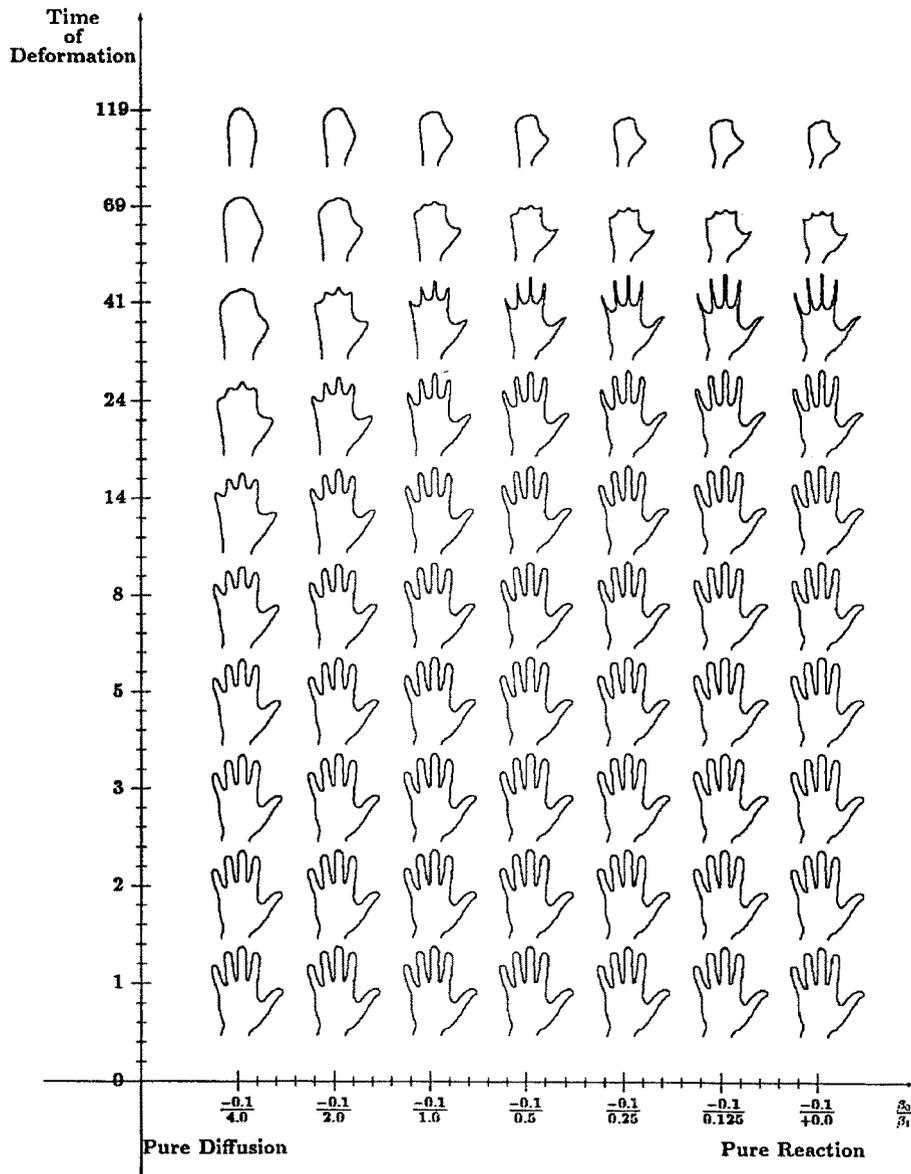


Fig. 21. The right half of the reaction-diffusion space for the HAND image. Note how each digit maps to a set of shocks. The shocks in the reaction-diffusion space support a description of a hand as five elongated blobs attached onto a central blob.

9 Conclusion

Determining whether two shapes are similar to one another is almost immediate for us, and certainly is related to our spectacular ability to recognize arbitrary objects. However, while computers can be easily programmed to determine whether two shapes are metrically identical (a task of significant difficulty for us!), they have been very poor at general object recognition. This dif-

ference is more than an academic curiosity, because of the role that recognition plays in applications of computer vision.

Central to recognition is determining which shapes are similar, and we have begun to develop a theory of shape from which this can be determined. The foundations for the theory stand on two basic ideas. First, we observe that, if a boundary were changed only slightly, then, in general, its shape would change only slightly.

This leads us to propose an operational theory of shape based on incremental contour deformations, and to an analysis of the associated differential geometry. Contour deformations are characterized via two special ones: a deformation constant along the normal and another one that varies in proportion to the curvature, and much of the paper was devoted to analyzing and understanding these two deformations.

As long as the normal exists at every point along a contour, the deformation is classical. However, discontinuities naturally arise, and, if evolution is unrestricted, contour segments can pass over one another. The first of these problems causes mathematical difficulties, and the latter is awkward from a shape perspective. Our second basic observation about shapes: that they are not arbitrary contours, but rather are those contours which can enclose “physical” material, is therefore introduced to guide the development. Conservation laws are obtained for shape in analogy to the way they operate in mathematical physics using ideas from (Osher and Sethian 1988; Sethian 1985a, c). Special care is taken to introduce a number of principles that we take to be intuitive and self evident, and which formally guide the development of the theory. The mathematical delicacies are organized by an entropy principle, and Hamilton-Jacobi theory is introduced to computer vision. The discontinuities, or shocks, are shown to be much more than mathematical difficulties—they reveal the singular events that can arise during the process of shape deformation. It is these shocks that are the computational elements of shape; they provide the language in which to describe shapes. They include (i) first-order shocks, or deformations; (ii) second-order shocks, or parts; (iii) third-order shocks, or bends; and, finally (iv) fourth-order shocks, or the seeds for shapes. The type of shocks, the “time” at which they arise during deformation (their significance), and their spatial relationships provide the description of shape that we shall use for recognition. The entire theory is consistent with our intuitive principles, and suggests a mechanism for capturing the categorical nature of generic shapes.

In this paper, we have attempted to unify several of the competing aspects of shape. Special cases of the deformation geometry correspond to (a type of) Gaussian blurring and mathematical morphology, and the loci of points traced out by the shocks are related to skeletons. Local and global aspects of shape, as well as boundary vs. interior issues, are brought together. Finally, we develop a space of shapes, the

reaction-diffusion space, in which “similar” shapes are arranged according to the different axes of deformation. In subsequent papers we shall show how these ideas lead to approximation and (non-linear) simplifications of shapes, and to a topology over shapes. We believe that this will begin to provide the foundation for general shape recognition.

Acknowledgments

This research was supported by grants from AFSOR, ARO, NSERC, and NSF. SNZ is a fellow of the Canadian Institute for Advanced Research.

We thank David Mumford for an early reference to the work of Grayson, and Allan Dobbins and Lee Iverson for discussions and technical help.

A Entropy and Viscosity Solutions

In this appendix, for the convenience of the reader, we collect some standard facts about entropy solutions to conservation laws and their relation to viscosity theory of the Hamilton-Jacobi equation. Full details can be found in (Barles 1985; Lions 1981; Crandall and Lions 1983; Fleming and Soner 1993).

A.1 Weak Solutions of Conservation Laws

In this section, we discuss a weak form of a solution for a hyperbolic conservation law of the form

$$u_t + f(u)_x = 0, \quad u(x, 0) = u_0. \quad (57)$$

Suppose that u in fact is a strong classical solution of (57). Let C_0^1 denote the set of continuously differentiable functions with compact support on $\mathbf{R} \times \mathbf{R}^+$. (Here \mathbf{R}^+ denotes the set of real numbers ≥ 0 .) Let $\psi \in C_0^1$. We suppose that the support of ψ is contained in the rectangle $R = [x_1, x_2] \times [0, T]$ with $|x_1|, |x_2|, T$ chosen sufficiently large so that

$$\psi(x_1, t) = \psi(x_2, t) = \psi(x, T) = 0, \quad \forall t \in \mathbf{R}^+, x \in \mathbf{R}.$$

Then multiplying (57) by ψ and integrating, we see that

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_0^{\infty} (u_t + f(u)_x) \psi dx dt \\ &= \int \int_R (u_t + f(u)_x) \psi dx dt = 0. \end{aligned}$$

The trick is now to integrate by parts:

$$\begin{aligned} & \int_{x_1}^{x_2} \int_0^T u_t \psi dx dt \\ &= \int_{x_1}^{x_2} [u(x, T)\psi(x, T) - u(x, 0)\psi(x, 0)] dx \\ & \quad - \int_{x_1}^{x_2} \int_0^T u \psi_t dx dt \\ &= \int_{-\infty}^{\infty} -u_0(x)\psi(x, 0) dx - \int_{x_1}^{x_2} \int_0^T u \psi_t dx dt. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{x_1}^{x_2} \int_0^T g(u)_x \psi dx dt &= \int_0^T [f(u(x_2, t))\psi(x_2, t) \\ & \quad - f(u(x_1, t))\psi(x_1, t)] dt \\ & \quad - \int_0^T \int_{x_1}^{x_2} f(u) \psi_x dx dt \\ &= - \int_0^T \int_{x_1}^{x_2} f(u) \psi_x dx dt. \end{aligned}$$

Hence, we derive the *weak form* of the conservation law (57):

$$\int_{-\infty}^{\infty} \int_0^{\infty} (u \psi_t + f(u) \psi_x) dx dt + \int_0^{\infty} u_0 \psi dx = 0. \quad (58)$$

The point is that (58) makes sense for $u(x, t)$ bounded and measurable (with bounded measurable initial data u_0), for all $\psi \in C_0^1$. Thus such functions $u(x, t)$ which satisfy (58) for all $\psi \in C_0^1$ are called *weak solutions*. Finally, it is easy to show that if a continuously differentiable function u satisfies (58) for all $\psi \in C_0^1$, then ψ is a classical solution of (57).

B Hamilton-Jacobi Equation

In this section, we review the connection of Hamilton-Jacobi theory with optimal control in order to motivate the definition of “viscosity solution.” The discussion we give below can be applied more generally to the Hamilton-Jacobi-Bellman equation, but we are only interested in outlining the main ideas here. For more details, see the recent book (Fleming and Soner 1993) as well as (Lions 1981).

Let us begin by studying the calculus of variations problem on the fixed interval $[t, t_1]$. Set

$$U := L^\infty([t, t_1]; \mathbf{R}^n).$$

Consider the dynamical system

$$\frac{dx}{ds} = u(s), \quad u \in U \quad (59)$$

that is $x(s)$ is the *state variable* and $u(s)$ is the *control*. We require the boundary conditions $x(t) = x$, and $x(t_1) \in \mathcal{Q}$ where $\mathcal{Q} \subset \mathbf{R}^n$ is closed. (We should note that one can take more general functions $g(s, x(s), u(s))$ on the right-hand side of (59). We have chosen just to consider the simplest classical case in order to motivate our treatment of the Hamilton-Jacobi equation via control principles.) Then the problem is to minimize the integral (*cost functional*)

$$J := \int_t^{t_1} \mathcal{L}(s, x(s), u(s)) ds + \phi(x(t_1)), \quad (60)$$

subject to the constraint (59) over all Lipschitz continuous curves $x: [t, t_1] \rightarrow \mathbf{R}^n$ which satisfy the endpoint conditions.

The associated Hamiltonian function is

$$\mathcal{H}(t, x, p) := \max_{v \in \mathbf{R}^n} \{-v \cdot p - \mathcal{L}(t, x, v)\}. \quad (61)$$

The dual formula is

$$\mathcal{L}(t, x, p) := \max_{p \in \mathbf{R}^n} \{-v \cdot p - \mathcal{H}(t, x, p)\}. \quad (62)$$

The extremal points where the maximum is taken on in (61) and (62) are related by

$$p = -\mathcal{L}_v, \quad v = -\mathcal{H}_p \quad (63)$$

(the Legendre transformation). This duality between \mathcal{L} and \mathcal{H} is basically that of the Lagrangian and Hamiltonian formulations of classical mechanics where v is velocity and p is momentum.

Define the *value function* or the *optimal return*

$$V(t, x) := \inf_{u \in U} J(t, x; u). \quad (64)$$

Then for continuously differentiable V , one may show using dynamic programming that V satisfies the Hamilton-Jacobi equation

$$-\frac{\partial V}{\partial t} + \mathcal{H}(t, x, \nabla_x V) = 0. \quad (65)$$

(Note that ∇_x denotes the gradient computed with respect to the space variables x .) Assuming that $\mathcal{Q} = \mathbf{R}^n$, we get the boundary condition

$$V(t_1, x) = \phi(x). \quad (66)$$

(See a sketch of the derivation of Eq. (65) below.)

One can also show that one gets the same Hamilton-Jacobi equation as (65) for the following *exit time problem*: Minimize

$$J := \int_t^{t_e} \mathcal{L}(s, x(s), \dot{x}(s)) ds + \Phi(t_e, x(t_e)),$$

where t_e is the exit time of the “state trajectory” $(s, x(s))$ from $[t_0, t_1] \times X$, where $X \in \mathbf{R}^n$ is closed, and Φ is a function such that

$$\Phi(t, x) := \begin{cases} f(t, x) & \text{if } (t, x) \in [t_0, t_1] \times \mathbf{R}^n, \\ \phi(x) & \text{if } (t, x) \in \{t_1\} \times \mathbf{R}^n. \end{cases} \quad (67)$$

REMARK. The problems considered above are called *finite horizon problems*. There is a related class of optimization problems called *infinite horizon*. Suppose that we have a dynamical system of the form

$$\begin{aligned} \frac{dx}{ds} &= g(x(s), u(s)), \\ x(0) &= x_0. \end{aligned}$$

The cost in this case is given by a functional of the form

$$J(x, u) := \int_0^t e^{-\alpha s} \mathcal{L}(x(s), u(s)) ds + e^{-\alpha t} \phi(x(t)).$$

A Hamilton-Jacobi theory with associated value functions can be developed in this setting as well.

One needs to introduce *generalized solutions* to the Hamilton-Jacobi Eq. (65) in case V is not differentiable. A standard example for the exit time problem may be constructed as follows: We let the “Lagrangian”

$$\mathcal{L}(t, x, v) := 1 + \frac{v^2}{4},$$

$\Phi \equiv 0$, $X := [-1, 1]$, $[t_0, t_1] = [0, 1]$. So we want to minimize

$$\int_t^{t_e} \left\{ 1 + \frac{\dot{x}(s)^2}{4} \right\} ds.$$

The optimal control in this case for given initial condition (t, x) can be computed to be

$$u_{\text{opt}} := \begin{cases} -2 & \text{if } x \leq -t, \\ 0 & \text{if } |x| < t, \\ 2 & \text{if } x \geq t. \end{cases}$$

The corresponding value function is

$$V(t, x) := \begin{cases} 1 - t & \text{if } |x| \leq -t, \\ 1 - |x| & \text{if } |x| \geq t. \end{cases}$$

V is not differentiable for $t = |x|$. Note however that V satisfies the corresponding Hamilton-Jacobi equation

$$-\frac{\partial V(t, x)}{\partial t} + \left(\frac{\partial V(t, x)}{\partial x} \right)^2 - 1 = 0$$

except when $t = |x|$.

In fact, this illustrates the general fact that the value function satisfies the Hamilton-Jacobi at all points where it is differentiable. Let us call \mathcal{V} a *generalized solution* of the Hamilton-Jacobi Eq. (65) if \mathcal{V} is locally Lipschitz and satisfies (65) almost everywhere. Then one may show that the value function is a generalized solution in this sense. (See our discussion in the next section.) The problem is that typically there are many generalized solutions. We therefore want to pick out one “natural” solution. We show how to do this next.

B.1 Viscosity Solutions

The above discussion motivates the introduction of *viscosity solutions*. Viscosity solutions were introduced by Crandall and Lions (1983). For another approach to the generalized solutions of Hamilton-Jacobi using techniques from *nonsmooth analysis*, see (Clarke 1989) and the references therein.

We set $C^\infty = C^\infty(\mathbf{R}^n \times (0, \infty))$. Let V_ϵ be a classical solution of (65) and let $\omega \in C^\infty$. Suppose that (x_ϵ, t_ϵ) is a local maximum of $V_\epsilon - \omega$. Then via the maximum principle,

$$\begin{aligned} \frac{\partial V_\epsilon}{\partial t}(x_\epsilon, t_\epsilon) &= \frac{\partial \omega}{\partial t}(x_\epsilon, t_\epsilon), \\ \nabla_x V_\epsilon(x_\epsilon, t_\epsilon) &= \nabla_x \omega(x_\epsilon, t_\epsilon), \\ \Delta V_\epsilon(x_\epsilon, t_\epsilon) &\leq \Delta \omega(x_\epsilon, t_\epsilon). \end{aligned}$$

Thus we see from (65) that for all $\omega \in C^\infty$, we have

$$\begin{aligned} -\frac{\partial \omega}{\partial t}(x_\epsilon, t_\epsilon) - \epsilon \Delta \omega(x_\epsilon, t_\epsilon) \\ + \mathcal{H}(x_\epsilon, t_\epsilon, \nabla_x \omega(x_\epsilon, t_\epsilon)) \leq 0. \end{aligned} \quad (68)$$

Similarly, playing the same game for a local minimum, we get that for $\omega \in C^\infty$,

$$\begin{aligned} -\frac{\partial \omega}{\partial t}(x_\epsilon, t_\epsilon) - \epsilon \Delta \omega(x_\epsilon, t_\epsilon) \\ + \mathcal{H}(x_\epsilon, t_\epsilon, \nabla_x \omega(x_\epsilon, t_\epsilon)) \geq 0. \end{aligned} \quad (69)$$

Now if V_ϵ converges uniformly on compact subsets to $V: \mathbf{R}^n \times (0, \infty) \rightarrow \mathbf{R}$ continuous, and if we take $\omega \in C^\infty$, and assume that $V - \omega$ has a strict local maximum (x_0, t_0) , then there exists (x_ϵ, t_ϵ) which converges to

(x_0, t_0) such that each (x_ϵ, t_ϵ) is a local maximum point of $V_\epsilon - \omega$. Hence (68) is satisfied, and we can pass to the limit. We therefore make the following definition:

DEFINITION 7. Let $V: \mathbf{R}^n \times (0, \infty) \rightarrow \mathbf{R}$ be continuous.

1. V is a *viscosity subsolution* of (65) if for any given $\omega \in C^\infty$, at each local maximum point of $V - \omega$, say (x_0, t_0) , we have

$$-\frac{\partial \omega}{\partial t}(x_0, t_0) + \mathcal{H}(x_0, t_0, \nabla_x \omega(x_0, t_0)) \leq 0.$$

2. V is a *viscosity supersolution* of (65) if for any given $\omega \in C^\infty$, at each local minimum point of $V - \omega$, say (x_0, t_0) , we have

$$-\frac{\partial \omega}{\partial t}(x_0, t_0) + \mathcal{H}(x_0, t_0, \nabla_x \omega(x_0, t_0)) \geq 0.$$

3. V is a *viscosity solution* of (65) if it is both a viscosity subsolution and supersolution.

A beautiful fact is that the value function defined by (64) is the unique **viscosity solution** of (65). For the eikonal equation (the “prairie fire” model of computer vision) this has an especially neat form.

B.2 Eikonal Equation

Barles (1985) noticed that the entropy solution of a hyperbolic conservation law is equivalent to the viscosity solution of the corresponding Hamilton-Jacobi equation. We will outline his argument for the classical eikonal equation which appears in geometric optics.

Consider the eikonal equation

$$-\frac{\partial V}{\partial t} + H\left(\frac{\partial V}{\partial x}\right) = 0, \tag{70}$$

$$V(x, 0) = V_0(x),$$

where the Hamiltonian is given by

$$H(p) := \sqrt{1 + p^2}.$$

Then one can show (Lions 1981) in this case that the value function which gives the unique viscosity solution of (70) is given by the Lax-Oleinik formula

$$V(t, x) = \min_y \left[V_0(y) + t H^* \left(\frac{x - y}{t} \right) \right]. \tag{71}$$

Here H^* is the *conjugate function* of H defined by

$$H^*(w) = \max[uw - H(u)].$$

Barles showed that this is closely related to the entropy condition introduced by Sethian in relation to the prairie fire model of flame propagation (Sethian 1985a). We work in \mathbf{R}^2 for simplicity.

Let

$$C = \{(x, y) \in \mathbf{R}^2: y = \phi(x)\},$$

where ϕ is any continuous function. We assume that the particles below C are burnt, and the region above C is filled with a combustible fluid. We assume that the front propagates in the normal direction with constant speed one. Sethian (1985a) describes the motion of the flame front even when singularities develop by introducing the following entropy condition:

“Once a particle is burnt, it remains burnt.”

Let

$$y = V(x, t), \quad t > 0$$

denote the position of the prairie fire when evolving according to this law. Then one may prove the following (Barles 1985):

PROPOSITION 1 (Barles). *Notation as above. Then the position of the prairie fire is given by $y = V(x, t)$, where V is the (unique) viscosity solution of the eikonal Eq. (70) with boundary condition*

$$V(x, 0) = \phi(x).$$

PROOF. The proof is very easy. One may show using the above description of the prairie fire evolution that the front is given by

$$V(x, t) = \max_{|z-x| \leq t} \{\phi(z) + \sqrt{t^2 - |z-x|^2}\}.$$

It is then easy to check that this is the Lax-Oleinik (value function) formula (71) given above.

C Generic Singularities

In this appendix, we would like to use Arnold’s (1989) classification of singularities to show that shocks of type 1 and 2 are generic. We will do this for the grass-fire evolution in the plane, given by the parallel evolution $C_t = -\vec{N}$, where as above \vec{N} denotes the outward normal. In order to do this, we must realize this flow as defining a (non-singular) surface in \mathbf{R}^4 .

Let $C \subset \mathbf{R}^2$ denote a plane curve. We associate to C , the curve $\tilde{C} \subset \mathbf{R}^4$ given by

$$\tilde{C} := \{(x, y, p_x, p_y) : (x, y) \in C, p_x^2 + p_y^2 = 1, p_x(\xi - x) + p_y(\eta - y) = 1\}$$

where (ξ, η) is the point on the (outward) normal to the curve at (x, y) , at distance 1. We now use the treatment of Hamilton's equations and the calculus of variations from Arnold (1989), Chapter 3 and the Huygens principle, Chapter 9 (see pages 248–252). Accordingly, we consider the Hamiltonian $H(x, y, p_x, p_y) := 1/2(p_x^2 + p_y^2)$, and using the equations from (Arnold 1989) page 65, we arrive at the following system of first-order equations:

$$\begin{aligned} \dot{x} &= p_x \\ \dot{y} &= p_y \\ \dot{p}_x &= 0 \\ \dot{p}_y &= 0. \end{aligned}$$

We evolve \tilde{C} under this system of equations. Let $S := \{\tilde{C}(t)\} \subset \mathbf{R}^4$ denote the resulting surface. Let $\pi: \mathbf{R}^4 \rightarrow \mathbf{R}^2$ denote the projection $\pi(x, y, p_x, p_y) := (x, y)$. Then $\pi(\tilde{C}(t)) := C(t)$ is the solution to

$$C_t = -\vec{N}$$

at time $t > 0$. Notice that surface S always remains smooth.

Now we can use Arnold's classification of generic singularities, i.e., singularities which are preserved under small perturbations. Indeed, let (x_1, x_2) denote local coordinates on the surface S . Then we can (locally) express $\pi: S \rightarrow \mathbf{R}^2$ as $\pi(x_1, x_2) = (y_1, y_2)$. One can write down normal forms for the mapping $\pi|_S$. Indeed, there are three normal forms. Only two are relevant for the classification of generic shocks. The first normal form is

$$y_1 = x_1, \quad y_2 = x_2.$$

This corresponds to a transversal intersection shock in the evolution $C_t = -\vec{N}$, that is, a shock of type 1. The second normal form is the *Whitney tuck*, that is,

$$y_1 = x_1x_2 - x_1^3, \quad y_2 = x_2.$$

This will give a semi-cubical cusp in the evolution $C_t = -\vec{N}$, that is a singularity of the form $y^3 = x^2$ which is a type 2 shock.

By the Arnold theory, these are the only possible generic shocks.

Notes

- i.e. Deformations that do not depend on the coordinate system.
- A classical or strong solution to a partial differential equation is one whose derivatives exists and satisfy the equality. A non-classical or weak solution requires a treatment in the sense of distributions and equality in the integral form of the equation since derivatives do not exist (see appendix).
- A function is said to be Hölder continuous of order α if there exists $M > 0$ such that

$$\|f(x) - f(y)\| \leq M\|x - y\|^\alpha.$$
- As a preview, see the shapes arranged in the entropy scale space illustrated in Fig. 19 (Kimia et al. 1994).

References

- Alvarez, L., Guichard, F., Lions, P., and Morel, J. 1992. Axiomatization et nouveaux operateurs de la morphologie mathematique. *C. R. Acad. Sci. Paris*, 315:265–268.
- Alvarez, L., Lions, P.-L., and Morel, J.-M. 1992. Image selective smoothing and edge detection by nonlinear diffusion: II. *SIAM Journal of Numerical Analysis*, 29(3):845–866.
- Angenent, S.B. 1988. The zero set of a solution of a parabolic equation. *J. für die Reine und Angewandte Mathematik*, 390:79–96.
- Arehart, A., Vincent, L., and Kimia, B.B. 1993. Mathematical morphology: The Hamilton-Jacobi connection. In *Fourth International Conference on Computer Vision* (Germany, Berlin), Washington, D.C. Computer Society Press.
- Arnold, V. 1989. *Mathematical Methods of Classical Mechanics*, Second Edition. Springer-Verlag.
- Asada, H. and Brady, M. 1983. The curvature primal sketch. *IEEE PAMI*, 8:2–14.
- Attneave, F. 1954. Some informational aspects of visual perception. *Psych. Review*, 61:183–193.
- Ballard, D. 1981. Strip trees: A hierarchical representation for curves. *Comm. ACM*, 24(5):310–321.
- Ballard, D.H. and Brown, C.M. 1982. *Computer Vision*. Prentice-Hall, New Jersey.
- Barles, G. 1985. Remarks on a flame propagation model. Technical Report No. 464, INRIA Rapports de Recherche.
- Biederman, I. 1987. Recognition by components. *Psych. Review*, 94:115–147.
- Binford, T. 1981. Inferring surfaces from images. *Artificial Intelligence*, 17:205–244.
- Blake, A. and Zisserman, A. 1987. *Visual Reconstruction*. MIT Press, Cambridge, MA.
- Blum, H. 1973. Biological shape and visual science. *J. Theor. Biol.*, 38:205–287.
- Bolles, R. and Cain, R. 1982. Recognizing and locating partially visible objects: The local-feature-focus method. *International Journal of Robotics Research*, 1(3):57–82.
- Brockett, R. and Maragos, P. 1992. Evolution equations for continuous-scale morphology. In *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, San Francisco, CA.
- Brooks, R. 1981. Symbolic reasoning among 3-d models and 2-d images. *Artificial Intelligence*, 17:285–348.

- Bulthoff, H.H. and Edelman, S. 1992. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci., USA*, 89:60–64.
- Clarke, F. 1989. *Optimization and Nonsmooth Analysis*. Montreal:Les publications CRM.
- Conway, E. and Hopf, E. 1964. Hamilton's theory and generalized solutions of the Hamilton-Jacobi equation. *Journal of Mathematics and Mechanics*, 13(6):939–986.
- Crandall, M.G., Ishii, H., and Lions, P.-L. 1992. User's guide to viscosity solutions of second-order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67.
- Crandall, M.G. and Lions, P.-L. 1983. Viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 277:1–42.
- Dill, A.R., Levine, M.D., and Noble, P.B. 1987. Multiple resolution:skeletons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):495–504.
- do Carmo, M.P. 1976. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, New Jersey.
- Dudek, G. 1990. Shape representation from curvature. Ph.D. dissertation, Dept. of Computer Science, University of Toronto, Toronto, Canada.
- Elder, J. and Zucker, S.W. 1993. The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research*, 33(7):981–991.
- Ferrie, F.P., Levine, M.D., and Zucker, S.W. 1982. Cell tracking: A modeling and minimization approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4:277–291.
- Firey, W.J. 1974. Shapes of worn stones. *Mathematika*, 21:1–11.
- Fischler, M. and Elschlager, R. 1973. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92.
- Fleming, W. and Soner, H. 1993. *Controlled Markov Processes and Viscosity Solutions*. Springer-Verlag.
- Freeman, H. 1974. Computer processing of line drawing images. *Computer Surveys*, 6(1):57–98.
- Gage, M. and Hamilton, R.S. 1986. The heat equation shrinking convex plane curves. *J. Differential Geometry*, 23:69–96.
- Garabedian, P. 1964. *Partial Differential Equations*. John Wiley and Sons.
- Gardenier, P., McCallum, B., and Bates, R. 1986. Fourier transform magnitudes are unique pattern recognition templates. *Biological Cybernetics*, 54:385–391.
- Gombrich, E. 1956. *Art and Illusion*. Princeton University Press.
- Granlund, G. 1972. Fourier processing for handprinted character recognition. *IEEE Transaction on Computing*, c-21:195–201.
- Grayson, M.A. 1989. Shortening embedded curves. *Annals of Mathematics*, 129:71–111.
- Hoffman, D.D. and Richards, W.A. 1985. Parts of recognition. *Cognition*, 18:65–96.
- Hopf, E. 1950. The partial differential equation $u_t + uu_x = \epsilon u_{xx}$. *Comm. Pure Appl. Math.*, 3:201–230.
- Kass, M., Witkin, A., and Terzopoulos, D. 1988. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331.
- Kimia, B.B. 1990. Conservation laws and a theory of shape. Ph.D. dissertation. McGill Centre for Intelligent Machines, McGill University, Montreal, Canada.
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1989. Toward a computational theory of shape: An overview. CIM-89-13, McGill Centre for Intelligent Machines, McGill University, Montreal, Canada.
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1990. Toward a computational theory of shape: An overview. In O. Faugeras, ed., *Lecture Notes in Computer Science*, Berlin, Springer Verlag, 427:402–407.
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1992. On the evolution of curves via a function of curvature, I: The classical case. *Journal of Mathematical Analysis and Applications*, 163(2).
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1993. Non-linear shape approximation via the entropy scale space. In *Proceedings of the SPIE's Geometric Methods in Computer Vision II*, volume 2031, San Diego, California, pp. 218–233.
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1994. On the evolution of curves via a function of curvature, II: Local properties. In Preparation.
- Kimia, B.B., Tannenbaum, A.R., and Zucker, S.W. 1994. Shapes, shocks, and deformations, II: The simplification of shape and the entropy scale-space. In preparation.
- Koenderink, J.J. 1984. The structure of images. *Biological Cybernetics*, 50:363–370.
- Koenderink, J.J. 1990. *Solid Shape*. MIT Press, Cambridge, Massachusetts.
- Koenderink, J.J. and van Doorn, A.J. 1986. Dynamic shape. *Biological Cybernetics*, 53:383–396.
- Langer, J. 1980. Instabilities and pattern formation in crystal growth. *Reviews of Modern Physics*, 52(1):1–28.
- Lax, P.D. 1957. Hyperbolic systems of conservation laws II. *Comm. Pure Appl. Math.*, 10:537–566.
- Lax, P.D. 1971. *Shock Waves and Entropy*, Academic Press, New York, pp. 603–634.
- Lax, P.D. 1973. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. SIAM Regional Conference series in Applied Mathematics, Philadelphia.
- Leclerc, Y.G. 1989. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102.
- Leyton, M. 1987. Symmetry-curvature duality. *Computer Vision, Graphics, and Image Processing*, 38:327–341.
- Leyton, M. 1988. A process grammar for shape. *Artificial Intelligence*, 34:213–247.
- Leyton, M. 1989. Inferring causal history from shape. *Cognitive Science*, 13:357–387.
- Link, N.K. and Zucker, S.W. 1987. Sensitivity to corners in flow patterns. *Spatial Vision*, 12(3):233–244.
- Lions, P. 1981. *Generalized Solutions of Hamilton-Jacobi Equations*. Pitman.
- Marr, D. and Nishihara, K.H. 1978. Representation and recognition of the spatial organization of three-dimensional structure. *Proceedings of the Royal Society of London, B*, 200:269–294.
- Mokhtarian, F. and Mackworth, A. 1986. Scale-based description of planar curves and two-dimensional shapes. *PAMI*, 8:34–43.
- Mumford, D. 1987. The problem of robust shape descriptors. In *Proceedings of the First International Conference on Computer Vision*, London, England, pp. 602–606.
- Oleinik, O. 1957. Discontinuous solutions of nonlinear differential equations. *Amer. Math. Soc. Transl. Ser. 2*, 26:95–172.
- Osher, S. and Sethian, J. 1988. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49.
- Pentland, A. 1987. Recognition by parts. In *First International Conference on Computer Vision*, (London, England), IEEE Computer Society Press, Washington, DC.

- Pentland, A. 1989. Part segmentation for object recognition. *Neural Computation*, 1:82–91.
- Perona, P. and Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):629–639.
- Persoon, E. and Fu, K. 1977. Shape discrimination using fourier descriptors. *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-7:170–179.
- Pizer, S.M., Oliver, W.R., and Bloomberg, S.H. 1987. Hierarchical shape description via the multiresolution symmetric axis transform. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4):505–511.
- Ramer, U. 1972. An iterative procedure for the polygonal approximation of plane curves. *Computer Vision, Graphics and Image Processing*, 1(3):244–256.
- Richards, W., Dawson, B., and Whittington, D. 1986. Encoding contour shape by curvature extrema. *Journal of Optical Society of America*, 3(9):1483–1489.
- Richards, W. and Hoffman, D.D. 1985. Codon constraints on closed 2-d shapes. *Computer Vision, Graphics and Image Processing*, 31(2):156–177.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439.
- Rosenfeld, A. and Kak, A.C. 1982. *Digital Picture Processing*. Academic Press.
- Samet, H. 1980. Region representation: Quad trees from boundary codes. *Comm. ACM*, 23(3):163–170.
- Sapiro, G., Kimia, B.B., Kimmel, R., Shaked, D., and Bruckstein, A. 1992. Implementing continuous-scale morphology. *Pattern Recognition*, 26(9).
- Sapiro, G. and Tannenbaum, A. 1993. Affine invariant scale-space. *International Journal of Computer Vision*, 10:25–44.
- Sapiro, G. and Tannenbaum, A. 1994. On affine plane curve evolution. *Journal of Functional Analysis*, 119:79–120.
- Serra, J. 1982. Ed. *Image Analysis and Mathematical Morphology*. Academic Press.
- Serra, J. 1988. Ed. *Image Analysis and Mathematical Morphology, Part II: Theoretical Advances*. Academic Press.
- Sethian, J. and Osher, S. 1988. *The Design of Algorithms for Hypersurfaces Moving with Curvature-Dependent Speed*. Springer Verlag.
- Sethian, J. 1985a. An analysis of flame propagation. Ph.D. dissertation, University of California, Berkeley, California.
- Sethian, J. 1985b. *Numerical Methods for Propagating Fronts*, Springer-Verlag, New York, pp. 155–164.
- Sethian, J.A. 1985c. Curvature and the evolution of fronts. *Comm. Math. Physics*, 101:487–499.
- Sethian, J. 1989. A review of recent numerical algorithms for hypersurfaces moving with curvature dependent speed. *J. Diff. Geom.*, (33):131–161.
- Sethian, J.A. 1990. Numerical algorithms for propagating interfaces: Hamilton-Jacobi equations and conservation laws. *J. Differential Geometry*, 31:131–161.
- Siddiqi, K. and Kimia, B.B. 1994. Parts of visual form: Computational aspects. *IEEE Trans. Pattern Analysis and Machine Intelligence*. To appear.
- Siddiqi, K., Manos, A., and Kimia, B.B. 1993. Colored skeletons in scale. Technical Report. In preparation, Brown University.
- Siddiqi, K., Tresness, K.J., and Kimia, B.B. 1992. Parts of visual form: Ecological and psychophysical aspects. Technical Report LEMS 104, LEMS, Brown University.
- Sivashinsky, G. 1977. Nonlinear analysis of hydrodynamic instability in laminar flames-i. derivation of basic equations. *Acta Astronautica*, 4:1177–1206.
- Smith, J. 1981. Shape instabilities and pattern formation in solidification: A new method for numerical solution of the moving boundary problem. *Journal of Computational Physics*, 39:112–127.
- Smoller, J. 1993. *Shock Waves and Reaction-Diffusion Equations*. Springer-Verlag, New York.
- Solina, F. and Bajcsy, R. 1990. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:131–147.
- Ulupinar, F. and Nevatia, R. 1990. Recovering shape from contour for constant cross section generalized cylinders. In *CVPR'91 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, HI)*, Washington, DC., Computer Society Press, pp. 674–676.
- Vincent, L. 1991. Morphological transformations of binary images with arbitrary structuring elements. *Signal Processing*, 22(1):3–23.
- Wallace, T. and Wintz, P. 1980. An efficient three-dimensional aircraft recognition algorithm using normalized fourier descriptors. *Computer Vision, Graphics, and Image Processing*, 13:99–126.
- Waltz, D. 1975. *Understanding Line Drawings of Scenes with Shadows*. McGraw-Hill.
- Widder, D. 1975. *The Heat Equation*. Academic Press.
- Witkin, A.P. 1983. Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 1019–1022.
- Zahn, C. and Roskies, R. 1972. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, C-21:269–281.
- Zucker, S.W., Dobbins, A., and Iverson, L. 1989. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68–81.