Qualitative Physics and the Shapes of Objects

Eric Saund
Department of Brain and Cognitive Sciences and
the Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

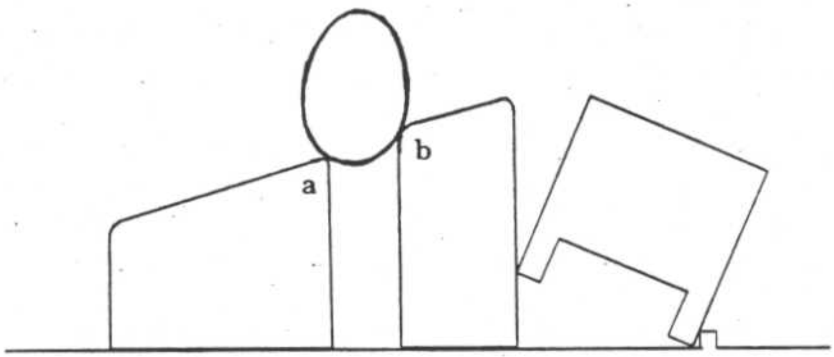

Figure 1. A qualitative physics problem.

In figure 1, does the egg break? Reasoning about situations such as this requires the manipulation of information about the geometry and spatial relationships among the objects involved—it requires a representation for describing shape. In this abstract I briefly discuss some weaknesses of current approaches to shape representation, as exemplified by shape representations of the generalized cylinder family, and I sketch what a shape representation better suited to supporting qualitative reasoning tasks might look like. The main point is that current shape representations fail because they are *knowledge-sparse*; what is needed is an approach allowing a representation to be *knowledge-rich*, in that the representation should have a large vocabulary for making explicit a wide range of spatial properties and events.

1 Knowledge-Sparse Representation

The predominant approach to designing representations for the shapes of objects reasons that a shape representation should consist of a set of building blocks which can be pieced together to approximate an object's shape. The building blocks should be designed to achieve a good approximation as efficiently as possible, and to effect meaningful segmentations of an object's form. For example, shape representations of the generalized cylinder family[1] typically define a part primitive endowed with a few parameters such as spine length and curvature, taper, and width, plus some data structure for describing the spatial relationships among parts. The motivation is that objects' shapes can often be segmented into meaningful parts. The symbolic description of many objects' shapes in terms of generalized cylinders is certainly much less cumbersome to store and work with than the iconic, "pixel by pixel," description afforded by a graphic picture.

A representation facilitates computation by virtue of the information it makes explicit. The question we must put to a shape representation intended to support qualitative physics is, does it make shape information readily available that is needed for computing things like the proximities, forces, and motions of blocks and eggs on a table?

Unfortunately, generalized cylinder representations do not fare well here, in fact, they often obscure the information most important to reasoning about the shape-dependent interactions of objects. Generalized cylinder representations make explicit only a few parameters computed from an object's shape. These may make it easy to determine, for example, an object's axis of elongation, its maximum width, and which end is pointed, but other information can be much more difficult to compute. For example, under a generalized cylinder description of the shape object shown in figure 2a, it can become extremely cumbersome even to locate the two most distant points on the object's contour. Note that what is important is *not* that the information is not lost (for the generalized cylinder representation may indeed be information-preserving), but that under this representation certain information may require an unduly complex computation to recover. Figures 2b through 2d offer additional illustrations of important spatial relationships and geometrical properties that are obscured by generalized cylinder representations. Among these, note that generalized cylinders tend to handle scale poorly; small features, if they are captured at all, are typically intended to be regarded as mere detail, and are relegated to obscure levels in a shape hierarchy, no matter what their functional or definitional significance to an object's shape.

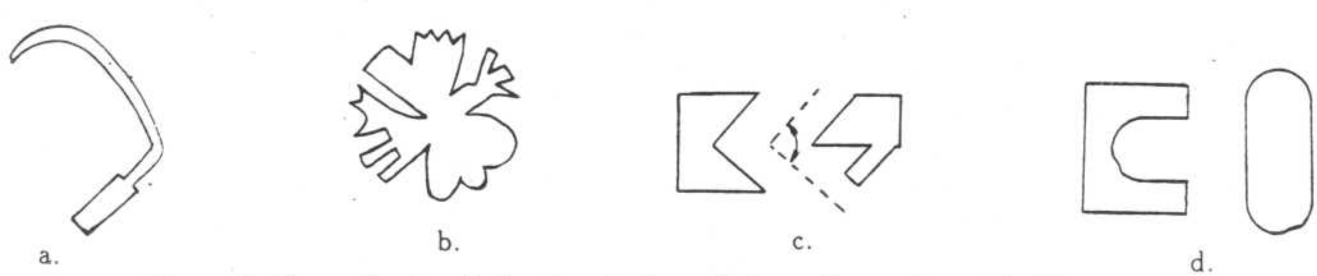

Figure 2. Generalized cylinder descriptions of these shapes do not facilitate the establishment of important spatial properties. a. Locate the two most distant points on the contour. b. At a coarse scale, this is a round object. c. Two subparts in the rightmost object create a 90° angle. d. Which end of the lozenge shape better fits into the slot?

[1]In referring to the family of generalized cylinder representations, I intend to include all parameterized part based representations, including superquadrics, volumetric primitives, Smoothed Local Symmetries, and so forth, as well as generalized cylinders proper.

The root of these difficulties lies not with any choice of volumetric primitive or part joining function, but with the assumption that a simple, universal vocabulary can be created to describe all shapes, in all situations, for all computational purposes. This is to say, generalized cylinder representations fail because they are *knowledge-sparse*. Because they supply only an impoverished vocabulary of shape descriptors, generalized cylinders provide no means for explicitly naming and describing specialized classes of shapes, configurations of features, and characteristic spatial relationships that may be significant to particular object domains, physical situations, or shape dependent task goals. The existence of a rich descriptive vocabulary for describing such information would constitute a body of *knowledge* about shape; what is needed is *knowledge-rich* shape representation offering a vocabulary of descriptors that can make explicit precisely the aspect of shape that is important to the situation at hand.

2 Knowledge-Rich Shape Representation

What would a knowledge-rich shape representation look like? First, in contrast to building block representations, it should be extensible. One should be able to define new vocabulary elements specialized for new shape domains and classes of physical situations. Furthermore, these vocabulary elements should provide parameters making explicit just that information pertinent to the shape's interaction with other shapes. For example, if a representation is to support qualitative reasoning in a blocks world in which bumps and notches may appear (see figure 3), then the various sorts of bumps and notches should have their own symbolic descriptors, and these descriptors should make explicit parameters that only have meaning with regard to bumps and notches, such as width, height or depth, wall angle, and "squaredoffness." Specialized bump descriptors and notch descriptors can simplify the computation of, for example, whether a particular bump is likely or not to catch and hold a block, because the determining factors of the bump's shape will be readily available as part of the bump's symbolic description.

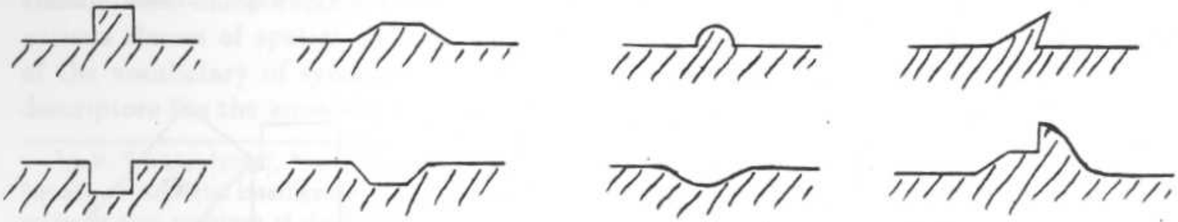

Figure 3. Important parameters of bumps and notches include height or depth, width, side angles, and squareness/roundedness. The computation of these properties becomes more difficult as bumps and notches become more complicated in shape.

Another important characteristic of knowledge-rich shape representation is that a shape should be describable at many levels of abstraction, and from alternative perspectives. Computations should be able to consider the egg shape in figure 1 as a whole, or in terms of various component aspects of its structure: the egg defines a region with a certain area; it defines a smooth, rounded contour with a certain local radius; the regions of contour where the egg is supported are oriented at certain angles with respect to the vertical; and so forth. For the purposes of determining the approximate weight of the egg, only the first of these properties is important; to determine whether the egg will roll, the second is important; and to evaluate the egg as a wedge, the third property is needed. Only a subset of the many potentially useful properties may be computed and used in the course of any particular task, but the apparatus should be at hand to establish alternative perspectives that may be useful for different purposes. A representation capable of making explicit various sorts of such information is in marked contrast to generalized cylinder representations which are intended to define one and only one "canonical" description for any object shape.

A shape representation serves to manage the volume of information that reasoning processes must deal with by grouping, chunking, segmenting, and categorizing raw data occurring in an image-like, pixel array. It is important that the categories provided by the representation reflect meaningful functional implications and relationships in the domain. Thus, in figure 4, object **A** might profitably be classified as a "leaning object" for the purposes of reasoning about the forces it exerts on other objects, its motion were the leaning support removed, etc. The mapping between configurations of matter in a scene and the chunks and categories into which they might be meaningfully organized can be quite complex. Object **B** is a leaning object although it's shape is manifestly different from that of object **A**, while object **C** is not a leaning object (all forces are vertical); a crucial factor apparently hinges upon the spatial configurations of an object's points of support, which can themselves be classified as "supported from below," "supported by a (near) vertical wall (at angle α)," and so forth. The computation of a shape description in a knowledge-rich shape representation may entail a great deal of effort as successively more abstract aspects of a shape scene are established. This process, of computing a description of shape in terms of a vocabulary suited to qualitative reasoning, becomes undifferentiatable from *shape recognition* as it is known in the field of computer vision. Configurations of primitives must be segmented, chunked, and analyzed in terms of abstract (perhaps parameterized) equivalence classes.

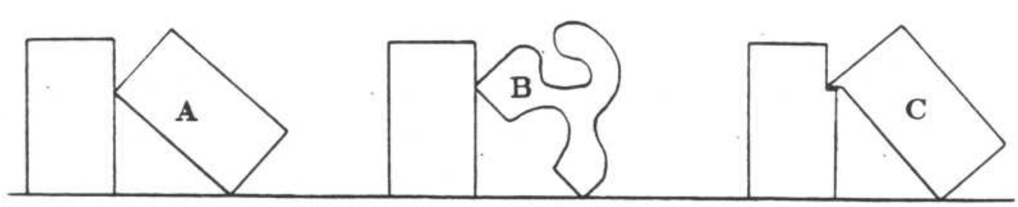

Figure 4. The spatial relationships determining a shape's membership in useful equivalence classes can be subtle. **A** and **B** may be classified as "leaning objects" despite the gross differences in their shapes. The leftmost corner of object **C**, however, is supported from below, so this object plays a qualitatively different role with respect to its partner block.

Finally, a shape representation should preserve global spatial relationships at all levels of abstraction. For example, the proximity of corners **a** and **b** in figure 1 should be easily noticeable despite the fact that they lie on different objects. Global spatial relationships are preserved in an iconic, array-like, data structure, while they are easily lost in symbolic data structures such as semantic networks. One important goal is to achieve chunking and compression of the volumous pixel by pixel data accomplished by symbolic encoding of shape information, yet to retain access to global spatial relationships. In addition, a shape representation should handle scale gracefully by segregating events occurring at different spatial resolutions, while not favoring any one scale over others, and by expressing not only the relative locations and orientations of shape features in space, but their relative sizes as well.

3 Tokens in Scale-Space

I suggest that a knowledge-rich approach to shape representation employ a semi-iconic scale-space[2] data structure. Scale-space is an augmentation of a (say) two-dimensional space by the addition of another, scale, dimension (see figure 5). The placement of a symbolic token in scale-space makes explicit a spatial location, and a scale. A token placed at a coarse scale indicates a large geometric feature or shape event, and a token placed at a fine scale denotes a small one. By "semi-iconic" data structure is meant that symbolic tokens may represent geometrical occurrences of considerable extent in the original data array, nonetheless, by their placement in scale-space, their locations and sizes are denoted explicitly in such a way that the spatial relationships and scale relationships among such tokens are maintained.

One scenario for computation under this approach to shape representation involves the incremental construction of a shape description by the placement of attributed tokens in a scale-space "blackboard." For example, pattern matching processes might examine the contents of the blackboard, and trigger on recognizable constellations of tokens as they appear. A triggered process then writes a new token onto the blackboard, serving to group, classify, and name explicitly the configuration of more primitive tokens it found. The various classes of spatial configurations that may be identified and named is a function of the vocabulary of symbolic descriptors available. Thus, in the open-ended library of descriptors lies the knowledge of a shape representation.

[2] A.P. Witkin, [1983], "Scale-Space Filtering," in the Proceedings of the Eighth International Joint Conference on Artificial Intelligence. Witkin's original presentation of scale-space dealt with the evolution across scales of zero crossings of the DOG convolution with the original signal. Of interest here is just the general notion of a separate scale dimension to represent size.

4 Conclusion

Knowledge-sparse shape representations, such as representations of the generalized cylinder family, do not adequately support qualitative reasoning about the interactions of physical objects because they do not make explicit, and they often obscure, the functionally significant geometrical properties and spatial relationships involved. What is needed is a knowledge-rich approach, in which a large vocabulary of descriptors can be made available for naming spatial configurations characteristic to particular shape domains, physical situations, or computational task goals. A scale-space data structure, in which symbolic tokens can be placed at appropriate spatial locations and scales, may provide a suitable framework for building such a shape representation.

A great many other issues remain to be addressed in the consideration of an approach to shape representation directed toward the support of qualitative reasoning processes. Among these are: 1. the representation must provide hooks for representing additional, not purely spatial information, including mass and density, friction, forces and torques, etc.; 2. means must be provided for directing attention to appropriate aspects of a shape scene, that is, for deciding what descriptors to compute; 3. to what extent must descriptive classes and properties be drawn from a pre-existing library, or, may they be generated on the fly? As these and other issues are faced it will become increasingly clear that the ability of a shape representation to make explicit the important spatial and geometrical information carries crucial significance.

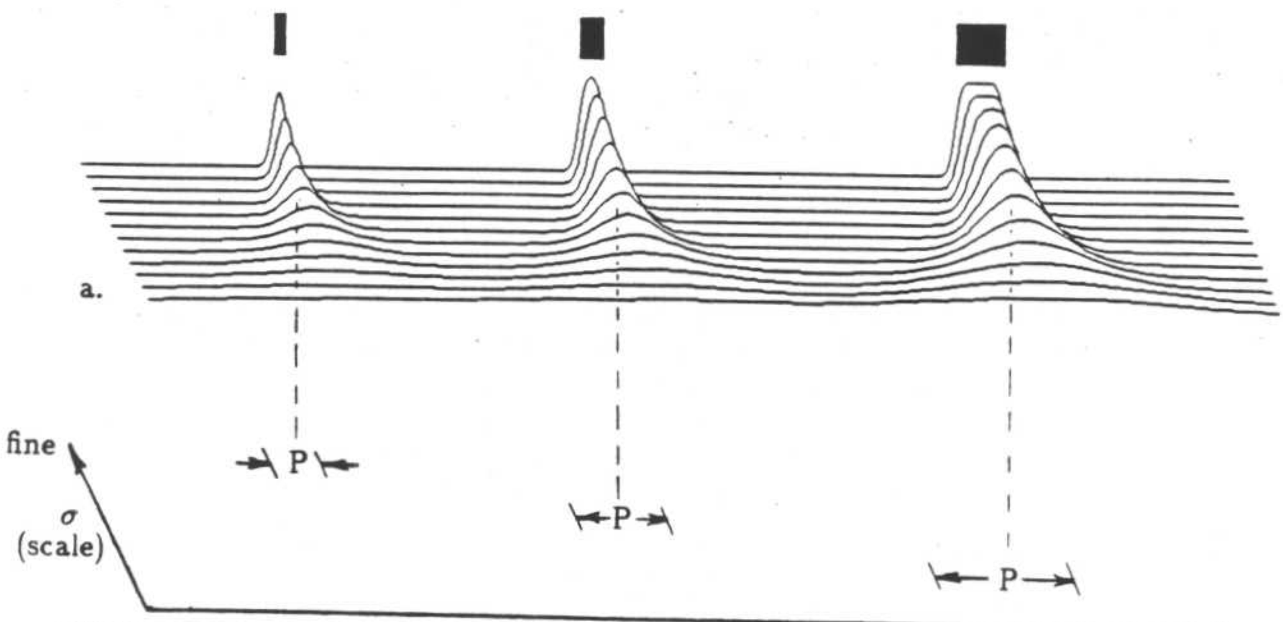

Figure 5 Scale-space. a. Scale-space is created by smoothing a signal (in this case binary pulses) with filters of different sizes: b. "P" tokens denote pulse features in scale-space. The size of a feature is indicated by its placement along the scale (σ) dimension.