On the Psychophysics of the Shape Triangle

Kaleem Siddiqi† **Benjamin B. Kimia**§ **Allen Tannenbaum**‡
Steven W. Zucker†

†Center for Computational Vision & Control
Department of Computer Science
Yale University, New Haven, CT 06520-8285
{siddiqi-kaleem,zucker-steven}@cs.yale.edu

§Laboratory for Engineering Man/Machine Systems
Division of Engineering
Brown University, Providence, RI 02912
kimia@lems.brown.edu

‡Department of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455
tannenba@ee.umn.edu

CVC TR-98-002/CS TR-1150

February, 1998

Abstract

We earlier introduced an approach to categorical shape description based on the singularities (shocks) of curve evolution equations. We now consider the simplest compositions of shocks, and show that they lead to three classes of parametrically ordered shape sequences, organized along the sides of a shape triangle. By conducting several psychophysical experiments we demonstrate that shock-based descriptions are predictive of performance in shape perception. Most significantly, the experiments reveal a fundamental difference between perceptual effects dominated by *when* shocks form with respect to one another, versus those dominated by *where* they form. The shock-based theory provides a foundation for unifying tasks as diverse as shape bisection, recognition, and categorization.

1 Introduction

Whether one views a colleague at a distance, a Dürer woodcut in a museum, or a *Koren* cartoon in a magazine, the human form is immediately recognizable despite the immense differences in photometric and geometric detail. This exemplifies our spectacular ability to infer the generic structure of object categories, and also to place specific instances within them. Such an ability supports a consistent interpretation across deformations in the retinal image, whether they arise from object growth, changes in viewpoint, or changes in illumination. It also structures the organization of knowledge about objects in our world into abstraction hierarchies, as is required for efficient computational access to memory. As observed by Rosch (Rosch, 1978):

> ...the world consists of a virtually infinite number of discriminably different stimuli. Since no organism can cope with infinite diversity, one of the most basic functions of all organisms is the cutting up of the environment into classifications by which nonidentical stimuli can be treated as equivalent...

How the human visual system accomplishes this task for object recognition remains an area of active research in psychology and neurophysiology, e.g., see (Logothetis and Sheinberg, 1996) for a recent review. Several questions are immediately raised: Are the internal classifications organized around solid, volumetric models, or are they organized around arrangements of features? If the former, what are the models and their invariances; how do they compose? If the latter, what are the features, and which arrangements are legitimate? What is the effect of viewpoint, and how do boundary and interior effects associate?

There are many different proposals on either side of these questions. One group suggests that objects are represented in a viewpoint independent fashion by a collection of volumetric parts derived from non-accidental properties of the retinal image (Biederman, 1987; Binford, 1971; Marr and Nishihara, 1978). Such approaches offer the advantage of computational efficiency, since a single model is stored, and can explain human performance in a variety of basic level recognition tasks. The classification into "basic level" derives from studies in cognitive psychology which indicate that, for a variety of tasks, there is a level of abstraction which is most significant with respect to several measures, such as ease of access (Rosch et al., 1976), and below which only fine distinctions are made. For example, in the abstraction hierarchy "furniture, chair, arm-chair", the classifications chair and arm-chair are at basic and subordinate levels, respectively. However, to date no complete theory exists for obtaining such volumetric descriptions from retinal images (except in highly constrained domains) and many computational issues remain to be addressed (Dickinson et al., 1997). Just laying out the mathematical form of such a theory is a significant challenge.

At the other extreme are theories motivated by viewpoint dependent performance, particularly in subordinate level recognition tasks (Tarr and Pinker, 1989; Edelman and Bulthoff, 1992; Bulthoff and Edelman, 1992; Logothetis et al., 1994). Such theories posit that objects are represented by a modest collection of 2D views, and that recognition is mediated by an interpolation between perspectives closest to the observed view. A typical computational account of this process provides a description of each stored view as a vector of image features; the vector of observed image features is then compared with the nearest stored views using a regularization network (Poggio and Edelman, 1990). Such a model is consistent with psychophysical data obtained for certain classes of unfamiliar objects (Bulthoff et al., 1995), but its extension to more complex domains is non-trivial. In particular: 1) how should relevant image features be determined?, 2) how should they be organized to allow for an alignment of an observation vector with vectors representing stored views?, 3) how should a small set of informative views be obtained? Once again, just laying out the mathematical form of a theory based on features is a serious challenge.

The debate between proponents of viewpoint invariant and viewpoint dependent representations remains an active one, e.g., see (Biederman and Gerhardstein, 1993; Biederman and Gerhardstein, 1995; Hummel and Stankiewicz, 1996; Kurbat, 1994; Humphrey and Khan, 1992; Tarr and Bulthoff, 1995; Tarr, 1995; Liu, 1996), at the heart of which are several subtle issues. For example, whereas viewpoint dependent performance has a default association with viewpoint dependent representations (Edelman and Bulthoff, 1992; Bulthoff and Edelman, 1992), the data could also be a reflection of the (viewpoint dependent) computational cost of obtaining a viewpoint invariant representation. Furthermore, it is possible that both types of representations co-exist, but for orthogonal purposes, i.e., for recognition tasks at different levels of abstraction. In particular, the experiments of (Logothetis et al.,

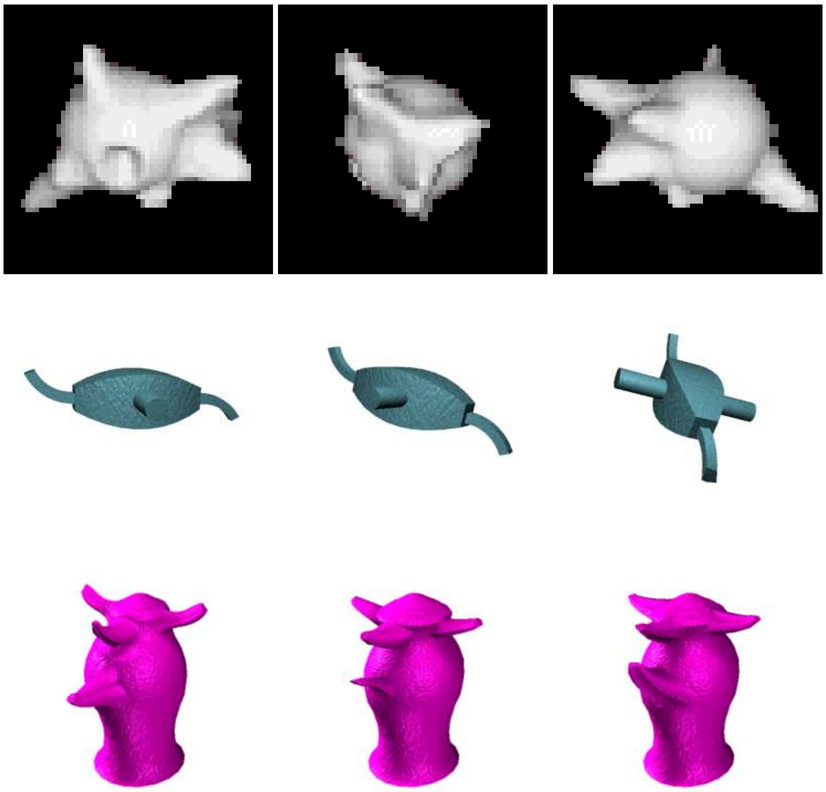


Figure 1: Representative sample stimuli upon which various psychophysical studies on object recognition have been based. Top Row: "Amoebas" used in (Bulthoff and Edelman, 1992; Logothetis et al., 1994). Middle Row: Multi-part objects with geon-like components used in (Hayward, 1998). Bottom Row: "Greebles" used in (Gauthier and Tarr, 1997).

1994, p. 405) and others have found subordinate level recognition to be viewpoint dependent but basic level recognition to be relatively robust to changes in viewpoint.

To motivate the need for both viewpoint dependent and viewpoint invariant properties, consider the stimuli in Figure 1, which are representative of those used in several psychophysical studies. Clearly, each row consists of a different generic class of objects, and the columns illustrate changes in viewpoint, or alterations of parts. Given the tremendous variation across different stimulus sets, how can the body of performance data be compared? Existing volumetric theories for basic level recognition, e.g. those based on *geons* (Biederman, 1987), are applicable to some of these images but not others, while feature-based approaches achieve generality only by using local structure that is restricted to a specific image class (following normalization). It is instructive, therefore, to look at those properties that might play a role in both basic and subordinate level classification. One necessary condition is clear: The answer must (at least) take into account the silhouettes obtained from the bounding

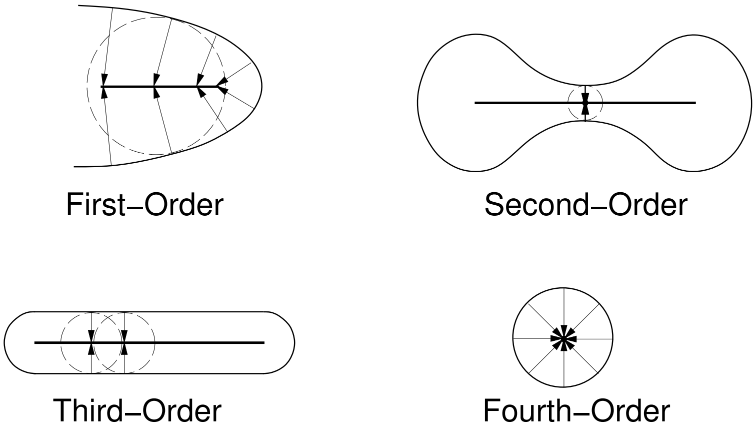


Figure 2: A coloring of shocks into four types (Kimia et al., 1995). A 1-shock derives from a *protrusion*, and traces out a curve segment of adjacent 1-shocks. A 2-shock arises at a *neck*, and is immediately followed by two 1-shocks flowing away from it in opposite directions. 3-shocks correspond to an annihilation into curve segment due to a *bend*, and a 4-shock an annihilation into a point or a *seed*. The loci of these shocks gives Blum's medial axis.

contours of the objects, which are an integral part of their descriptions. Indeed, the recent investigation by (Hayward, 1998) confirms that performance in a variety of recognition tasks is predicted by changes to the outline shape, seen under projection. This is consistent with findings that, in the human visual system, the degree of generalization to novel views when three-dimensional objects undergo deformations is primarily dependent on the amount of deformation in the image-plane (Sklar et al., 1993).

We focus on this necessary part of the description, and begin by considering 2D shapes, and deformations of their boundaries. In this domain it is possible to develop a mathematical theory that provides a formal answer to several of the above questions, and which leads to parametric classes of stimuli against which shape perception can be analyzed. The key point about outlines is that they change only slightly if a shape is deformed only slightly. This is clear unless a catastrophic change occurs, such as the nose appearing on the profile of a face. Such changes are singular events, and they organize different shapes into equivalence classes. Within each class is a generic category, and class transitions are signaled by the singularity. We submit that for shape analysis, such equivalence classes are, in mathematical terms, what Rosch was seeking. The feature-based geometric detail, so important for subordinate level recognition, is the quantitative information within each equivalence class. In this paper we consider the perceptual consequences of both qualitative and quantitative changes.

The fundamentals of our theory are reviewed in Appendix A. A structural description is derived from the shocks (singularities) of a curve evolution process acting on the bounding

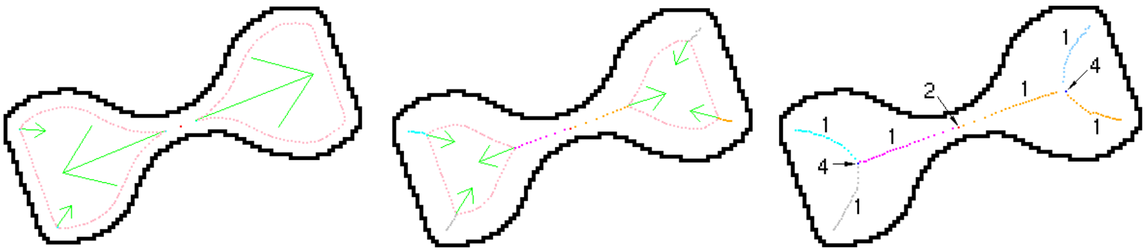


Figure 3: The detection of shocks for a dumbbell shape undergoing constant inward motion. Each sub-figure is a snapshot of the evolution in time, with the outline of the original shape shown in black, the evolved curve overlaid within, and the arrows representing velocity vectors for the current 1-shocks. Note the structural description of the shape as two parts connected at a neck, with each part described by three protrusions merging onto a seed. Adapted from (Siddiqi and Kimia, 1996).

contour of an object (Kimia et al., 1995). Four types of shocks arise, each of which has a direct perceptual correlate, as illustrated in Figure 2. Specifically, a connected segment of 1-shocks corresponds to a *protrusion*, a 2-shock corresponds to the partitioning of a shape at a *neck*, a 3-shock corresponds to an extended region with parallel sides or a *bend*, and a 4-shock corresponds to a local center of mass or a *seed*. To illustrate the formation of shocks, consider the numerical simulation of a dumbbell shape, evolving under constant inward motion, shown in Figure 3. Note the emergence of a qualitative description of the shape as that of two parts separated at a neck (2-shock), with each part consisting of three protrusions (1-shock groups) merging onto a seed (4-shock). An abstraction of this description into a graph of shock groups provides a mathematical framework in which to investigate shape recognition at different levels of abstraction (Siddiqi et al., 1998; Pelillo et al., 1998).

Returning to our earlier discussion, it is worth noting that our theory can be applied to the (2D) outlines of *all* of the stimuli in Figure 1. The shock-based descriptions are shown in Figure 4, with the associated shock graphs in Figure 5. The representation in Figure 4 illustrates the spatial configuration of shocks: where they form and the positions through which they migrate. The abstraction in Figure 5 illustrates the (reverse) order of their formation in time. Views which are qualitatively similar, e.g., Figure 1 (second row, left and middle), produce qualitatively similar shock graphs, Figure 5 (second row, left and middle). Observe that the same property holds for all the "greeble" objects, despite variations in the shapes of the individual parts. On the other hand, the emergence of new parts and thus a qualitatively different view, e.g., Figure 1 (second row, right), results in drastic modifications to the underlying shock graph, Figure 5 (second row, right). Thus, the shock graph encodes the spatial relationship between shock groups and may be viewed as a 2D analog of Biederman's geon-structural descriptions. It provides the requisite hierarchical structure

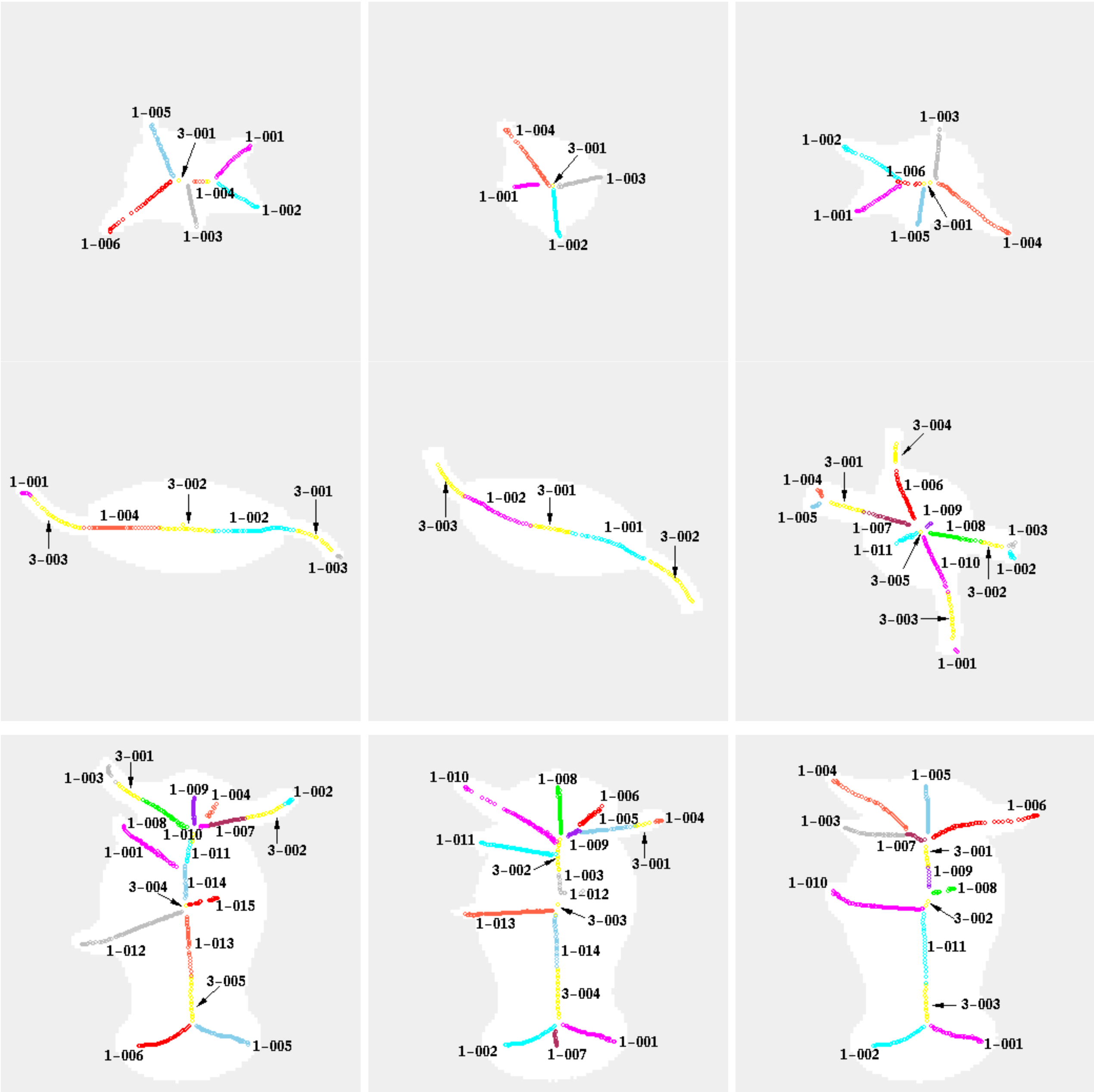


Figure 4: The shock-based descriptions of the profiles of stimuli in Figure 1: "amoebas" (top row); multi-part geons (middle row); "greebles" (bottom row). The representation illustrates where shocks form, and the locus of positions through which they migrate. The notation associated with each shock group is of the form shock_type-identifier; see Appendix A for a discussion of shock types. The corresponding shock graphs are shown in Figure 5.

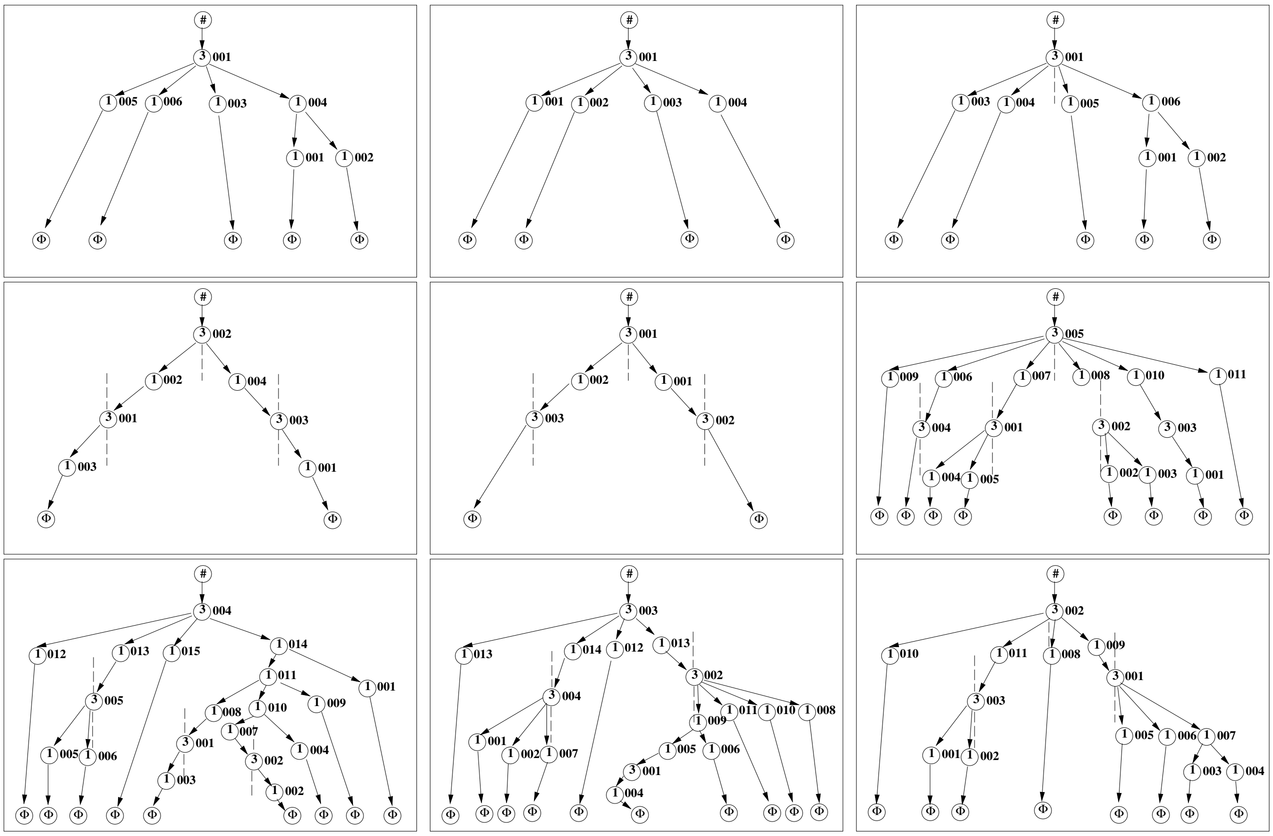


Figure 5: The shock graphs (Siddiqi et al., 1998) of the shock-based descriptions in Figure 4: "amoebas" (top row); multi-part geons (middle row); "greebles" (bottom row). Each distinct shock group comprises a different node. The nodes are ordered in reverse time, with edges indicating parent-child relationships, such that the last shock groups to form are highest in the hierarchy. Although matching is not discussed in this paper, the similarity in graph structure for similar views along each row (see Figure 1) is obvious.

for basic level recognition when objects are seen from similar views; qualitatively different views are signaled by changes to its topology. For finer shape discrimination tasks, such as those at subordinate levels, viewpoint dependent variations in the geometric properties of the underlying shocks (e.g. their positions and formation times) will also play a significant role. The key advantage of working with 2D shapes and their shock-based representations is that, in contrast to the two classes of theories discussed earlier, a precise mathematical framework has been developed which provides predictions against which psychophysical data can be tested.

To develop this link between our computational model and the perception of 2D shapes, we begin by noting that whereas the interpretation of each shock type in isolation is clear, Figure 2, the morphogenesis of even simple shapes can be ambiguous when different shocks

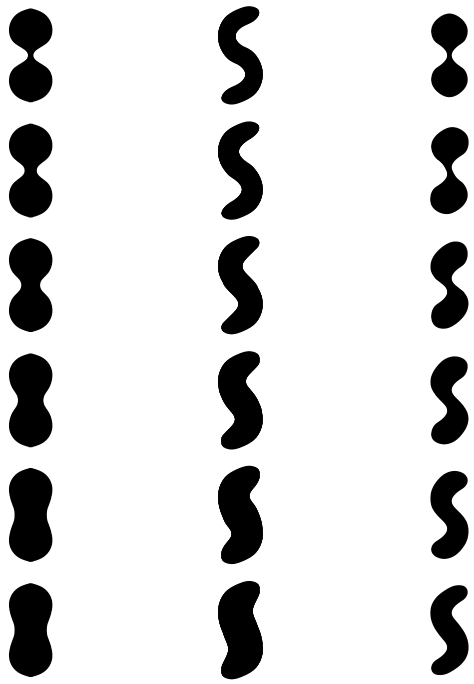


Figure 6: Three families of shapes that illustrate both the categorical nature of shape descriptions, and the delicate crossover between categories. Notice, e.g., how the upper left shape is perceived as two parts connected by a thin neck, while the lower left shape appears to have two indentations (or four protrusions). The middle column starts as a bend (top) but ends with indentations (bottom). The right column starts as two parts (top) but ends as a single bend (bottom). At which point along each progression does the categorical shift occur? The left and middle sequences we created by adding material on each side, the sequence on the right was created by shearing the left boundary with respect to the right one.

participate in their description. To illustrate, are shapes in the left and right columns of Figure 6 composed of one part or two? Do shapes in the middle column arise by bending an extended region along its central axis, or by carving out undulations on the boundary of a rectangle? The answer to such questions must consider the role of structural context, since the shocks arise from a global evolution. Stated differently, categories for shape are not rigid, but rather are subject to shifts under the action of deformations to the bounding contour. When such a deformation alters the topological shock structure sufficiently, new perceptual categories can arise.

To order the possibilities, we arrange the earlier sequences along the sides of a *shape triangle* (Kimia, 1990). Informally, the idea is that the interpretation of each shape lies on a continuum between distinct categories represented by the *parts*, *protrusions* and *bends* nodes, Figure 7. For example, along the parts-bends axis the percept changes from that of

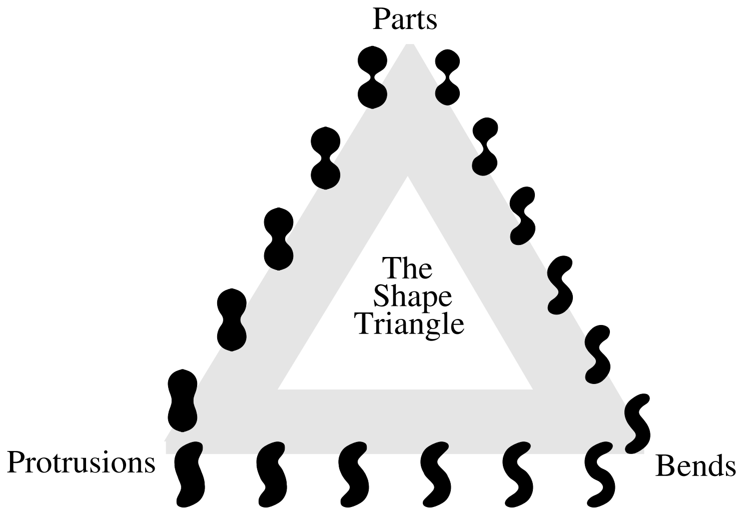


Figure 7: The sides of the shape triangle represent continua of shapes; the extremes correspond to the "parts", "protrusions" and "bends" nodes (Kimia, 1990).

a dumbbell shape with two distinct parts to that of a worm shape with a single part. The sides of the triangle reflect these continua and capture the tension between object composition (parts), boundary deformation (protrusions) and region deformation (bends). Most importantly, we know from the theory that, for such simple shapes, the shocks in Figure 2 are the only possible types. In analogy to Leyton's process grammar for shape (Leyton, 1988), the composition of a more complex shape from these shock types is characterized by a small number of rewrite rules (Siddiqi et al., 1998).

The stimuli provide a class of parametrically ordered shapes which can be used for psychophysical studies: the *parts-protrusions* and *bends-protrusions* sequences were created by adding material to each side; the *parts-bends* sequence was created by shearing the left boundary with respect to the right one. Thus, while the shapes have similar shock structures, the shock formation times and positions are gradually varied. Nevertheless, it is possible to interpret each sequence according to whether shock formation times are perceptually dominant over shock positions, or vice versa.

First, the shock formation times, or equivalently their distance to the boundary of the shape, are related to their classification into distinct types (see Appendix A). Our prediction is that non-uniform alterations to shock formation times, which result in qualitative changes to the shock structure, are dominant over changes to shock locations. This phenomena occurs along both the *parts-protrusions* and *parts-bends* axes. We hypothesize that an encoding of this non-uniformity is at the heart of an observer's ability to discriminate between any two shapes from within each sequence. Specifically,

Hypothesis 1 *A human observer's ability to discriminate between shapes lying on the parts-protrusions or parts-bends axes is primarily determined by their minimum (local) width to maximum (local) width ratios.*

Regions of minimum local width (necks) correspond to 2-shocks and regions of maximal local width (seeds) to 4-shocks, see Figure 2.

Second, the shock positions provide quantitative geometric information, which is orthogonal to their classification into types. We expect that such geometric properties are dominant only when the qualitative shock structure is preserved, as is the case along the *bends-protrusions* axis where there is a uniform addition of mass. Observe that these shapes closely resemble wiggles (Burbeck et al., 1996). For thin objects placed close to the *bends* node the central axis is seen to wiggle, for thicker objects placed close to the *protrusions* node the central axis is perceived to be straight. Our second prediction relates such effects to the positions of high-order shocks which arise, Figure 26. Specifically,

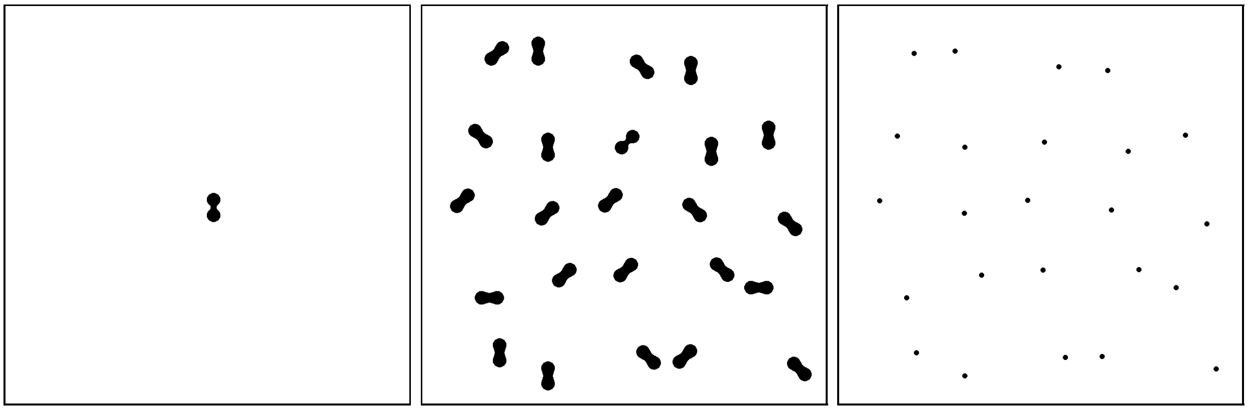
Hypothesis 2 *For a "wiggle" shape taken from the bends-protrusions axis, the perceived center along a horizontal line in alignment with a sinusoidal peak coincides with a high-order shock.*

These will be type 3 shocks if the opposing boundaries are exactly parallel, and type 4 shocks otherwise, see Appendix A.

Thus, considerations of *when* shocks form versus *where* they form comprise the two main lines of investigation in this paper. In Section 2 we use a visual search paradigm to test the first hypothesis, which is related to "when"-dominated perceptual effects. In Section 3 we review the psychophysical experiments of (Burbeck et al., 1996), in which subjects were required to bisect wiggle like stimuli. We demonstrate that the results are consistent with our second hypothesis, which is related to "where"-dominated effects. This is followed by a general discussion in Section 4.

2 The Effects of "When" Shocks Form

We used a visual search paradigm to test our first hypothesis that the role of context, i.e., the relative formation times of different shocks, is the dominant factor affecting shape discrimination along the *parts-protrusions* and *parts-bends* axes. The speed with which a subject located a target shape, randomly placed in a field comprised of repeated copies of a distractor shape, was recorded. The results were then compared against differences between the relative shock formation times (specifically the minimum/maximum width ratios) of



(a) (b) (c)

Figure 8: A visual search sequence. First an example of the target is shown in the center of the display (a). In each trial a display containing exactly one target embedded in a field composed of several copies of a distractor element is presented (b). The subject is asked to press a mouse button upon finding the target, and then identify its location in a validation display composed of small dots (c).

the target and distractor elements. We first describe the experiments and then provide a discussion of the results.

2.1 Methods

Our experimental procedure closely follows that introduced in (Elder and Zucker, 1993), with only slight modifications. Visual search displays were created on a 60 Hz color SUN monitor, driven by a SPARC 10 computer. Subjects sat in a dimly lit room, 2 meters from the screen. A $7^\circ \times 7^\circ$ square display window of luminance 2.3 cd/m^2 was positioned in the center of the screen against a background luminance of 0.9 cd/m^2. Stimuli were drawn in the display window with a luminance of 36.7 cd/m^2.

The stimuli were approximately $0.5^\circ \times 0.5^\circ$ in size. Their placement in the display window was based on a regular 5×5 grid, with nodes spaced 1.4° apart in the vertical and horizontal directions. A node was selected using a psuedorandom number generator and the precise location of each stimulus was then chosen pseudorandomly from the set of positions within a 0.3° horizontal and vertical distance from the selected node. Each stimulus appeared oriented at 0, 45, 90 or 135 degrees, with equal probability.

Displays contained either 7, 15 or 23 distractor stimuli and one target stimulus. First an example of the target is shown in the center of the display (Figure 8(a)). The subject then presses a mouse button to trigger a sequence of 30 visual search trials (10 for each display size, randomly interleaved). In each trial a display is presented which always contains exactly one target (Figure 8(b)). The subject presses a mouse button when the target is detected. The response time for detection is recorded and the visual search display is immediately replaced with a validation display, where the stimulus positions are represented by small reference dots (Figure 8(c)). The subject must identify the target location by moving the mouse to and clicking on the appropriate dot. If an error is made, the trial is considered invalid and another trial with the same display size and stimulus type is randomly inserted into the sequence as a replacement.

Before each session, subjects completed a practice sequence identical to the recorded one, but including only four trials for each display condition. Note that the method differs from traditional approaches, in which only half of the displays actually contain a target, and one of two mouse buttons is pressed depending upon whether the subject perceives the target to be present or absent (Triesman and Gelade, 1980). This modification avoids a systematic bias that the traditional method suffers from, and typically gives lower error rates (Elder and Zucker, 1993).

2.2 Subjects

Three subjects participated in each of the main experiments in this study (all male). There was sufficient consistency in results so that only one subject was used for the control experiments. Each subject participated voluntarily and reported normal or corrected vision. One subject was completely aware of the goals of the study; the other two had only limited awareness. Results are averaged over all participating subjects, with error bars indicating standard deviation from the mean.

2.3 Stimuli

Magnified versions of the stimuli were drawn at each of the four orientations (0, 45, 90 and 135 degrees) using spline functions of the IDRAW computer program. This allowed for the precise placement and manipulation of control points, and a minimization of discretization artifacts when the stimuli were digitized and scaled down to $0.5° \times 0.5°$ (70 x 70 pixels at a viewing distance of 2 meters). Details of each stimulus set are presented along with the experiments. Note that the dashed rectangles in Figures 9, 11 and 13 were not present in the psychophysical displays.

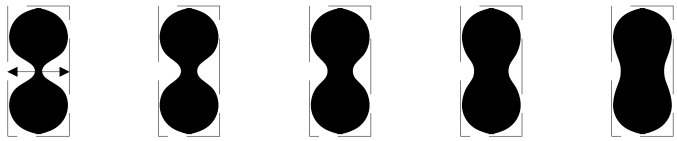


Figure 9: The stimuli used in the *parts-protrusions* visual search experiment were created by pulling apart the opposing boundaries of the bowtie shape (on the left) at its neck. Each stimulus had the same overall size, as indicated by the dashed rectangle. The dashed rectangle was not visible in the psychophysical display.

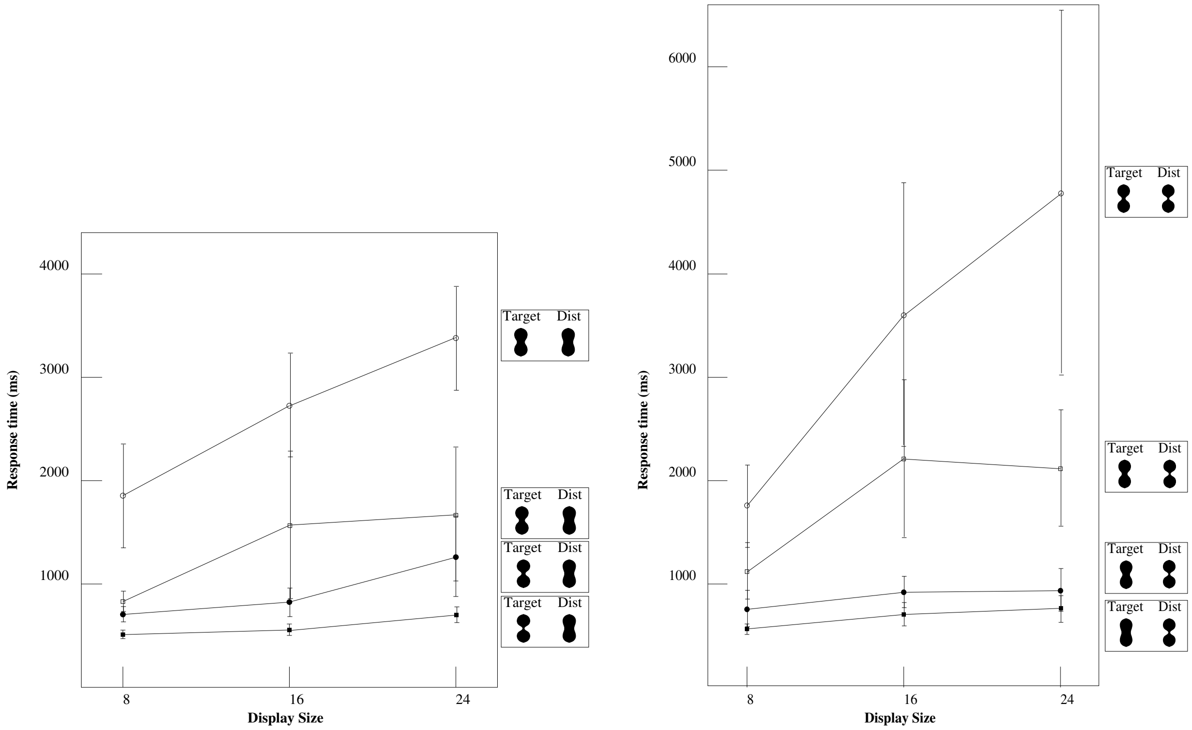


Figure 10: Results for the *parts-protrusions* visual search experiment.

2.4 Experiment I: Parts Versus Protrusions

Stimuli: The stimuli for the first experiment consisted of shapes along the *parts-protrusions* axis. The sequence was created by pulling apart the opposing boundaries of a bowtie shape in the region of the neck, Figure 9. Each stimulus had the same overall size, as indicated by the dashed rectangle in the figure, but not the same overall area.

For each session, the distractor stimulus was taken from one of the extremes, with the target stimulus being one of the four other shapes. Thus there were a total of 2×4 sessions for each subject (each session consisting of 30 trials in total). The results are summarized in

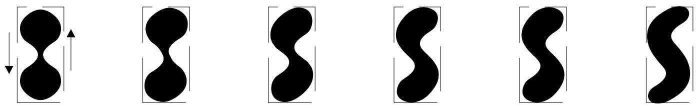


Figure 11: The stimuli used in the *parts-bends* visual search experiment were created by shearing the opposing boundaries of the bowtie shape (on the left) with respect to each other. Each stimulus had the same overall size, as indicated by the dashed rectangle, as well as the same total area. The dashed rectangle was not visible in the psychophysical display.

Figure 10. The search time decreases monotonically as the two stimuli are further separated along the *parts-protrusions* axis.

2.5 Experiment II: Parts Versus Bends

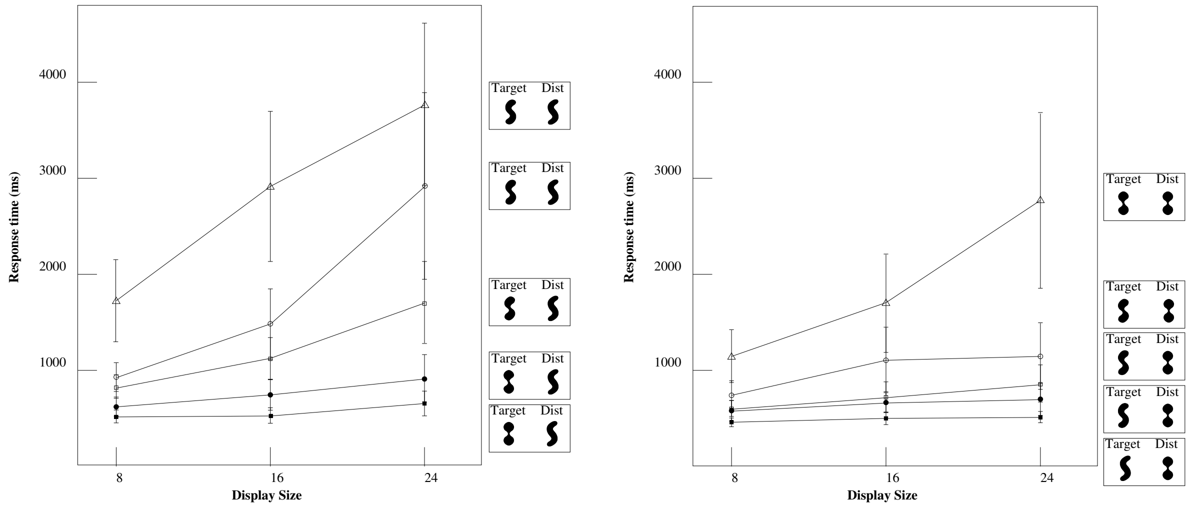


Figure 12: Results for the *parts-bends* visual search experiment.

Stimuli: The stimuli for the second experiment consisted of shapes along the *parts-bends* axis. The sequence was created by incrementally shearing one side of a bowtie shape with respect to the other, Figure 11. Each stimulus had the same overall size, as indicated by the dashed rectangle, as well as the same total area.

For each session, the distractor stimulus was taken from one of the extremes, with the target stimulus being one of the five other shapes. Thus there were a total of 2×5 sessions for each subject (each session consisting of 30 trials in total). The results are summarized in Figure 12. Once again, the search time is a monotonically decreasing function of the separation between the two stimuli along the *parts-bends* axis.

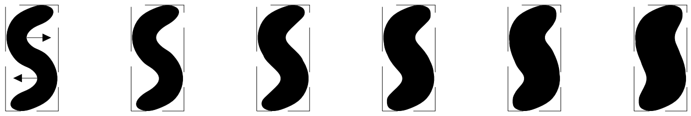


Figure 13: The stimuli used in the *bends-protrusions* visual search experiment were created by incrementally adding mass to the concave regions of the worm shape on the left. Each stimulus had the same overall size, as indicated by the dashed rectangle. The dashed rectangle was not visible in the psychophysical display.

2.6 Experiment III: Bends Versus Protrusions

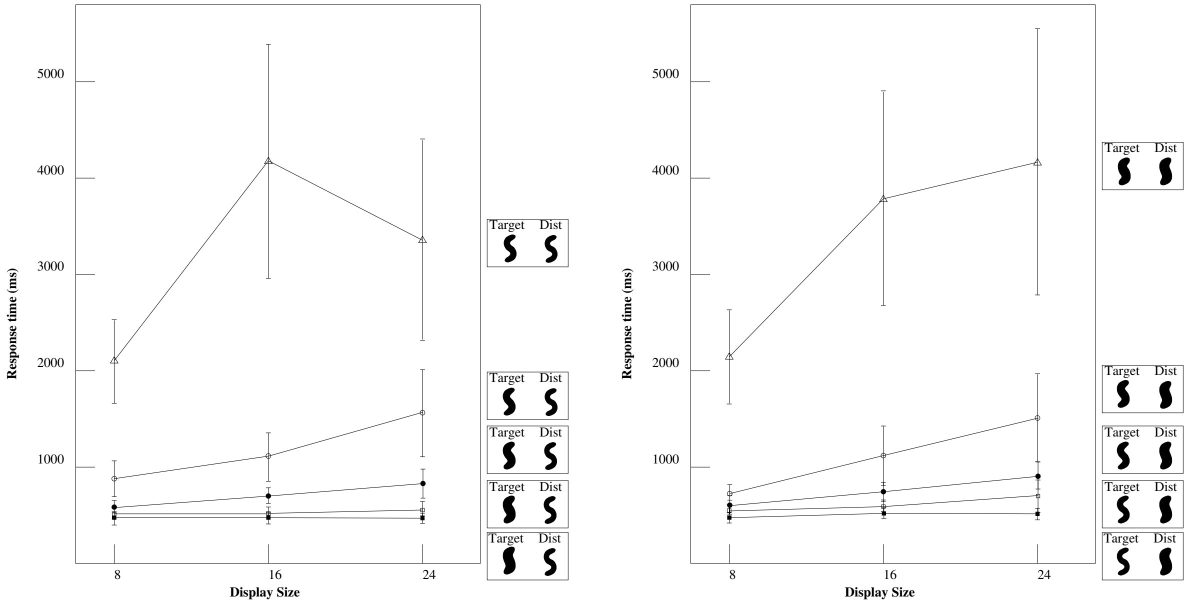


Figure 14: Results for the *bends-protrusions* visual search experiment.

Stimuli: The stimuli for the third experiment consisted of shapes along the *bends-protrusions* axis. The sequence was created by incrementally adding mass to the concave regions of the worm shape, Figure 13. Each stimulus had the same overall size, as indicated by the dashed rectangle in the figure, but not the same total area.

For each session, the distractor stimulus was taken from one of the extremes, with the target stimulus being one of the five other shapes. Thus there were a total of 2×5 sessions for each subject (each session consisting of 30 trials in total). The results are summarized in Figure 14. It is clear that the search time is a monotonic function of the separation between the two stimuli along the *bends-protrusions* axis.

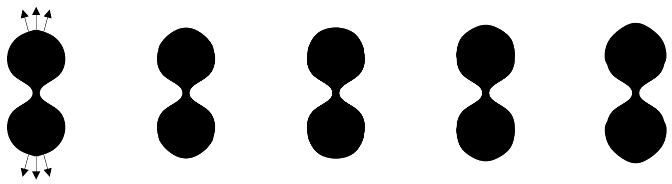


Figure 15: The control stimuli used in the *parts-protrusions* visual search experiment. The same amount of mass incrementally added to the neck region in Figure 9 is now added to the top and bottom of the bowtie shape.

2.7 Control Experiments

We note that care was taken to preserve local properties of the stimuli, such as the number of curvature extrema (Hoffman and Richards, 1985), as well as global ones, such as overall size. However, total area was preserved only in Experiment II, since mass was added to create the sequences in Figures 9 and 13. Therefore, it is possible that discrimination in Experiments I and III was based on total area. To test for this we ran the following two control experiments.

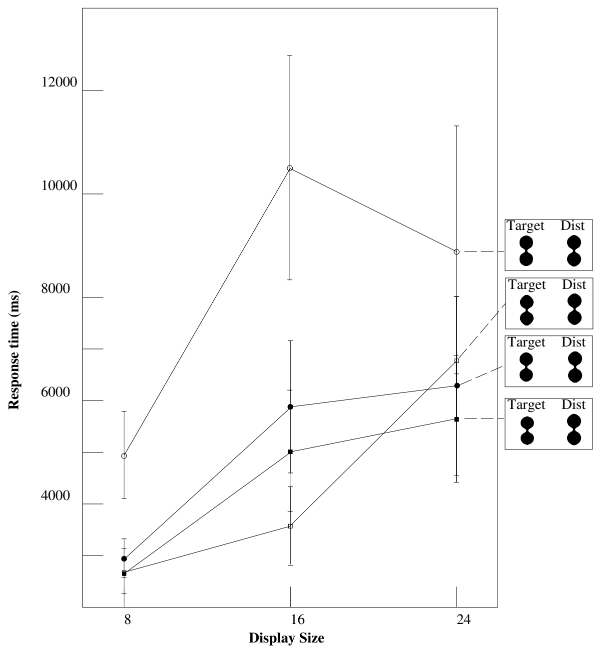


Figure 16: Results for the *parts-protrusions* control experiment. The significantly larger search times suggest that the localization of additional mass to the neck region is critical for discrimination in the original experiment.

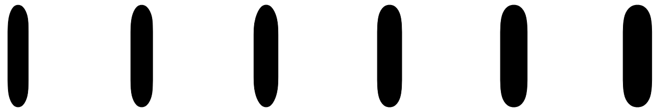


Figure 17: The control stimuli used in the *protrusions-bends* visual search experiment. The thickened rectangles have the same relative areas as the corresponding stimuli in Figure 13.

Experiment I: Parts Versus Protrusions (Control) The control sequence was created by incrementally adding the same amount of mass as in Figure 9, but now distributing it over the top and bottom of the bowtie shape, Figure 15. We ran a set of search trials on a single subject, the results of which are summarized in Figure 16. The large search times and error bars indicate that discrimination between the target and distractor stimuli is significantly more difficult. We conclude that the results of the original experiment are not due to a total area effect, since the localization of added mass to the neck region is critical.

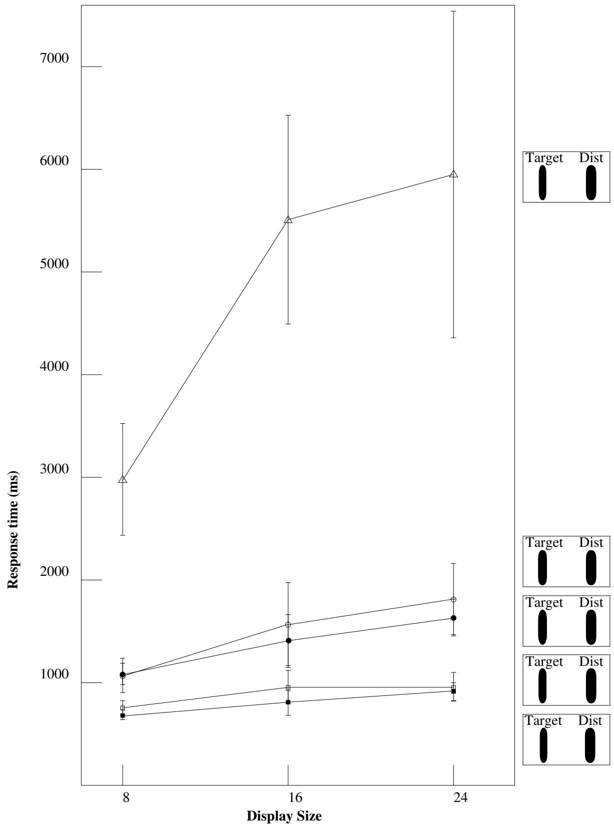
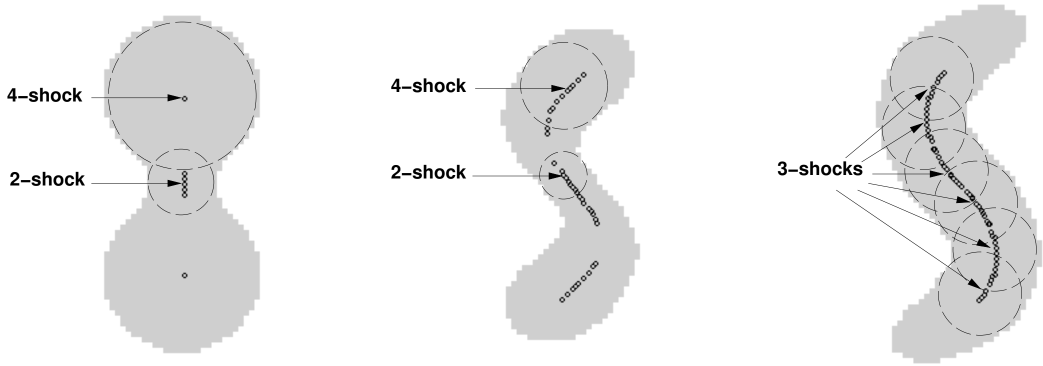


Figure 18: Results for the *bends-protrusions* control experiment. The search times compare with those in Figure 14 suggesting that discrimination for the original stimuli is primarily based on total area.



(a) (b) (c)

Figure 19: The high-order shocks of samples taken from (a) the *parts-protrusions* axis, (b) the *parts-bends* axis and (c) the *protrusions-bends* axis, in Figure 7. These were computed using the algorithm developed in (Siddiqi and Kimia, 1996). 2-shocks occur at (local) minima in object width, 3-shocks at regions of constant width and 4-shocks at (local) maxima in width, as indicated by the inscribed discs, see Appendix A.

Experiment III: Bends Versus Protrusions (Control) The control sequence consisted of incrementally thickened rectangles with rounded corners, Figure 17, with the area ratios between successive stimuli the same as that between successive members of the original sequence in Figure 13. We ran a set of search trials on a single subject, the results of which are summarized in Figure 18. The search times are comparable to those in Figure 14[1], suggesting that discrimination in the original experiment may be based on total area.

2.8 Analysis and Discussion

The results thus far can be summarized as follows: 1) There is a partial ordering of visual search times for all axes of the shape triangle, and 2) only the data along the *bends-protrusions* axis can be accounted for by a change in total area. In order to test our first hypothesis we now examine the shock-based descriptions of the stimuli used. Figure 19 depicts the high order shocks computed for one representative sample from each axis of the shape triangle (Figures 9, 11 and 13). The ratio of minimum (local) width to maximum (local) width is obtained from the ratio of the radii of the maximal inscribed discs at the underlying 2-shocks and 4-shocks, respectively. We plot this ratio for each of the three sequences in Figure 20.

The partial ordering of the 2-shock/4-shock ratios along the *parts-protrusions* and *parts-*

[1]With the exception of the topmost target-distractor pair.

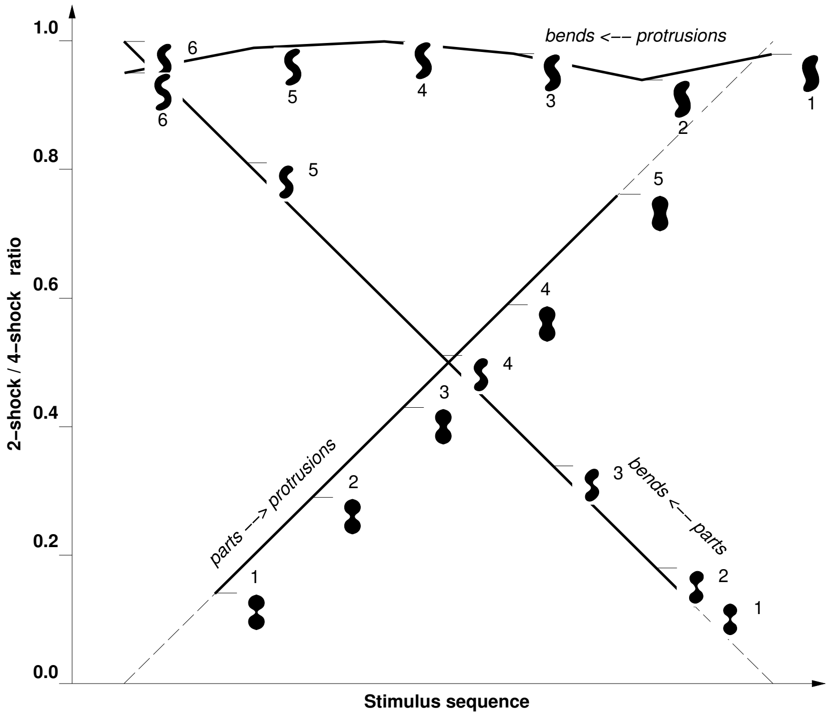


Figure 20: The ratio of radii of the maximal inscribed discs at the underlying 2-shocks and 4-shocks is plotted for the three sequences in Figures 9, 11 and 13.

bends axes reflects the nature of the visual search results. In Figure 21 we plot the magnitude of the 2-shock/4-shock ratio differences between successive target-distractor pairs. These differences are consistent with the separation of the visual search data along any column (Figures 10 and 12), supporting our hypothesis that discrimination along the *parts-protrusions* and *parts-bends* axes is primarily based on differences between the minimum (local) width to maximum (local) width ratios of the underlying stimuli.

A necessary condition for this hypothesis to be correct is that the partial ordering of visual search times should be preserved when the target and distractors are exchanged. To test for this we ran a new set of experiments on a single subject that included complementary trials in which the target and distractors were interchanged. The results, shown in Figures 22 and 23, verify the prediction.

On the other hand, for the *protrusions-bends* sequence the minimum width to maximum width ratio approaches 1, Figure 20, and cannot explain the partial ordering of visual search times. As indicated earlier, the search data can be considered a total area effect. Similar results have been reported in visual search tasks using a sequence of scaled ellipses (Triesman and Gelade, 1980), as well as in apparent motion (Anstis, 1978). The visual search paradigm confounds such area effects with the effects of *where* shocks form; i.e., with the

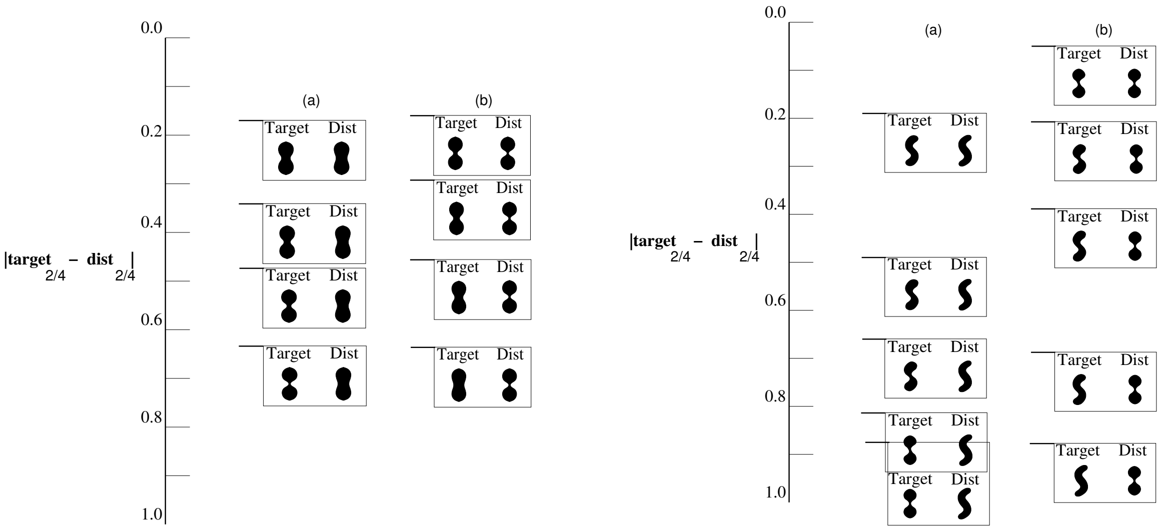


Figure 21: The magnitude of the 2-shock/4-shock ratio differences between target-distractor pairs for Experiment I (left) and Experiment II (right).

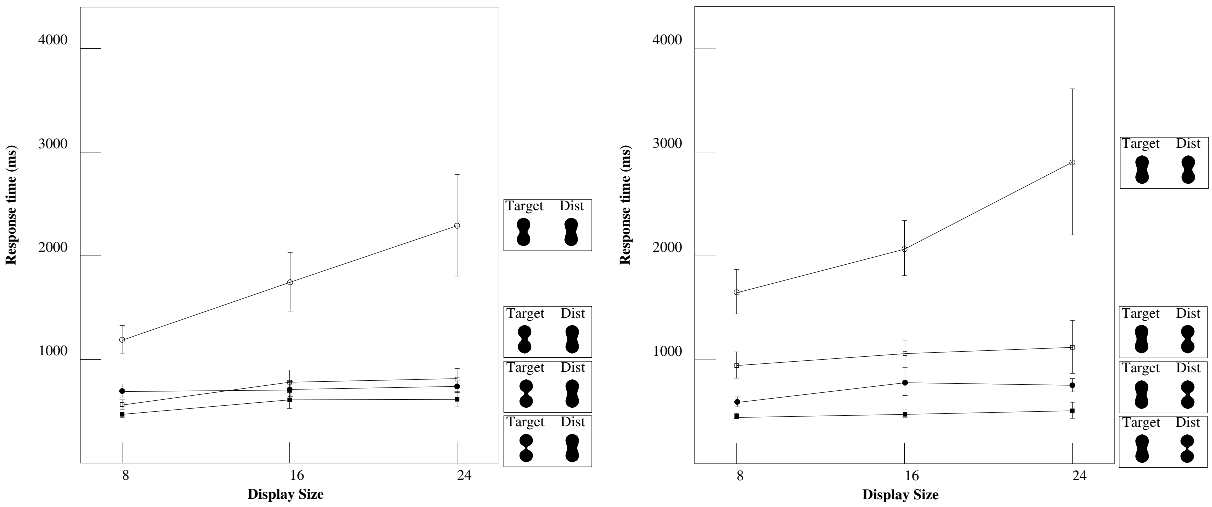


Figure 22: Testing for symmetry along the *parts-protrusions axis*: complementary trials are included in which the target and distractors are interchanged (right).

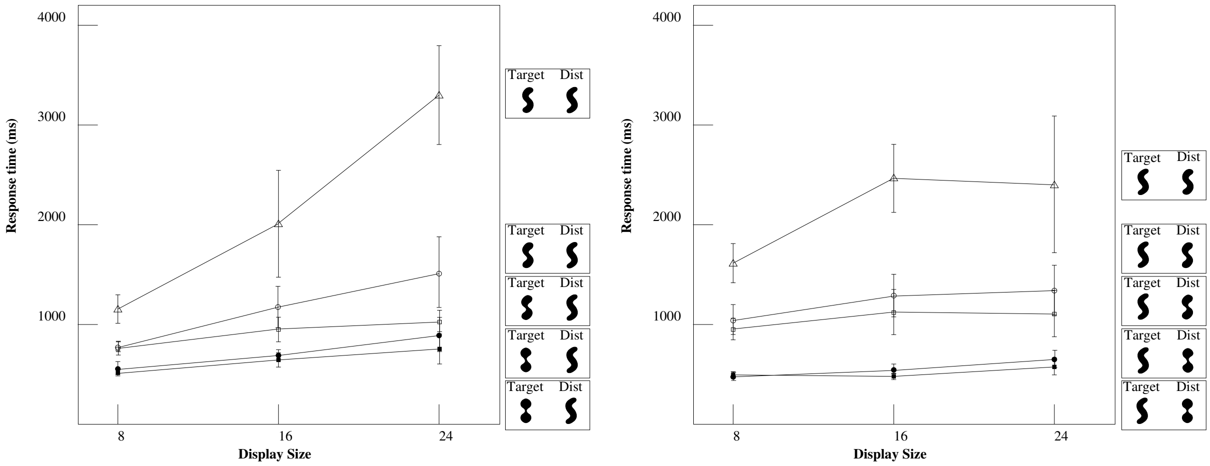


Figure 23: Testing for symmetry along the *parts-bends axis*: complementary trials are included in which the target and distractors are interchanged (right).

analysis of their spatial positions. What is required is a task that is normalized with respect to area, and this is precisely the case for one class of experiments conducted by Burbeck and Pizer (Burbeck et al., 1996). Although their motivation was quite different from ours theoretically, it is pleasing that the elongated stimuli with sinusoidal edge modulation they used (wiggles) closely resemble members of the *bends-protrusions* sequence in Figure 14. Observe that these shapes appear to be not only uniformly thickened but also increasingly straightened versions of one another. In the following Section we review the experiments of (Burbeck et al., 1996), and provide evidence in favor of our second hypothesis that the perceived straightening is related to the loci of high-order shocks.

3 The Effects of "Where" Shocks Form

Pizer *etal.* (Morse et al., 1994; Burbeck and Pizer, 1995; Burbeck et al., 1996) have developed an alternative approach to visual shape analysis called the *core* model. Underlying its formulation is the hypothesis that the scale at which the human visual system integrates local boundary information towards the formation of more global object representations is proportional to object width. Psychophysical examinations of Weber's Law for separation discrimination support this proposal (Burbeck and Hadden, 1993). Arguing that the same mechanism explains the attenuation of edge modulation effects with width, Burbeck *et al.* have recently described a set of psychophysical experiments where subjects were required to bisect elongated stimuli with wiggly sides (Burbeck et al., 1996). These stimuli closely resemble shapes from the *bends-protrusions* sequence in Figure 14. In the following we present

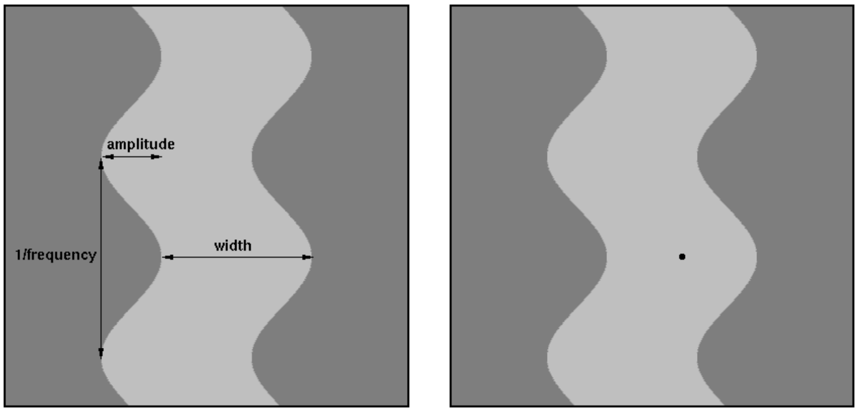


Figure 24: LEFT: The geometry of a "wiggle" stimulus. RIGHT: Is the dot to the left or to the right of the object's center?

their main findings and show that they are consistent with our second hypothesis that the perceived centers of wiggle shapes coincide with high-order shocks.

3.1 Summary of Experimental Findings

The stimuli consisted of rectangles subtending 4 degrees of visual arc in height, with sinusoidal edge modulation, Figure 24 (left). Two widths were considered ($0.75°$ and $1.5°$) and for each width there were 6 edge modulation frequencies ($0.25, 0.5, 1, 2, 4, 8$ cycles/$\circ$) and 2 edge modulation amplitudes (20% and 40% of object width). A black probe dot appeared near the center of each stimulus, in horizontal alignment with a sinusoidal peak. The subject was asked to indicate "whether the probe dot appeared to be left or right of the center of the object, as measured along a horizontal line through the dot."[2]

As a sample trial, view the stimulus on the right of Figure 24 for a period of one second from a distance of 1.5 meters. You are likely to judge the dot to be to the right of the object's center. It may surprise you to find that it actually lies midway between the boundaries on either side, as can be verified by placing a ruler across the figure. In fact, despite instructions to make a local judgement your visual system is biased towards acquiring edge information across a more global spatial extent.

Burbeck and Pizer quantified this effect of edge modulation on the perceived center by varying the horizontal position of the probe dot and subjecting the data to probit analysis. The center of the object was inferred as the 50% point on the best-fitting probit function[3],

[2]See (Burbeck et al., 1996) for further details.

[3]The location at which a subject is statistically equally likely to judge the probe dot to be to the left or

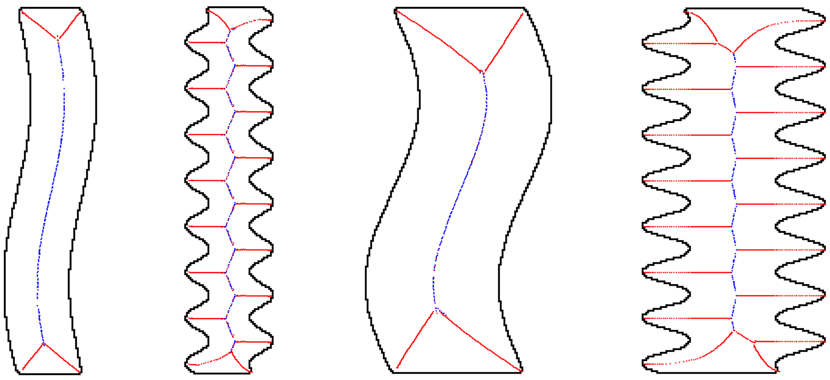


Figure 25: The shock-based description of selected 40% amplitude modulation stimuli used in (Burbeck et al., 1996), computed using the algorithm in (Siddiqi and Kimia, 1996). a) 0.75 degree width, 0.25 cycles/degree edge modulation; b) 0.75 degree object, 2.0 cycles/degree edge modulation; c) 1.5 degree object, 0.25 cycles/degree edge modulation; d) 1.5 degree object, 2.0 cycles/degree edge modulation.

and the *bisection threshold* was defined as the variance of this function. The *perceived central modulation* was then obtained as the horizontal displacement between the perceived centers in alignment with left and right sinusoidal peaks. The main findings were:

Result 1 *For a fixed edge modulation frequency the perceived central modulation decreases with increasing object width.*

Result 2 *For a fixed object width the perceived central modulation decreases with increasing edge modulation frequency.*

These results appear to be consistent with our second hypothesis. Specifically, if the perceived centers of the wiggle stimuli inferred by Burbeck *et al.* coincide with high-order shocks, the central modulation computed as the horizontal displacement between fourth-order shocks in alignment with successive left and right sinusoidal peaks, Figure 26, should agree with the psychophysical data. Thus in the following section we compare computational results obtained from shock-based descriptions with observer data from (Burbeck et al., 1996).

3.2 Analysis and Discussion

We computed shock-based representations for all 24 wiggles using the algorithm in (Siddiqi and Kimia, 1996). Results for selected stimuli are shown in Figure 25, with the geometry

to the right of the object's center.

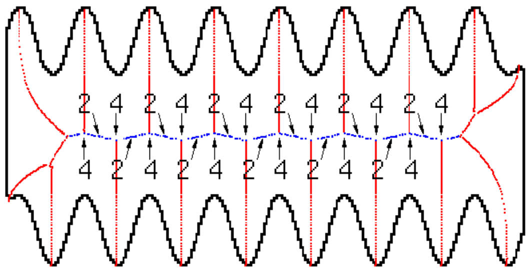


Figure 26: Shape (d) from Figure 25 is rotated and second-order and fourth-order shocks are labeled (all other shocks are first-order). Note that the fourth-order shocks are in alignment with the sinusoidal peaks.

of the high-order shocks explained in Figure 26. As evidence in support of our hypothesis, consider the computed central modulations overlaid as solid lines on the original observer data taken from (Burbeck et al., 1996), in Figure 27. The predicted central modulations are clearly consistent with an "average" observer's data.

Whereas the core model and the shock-based representation are motivated from quite different points of view, the strong overlap between computational and psychophysical results for each model points to a close relationship between the two. We identify two significant qualitative connections. First, the centers of cores (in horizontal alignment with sinusoidal peaks) for the wiggle stimuli coincide with high-order shocks. Second, the "fuzziness" of the core model, whereby the width of the core scales with object width, is paralleled by the ratio of a shock's formation time to the lifetime of the shape, a measure of local width/global width. This property is also reflected in the "bisection-threshold" or variance of the perceived centers in (Burbeck et al., 1996). Underlying this notion is the concept that the scale at which boundaries should interact to form more global object models is proportional to the spatial extent across which they communicate.

4 General Discussion

The problem of shape representation is an extremely subtle one. Different tasks exercise different aspects of it; at certain times more qualitative, generic (or basic level) effects appear to dominate, while at others more quantitative, geometric (or subordinate level) effects dominate. Medial axis style representations of shape have been popular for several decades as a candidate description, in part because they carry both qualitative as well as quantitative information. Blum placed an emphasis on the former type of information early on in his classic work on axis-morphologies (Blum, 1973), where he attempted to interpret

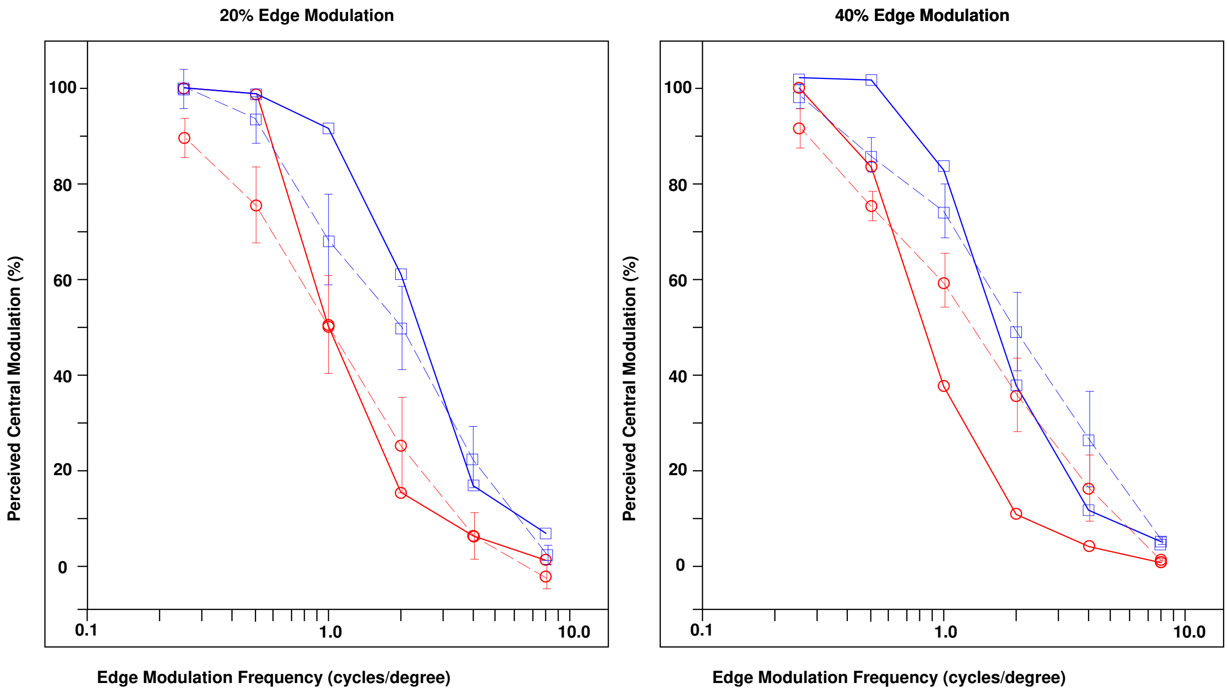


Figure 27: Central modulations computed from shock-based descriptions (solid lines) are overlaid on the data averaged over the two observers in (Burbeck et al., 1996) (dashed lines). The central modulations are expressed as a percentage of the edge modulation amplitude and are plotted against edge modulation frequency for amplitudes of 20% of the object width (left) and 40% of the object width (right). Results for the wider 1.5° object are depicted by the circles and for the narrower 0.75° object by the squares.

the skeleton as a directed graph. Curiously, the wealth of computational literature on the medial axis, with the exception of (Leyton, 1987; Leyton, 1988), has focussed primarily on the latter quantitative information, e.g., (Arcelli et al., 1981; Arcelli and di Baja, 1993; Brady and Asada, 1984; Pizer et al., 1987; Leymarie and Levine, 1992; Ogniewicz, 1993; Rom and Medioni, 1993; Kelly and Levine, 1995). In addition, some physiological evidence (Lamme, 1995; Lee, 1996; Lee et al., 1995) as well as psychophysical evidence (Kovacs and Julesz, 1993; Kovacs and Julesz, 1994) has begun to emerge which suggests a role for medial axes in sensitivity maps. In particular, Kovács and Julesz have shown that in displays composed of Gabor patches (Gaussian-modulated sinusoids), contrast sensitivity can be enhanced within a figure due to long range effects from an enclosing configuration of boundary elements. The location of points of maximal contrast sensitivity enhancement have been shown to coincide with certain special points of the medial axis.

This composite experience implies a role for skeletons in shape, and even suggests how the different performance criteria can be met. We have provided several examples to support the argument that generic recognition is mediated by the topology of the skeletal branches, while subordinate level differences are based primarily on their geometry (Section 1). However, topological information is known to be sensitive to boundary noise (Serra, 1982), a property at odds with requirements for generic recognition. To complicate matters, variations on the skeleton theme, such as the core model (Morse et al., 1994; Burbeck and Pizer, 1995; Burbeck et al., 1996) introduce new issues of scale. While this has led to novel psychophysical experiments (such as shape bisection, Section 3), it is not clear how to compare these results against those obtained in more classical object recognition experiments. Moreover, the relationship between bisection performance and sensitivity maps remains obscure. There appears to be little agreement concerning even the type of data set that might provide a parametric variation relevant to each of these tasks.

We believe that our shock-based theory addresses many of these dilemmas. The theory derives from a position significantly more abstract than the medial axis skeleton, and provides a natural mathematical structure on which shape descriptions can be based: the singularities of a curve evolution process. While we have focussed on the projection of the occluding contour of an object, and hence the boundary evolution of 2D shapes, the descriptions are clearly relevant to the larger problem of 3D object recognition. One of the key features of the mathematics is that it provides a means to mix such boundary effects with area (or region) effects; to our knowledge this is a unique feature of such theories.

In previous research we have been able to articulate the singularities of this curve evolution process, and to relate them to the natural components of shape. One key formal result is that the locus of positions through which the shocks migrate corresponds to the classical Blum skeleton, which relates this class of theory to those mentioned above. However, the shocks carry significantly more information than that obtained from the locus of their positions: we calculate their type as well as the order in which they form. This additional information is instrumental, we believe, in making the computation robust, in separating entry-level from subordinate-level information, and in supporting tasks as diverse as bisection, recognition, and categorization.

In this paper we have considered some of the simplest possible compositions of shock types, and have shown how they lead to explicit classes of parametrically ordered shapes. We have denoted these by a shape triangle, and have examined several of their psychophysical properties. These experiments not only unified visual search and bisection tasks over comparable data, but more importantly revealed differences in tasks related to *when* shocks form as opposed to *where* they form. There also appears to be a fundamental connection to the sensitivity maps: the points of maximal sensitivity for the class of shapes studied in (Kovács et al., 1997) bear a close resemblance to high-order (3 and 4) shocks.

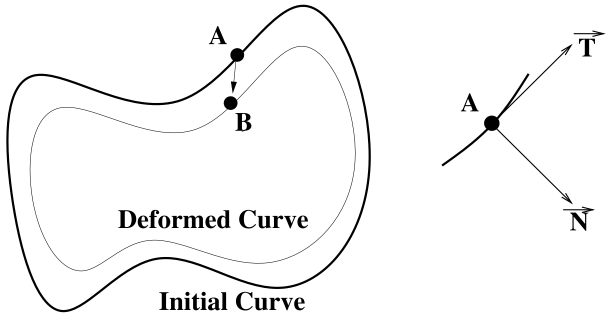


Figure 28: The deformation of an initial curve is described by the displacement of each point in the tangential and normal directions (Kimia et al., 1990; Kimia et al., 1995).

We expect other performance differences to emerge that can also be related to shocks. Computational experiments in shape recognition (Siddiqi et al., 1998; Pelillo et al., 1998) separate the topology of the shock graph and the shock types, from the quantitative information associated with the shock geometry. We hope to extend these results to the psychophysics of recognition as well.

A Curve Evolution And Shocks

The mathematical theory underlying the investigations in this paper is built on the insight that shapes which are slight deformations of one another appear similar. An arbitrary deformation is illustrated in Figure 28, where each point on an initial curve is displaced by a velocity vector with components in the tangential and normal directions. Without loss of generality, it is possible to drop the tangential component (by a reparametrization of the evolved curve). Kimia *et al.* proposed the following evolution equation for 2D shape analysis (Kimia et al., 1990; Kimia et al., 1995):

$$\begin{aligned} \mathcal{C}_t &= (1+\alpha\kappa)\mathcal{N} \\ \mathcal{C}(p,0) &= \mathcal{C}_0(p). \end{aligned} \tag{1}$$

Here $\mathcal{C}(p,t)$ is the vector of curve coordinates, $\mathcal{N}(p,t)$ is the inward normal, p is the curve parameter, and t is the evolutionary time of the deformation. The constant $\alpha \geq 0$ controls the regularizing effects of curvature κ. When α is large, the equation becomes a geometric heat equation which smooths the curve. In this paper we focus on the case where $\alpha = 0$, for which the equation is hyperbolic and *shocks* (Lax, 1971), or entropy-satisfying singularities, can form. The locus of points through which the shocks migrate is related to Blum's grassfire transformation (Brockett and Maragos, 1992; Kimia et al., 1995), although significantly more information is available via a "coloring" of these positions. Four types can arise, according to the local variation of the radius function along the medial axis (Figure 2). Intuitively, the radius function varies monotonically at a type 1, reaches a strict local minimum at a type 2, is constant at a type 3[4] and reaches a strict local maximum at a type 4. The classification of shock positions according to their colors is at the heart of the results in this paper.

The numerical simulation of equation (1) is based on level set methods developed by Osher and Sethian (Osher and Sethian, 1988; Sethian, 1996). The essential idea is to represent the curve $\mathcal{C}(p,t)$ as the zero level set of a smooth and Lipschitz continuous function $\Psi : \mathbf{R}^2 \times [0,\tau) \rightarrow \mathbf{R}$, given by $\{X \in \mathbf{R}^2 : \Psi(X,t) = 0\}$. Since $\mathcal{C}(p,t)$ is on the zero level set, it satisfies

$$\Psi(\mathcal{C},t) = 0\,. \tag{2}$$

By differentiating equation (2) with respect to t, and then with respect to the curve parameter p, it can be shown that

$$\Psi_t = (1+\alpha\kappa)||\nabla\Psi||. \tag{3}$$

Equation (3) is solved using a combination of discretization and numerical techniques derived from hyperbolic conservation laws. The curve $\mathcal{C}$, evolving according to equation (1), is then obtained as the zero level set of Ψ.

A convenient choice for Ψ is the signed distance function to the shape, which can be computed efficiently (Danielsson, 1980). The real-valued embedding surface also provides high resolution information for detecting and localizing shocks. In particular, the four shock types can be classified using differential properties of Ψ (see Figure 29). These ideas have been developed to provide an algorithm for shock detection (Siddiqi and Kimia, 1996), which was used to compute the shock-based descriptions in Figures 3, 4 and 25. An abstraction into a graph of shock groups provides a structural description for shape matching using graph theory (Siddiqi et al., 1998; Pelillo et al., 1998), see Figure 5.

Acknowledgements This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, from the National Science Foundation, from the Air Force Office of Scientific Research and from the Army Research Office. We thank Christina Burbeck, Steve Pizer and Xiaofei Wang for fruitful discussions and for supplying the wiggle stimuli, and Michael Tarr for providing the stimuli in Figure 1.

[4] Although this condition reflects the non-genericity of 3-shocks, "bend"-like structures are abundant in the world and are perceptually salient.

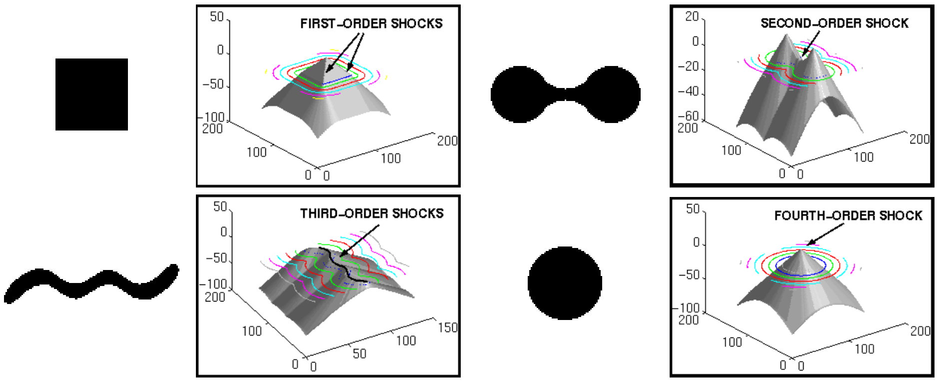


Figure 29: A classification of shock types based on properties of the embedding surface Ψ (in this case the signed distance function, shown as a shaded surface). TOP LEFT: First-order shocks occur at corners of the square shape, corresponding to creases on the surface with $|\nabla\Psi| > 0$. TOP RIGHT: A second-order shock forms at the "neck" of the peanut shape, corresponding to a hyperbolic point with $|\nabla\Psi| = 0$. BOTTOM LEFT: A set of third-order shocks forms along the central axis of the worm shape, where $\kappa_1\kappa_2 = |\nabla\Psi| = 0$. BOTTOM RIGHT: A fourth-order shock forms in the center of the circle shape, where $\kappa_1\kappa_2 > 0$ and $|\nabla\Psi| = 0$. Adapted from (Siddiqi and Kimia, 1996).

References

Anstis, S. M. (1978). Apparent movement. In Held, R., Leibowitz, H., and Teuber, H.-L., editors, *Handbook of Sensory Physiology*, volume 8. Springer-Verlag.

Arcelli, C., Cordella, L. P., and Leviadi, S. (1981). From local maxima to connected skeletons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(2):134–143.

Arcelli, C. and di Baja, G. S. (1993). Euclidean skeleton via centre-of-maximal-disc extraction. *Image and Vision Computing*, 11(3):163–173.

Biederman, I. (1987). Recognition by components. *Psychological Review*, 94:115–147.

Biederman, I. and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, 19:1162–1182.

Biederman, I. and Gerhardstein, P. C. (1995). Viewpoint dependent mechanisms in visual object recognition: A critical analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 21:1506–1514.

Binford, T. O. (1971). Visual perception by computer. In *IEEE Conference on Systems and Control.*

Blum, H. (1973). Biological shape and visual science. *J. Theor. Biol.*, 38:205–287.

Brady, M. and Asada, H. (1984). Smoothed local symmetries and their implementation. *International Journal of Robotics Research*, 3(3):36–61.

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.

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.

Bulthoff, H. H., Edelman, S., and Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 5:247–260.

Burbeck, C. and Hadden, S. (1993). Scaled position integration areas: Accounting for weber's law for separation. *Journal of the Optical Society of America A*, 10(1):5–15.

Burbeck, C. A. and Pizer, S. M. (1995). Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research*, 35:1917–1930.

Burbeck, C. A., Pizer, S. M., Morse, B. S., Ariely, D., Zauberman, G. S., and Rolland, J. (1996). Linking object boundaries at scale: A common mechanism for size and shape judgements. *Vision Research*, In press.

Danielsson, P. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, 14:227–248.

Dickinson, S. J., Bergevin, R., Biederman, I., Eklundh, J.-O., Munck-Fairwood, R., Jain, A. K., and Pentland, A. (1997). Panel report: The potential of geons for generic 3-d object recognition. *Image and Vision Computing*, 15(4):277–292.

Edelman, S. and Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32:2385–2400.

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.

Gauthier, I. and Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, 37(12):1673–1682.

Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, In press.

Hoffman, D. D. and Richards, W. A. (1985). Parts of recognition. *Cognition*, 18:65–96.

Hummel, J. E. and Stankiewicz, J. (1996). Categorical relations in shape perception. *Spatial Vision*, 10(3):201–236.

Humphrey, G. K. and Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, 46:170–190.

Kelly, M. and Levine, M. D. (1995). Annular symmetry operators: A method for locating and describing objects. In *Fifth International Conference on Computer Vision*, pages 1016–1021.

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., and Zucker, S. W. (1990). Toward a computational theory of shape: An overview. *Proceedings of the First European Conference on Computer Vision*, pages 402–407.

Kimia, B. B., Tannenbaum, A., and Zucker, S. W. (1995). Shape, shocks, and deformations I: The components of two-dimensional shape and the reaction-diffusion space. *Int. Journal of Computer Vision*, 15:189–224.

Kovács, I., Fehér, A., and Julesz, B. (1997). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, submitted.

Kovacs, I. and Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences, USA*, 90:7495–7497.

Kovacs, I. and Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370:644–646.

Kurbat, M. A. (1994). Structural descriptions theories: Is rbc/jim a general-purpose theory of human entry-level object recognition? *Perception*, 23:1339–1368.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segmentation in primary visual cortex. *Journal of Neuroscience*, 15:1605–1615.

Lax, P. D. (1971). Shock waves and entropy. In *Contributions to Nonlinear Functional Analysis*, pages 603–634, New York. Academic Press.

Lee, T. S. (1996). Neurophysiological evidence for image segmentation and medial axis computation in primate v1. In *Computation and Neural Systems: Proceedings of the Fourth Annual Computational Neuroscience Conference*. Kluwer Academic.

Lee, T. S., Mumford, D., and Schiller, P. H. (1995). Neuronal correlates of boundary and medial axis representations in primate visual cortex. *Investigative Opthalmology and Visual Science*, 36:477.

Leymarie, F. and Levine, M. D. (1992). Simulating the grassfire transform using an active contour model. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 14(1):56–75.

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.

Liu, Z. (1996). Viewpoint dependency in object representation and recognition. *Spatial Vision*, 9(4):491–521.

Logothetis, N. K., Pauls, J., Bulthoff, H. H., and Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, 4(5):401–414.

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. *Annu. Rev. Neurosci.*, 19:577–621.

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.

Morse, B. S., Pizer, S. M., and Burbeck, C. A. (1994). General shape and specific detail: Context-dependent use of scale in determining visual form. In *Aspects of Visual Form Processing*, pages 374–383. World Scientific.

Ogniewicz, R. L. (1993). *Discrete Voronoi Skeletons.* Hartung-Gorre.

Osher, S. J. and Sethian, J. A. (1988). Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49.

Pelillo, M., Siddiqi, K., and Zucker, S. W. (1998). Matching hierarchical structures using association graphs. In *Proceedings of the European Conference on Computer Vision*, Freiburg, Germany.

Pizer, S. M., Oliver, W. R., and Bloomberg, S. H. (1987). Hierarchical shape description via the multiresolution symmetric axis transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):505–511.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.

Rom, H. and Medioni, G. (1993). Hierarchical decomposition and axial shape description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):973–981.

Rosch, E. (1978). Principles of categorization. In *Cognition and Categorization.* L. Erlbaum Associates.

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.

Serra, J., editor (1982). *Image Analysis and Mathematical Morpholpogy*. Academic Press.

Sethian, J. A. (1996). *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press.

Siddiqi, K. and Kimia, B. B. (1996). A shock grammar for recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 507–513. IEEE Computer Society Press.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., and Zucker, S. W. (1998). Shock graphs and shape matching. In *Proceedings of the International Conference on Computer Vision*, Bombay, India.

Sklar, E., Bulthoff, H. H., Edelman, S., and Basri, R. (1993). Generalization of object recognition across stimulus rotation and deformation. In *Investigative Opthalmology and Visual Science*, volume 34.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, 2(1):55–82.

Tarr, M. J. and Bulthoff, H. H. (1995). Is human object recognition better described by geon-structural-descriptions or by multiple views? *Journal of Experimental Psychology: Human Perception and Performance*, 21:1494–1505.

Tarr, M. J. and Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, 21:233–283.

Triesman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12:97–136.