PERGAMON

Vision Research 41 (2001) 1153–1178

Vision Research

www.elsevier.com/locate/visres

# On the psychophysics of the shape triangle☆

Kaleem Siddiqi [a,*], Benjamin B. Kimia [b], Allen Tannenbaum [c], Steven W. Zucker [d]

[a] *Center for Intelligent Machines & School of Computer Science, McGill University, 3480 University Street, Montréal, Québec, Canada QC H3A 2A7*

[b] *Laboratory for Engineering Man/Machine Systems, Division of Engineering, Brown University, Providence, RI 02912, USA*

[c] *Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30082-0250, USA*

[d] *Department of Computer Science, Center for Computational Vision & Control, Yale University, New Haven, CT 06520-8285, USA*

Received 29 April 1998; received in revised form 22 June 1999

**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. © 2001 Elsevier Science Ltd. All rights reserved.

*Keywords:* Shocks; Shape perception; Visual search; Shape bisection

## 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 (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-acci-

☆ Portions of this material were presented at the meeting of the Eastern Psychological Association (April 1999) and at the Association for Research in Vision and Opthalmology Conference (May 1999).

* Corresponding author. Tel.: +1-514-398-3371; fax: +1-514-398-7348.

*E-mail address:* siddiqi@cim.mcgill.ca (K. Siddiqi).

0042-6989/01/$ - see front matter 
PII: S0042-6989(00)00274-1

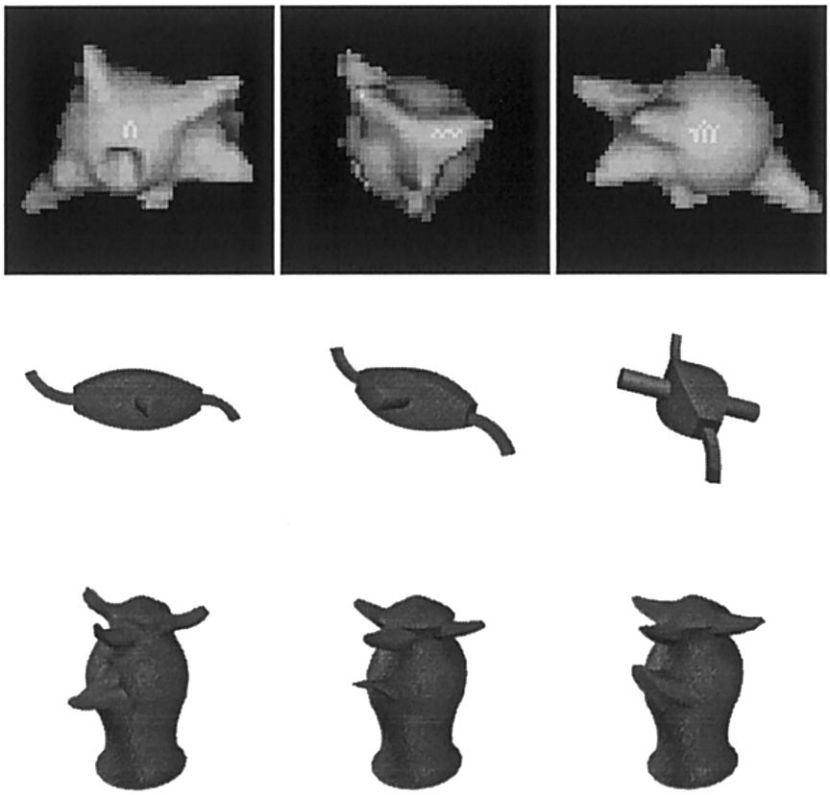

Fig. 1. Representative sample stimuli upon which various psychophysical studies on object recognition have been used. Top row: 'amoebas' used in Bulthoff and Edelman (1992) and Logothetis et al. (1994). Middle row: multi-part objects with geon-like components used in Hayward (1998). Bottom row: 'greebles' used in Gauthier and Tarr (1997).

dental properties of the retinal image (Binford, 1971; Marr & Nishihara, 1978; Biederman, 1987). 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, Mervis, Gray, Johnson, & Boyes-Braem, 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 & Pinker, 1989; Bulthoff & Edelman, 1992; Edelman & Bulthoff, 1992; Logothetis, Pauls, Bulthoff, & Poggio, 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 & Edelman, 1990). Such a model is consistent with psychophysical

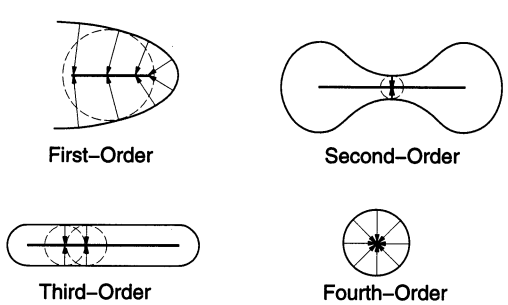


Fig. 2. A coloring of shocks into four types (Kimia et al., 1995). A 1-shock derives from a *protrusion*, and traces out a curve segment of adjacent 1-shocks. A 2-shock arises at a *neck*, and is immediately followed by two 1-shocks flowing away from it in opposite directions. 3-shocks correspond to an annihilation into a 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.

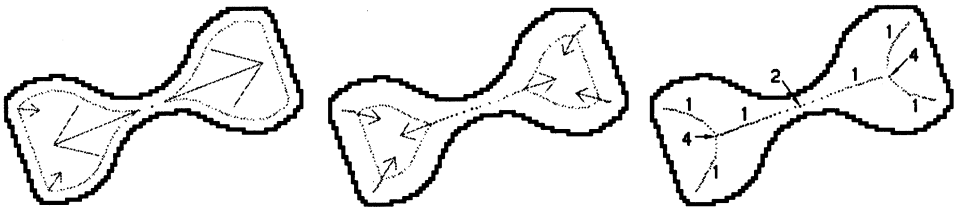

Fig. 3. The detection of shocks for a dumbbell shape undergoing constant inward motion. Each sub-figure is a snapshot of the evolution in time, with the outline of the original shape shown in black, the evolved curve overlaid within, and the arrows representing velocity vectors for the current 1-shocks. 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).

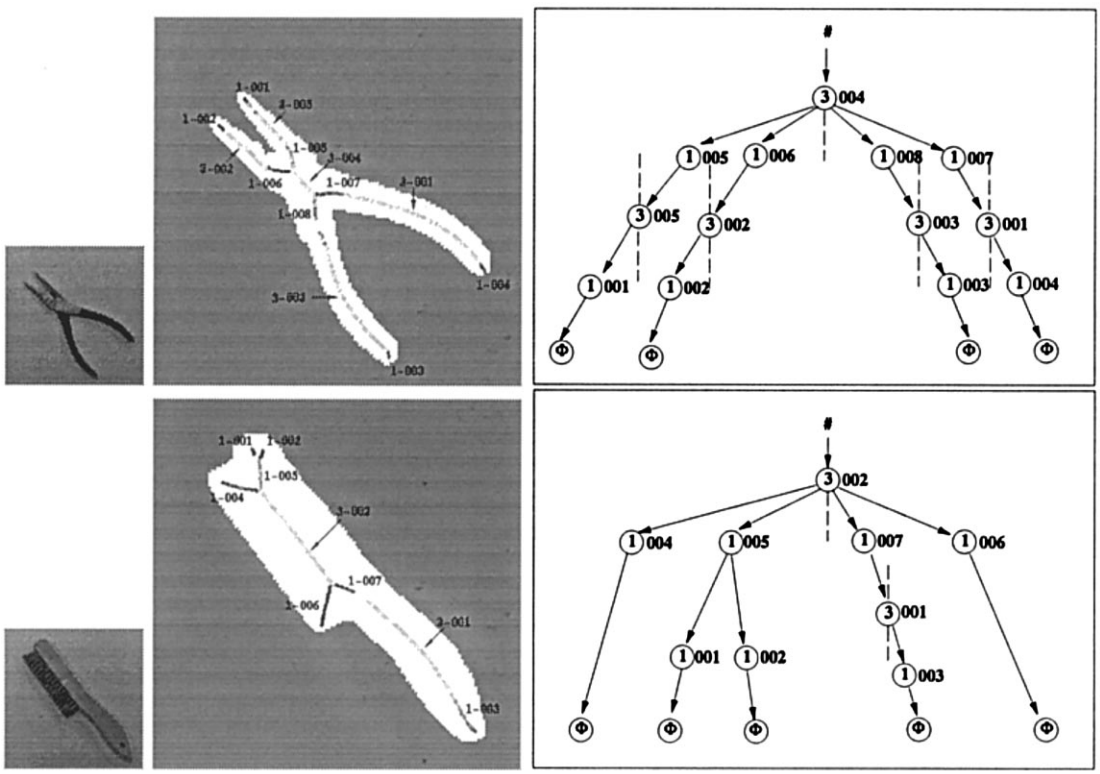

Fig. 4. The shock-based representations and shock graphs for a plier (top) and a brush (bottom). The representation illustrates where shocks form. The notation associated with each shock group is of the form shock-type-identifier. Each shock group leads to a distinct node in the shock graph. These nodes are ordered in reverse time, with edges indicating parent-child relationships. Adapted from Siddiqi et al. (1999b). Reproduced by permission of Kluwer Academic/Plenum Publisher.

data obtained for certain classes of unfamiliar objects (Bulthoff, Edelman, & Tarr, 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?; and (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 (Humphrey & Khan, 1992; Biederman & Gerhardstein, 1993; Kurbat, 1994; Biederman & Gerhardstein, 1995; Tarr, 1995; Tarr & Bulthoff, 1995; Hummel & Stankiewicz, 1996; 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 (Bulthoff & Edelman, 1992; Edelman & Bulthoff, 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, e.g. for recognition tasks at different levels of abstraction. In particular, the experiments of (Logothetis et al., 1994, p. 405) and others have found subordinate level recog-

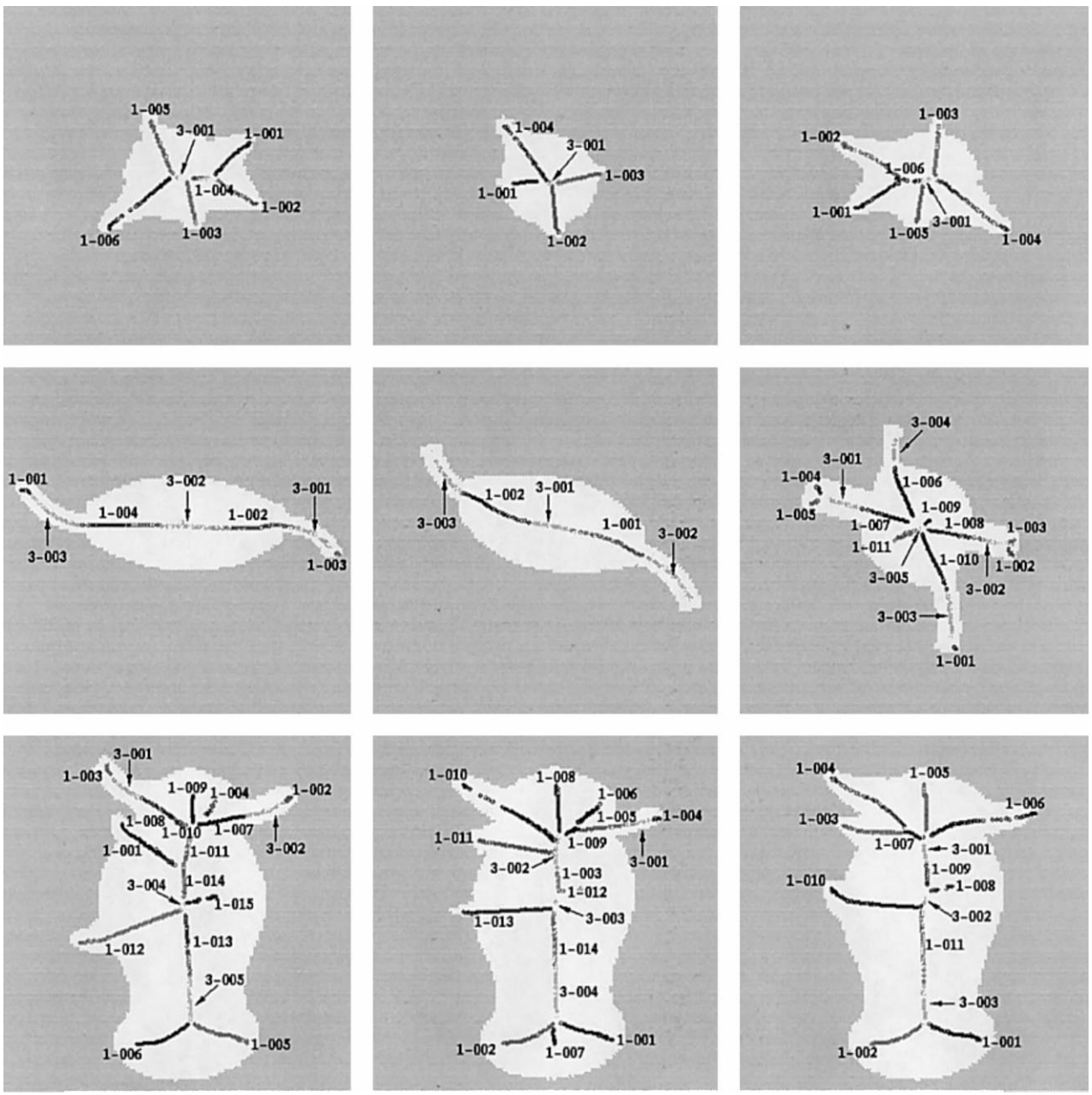


Fig. 5. The shock-based descriptions of the profiles of stimuli in Fig. 1: 'amoebas' (top row); multi-part geons (middle row); 'greebles' (bottom row). The representation illustrates where shocks form. The notation associated with each shock group is of the form shock-type-identifier; see Appendix A for a discussion of shock types. The corresponding shock graphs are shown in Fig. 6.

nition to be viewpoint dependent but basic level recognition to be relatively robust to changes in viewpoint. In order to provide a computational explanation for such results, Tjan and Legge have recently proposed a measure of viewpoint complexity (Tjan & Legge, 1998). Here, the basic idea is to characterize factors which are due to the nature of the object recognition task (the stimuli used, and the judgment required of the observer) but are independent of the nature of visual representation or processing. They have argued that the published data on viewpoint dependency corresponds with the viewpoint complexity of the underlying task.

To motivate the need for both viewpoint dependent and viewpoint invariant properties, consider the stimuli in Fig. 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

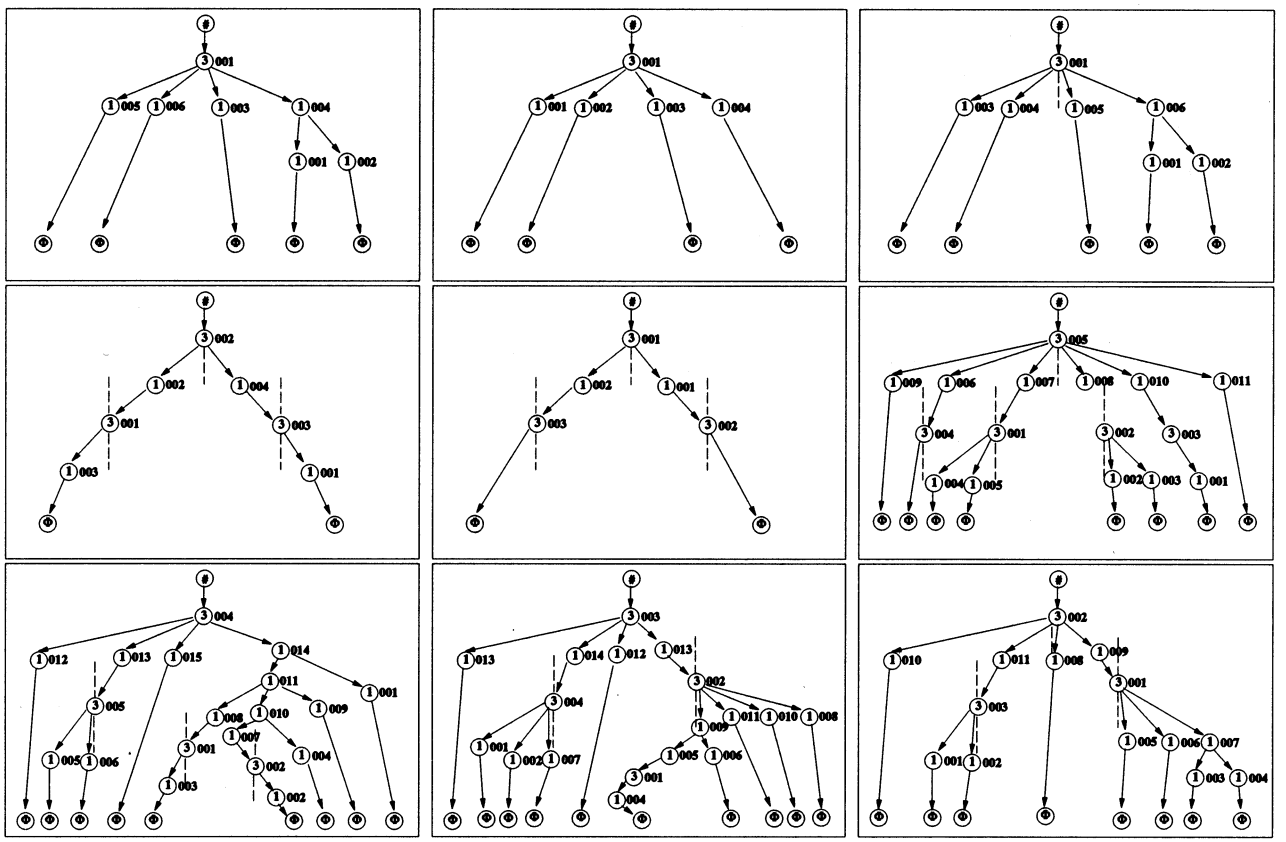

Fig. 6. The shock graphs (Siddiqi et al., 1999b) of the shock-based descriptions in Fig. 5: 'amoebas' (top row); multi-part geons (middle row); 'greebles' (bottom row). Each distinct shock group comprises a different node. The nodes are ordered in reverse time, with edges indicating parent-child relationships, 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 Fig. 1) is obvious.

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 contours of the objects, which are an integral part of their descriptions. In fact, a recent study functionally articulate sub-areas within the IT cortex which are selective to 2D rather than 3D shape (Janssen, Vogels, & Orban, 2000). The investigation by Hayward (1998) also 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, Bulthoff, Edelman, & Basri, 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 that 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 contour of an object (Kimia, Tannenbaum, & Zucker, 1995). Four types of shocks arise, each of which has a direct perceptual correlate, as illustrated in Fig. 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 Fig. 3. A qualitative description of the shape emerges, 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). To further this intuition, in the side view of a cat the neck would correspond to a 2-shock, the tail and the limbs to 3-shocks, and the ears to 1-shocks. These would be connected by a hierarchy of 1-shocks to two 4-shocks, one at the center of the belly and one in the center of the head. Similarly, for a hand silhouette the fingers (3-shocks) would be connected to the central palm (a 4-shock) by a hierarchy of 1-shocks. 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. In related work we have generated shock graphs for various classes of real-world objects and have investigated shape matching using graph theoretic concepts (Sharvit, Chan, Tek, & Kimia, 1998; Tirthapura, Sharvit, Klein, & Kimia, 1998; Pelillo, Siddiqi, & Zucker, 1999; Siddiqi, Shokoufandeh, Dickinson, & Zucker, 1999b). The results indicate that objects with similar part structures have similar shock graph topologies, while geometric information (contained within each node) plays a greater role in accounting for variations at a subordinate level. Representative computational examples are shown in Fig. 4. We note that a complementary shock-based representation exists for the background of a shape, which can be obtained by evolving the initial closed contour outwards. In general, the shock groups would have infinite spatial extent since the background is unbounded. In the current paper we focus on the shocks in the interior of a shape.

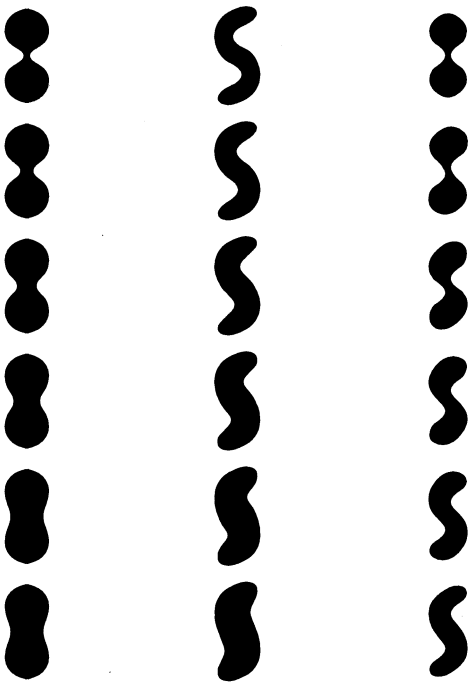

Fig. 7. Three families of shapes that illustrate both the categorical nature of shape descriptions, and the delicate crossover between categories. Notice, e.g. how the upper left shape is perceived as two parts connected by a thin neck, while the lower left shape appears to have two indentations (or 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.

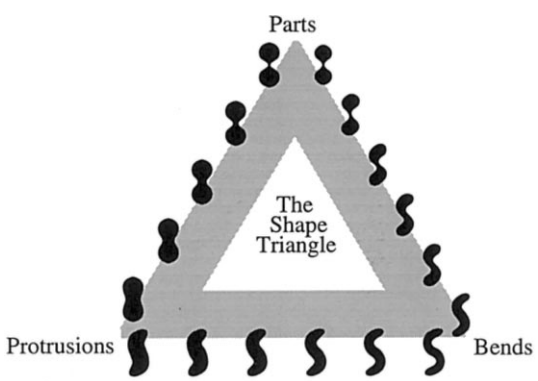


Fig. 8. The sides of the shape triangle represent continua of shapes; the extremes correspond to the 'parts', 'protrusions' and 'bends' nodes (Kimia, 1990).

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 Fig. 1. The shock-based descriptions are shown in Fig. 5, with the associated shock graphs in Fig. 6. The representation in Fig. 5 illustrates the spatial configuration of shocks. The abstraction in Fig. 6 illustrates the (reverse) order of their formation in time. Views which are qualitatively similar, e.g. Fig. 1 (second row, left and middle), produce qualitatively similar shock graphs, Fig. 6 (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. Fig. 1 (second row, right), results in drastic modifications to the underlying shock graph, Fig. 6 (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 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, view-

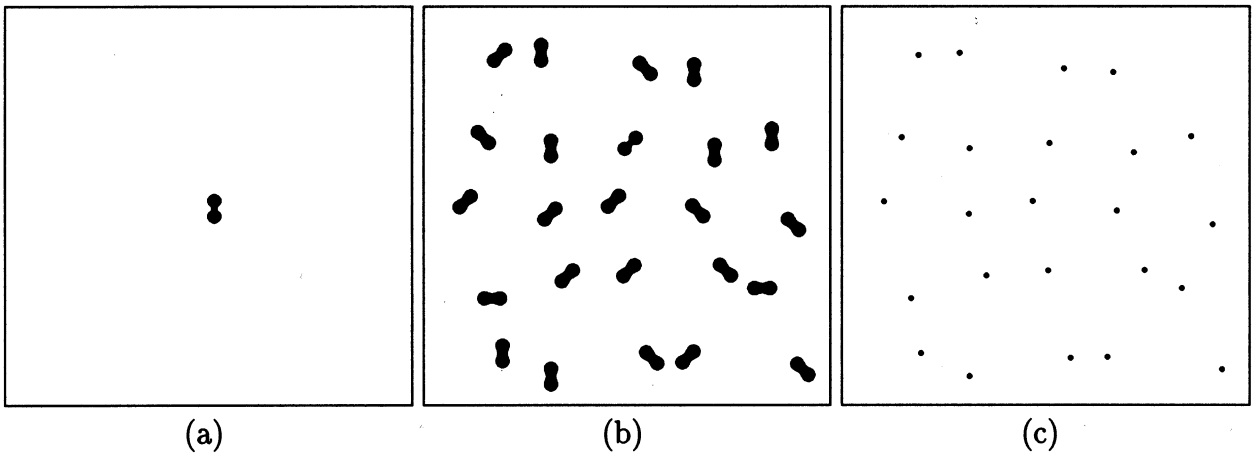


Fig. 9. A visual search sequence. First an example of the target is shown in the center of the display (a). In each trial a display containing exactly one target embedded in a field composed of several copies of a distractor element is presented (b). The subject is asked to press a mouse button upon finding the target, and then identify its location in a validation display composed of small dots (c).

point 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, Fig. 2, the morphogenesis of even simple shapes can be ambiguous when different shocks participate in their description. To illustrate, are shapes in the left and right columns of Fig. 7 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, Fig. 8. For example, along the parts-bends axis the percept changes from that of 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 Fig. 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., 1999b).

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, from one element to the next within the same sequence, the shock formation times and positions may gradually vary. 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

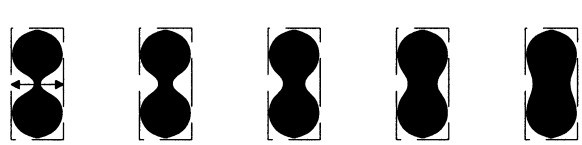

Fig. 10. The stimuli used in the *parts-protrusions* visual search experiment were created by pulling apart the opposing boundaries of the bowtie shape (on the left) at its neck. Each stimulus had the same overall size, as indicated by the dashed rectangle. The dashed rectangle was not visible in the psychophysical display.

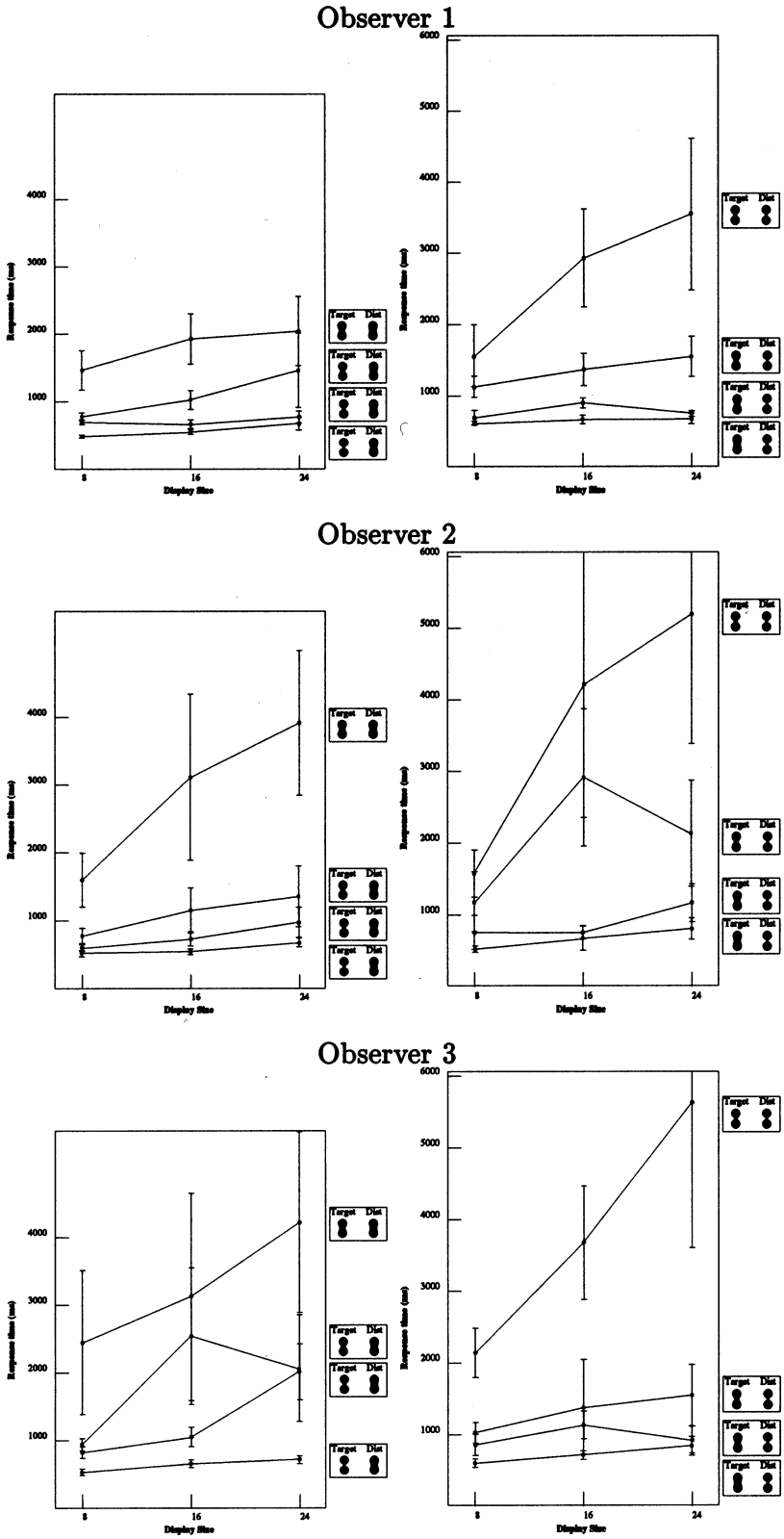


Fig. 11. Results for the *parts-protrusions* visual search experiment.

shock locations. The former phenomenon 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,

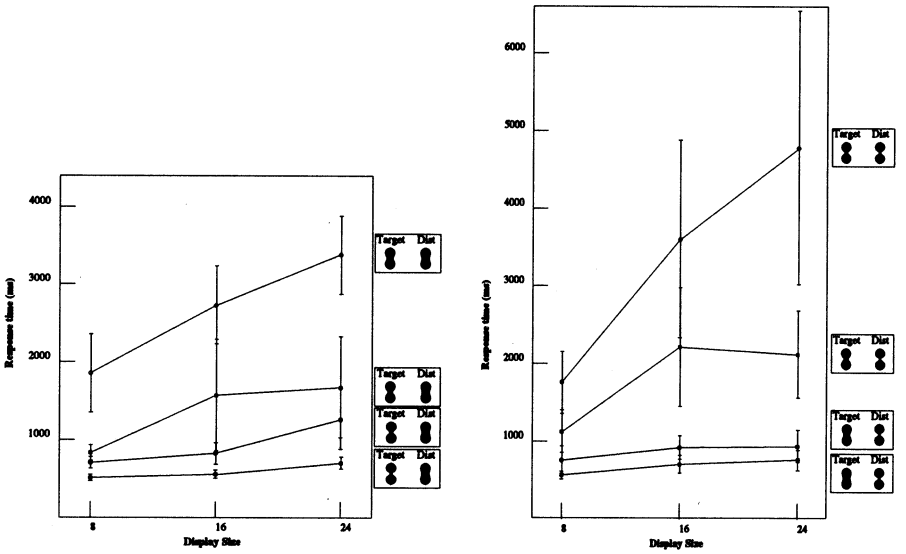


Fig. 12. Results for the *parts-protrusions* visual search experiment, averaged over the three observers (see Fig. 11).

**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 Fig. 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, Fig. 36. 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 the target and distractor elements. We first describe the experiments and then provide a discussion of the results.

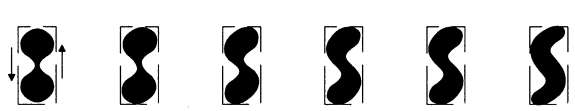

Fig. 13. The stimuli used in the *parts-bends* visual search experiment were created by shearing the opposing boundaries of the bowtie shape (on the left) with respect to each other. Each stimulus had the same overall size, as indicated by the dashed rectangle, as well as the same total area. The dashed rectangle was not visible in the psychophysical display.

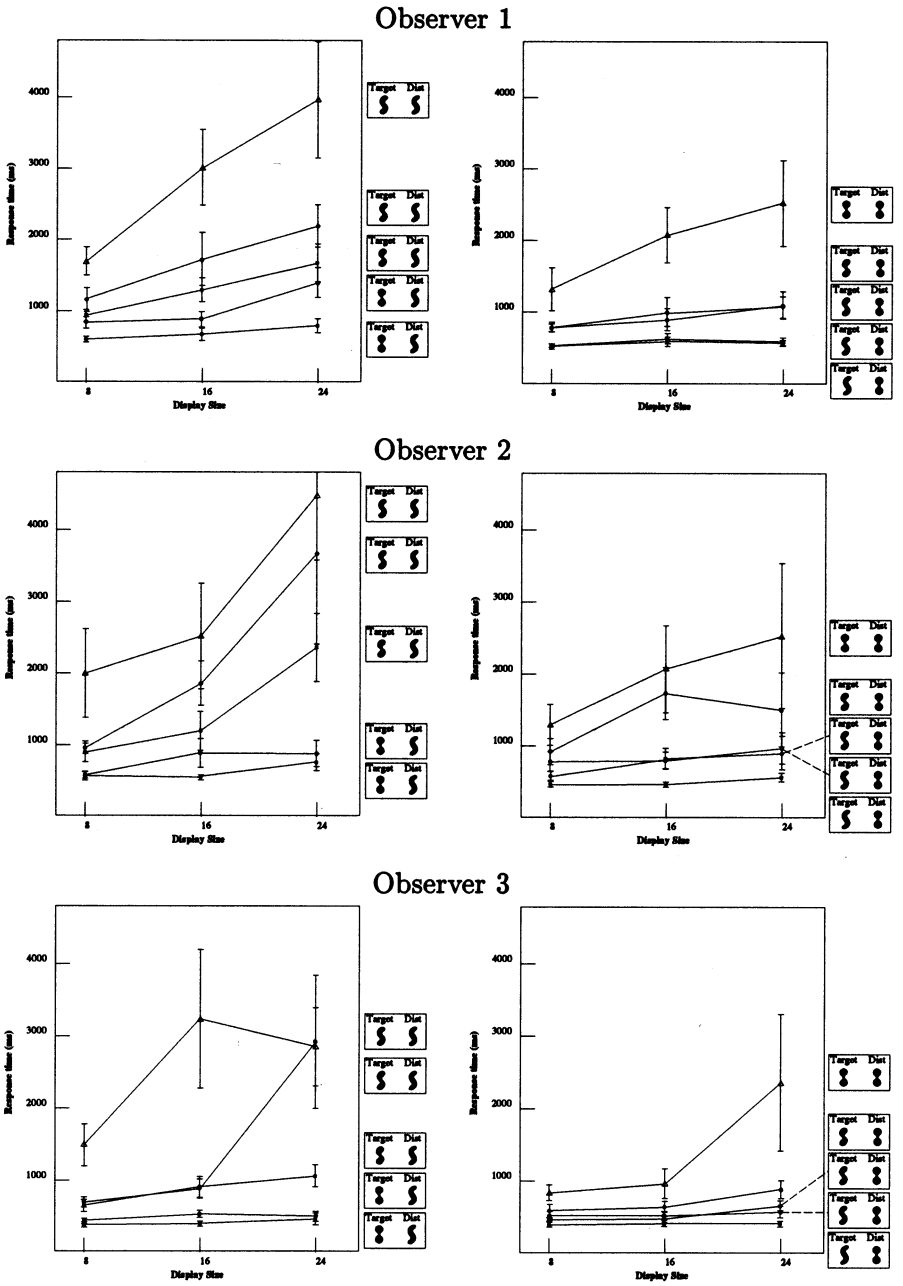


Fig. 14. Results for the *parts-bends* visual search experiment.

## 2.1. Methods

Our experimental procedure closely follows that introduced in (Elder & 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 m from the screen. A $7° \times 7°$ 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 $0.5° \times 0.5°$ in size. Their placement in the display window was based on a regular $5 \times 5$ 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°, with equal probability.

Displays contained either 7, 15 or 23 distractor stimuli and one target stimulus. First an example of the

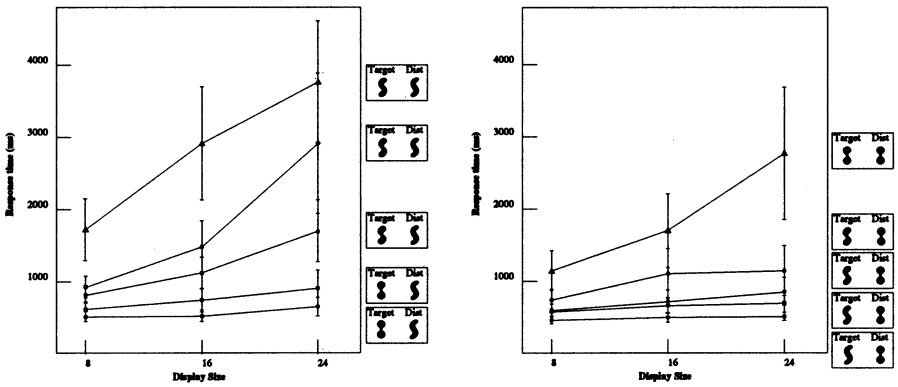


Fig. 15. Results for the *parts-bends* visual search experiment, averaged over the three observers (see Fig. 14).

target is shown in the center of the display (Fig. 9a). 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 (Fig. 9b). 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 (Fig. 9c). 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 & Gelade, 1980). This modification avoids a systematic bias that the traditional method suffers from, and typically gives lower error rates (Elder & 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°) 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 $0.5° \times 0.5°$ ($70 \times 70$ pixels at a viewing distance of 2 m). Details of each stimulus set are presented along with the experiments. Note that the dashed rectangles in Figs. 10, 13 and 16 were not present in the psychophysical displays.

### 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, Fig. 10. 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 $2 \times 4$ sessions for each subject (each session consisting of 30 trials in total). The results are summarized in Figs. 11 and 12. The search time decreases monotonically as the two stimuli are further separated along the *parts-protrusions* axis.

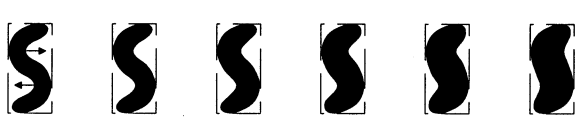

Fig. 16. The stimuli used in the *bends-protrusions* visual search experiment were created by incrementally adding mass to the concave regions of the worm shape on the left. Each stimulus had the same overall size, as indicated by the dashed rectangle. The dashed rectangle was not visible in the psychophysical display.

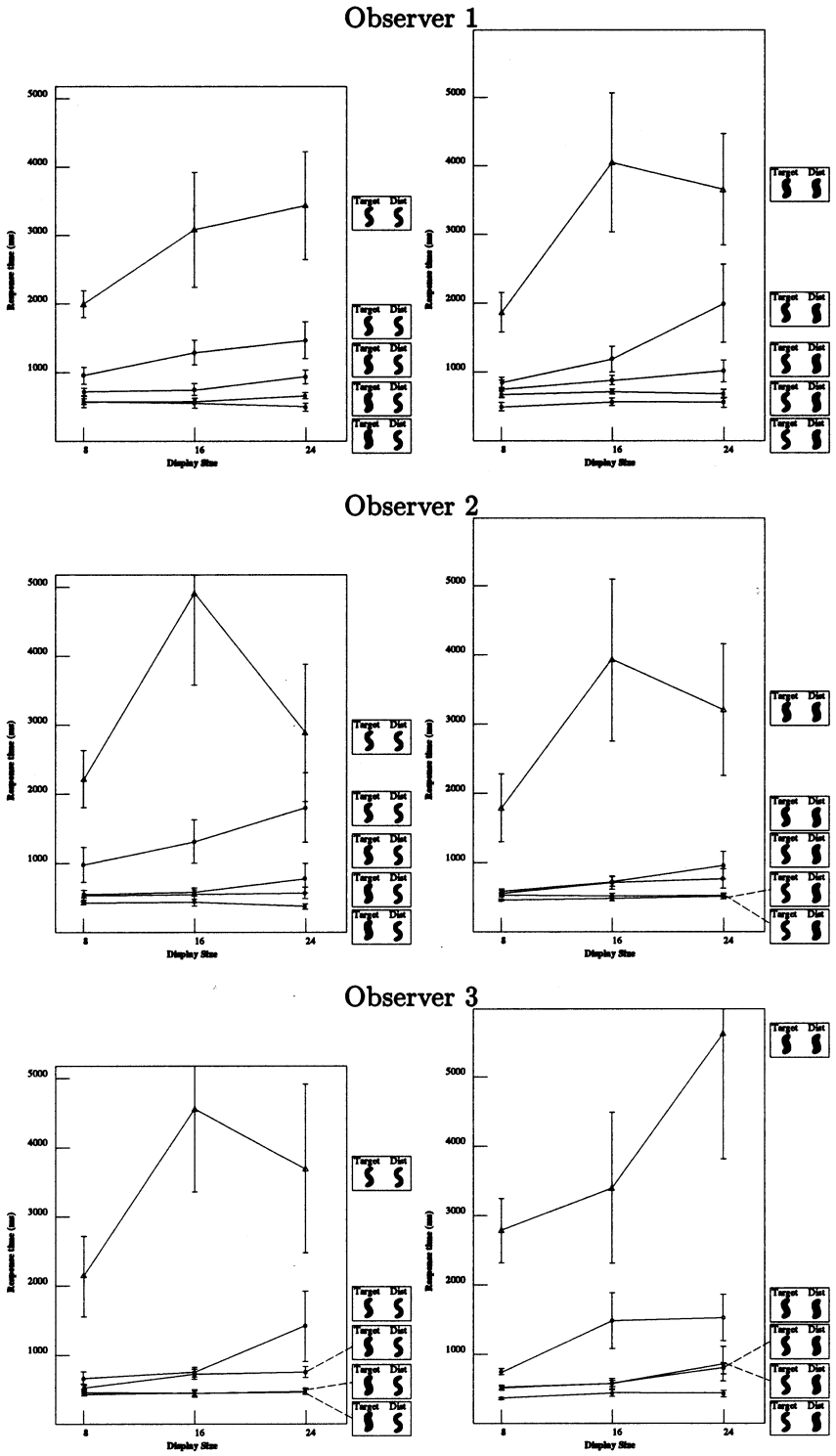


Fig. 17. Results for the *bends-protrusions* visual search experiment.

### 2.5. Experiment II: parts versus bends

**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, Fig. 13. 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 $2 \times 5$ sessions for each subject (each session consisting

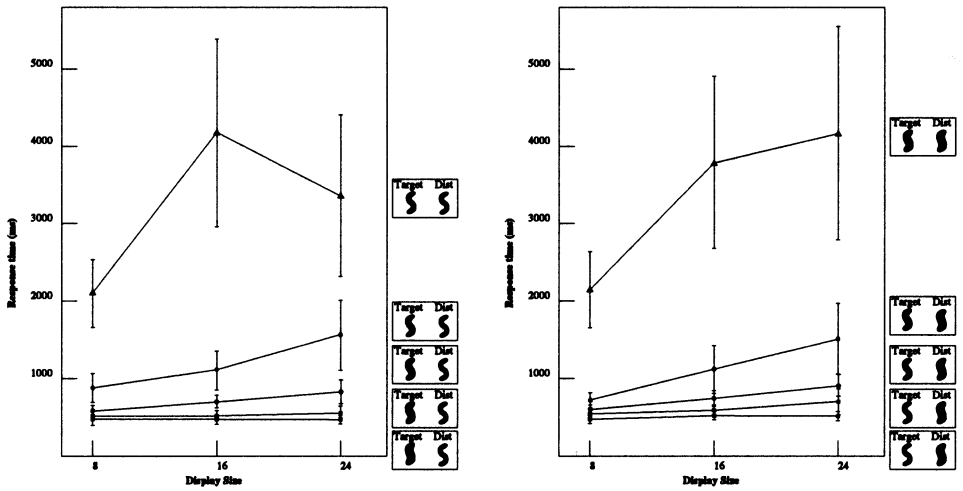


Fig. 18. Results for the *bends-protrusions* visual search experiment, averaged over the three observers (see Fig. 17).

of 30 trials in total). The results are summarized in Figs. 14 and 15. Once again, the search time is a monotonically decreasing function of the separation between the two stimuli along the *parts-bends* axis.

### 2.6. Experiment III: bends versus protrusions

**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, Fig. 16. 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 $2 \times 5$ sessions for each subject (each session consisting of 30 trials in total). The results are summarized in Figs. 17 and 18. It is clear that the search time is a monotonic function of the separation between the two stimuli along the *bends-protrusions* axis.

### 2.7. Effects of size and area

We note that care was taken to preserve local properties of the stimuli, such as the number and order of curvature extrema along the boundary (Hoffman & Richards, 1985), as well as certain global ones, such as overall size. However, total area was preserved only in Experiment II, since mass was added to create the sequences in Figs. 10 and 16. We shall now consider the effects of overall size and area in interpreting the results. First, we shall examine the extent to which the results are a property of object shape, by allowing the overall size of both the targets and the distractors to vary during a trial. Second, we shall consider the possibility that discrimination in Experiments I and III is based on total area.

#### 2.7.1. Overall size

To examine the effect of overall size, we repeated Experiments I, II, and III, but now allowing the target and the distractors to vary in size (independently). For each trial, the size of a target or distractor element was chosen randomly to be either $0.407° \times 0.407°$, $0.443° \times 0.443°$, or $0.5° \times 0.5°$, with the constraint that across the sequence of 30 trials, the target appeared at each of the three sizes an equal number of times. The size

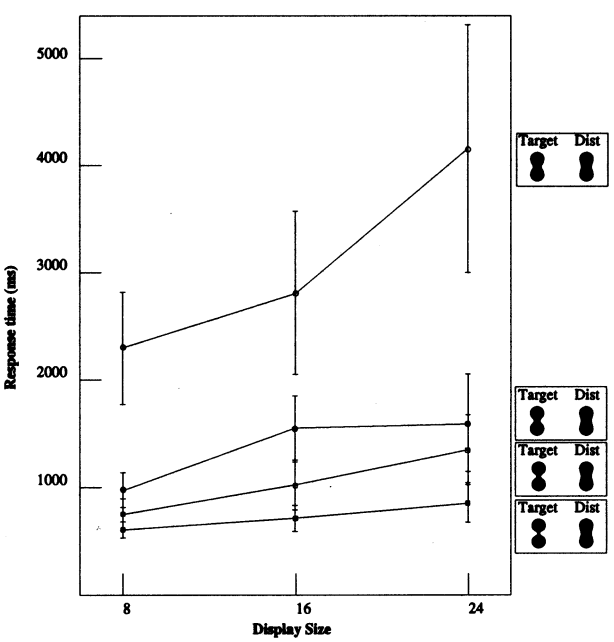


Fig. 19. Results for the *parts-protrusions* visual search experiment, averaged over the three subjects, but now with the target/distractor sizes varying independently (see text). Compare with the left hand side of Fig. 12.

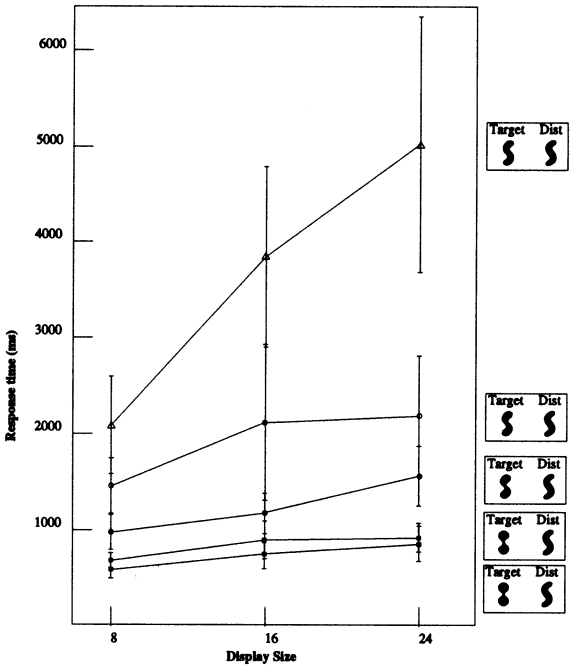


Fig. 20. Results for the *parts-bends* visual search experiment, averaged over the three subjects, but now with the target/distractor sizes varying independently (see text). Compare with the left hand side of Fig. 15.

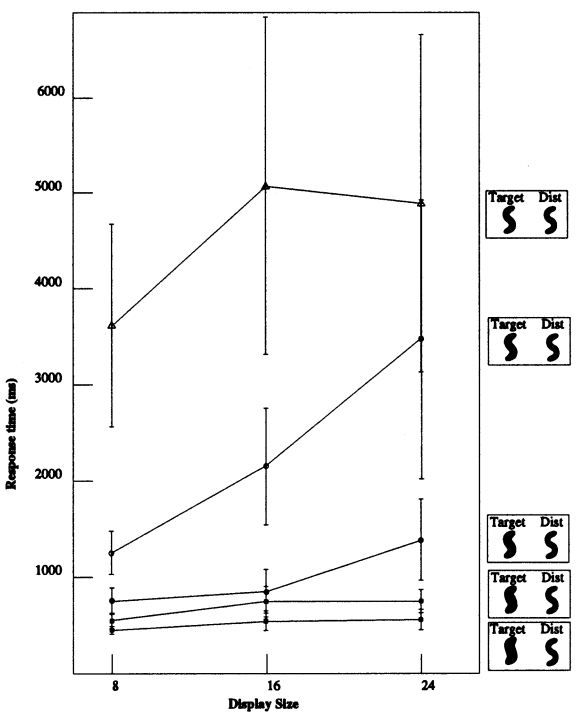


Fig. 21. Results for the *bends-protrusions* visual search experiment, averaged over the three subjects, but now with the target/distractor sizes varying independently (see text). Compare with the left hand side of Fig. 18.

variation was $-11.4\%$ and $-18.6\%$ about the maximum size. The results, averaged over the three subjects, appear in Figs. 19–21. Comparing with the left hand sides of Figs. 12, 15 and 18, respectively, the partial ordering of the data is preserved, although on average there is slight increase in the search times. Hence, the results may be interpreted to be a function of object shape, independent of magnification (or, equivalently, viewing distance).

### 2.7.2. Overall area

To test the possibility that discrimination in Experiments I and III was based on total area, we ran the following two control experiments.

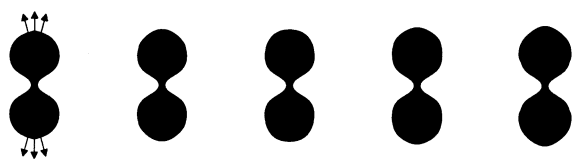

Fig. 22. The control stimuli used in the *parts-protrusions* visual search experiment. The same amount of mass incrementally added to the neck region in Fig. 10 is now added to the top and bottom of the bowtie shape.

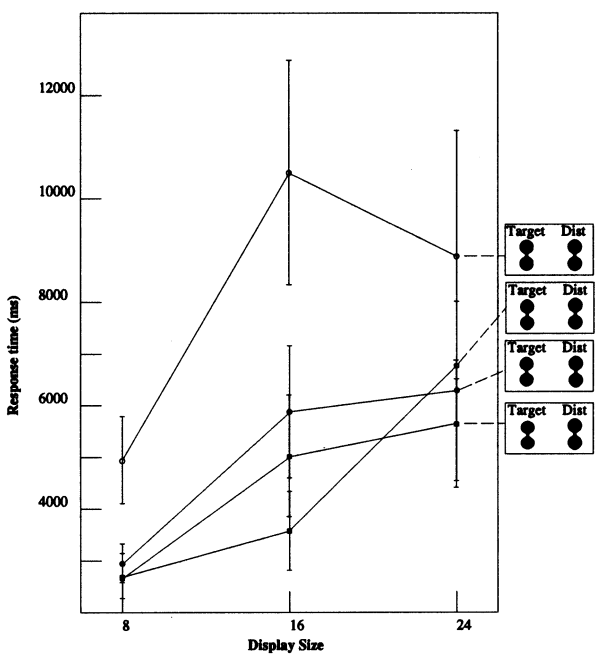


Fig. 23. Results for the *parts-protrusions* control experiment for a single subject (Observer 1). The significantly larger search times suggest that the localization of additional mass to the neck region is critical for discrimination in the original experiment (see Fig. 11).

Fig. 24. The control stimuli used in the *protrusions-bends* visual search experiment. the thickened rectangles have the same relative areas as the corresponding stimuli in Fig. 16.

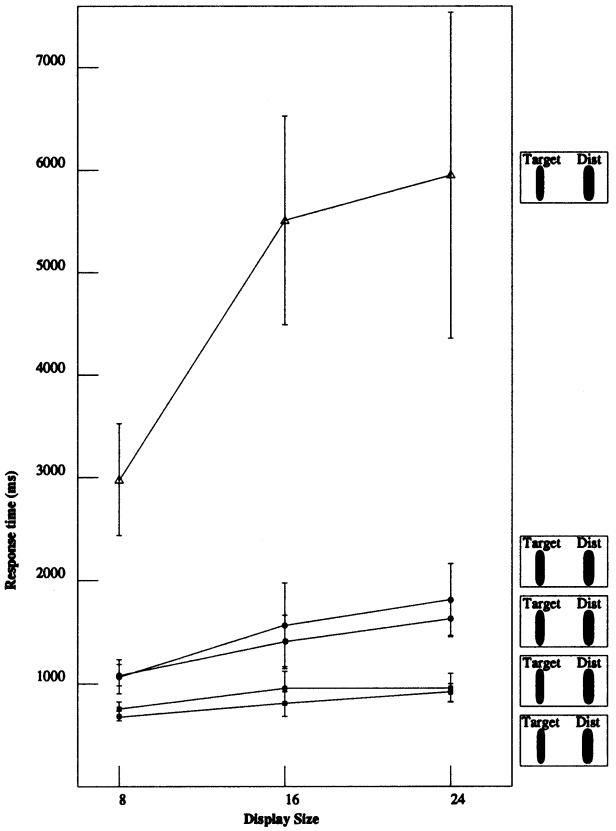


Fig. 25. Results for the *bends-protrusions* control experiment for a single subject (Observer 1). The search times compare with those in Fig. 17 suggesting that discrimination for the original stimuli is primarily based on total area.

**Experiment I: parts versus protrusions (control)** The control sequence was created by incrementally adding the same amount of mass as in Fig. 10, but now distributing it over the top and bottom of the bowtie shape, Fig. 22. We ran a set of search trials on a single subject (Observer 1), the results of which are summarized in Fig. 23. 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.

**Experiment III: bends versus protrusions (control)** The control sequence consisted of incrementally thickened rectangles with rounded corners, Fig. 24, with the area ratios between successive stimuli the same as that between successive members of the original sequence in Fig. 16. We ran a set of search trials on a single subject (Observer 1), the results of which are summarized in Fig. 25. The search times are comparable to those in Fig. 17[1], suggesting that discrimination in the original experiment may be based on total area. However, it is also possible to attribute the results to width, since area and width co-vary for this stimulus set.

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

### 2.8. Independence of axes

In our computational work (Sharvit et al., 1998; Tirthapura et al., 1998; Pelillo, Siddiqi, & Zucker, 1999; Siddiqi et al., 1999b) we have found that algorithms for matching two shapes converge rapidly when the two underlying shock graphs have very different topologies. Intuitively, in such cases the two shapes have few part structures in common and are likely to arise from different entry level categories. In order to test the hypothesis that discrimination for such a pair should be rapid, we carried out a set of visual search trials on a single subject in which a fixed target was chosen from the middle of one axis and the distractor was selected from one of the other two axes. Thus, the shock graph topologies were distinct. The results, for all target-distractor pairs, are shown in Figs. 26–28. It is evident that the visual search curves are relatively flat.

### 2.9. 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; (2) this partial ordering is independent of overall size, and hence may be interpreted primarily as a function of object shape; and (3) only the data along the *bends-protrusions* axis can be accounted for by a change in total area. Further, when the target/distractor pairs are taken from different axes, discrimination is rapid.

In order to test our first hypothesis we now examine the shock-based descriptions of the stimuli used. Fig. 29 depicts the high order shocks computed for one representative sample from each axis of the shape triangle (Figs. 10, 13 and 16). 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 Fig. 30.

It is evident that the partial ordering of the 2-shock/4-shock ratios along the *parts-protrusions* and *parts-bends* axes reflects the nature of the visual search results. In order to quantify this relationship, in Fig. 31 we plot the visual search times in Experiments I and II, averaged over the three subjects for a fixed set size of 16 distractors (see Figs. 12 and 15), but now as a function of the 2-shock/4-shock ratio differences between each target-distractor pair. The Pearson correlation coefficients for the upper and lower curves are $r = -0.965$, $-0.930$ (top) and $r = -0.967$, $-0.914$ (bottom). The strong (negative) correlation supports our hypothesis that discrimination along the *parts-protrusions* and *parts-bends* axes is primarily based on differences be-

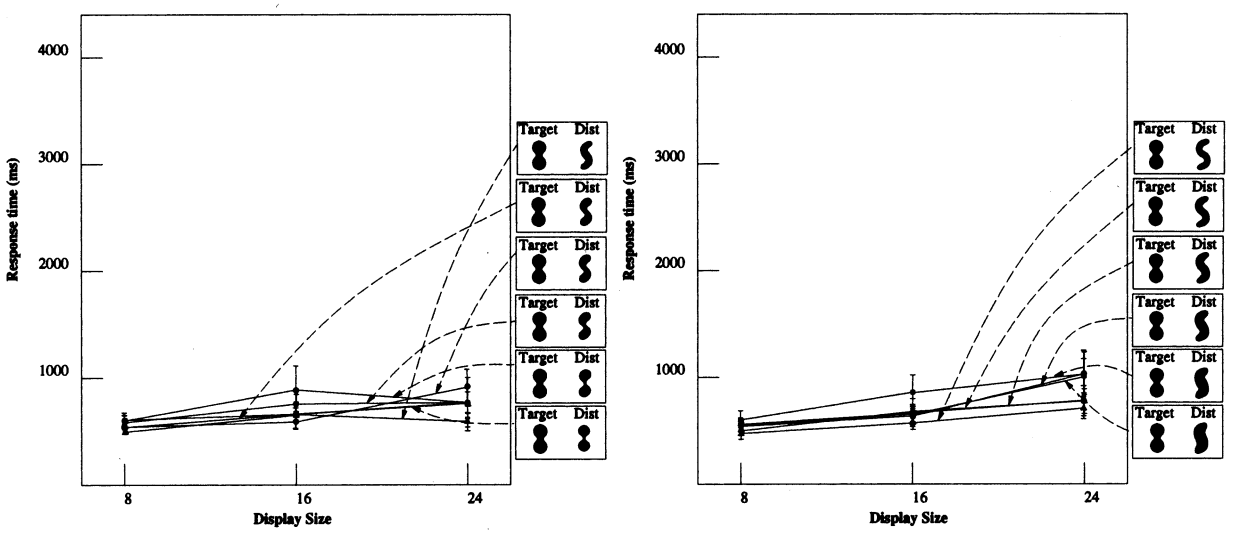


Fig. 26. The target was selected from the middle of the *parts-protrusions* axis, and the distractor was selected from the *parts-bends* axis (left) or the *bends-protrusions* axis (right).

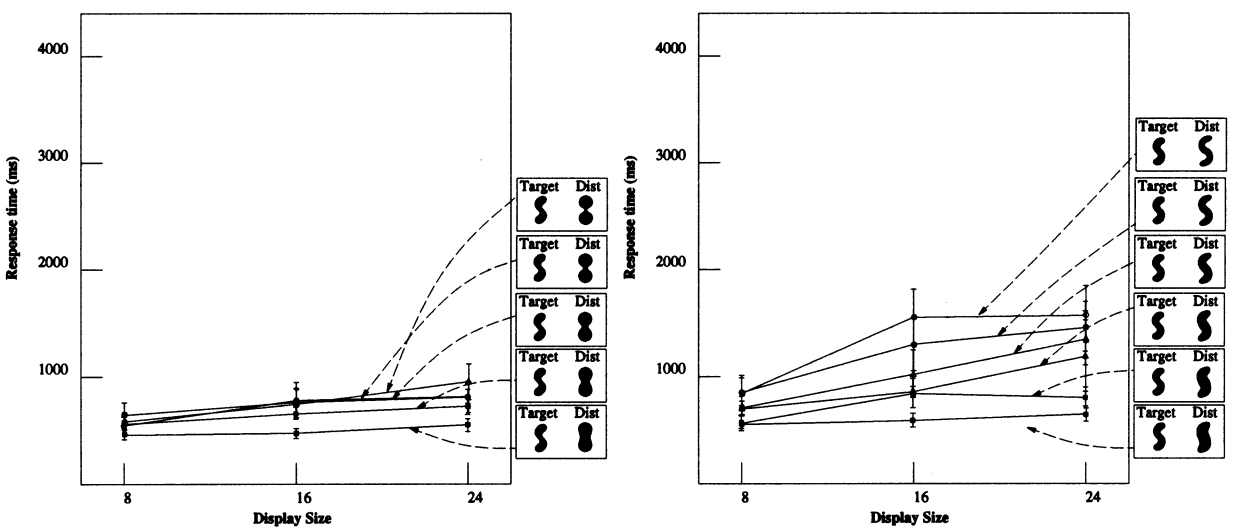


Fig. 27. The target was selected from the middle of the *parts-bends* axis, and the distractor was selected from the *parts-protrusions* axis (left) or the *bends-protrusions* axis (right).

tween the minimum (local) width to maximum (local) width ratios of the underlying stimuli. In particular, as the difference in this ratio increases, discrimination becomes easier and the visual search time decreases.

**Experiment IV: symmetry between targets and distractors** 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 (Observer 1) that included complementary trials in which the target and distractors were interchanged. The results, shown in Figs. 32 and 33, verify the prediction.

On the other hand, for the *protrusions-bends* sequence the minimum width to maximum width ratio approaches 1, Fig. 30, 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 & 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 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 (wig-

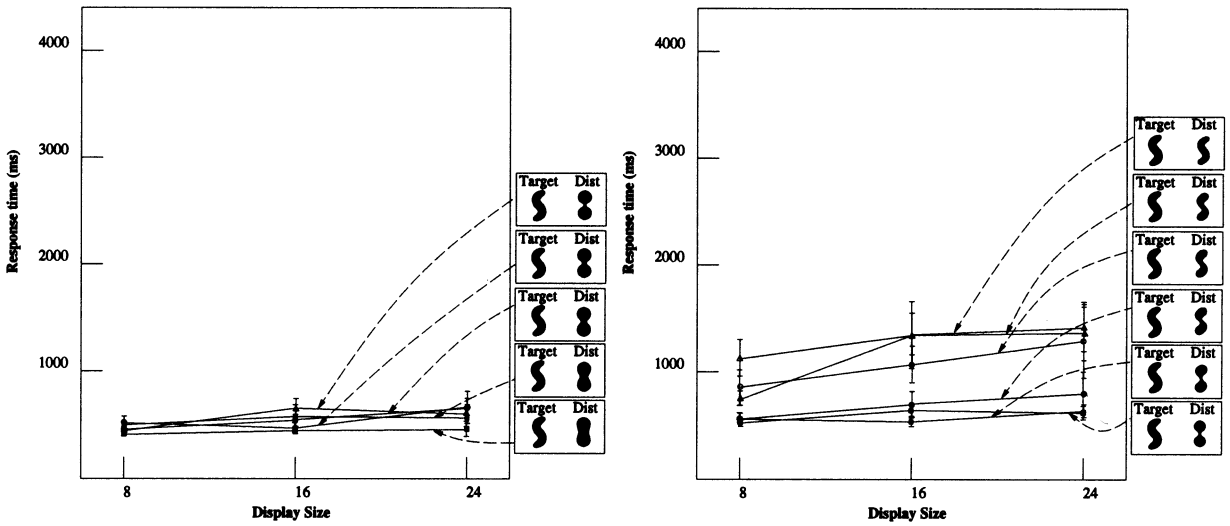


Fig. 28. The target was selected from the middle of the *bends-protrusions* axis, and the distractor was selected from the *parts-protrusions* axis (left) or the *parts-bends* axis (right).

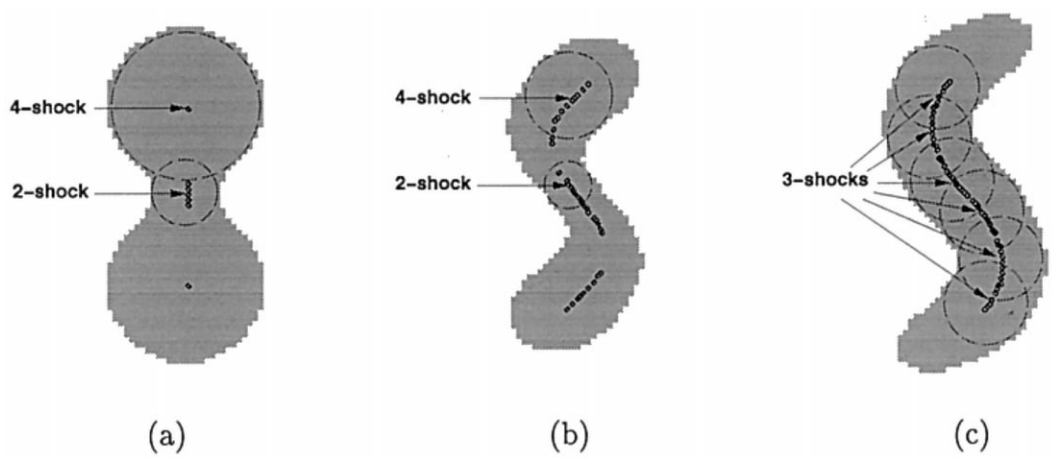


Fig. 29. The high-order shocks of samples taken from (a) the *parts-protrusions* axis; (b) the *parts-bends* axis; and (c) the *protrusions-bends* axis, in Fig. 8. These were computed using the algorithm developed in Siddiqi and Kimia (1996). 2-shocks occur at (local) minima in object width, 3-shocks at regions of constant width and 4-shocks at (local) maxima in width, as indicated by the inscribed discs, see Appendix A.

gles) closely resemble members of the *bends-protrusions* sequence in Fig. 17. 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 et al. (Morse, Pizer, & Burbeck, 1994;Burbeck & 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 & 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 Fig. 17. In the following we present 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.

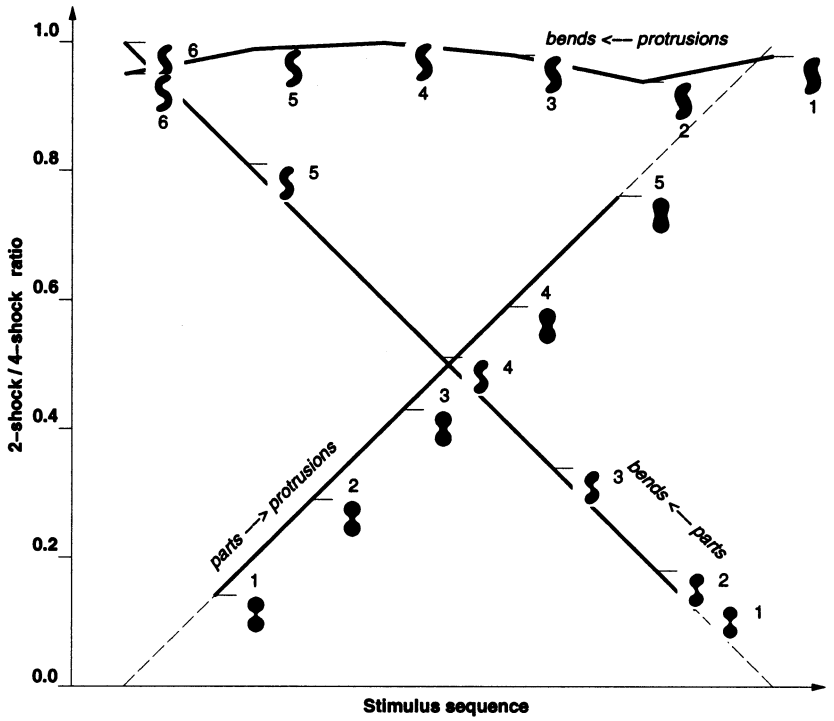


Fig. 30. The ratio of radii of the maximal inscribed discs at the underlying 2-shocks and 4-shocks is plotted for the three sequences in Figs. 10, 13 and 16.

### 3.1. Summary of experimental findings

The stimuli consisted of rectangles subtending 4° of visual arc in height, with sinusoidal edge modulation, Fig. 34 (left). Two widths were considered (0.75° and 1.5°) and for each width there were six edge modulation frequencies (0.25, 0.5, 1, 2, 4, 8 cycles/degree) and two 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 Fig. 34 for a period of one second from a distance of 1.5 m. 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 judgment your visual system is biased towards acquiring edge information across a more global spatial extent.

Burbeck and Pizer (1995) 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], 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, Fig. 36, 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).

[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 to the right of the object's center.

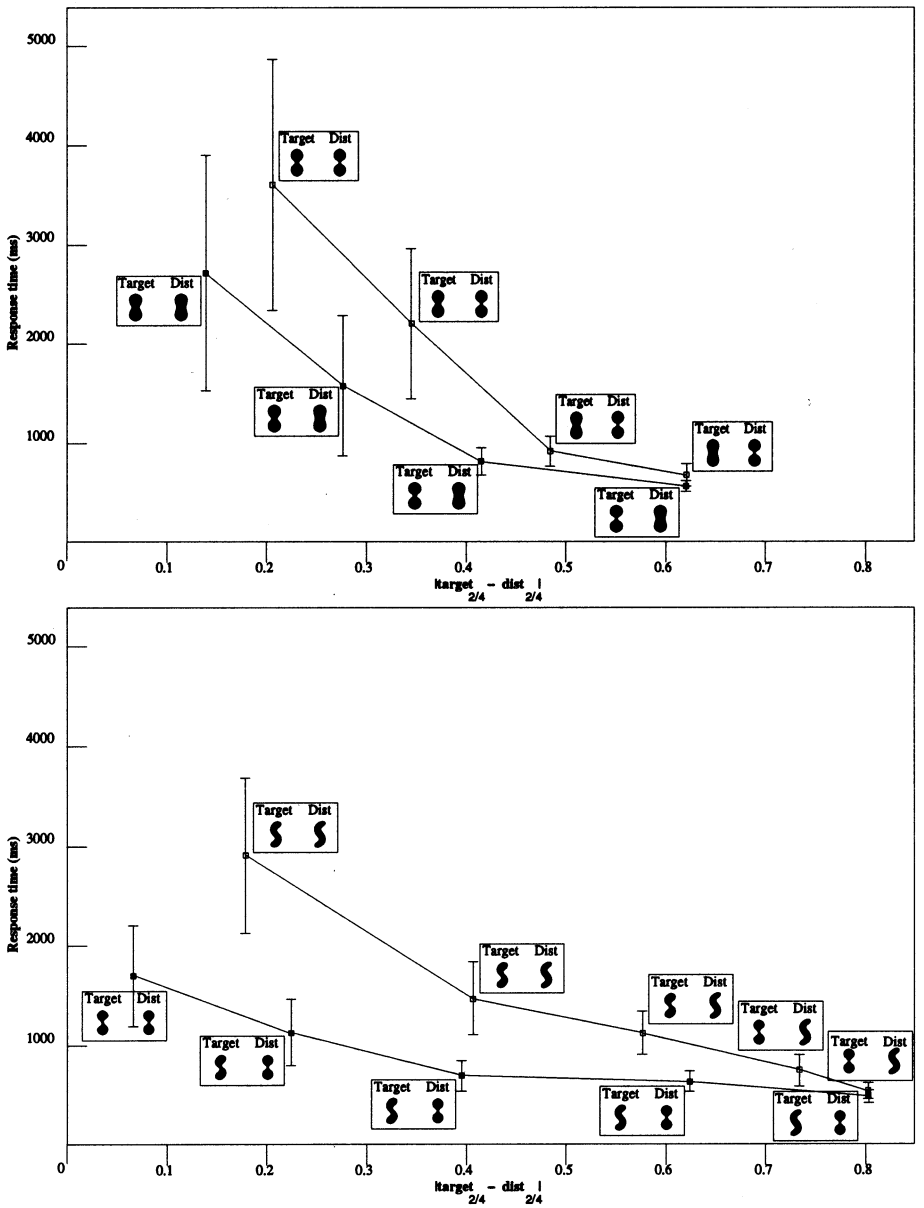


Fig. 31. The visual search times in Experiments I and II, averaged over the three subjects for a fixed set size of 16 distractors (see Figs 12 and 15), are now plotted as a function of the 2-shock/4-shock ratio differences between each target-distractor pair. Top: data for the *parts-protrusions* experiment; the Pearson correlation coefficients for the top and bottom curves are $r = -0.965$ and $r = -0.930$. Bottom: data for the *parts-bends* experiment; the Pearson correlation coefficients for the top and bottom curves are $r = -0.967$ and $r = -0.914$.

### 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 Fig. 35, with the geometry of the high-order shocks explained in Fig. 36. As evidence in support of our hypothesis, consider the computed central modulations overlaid as solid lines on the individual observer data taken from Burbeck et al. (1996), in Fig. 37. The Pearson correlation coefficients, shown in Table 1, indicate that the psychophysical and computational data are strongly correlated.

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, which corresponds to width of the object at that location, to the time of

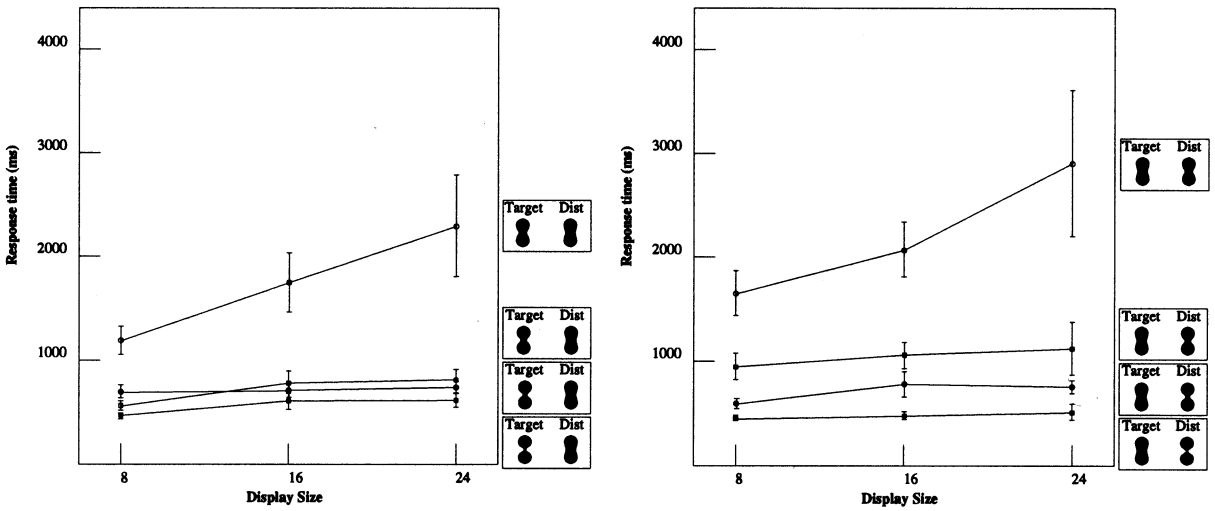


Fig. 32. Testing for symmetry along the *parts-protrusions* axis for a single subject (Observer 1): complementary trials are included in which the target and distractors are interchanged (right).

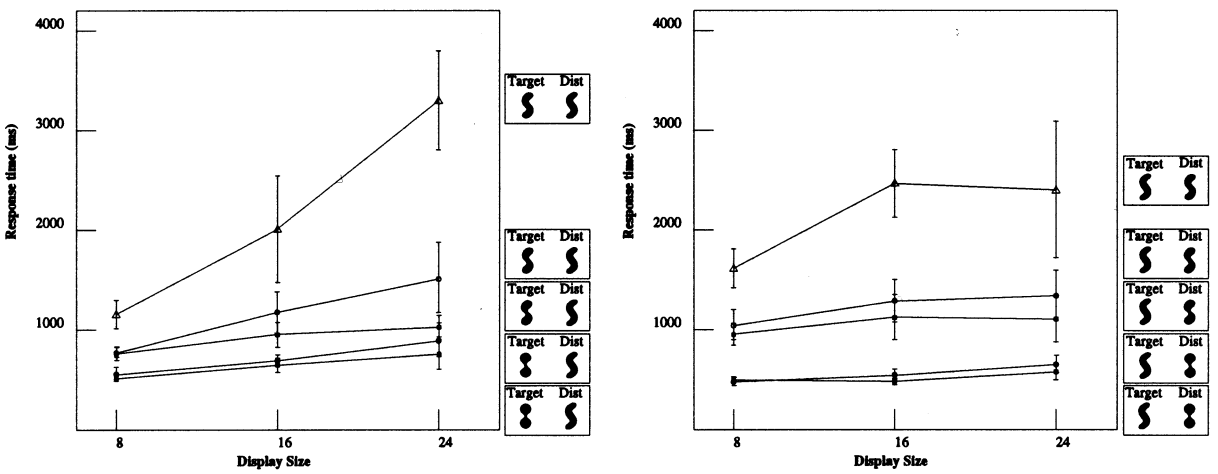


Fig. 33. Testing for symmetry along the *parts-bends* axis for a single subject (Observer 1): complementary trials are included in which the target and distractors are interchanged (right).

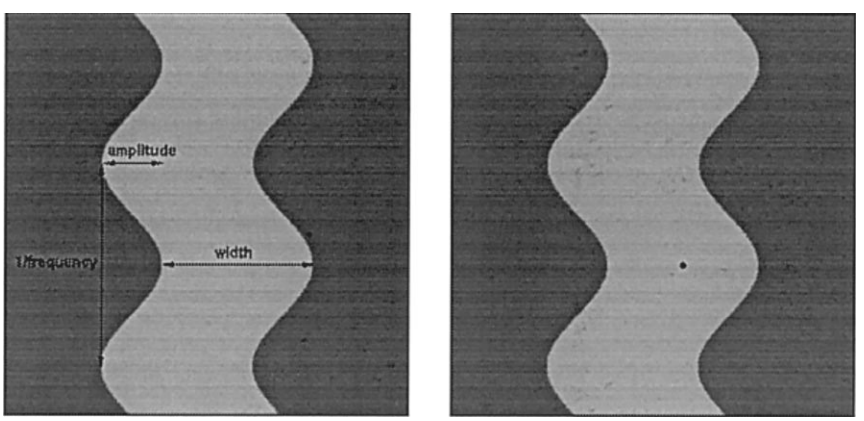


Fig. 34. The geometry of a 'wiggle' stimulus. Right: is the dot to the left or to the right of the object's center?

formation of its last 4-shock, which coincides with the maximal width of the object. 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 the skeleton as a directed graph. Curiously, the wealth of computational literature on the medial axis, with the exception of (Leyton, 1987, 1988), has focussed primarily on the latter quantitative information, (e.g. Arcelli, Cordella, & Leviadi, 1981; Brady & Asada, 1984; Pizer, Oliver, & Bloomberg, 1987; Leymarie & Levine, 1992; Arcelli & di Baja, 1993; Ogniewicz, 1993; Rom & Medioni, 1993; Kelly & Levine, 1995). In addition, some physiological evidence (Lamme, 1995; Lee, Mumford, & Schiller, 1995; Lee, 1996) as well as psychophysical evidence (Kovács & Julesz, 1993, 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.

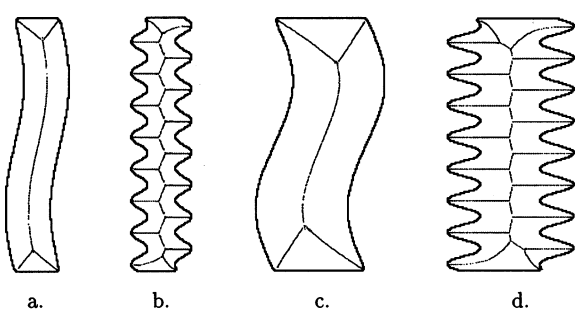


Fig. 35. The shock-based description of selected 40% amplitude modulation stimuli used in Burbeck et al. (1996), computed using the algorithm in Siddiqi and Kimia (1996). (a) 0.75° width, 0.25 cycles/degree edge modulation; (b) 0.75° object, 2.0 cycles/degree edge modulation; (c) 1.5° object, 0.25 cycles/degree edge modulation; and (d) 1.5° object, 2.0 cycles/degree edge modulation.

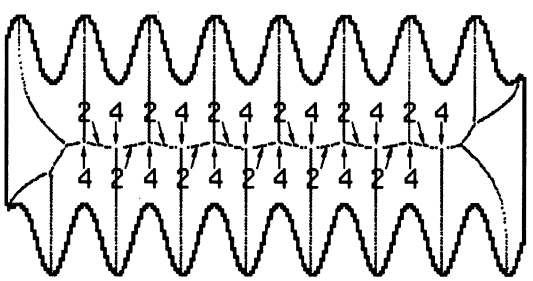


Fig. 36. Shape (d) from Fig. 35 is rotated and second-order and fourth-order shocks are labeled (all other shocks are first-order). Note that the fourth-order shocks are in alignment with the sinusoidal peaks.

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 & 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

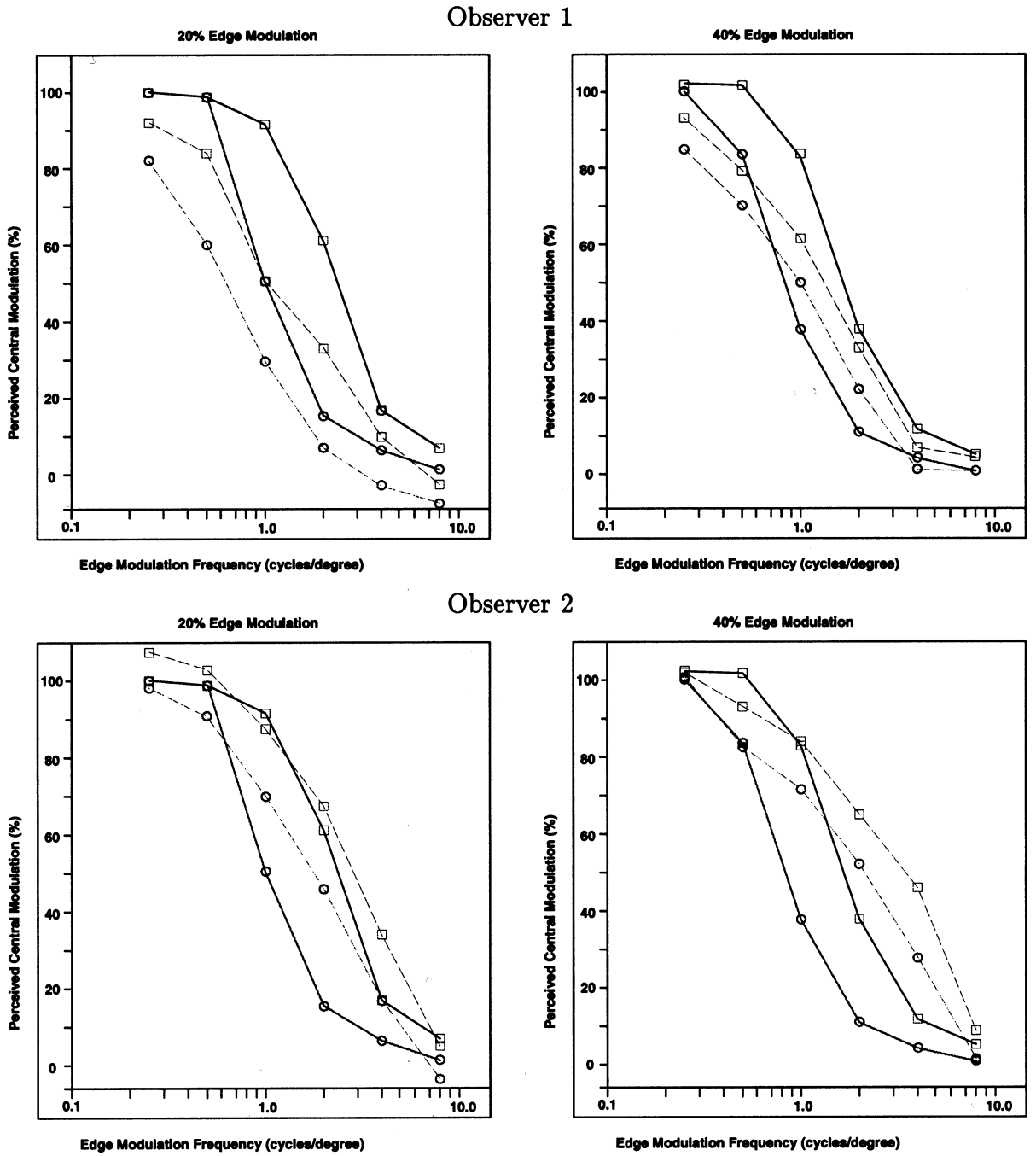


Fig. 37. Central modulations computed from shock-based descriptions (solid lines) are overlayed on the individual observer data (dashed lines) reproduced from Burbeck et al. (1996). The central modulations are expressed as a percentage of the edge modulation amplitude and are plotted against edge modulation 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 Pearson correlation coefficients between the psychophysical and computational data are shown in Table 1.

relate them to the natural components of shape. One key formal result is that the locus of positions at which shocks form 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

Table 1
The Pearson correlation coefficients between the individual observer data and the shock-based predictions in Fig. 37. It is evident that the psychophysical and computational data are strongly correlated.

| | 20% edge modulation | 40% edge modulation |
|---|---|---|
| Observer 1, 0.75° object | $r = 0.942$ | $r = 0.989$ |
| Observer 1, 1.5° object | $r = 0.983$ | $r = 0.972$ |
| Observer 2, 0.75° object | $r = 0.983$ | $r = 0.936$ |
| Observer 2, 1.5° object | $r = 0.950$ | $r = 0.897$ |

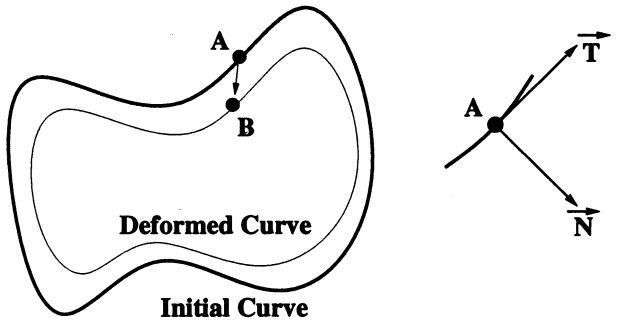


Fig. 39. The deformation of an initial curve is described by the displacement of each point in the tangential and normal directions (Kimia et al., 1990, 1995).

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, Fehér, and Julesz (1998) bear a close resemblance to high-order (3 and 4) shocks.

The psychophysical predictive power of shock-based theories for shape raises the question of whether there is any natural substrate for computing them physiologically. While it is clearly too early to answer such a question completely, we can now report a first cut at a preliminary, but necessary, part of the answer. If complex numerical routines were required, such as the Osher-Sethian algorithm (Osher & Sethian, 1988) which is now used for most simulations of curve evolution, then it would be difficult to argue biological plausibility. Similar statements could be made about the matching portion of the task for shape recognition: if delicate combinatorial algorithms were required for tree isomorphism, then again biological plausibility would seem remote. However, we have recently been able to formulate both the curve evolution (Siddiqi, Bouix, Tannenbau, & Zucker, 1999a) and the matching (Pelillo et al., 1999) in terms of simple dynamical systems, which at least suggests that a biological substrate is not totally implausible; for implementation of related equations that have been developed in some detail, see (Miller & Zucker, 1999). Much remains to be done along these lines, especially on how such computations might be implemented in higher visual areas such as IT. The psychophysical results in this paper should stimulate more active research on these fronts.

As a final example, consider the set of ellipses shown in Fig. 38. These stimuli are representative of a class of simple symmetrical patterns that have been used by Li and Westheimer (1997) to investigate orientation discrimination in humans. The overall finding is that hyperacuity effects in thresholds for detecting the angle of rotation about the vertical, shown earlier for straight lines (Westheimer, Shimamura, & McKee, 1976), apply also to stimuli whose implicit orientation is not primarily reflected by the contours that outline them. Specifically, the data shows that orientation discrimination is

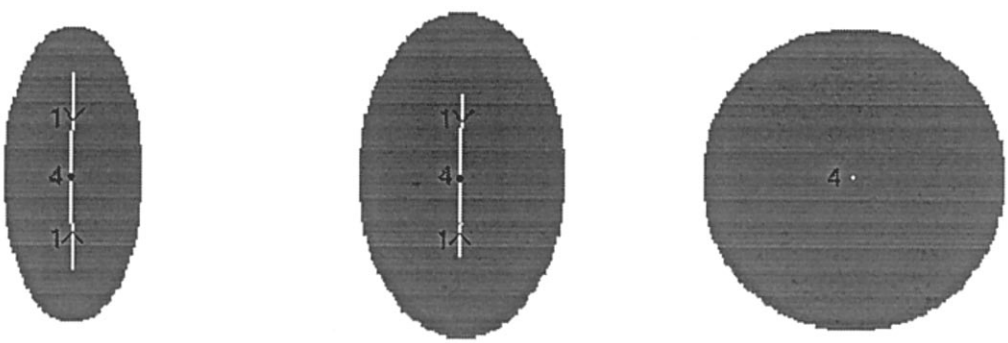


Fig. 38. The shock-based descriptions of three ellipses, adapted from Li and Westheimer (1997). The first two ellipses exhibit hyperacuity effects in orientation discrimination tasks about the vertical axis (see text). Note that these two stimuli also have significant 1-shock or *protrusion* branches driving into the central 4-shock; for the perfect circle the 1-shock branches are absent.

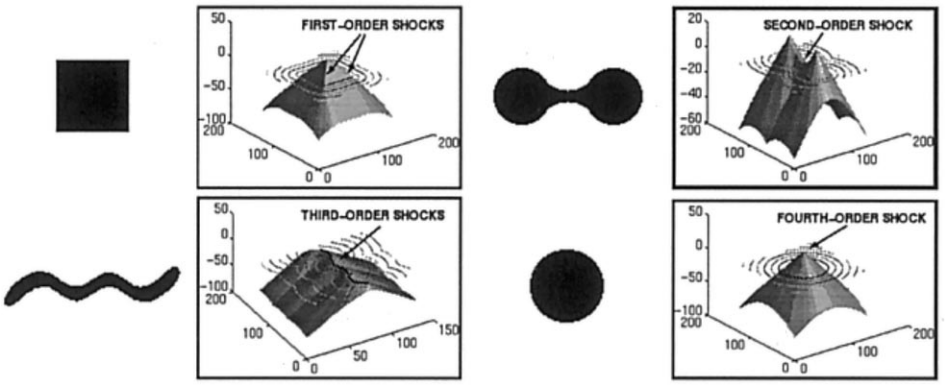


Fig. 40. A classification of shock types based on properties of the embedding surface $\Psi$ (in this case the signed distance function, shown as a shaded surface). Top left: first-order shocks occur at corners of the square shape, corresponding to creases on the surface with $|\nabla\Psi| > 0$. Top 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 & Kimia, 1996)

identical[4] for ellipses with aspect ratios between 0 and 0.5 (e.g. the first two ellipses in Fig. 38), and that of a vertical line, and then continuously degrades (there is no orientation discrimination for a perfect circle). We note that this is exactly what a shock-based representation would predict, where the implicit orientation would be determined by a measure of the length of the 1-shock branches driving into the central 4-shock; for the perfect circle these are absent. This view is consistent with the finding that orientation discrimination for lines improves with line length (Westheimer, 1981).

One of the motivations for choosing the stimuli along the axes of a shape triangle is that for these sequences the shock-based predictions can be related to easily measurable quantities such as the ratio of minimum to maximum local width. For other stimulus sets we expect new performance differences to emerge that can also be related to shocks. Computational experiments in shape recognition (Sharvit et al., 1998; Pelillo et al., 1999; Siddiqi et al., 1999b) 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.

## Acknowledgements

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, from FCAR Québec, from the National Science Foundation, from the Air Force Office of Scientific Research and from the Army Research Office. We are grateful to the reviewers for many helpful comments which have greatly improved the quality of the manuscript. 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 Fig. 1.

[4] The threshold for change detection for rotations about the vertical is less than 30 min of arc.

## Appendix 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 Fig. 39, 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, Tannenbaum, & Zucker, 1990;Kimia et al., 1995):

$$\mathscr{C}_t = (1 + \alpha\kappa)\mathscr{N} \qquad (1)$$

$$\mathscr{C}(p, 0) = \mathscr{C}_0(p).$$

Here $\mathscr{C}(p, t)$ is the vector of curve coordinates, $\mathscr{N}(p, t)$ 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 $\kappa$. When $\alpha$ 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 at which shocks form is related to Blum's grassfire transformation (Brockett & 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 (Fig. 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[5] 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 Eq. (1) is based on level set methods developed by Osher and Sethian (1988) (Sethian, 1996). The essential idea is to represent the curve $\mathscr{C}(p, t)$ 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 $\mathscr{C}(p, t)$ is on the zero level set, it satisfies

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

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

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

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

A convenient choice for $\Psi$ 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 $\Psi$ (see Fig. 40). These ideas have been developed to provide an algorithm for shock detection (Siddiqi & Kimia, 1996), which was used to compute the shock-based descriptions in Figs. 3, 5 and 35. An abstraction into a graph of shock groups provides a structural description for shape matching using graph theory (Sharvit et al., 1998; Tirthapura et al., 1998; Pelillo et al., 1999; Siddiqi et al., 1999b), see Fig. 6.

## References


Anstis, S. M. (1978). Apparent movement. In R. Held, H. Leibowitz, & H.-L. Teuber, *Handbook of sensory physiology*, vol. 8. Springer-Verlag.

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

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

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

Biederman, I., & 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., & 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. *Journal of Theoretical Biology*, *38*, 205–287.

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

Brockett, R., & 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., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of National Academy of Sicence USA*, *89*, 60–64.

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

Burbeck, C., & 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., & 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., & Rolland, J. (1996). Linking object boundaries at scale: a common mechanism for size and shape judgements. *Vision Research*, *36*(3), 361–372.

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., & 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., & 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., & 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., & 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*, *24*, 427–440.

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

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

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

Janssen, P., Vogels, R., & Orban, G. A. (2000). Selectivity for 3d shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, *288*, 2054–2056.

Kelly, M., & Levine, M.D. (1995). Annular symmetry operators: a method for locating and describing objects. In *Fifth International Conference on Computer Vision*, IEEE Computer Science Press, Cambridge, Massachusetts, 1016–1021.

Kimia, B. B. (1990). *Conservation laws and a theory of shape. Ph.D. dissertation*. McGill University, Montreal, Canada: McGill Centre for Intelligent Machines.


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

Kimia, B.B., Tannenbaum, A., & Zucker, S.W. (1990). Toward a computational theory of shape: an overview. *Proceedings of the First European Conference on Computer Vision*, Lecture notes in Computer Science (LNCS) Vol. 427, Antibes, France, 402–407.

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

Kovács, I., & 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.

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

Kovács, I., Fehér, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, *38*(15–16), 2323–2333.

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 E. H. Zarantonello (Ed.), *Contributions to Nonlinear Functional Analysis* (pp. 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., & 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., & 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.

Li, W., & Westheimer, G. (1997). Human discrimination of the implicit orientation of simple symmetrical patterns. *Vision Research*, *37*(5), 565–572.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

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

Marr, D., & 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.

Miller, D., & Zucker, S. W. (1999). Computing with self-excitatory cliques: a model and an application to hyperacuity-scale computation in visual cortex. *Neural Computation*, *11*(1), 21–66.

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

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

Osher, S., & Sethian, J. (1988). Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, *79*, 12–49.

Pelillo, M., Siddiqi, K., & Zucker, S. W. (1999). Matching hierarchical structures using association graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(11), 1105–1120.

Pizer, S. M., Oliver, W. R., & 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., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.

Rom, H., & 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 E. Rosch, & B. B. Lloyd (Ed.), *Cognition and categorization*. L. Erlbaum Associates.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & 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.

Sharvit, D., Chan, J., Tek, H., & Kimia, B. B. (1998). Symmetry-based indexing of image databases. *Journal of Visual Communication and Image Representation*, *9*(4), 366–380.

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

Siddiqi, K., Bouix, S., Tannenbaum, A., & Zucker, S. W. (1999a). The hamilton-jacobi skeleton. In *Proceedings of the Seventh International Conference on Computer Vision*, *vol. 2* (pp. 828–834) Kerkyra, Greece.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999b). Shock graphs and shape matching. *International Journal of Computer Vision*, *35*(1), 13–32.

Sklar, E., Bulthoff, H.H., Edelman, S., & Basri, R. (1993). Generalization of object recognition across stimulus rotation and deformation. In *Investigative opthalmology and visual science*, *vol. 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., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, *21*, 233–283.

Tarr, M. J., & 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.

Tirthapura, S., Sharvit, D., Klein, P., & Kimia, B.B. (1998). Indexing based on edit-distance matching of shape graphs. In *SPIE Proceedings on Multimedia Storage and Archiving Systems III*, 25–36.

Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object recognition taks. *Vision Research*, *38*, 2335–2350.

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

Westheimer, G. (1981). Visual hyperacuity. In H. Autrum (Ed.), *Progress in sensory physiology*, vol. 1 (pp. 1–30). Springer.

Westheimer, G., Shimamura, K., & Mckee, S. P. (1976). Interference with line orientation sensitivity. *Journal of the Optical Society of America*, *66*, 332–338.