Preattentive Recovery of Three-Dimensional Orientation From Line Drawings

James T. Enns and Ronald A. Rensink
University of British Columbia, Vancouver BC, Canada

It has generally been assumed that rapid visual search is based on simple features and that spatial relations between features are irrelevant for this task. Seven experiments involving search for line drawings contradict this assumption; a major determinant of search is the presence of line junctions. Arrow- and Y-junctions were detected rapidly in isolation and when they were embedded in drawings of rectangular polyhedra. Search for T-junctions was considerably slower. Drawings containing T-junctions often gave rise to very slow search even when distinguishing arrow- or Y-junctions were present. This sensitivity to line relations suggests that preattentive processes can extract 3-dimensional orientation from line drawings. A computational model is outlined for how this may be accomplished in early human vision.

Although we are still a long way from a complete understanding of visual perception, considerable progress has been made in our understanding of its earliest stages (see Zucker, 1987). These stages are concerned with the extrac-tion of information from the retinal image, and as such are generally assumed to be carried out by processes operating in parallel across the visual field. They are also generally assumed to be based on simple geometric elements such as points and oriented bars.

In this article, we will show that this second assumption is too restrictive and must be replaced. Although early visual processes can make use of geometrically simple properties, we will show that they can also make use of more complex properties, provided that these are environmentally relevant and can be computed rapidly. This allows early visual processes to recover a number of properties of the three-dimensional scene, thereby facilitating the operation of processes further down the visual stream. To support this view, we will demonstrate that early vision can be influenced by spatial relations that are present in line drawings of simple objects. We then show that this sensitivity can be explained by a process that rapidly recovers three-dimensional orientation from the image.

Preattentive Vision

For the past two decades, most theories of vision have postulated the existence of two subsystems (Beck, 1982; Julesz, 1984; Neisser, 1967; Treisman, 1986; Treisman, Cavanagh, Fischer, Ramachandran, & von der Heydt, 1990). The first is a *preattentive* system that registers simple features of the two-dimensional image (e.g., orientation, length, curvature, and color) rapidly and in parallel across the visual field. These features are often taken to be the basic elements of human vision. Indeed, it has been suggested that the registration of such features is carried out at the earliest stages of visual processing in the cortex (Treisman et al., 1990).

The high speed of the preattentive system, however, is obtained at the cost of a fragmented representation, i.e., one in which different features are represented in different spatio-topic maps. Among other things, this rules out any explicit representation of the spatial relations among the different features in an image. This in turn prohibits the preattentive computation of any property dependent on these relations.

In order to overcome these limitations, a second subsystem of *attentive* vision must be postulated. This system is capable of applying a large set of operations to the representations at the preattentive level. For instance, it can establish the spatial relations that exist between different features. The cost of this flexibility, however, is that only a small region of space can be examined at any given time and so serial inspection is required to process all the information in an image.

Although there is general consensus on the existence of these two systems, there is less agreement on the specific mechanisms used. For example, proposals for the operation of the preattentive system range from spatial filtering (Fogel & Sagi, 1989; Gurnsey & Browse, 1989; Sutter, Beck, & Graham, 1989; Watt, 1987), to local detection of particular features (Treisman, 1986), to statistical measures of feature density in image regions (Julesz, 1984).

One of the main psychophysical tasks used to explore preattentive vision is visual search (Neisser, 1967; Schneider & Shiffrin, 1977; Treisman, 1986). In this task, observers try

to determine as rapidly as possible whether a target item is present or absent in a display. Target items that are detected rapidly and show little dependence on the total number of items in the display are assumed to contain a distinctive feature at the preattentive level. No attentive operations are required for their detection—the target simply "pops out" of the display. In contrast, other targets are more difficult to find, with search time depending strongly on the total number of items in the display. These targets are considered to be conjunctions of elementary features, requiring the serial operations of the attentive system for their detection.

Recent findings have shown that this picture of visual processing is too simple and must be revised in several important ways. To begin with, the dichotomy of serial and parallel search is challenged by the observation that a continuum of search rates exists, ranging from very fast (i. e., less than 10 ms per item) to very slow (i. e., more than 100 ms per item). Several attempts have been made to account for this finding while still holding to two separate subsystems (Julesz, 1986; Treisman & Souther, 1985). These efforts are now leading to proposals that search rates reflect processes that vary in speed as a function of target and distractor similarity (Duncan & Humphreys, 1989; Humphreys, Quinlan, & Riddoch, 1989; Treisman & Gormican, 1988).

Recent reports also argue that the representations at preattentive levels are much more complex than suggested by the conventional view. For example, rapid search is possible for targets defined by the conjunction of binocular disparity and motion (Nakayama & Silverman, 1986); by the conjunction of motion and form (McLeod, Driver, & Crisp, 1988); and by the conjunction of color, form, and orientation, provided that maximally different values are chosen within a dimension (Treisman, 1988; Wolfe, Cave, & Franzel, 1989). Furthermore, rapid detection of simple line relations is sometimes possible, provided that the lines are sufficiently long (Duncan & Humphreys, 1989; Humphreys et al., 1989).

Another recent challenge to the conventional view is the discovery that preattentive vision is sensitive to aspects of the three-dimensional scene corresponding to the two-dimensional image (Enns, 1990; Enns & Rensink, 1990a, 1990b; Epstein & Babler, 1990; Holliday & Braddick, 1989; Ramachandran, 1988; Ramachandran & Plummer, 1989). For example, subjects in the Enns and Rensink (1990b) study searched for target items defined only by the spatial relations between the constituent lines of simple drawings. Search was rapid when items could be interpreted as simple blocks with different three-dimensional orientations in the scene. Moreover, three-dimensional orientation was just as effective as two-dimensional orientation in directing search.

One puzzling aspect of these data was that rapid search occurred for some line drawings but not for others. Search was rapid for drawings of simple convex blocks, but not for drawings of U-shaped brackets that had equivalent differences in three-dimensional orientation. Why should this be? Is this a result of the interpretation given to the depicted objects, or is it a result of simple geometric operations performed on the image itself?

Scene-Based Versus Image-Based Properties

An important distinction for what follows is that between the world of objects in three-dimensional space (hereafter called the scene) and the two-dimensional array of light intensities projected from the scene to an observer (the image). In general, if a set of opaque objects is illuminated by a distant point source, the two-dimensional array of image intensities is completely determined by four factors: (a) direction of lighting, (b) surface orientation and location, (c) surface reflectance, and (d) viewing direction.

Strictly speaking, the complete recovery of all these quantities from a single image is impossible, because there exists a large equivalence class of scenes that gives rise to any particular image. In order to recover a unique set of values it is necessary to impose a set of constraints on the set of acceptable scenes. If the constraints are well chosen, they will cause one candidate to be selected from each equivalence class, thereby establishing a one-to-one correspondence between scene and image.

The scene-recovery problem can also be simplified by reducing the number of scene properties to be recovered. In the simplest case, all but a few scene quantities are fixed at a known value and the recovery process reduces to the calculation of the remaining parameters. Constraints of this kind translate directly into constraints on the kinds of images possible. Thus, any analysis of scene recovery must begin by describing (a) the constraints on the scene domain, (b) the constraints on the image domain, and (c) the correspondences between image and scene properties.

The Blocks World

Given this framework, what is the best choice of scene and image domain for studying the recovery of three-dimensional structure from line drawings? This question has received a great deal of attention in the field of computational vision. Early work (e.g., Roberts, 1965) attempted to analyze scenes containing a small set of known polyhedra. Recognition proceeded by using knowledge of these polyhedra to determine which image regions corresponded to which surfaces. Guzman (1968) showed that a priori knowledge of object shape was not always required to extract three-dimensional structure—this could be recovered using only the structural relations existing among lines in the image.

Subsequent work (Clowes, 1971; Huffman, 1971; Mackworth, 1973; Waltz, 1972) put Guzman's approach on a more rigorous footing. These studies were based on the blocks world, a scene domain of polyhedral objects consisting only of trihedral corners (i. e., corners formed from three polygonal faces). The corresponding image domain was formed by the orthographic projection of these objects onto the image plane. As such, it consisted of straight-line segments connected by dilinear or trilinear junctions. By using line drawings to represent the objects in the image, objects were assumed to have uniform reflectances on all visible surfaces. Furthermore, viewing direction and the direction of lighting were held constant, with the two directions being made coincident in order to avoid shadows. This left surface orientations and locations as the only variable scene properties.

Studies of the blocks world all began with the observation that each line in the image corresponds to one of three different kinds of edge in the scene: convex, concave, or object boundary. The first two kinds are formed by the intersection of two adjacent planar faces, whereas the third is formed from the boundary of a face that occludes a second, noncontiguous surface or background. To interpret a line drawing correctly, then, it is necessary to label each line as corresponding to a particular edge type, with the labeling being consistent for all lines in the image.

Several algorithms to carry out the line-labeling process have been developed (e.g., Horn, 1986; Mulder & Dawson, 1990; Waltz, 1972). These all rely on the fact that three kinds of trilinear junctions are possible in an image: arrow-junctions, in which the greatest angle between two lines is greater than 180°; Y-junctions, in which the greatest angle is less than 180°; and T-junctions, in which this angle is exactly 180° (see Figure 1). These junctions correspond to different aspects of the scene: Arrow-junctions correspond to corners with two visible faces; Y-junctions, to corners formed from two or three visible faces; and T-junctions, to surface relations of occlusion or accidental alignment (i.e., where one face is parallel to the line of sight). There also exists a fourth class of dilinear junctions, L-junctions, that correspond to corners of single visible faces.

As is evident in Figure 1, each   trilinear junction may correspond to more than one kind of corner in the scene. The interpretation process proceeds by incrementally eliminating junction interpretations that are inconsistent with those of their neighbors.   This process is iterated until the interpretation at each junction in the drawing is consistent with those at all other junctions.

The three trilinear junctions differ in the kind of   information they carry about the three-dimensional orientation of the corresponding corner. T-junctions most often correspond to

occlusion, and as such, will signal only that the two corresponding surfaces differ in their relative depth. Consequently, they carry no quantitative information about surface orientation. However, arrow- and Y-junctions can be used to recover the orientations of the surfaces at the corresponding corner, provided that the surfaces are mutually orthogonal to one another. Perkins's (1968) law states that for an arrow-junction corresponding to an orthogonal corner, no angle can be greater than 270°; for Y-junctions, no angle can be less than 90°. Perkins (1968) also showed that if corners are assumed to be orthogonal, their three-dimensional orientations can be calculated from the angles about the arrow- and Y-junctions (see also the Appendix; Mackworth, 1976). Mulder and Dawson (1990) have extended these ideas recently, showing that estimates based on orthogonal corners sometimes determine the three-dimensional orientations of nonorthogonal corners as well.

It is important to note that although the foregoing constraints are necessary for the recovery of three-dimensional orientation from a junction, they are not sufficient. This is well-illustrated by the Y-junction in the pyramid (Figure 1). This junction is consistent with Perkins's (1968) laws but does not, in fact, correspond to an orthogonal corner. Similar considerations apply to arrow junctions. For the complete recovery of object structure, the whole system of line relations must be examined.

Scope of the Study

We now examine the sensitivity of preattentive vision to the kinds of spatial relations found in line drawings. In particular, we will consider drawings composed of straight-line segments in which no more than three lines meet at any junction. This scene domain is somewhat larger than the world of polyhedral objects. Among other things, it contains objects formed from polygonal plates (e.g., the bracket in Figure 1) and objects with free ends (e.g., the isolated junctions in Figure 3). Although still a relatively restricted domain, it is larger than the polyhedral world, allowing more scope to explore issues such as the role of individual line junctions in the recovery process.

Our preceding discussion of the blocks world has suggested that there are at least two abilities the preattentive system must have if it is able to recover three-dimensional orientation from a line drawing. First, as described in the previous section, the correct interpretation of any junction in an item cannot be determined locally—the entire system of line relations must be considered. This is illustrated in Figure 1. Although three distinct classes of junctions can be seen, there is no unique correspondence between a given junction and its three-dimensional interpretation. Thus, one issue to be examined is whether preattentive vision is sensitive to the system of relations within an item.

The second ability concerns the recovery of   three dimensional orientation from individual corners. We have seen that it is formally possible to recover a unique orientation from any arrow- or Y-junction, provided that it corresponds to a corner composed of three orthogonal surfaces (see Appendix; Perkins, 1968). However, it remains to be seen whether the preattentive system can do so rapidly by taking advantage of the orthogonality constraint.
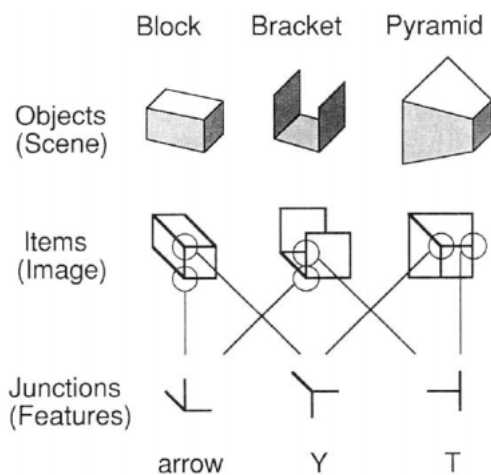

Figure 1. In line drawings of polyhedral objects, lines represent edges and bounded regions represent planar surfaces. (Junctions involving three lines fall into three classes: arrow-, Y-, and T-junctions. However, there is no unique correspondence between a given junction and its correct three-dimensional interpretation. This can only be determined by considering the system of junctions in an item.)

We will examine these questions by measuring observers' response times (RTs) to the presence and absence of targets in a visual search task. Following additive factors logic (Sternberg, 1969), we assume that the subjects' decision and response processes are reflected in the intercept of an RT function plotted against display size; the RT slope therefore is a measure of encoding and comparison processes. Although many models of visual search make detailed and competing claims about these processes (e.g., Duncan & Humphreys, 1989; Pashler, 1987; Townsend & Ashby, 1983; Treisman & Gormican, 1988), our primary questions are not dependent on any particular model. Rather, they can be answered quite simply by examining which spatial rela-tions give rise to relatively rapid search rates. As pointed out earlier, there is no sharp boundary between fast and slow rates of search. In this article, we will follow the convention of using "rapid search" to refer to target-present search rates (RT slopes) of faster than 10-15 ms per item. This speed is well below accepted estimates of attentional movement across the visual field (Jolicoeur, Ullman, & MacKay, 1986; Julesz, 1984; Treisman & Gormican, 1988).

Method

In each set of experiments, target and distractor items were composed of identical line segments that differed only in their spatial arrangement (see Figures 1-8). The methods used in the visual search task were similar to others in the literature (e.g., Treisman, 1988; Treisman & Gormican, 1988; Wolfe et al. , 1989), with observers searching for a single target item among a total of 1, 6, or 12 items. The target was present on half the trials, randomly distributed throughout the trial sequence. On each trial, items were dis-tributed randomly on an imaginary 4 x 6 grid subtending 10° x 15°. Each item subtended less than 1. 5° in any direction and was randomly jittered in its grid locations by ± 0.5° to prevent search being based on item collinearity.

A Macintosh computer was used to generate the displays, control the experiments, and collect the data (Enns, Ochs, & Rensink, 1990). Each trial began with a fixation symbol lit for 500 ms, followed by the display, which remained visible until the observer responded. The display was followed by a feedback symbol (plus or minus sign), which served as the fixation point for the next trial. Target presence or absence was reported as rapidly as possible by pressing one of two response keys. Observers were instructed to maintain fixation and to keep errors below 10%.

Ten observers with normal or corrected-to-normal vision completed four to six sets of 60 test trials in each condition of a given experiment. These observers were equally divided between those who had no previous experience in visual search tasks and those who ran routinely in other search experiments.

Data Analyses

Although observers were quite accurate overall (each made fewer than 10% errors on average), there were systematic differences in accuracy. Consistent with other reports, target-present trials led to more errors than target-absent trials (Humphreys et al., 1989; Klein & Farrell, 1989). Most im-portant for present purposes, however, was the observation that errors tended to increase with RT, indicating that observers were not simply trading accuracy for speed.

Only correct RT data were analyzed, and these were treated the same way in each experiment. First, simple regression lines were fit to the target-present and target-absent data for each observer (the average fit of these lines ranged from $r = .53$ to 1. 00 across conditions and experiments). Second, the estimated slope parameters were submitted to analyses of variance in which condition (A, B, etc.) and trial type (present or absent) were the effects of interest. Finally, Fisher's LSD tests determined the reliability of pairwise slope differences in the context of significant main effects and interactions. The reported t-tests, therefore, are tests of differences in RT slope based on the pooled error variance and degrees of freedom from the main analysis.

Experiment 1

The first experiment examined whether visual search was sensitive only to the most general kind of spatial relation among line segments—that of topology. It has been sugges-ted that topological relations between features can influence processing at the preattentive stage, and that these are the only kinds of relations to do so (Chen, 1982; 1990). If this is the case, rapid detection of particular line relations should be explicable purely in terms of topological considerations.

The items in Condition A of Experiment 1 (see Figure 2A) corresponded to simple blocks of different three-dimensional orientation, whereas the items in Condition B (Figure 2B) corresponded to truncated pyramids in which the line of sight was accidentally aligned with two of the surfaces. Quantitatively, the items in the two conditions differed considerably: The lines forming the L-junctions in Condition B were twice as long as those in Condition A, and two of the arrow-junctions in Condition A were replaced by T-junctions. Topologically, however, items in both conditions were the same. Thus, if topological relations alone are of relevance, the two conditions should give rise to similar search rates. If quantitative factors are also involved, one could expect a difference.
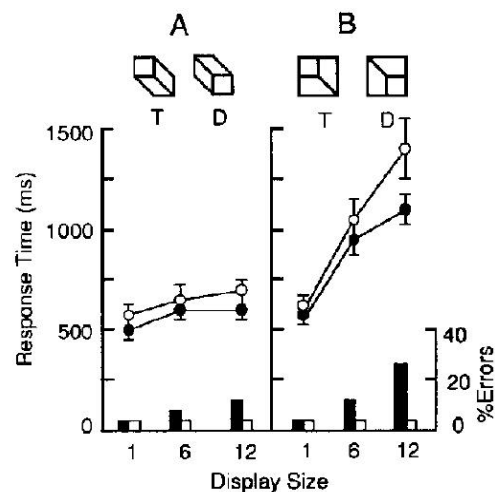

Figure 2. The target items (T), distractor items (D), and results in Experiment 1. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)

Consistent with the results of previous studies (Enns & Rensink, 1990b), search in Condition A was quite rapid (mean RT slopes 7 ms per item for target-present trials and 12 ms per item for target-absent trials). In contrast, search in Condition B was much slower: Mean RT slopes were 51 and 96 ms per item, t(18) = 8.19, *p* < .001, and t(18) = 15.63, p < . 001, respectively MSe = 144). This demonstrates that early vision is sensitive to more than just topological relations among line segments. Quantitative factors, such as the angles between connected lines, are also important.

Experiment 2

Are the differences between Conditions A and B of Experiment 1 attributable to the different kinds of line junctions in the items? To find out, we measured search rates for each of the trilinear junctions contained in these items. As shown in Figure 3, each pair of items consisted of the same line segments, so that the target and distractor differed only in the relations between segments. Note that the size of the stimulus was fairly small (1.5° arc) relative to the size of the overall array (10° x 15° arc), making it unlikely that rapid detection for spatial relations could be based on the size-eccentricity ratios found optimal by Humphreys et al. (1989).

Figure 3 shows the results. Arrow-junctions in Condition A yielded the fastest search rates (mean RT slopes were 10 and 13 ms per item). The Y-junctions in Condition B yielded significantly slower search rates: 18 and 24 ms per item, t(27) = 2.84, p < . 01, and t(27) = 3.91, p < .01, respectively MSe = 40). T-junctions in Condition C led to the slowest search rates of all: 37 and 66 ms per item, t(27) =6.75, p < . 001, and t(27) = 14.93, p < .001, respectively MSe = 40).

These results suggest that the difficulty of search in Experiment 1 was related directly to the kinds of junctions present in the items. The very slow search for the items in Condition B of Experiment 1 could therefore be attributed to the replacement of two arrow-junctions by two T-junctions. Curiously, search for these items was even slower than for any individual junction. This shows the existence of a strong context dependency for line junctions.
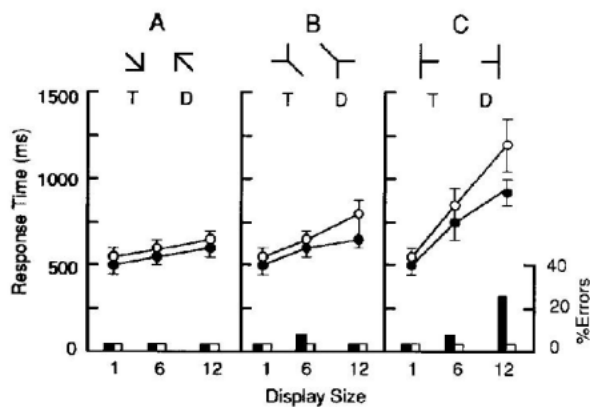
Experiment 3

To explore this context dependency more thoroughly, search rates were measured for the series of line drawings depicted in Figure 4. Here, targets always differed from distractors by a 180° rotation of the central Y-junction. These junctions were embedded in several different contexts in order to generate items that varied in the number and type of other junctions present.

In Condition A, Y-junctions were presented in isolation. Search rates were similar to those for the Y-junctions in Experiment 2 (20 and 29 ms per item, *p* > .05). In Condition B, one arrow-junction was added, resulting in a reliable speed-up in search on absent trials of 16 and 19 ms per item, t(45) = 0.82 and t(45) = 2.04, p < 05, respectively Mse = 121). In Condition C, two more arrow-junctions were added, but these did not alter search speed significantly: 16 and 21 ms per item, t(45) = 0.02 and t(45) = 0.41, respectively MSe = 121).

Interestingly, the presence of a single T-junction in the item slowed search dramatically. Condition D used items similar to those in Condition C, but with one of the arrow-junctions replaced by a T-junction. This replacement resulted in search speed being reduced by a factor of three: 42 and 67 ms per item, t(45) = 5.31, p < .001, and t(45) = 9.39, p < .001, respectively MSe = 121). Furthermore, when all three arrow-junctions were replaced by T-junctions in Condition E, search was even slower on absent trials: 48 and 77 ms per item, t(45) = 1. 22 and t(45) = 2. 04, p < . 05, respectively MSe = 121).

Taken together, these results confirm that the embedding context strongly affects the speed of search for particular line relations. Conditions B and C, along with Condition A of Experiment 1, show that if arrow- and Y-junctions are connected together, search is no slower than for any of the individual junctions. In contrast, the presence of a T-junction causes a striking slowdown in search rate, even though the items also contain arrow- and Y-junctions that by themselves distinguish the target from the distractor. This shows that search rates are influenced greatly by the entire system of line relations in an item.

Experiment 4

In the previous experiments, T-junctions usually corresponded to trihedral corners in which one or more of the faces were parallel to the line of sight. However, there also exist T-junctions that correspond to occlusions of one surface by another (see Figure 1). Do these different kinds of T-junctions have different effects on visual search? To answer this question, we used items corresponding to U-shaped brackets formed from three orthogonal plates, as shown in Figure 5.

In Condition A, targets and distractors differed only in the orientations of their arrow-junctions. In spite of a difference in the overall outline between target and distractor, search for these items was relatively slow (26 and 67 ms per item). However, these search rates were still considerably faster than for all previous items that contained T-junctions (Condition B in Experiment 1, target-present and target-absent p values < . 01; Conditions D and E in Experiment 3, target-present p values < . 01 and target-absent p values > . 05).

Figure 3. The target items (T), distractor items (D), and results in Experiment 2. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)
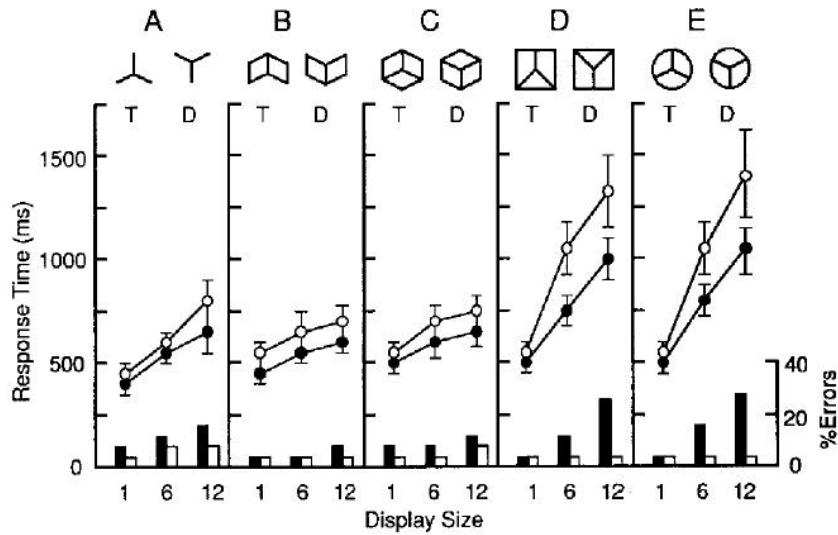
Figure 4. The target items (T), distractor items (D), and results in Experiment 3. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)

In Condition B, items differed in the orientations of their arrow- and T-junctions, increasing search speed by a factor of two: 15 and 32 ms per item, $t(18) = 3.01$, $p < .01$, and $t(18) = 9.59$, $p < .001$, respectively MSe = 67). The finding that T-junctions formed by occlusions can influence search in this way suggests that they have an effect markedly different from that of T-junctions formed by accidental alignment, which appear to actively interfere with search. We note also that search for occluding T-junctions, as for other junctions, is strongly influenced by the entire system of line relations in the item.

Figure 5. The target items (T), distractor items (D), and results in Experiment 4. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)

Experiment 5

Almost all of the arrow- and Y-junctions used so far have corresponded to corners formed from three orthogonal surfaces (see Figures 2, 3, 4 and 5). It can be shown mathematically that the three-dimensional orientation of these surfaces is recoverable from the two-dimensional orientations of lines about the junction (see the Appendix; Perkins, 1968). Interestingly, the items that we have found easiest to detect (e.g., Condition A in Experiment 1 and Conditions B and C in Experiment 3) have corresponded to objects with such corners. Furthermore, the truncated pyramid in Experiment 1 has a Y-junction that cannot correspond to such a corner, and search for that item is difficult. This suggests that the orthogonality of corners may be another factor influencing search rates.

We tested this hypothesis by using the items shown in Figure 6. In Condition A, items had the same outline as those in Experiment 1 (Condition A), but the smallest angle of the internal Y-junction was made less than 90°. This ruled out the possibility that the corresponding corner could be orthogonal. To control for the possible effects of the nonparallel orientations of the resulting lines, Condition B used drawings with similar Y-junctions, but in which parallel line orientations were maintained. These items had the same internal structure as those in Experiment 3 (Condition C), but the small angle of the Y-junction and the two wings of the arrow-junctions were both made less than 90°. As such, all items in Conditions A and B violated the orthogonality constraint.

In both conditions search was quite slow. Condition A resulted in mean search rates of 35 and 65 ms per item, whereas Condition B produced similar mean search rates of 37 and 66 ms per item, $t(18) = 0.15$ and $t(18) = 0.30$, respectively MSe = 266). Thus, it appears that the preattentive processes that extract orientation from line drawings can also detect when arrow- and Y-junctions violate the orthogonality constraint. In such a case, the slowdown in search is
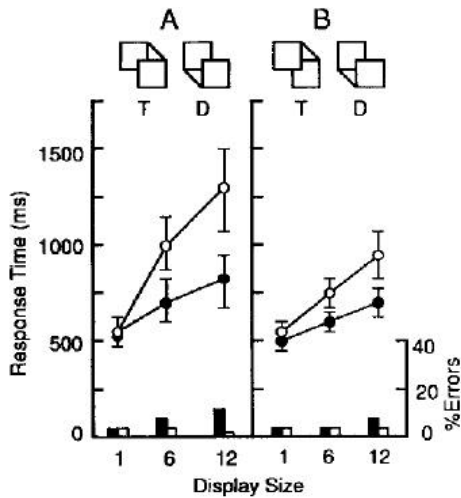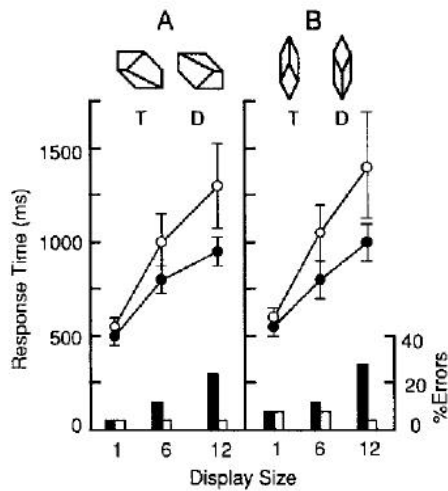
Figure 6. The target items (T), distractor items (D), and results in Experiment 5. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)

similar to that caused by T-junctions that correspond to accidental alignments.

Experiment 6

Up to this point, our discussion of line relations has focused on the role of junctions. However, it has also become clear that search for items cannot be based simply on local cues—the entire system of line relations in an item must be taken into account. This raises an important issue: Must line junctions be present explicitly for rapid search? Or is the perceptual organization induced by the line relations sufficient on its own?

In this experiment we first tested the necessity of explicitly represented junctions by removing them from the items used in Experiment 1 (Condition A). As shown in Figure 7A, these items still looked like oriented blocks when viewed individually. However, search for these items in a field of distractors was very slow (mean RT slopes were 52 and 80 ms per item). Clearly, the junctions must be represented explicitly in the items if search is to be rapid.

We next tested the sufficiency of the junctions by erasing a portion of each line in the items (see Figure 7B). This resulted in a set of junctions with the same spatial arrangement as before, except that now all junctions were disconnected from one another. This resulted in even slower search rates on target-absent trials: Mean RT slopes were 63 and 101 ms per item, t(18) = 1.48 and t(18) = 2.83, p < .05 MSe = 275).

These findings show that junctions are necessary for rapid search, but that they are not sufficient. In particular, it appears that a collection of junctions in an item must be connected by lines if they are to be detected rapidly. This finding has an interesting parallel in computational models of line drawing interpretation, where interpretation depends on assigning unique labels to lines that connect pairs of junctions (Horn, 1986; Mackworth, 1976).

The results of Conditions A and B also show that the preattentive system cannot make use of virtual lines to bridge the gaps between disconnected line segments. Although virtual lines have been used to explain some grouping and texture phenomena in visual perception (Beck, Rosenfeld, & Ivry, 1989; Stevens, 1978), they are apparently too abstract to be used for preattentive recovery of three-dimensional orientation.

Experiment 7

The failure of the preattentive system to make use of virtual lines raises questions about the generality of the line relation effects found in our experiments. Line drawings constitute a somewhat restricted domain, and it is therefore important to determine whether our findings are relevant only in this domain, or whether they touch on more general issues of representation at preattentive levels.

To this end, we repeated the tests of Experiment 1, replacing the lines in each item by luminance contours (see Figure 8). Several studies have shown that luminance contours behave like lines under some conditions (Cavanagh, Arguin, & Treisman, 1990; Enns & Wig, 1989). We asked whether this was also true for the spatial relations among luminance contours.

The pattern of results shown in Figure 8 was a strong replication of Experiment 1. When the contours were those of rectangular blocks, search was quite rapid (mean RT slopes of 9 and 20 ms per item). In contrast, contours of truncated pyramids gave rise to much slower search: mean RT slopes of 37 and 69 ms per item, t(18) = 4.13, p < .01, and t(18) = 7.24, p < .001, respectively MSe = 229). The similar pattern of results in Experiments 1 and 7 shows, therefore, that the influence of spatial relations generalizes from lines to luminance contours.

Discussion

The experiments described above show clearly that visual search can be sensitive to the spatial relations between the lines of a target item. Isolated T-junctions that differed from
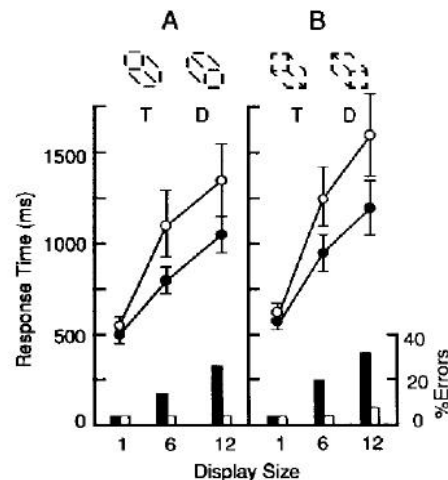

Figure 7. The target items (T), distractor items (D), and results in Experiment 6. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)
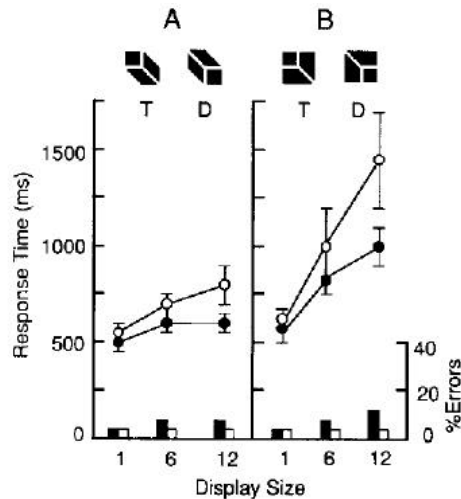
Figure 8. The target items (T), distractor items (D), and results in Experiment 7. (Closed circles and bars represent target-present trials; open circles and bars represent target-absent trials. Response time values are M ± SEM.)

distractors by a 180° rotation gave rise to slow, serial search, as predicted by conventional theories of early visual processing (Beck, 1982; Julesz, 1984; Treisman, 1986). However, Y-junctions that differed in the same way led to much faster search rates. When arrow-junctions were used, search was as rapid as for any simple feature. Thus, it is no longer possible to claim that all line relations require serial scanning for their detection (Beck, 1982; Julesz, 1984; Treisman, 1986).

The overall system of line relations in an item was found to be very important. Search for arrow- and Y-junctions could be sped up considerably if they were embedded in an item corresponding to a three-dimensional block. However, the presence of a T-junction in such an item slowed down the search rate dramatically. Strong context dependency has been shown previously for the detection of a single line in a drawing (e.g., Weisstein & Harris, 1974), but our findings demonstrate that such context effects also arise in the detection of line relations.

We have also shown that the orthogonality constraint has a strong influence on the speed of visual search. When items corresponded to blocks with orthogonal corners, search was rapid. When the blocks were distorted so as to contain no orthogonal corners, search was slowed down considerably.

It would be parsimonious to explain these findings purely in terms of local spatial operations on the two-dimensional image. For example, a degree of context dependency could be induced by spatially filtering the image (Fogel & Sagi, 1989; Gurnsey & Browse, 1989; Sutter et al. , 1989). This would cause corners to become rounded and free ends to become blurred. Indeed, such a mechanism has been proposed for the preattentive discrimination of Xs from Ts without recourse to special crossing detectors—the blurred image of a T simply covers a larger area than a blurred X made of same-length segments (Gurnsey & Browse, 1989). It may also be the reason why Humphreys et al. (1989) found that sufficiently large Ts could be discriminated rapidly from Ts rotated 180°—the filtered images contain distinctive conjunctions of orientation and curvature.

But could such an explanation also account for the present findings? Consider first the results of Experiment 2 (Figure 3), in which search was slower for T-junctions than for arrow- or Y-junctions. This is difficult to explain, because the same filtering would presumably be applied to all items, thereby inducing the same kinds of distinguishing features. Next, consider the results of Experiment 3 (Figure 4). There, search was slow for a Y-junction surrounded by a square or circle, but was almost three times as fast when the junction was surrounded by a hexagon. If spatial filtering is the relevant mechanism, the blurring required to induce sufficient context dependency would cause all three outlines to appear more or less the same. But similar search rates were not found. A similar point can be made by comparing Experiments 1 and 7. The same set of relations in each experiment resulted in nearly identical search rates, even though the features were lines in one case and luminance edges in the other. To a high degree, then, it is the spatial relations themselves that determine search rates.

Preattentive Recovery of Three-Dimensional Orientation

To explain how search can be influenced by line relations, and why only certain relations have an effect, we propose the following hypothesis: Relations between lines in the image are used by preattentive processes to determine the three-dimensional orientations of the corresponding objects in the scene.

At the most general level, there already exists strong support for this hypothesis. Enns and Rensink (1990b), along with the present study, have shown that three-dimensional orientation can be used directly as the basis for rapid visual search, and therefore must be represented at preattentive levels. But these findings in themselves do not show how this is accomplished.

How might three-dimensional structure be recovered rapidly from line relations? As discussed in the introduction, constraints on the different kinds of junctions are sufficient to yield interpretations of the lines as convex, concave, or boundary edges. Furthermore, the three-dimensional orientation of these edges can be recovered if junctions correspond to orthogonal corners. However, the computational complexity of line-labeling alone makes it very unlikely that it is actually used by the human visual system. Line-labeling of a polyhedral scene is a nondeterministic polynomial (NP) complete problem (Kirousis & Papadimitriou, 1988). Specifically, the time required for consistent labeling grows exponentially with the number of lines and junctions in the image (Garey & Johnson, 1979). Any such algorithm is therefore impractical for a real-time vision system (Tsotsos, 1988). If line relations are being used to determine three-dimensional structure in early human vision, it must be by way of "quick and dirty" estimates that give up perfect interpretation for an increase in speed.

A Model for the Rapid Recovery of Three-Dimensional Orientation

What would be required of a system that provides rapid estimates of three-dimensional orientation at all locations in the visual field? To begin with, it should make extensive use of local measurements, because these can be computed in parallel across the image. Second, because the NP-completeness of line-labeling comes about from the need to consider

all possible combinations of all possible local interpretations, this requirement must be dropped in favor of a process that examines relatively few candidates. Such a system could provide a rapid "first pass" at line interpretation, picking out a small set of interpretations at each location and passing the rest of the two-dimensional descriptions on to higher-level processes. The interpretations formed in this way are unlikely to form a complete reconstruction of the scene, because the relatively small number of interpretations considered at each location will often fail to match the physical world. What can be expected, however, is that these matches will occur at least some of the time, so that scene-based properties can be recovered at a relatively sparse set of locations in the visual field. Although incomplete, such a description would still be useful for processes further along the visual stream. We will refer to this proposal as a PRISM model, because it is intended to provide *parallel and rapid inference of scene magnitudes.*

With these considerations in mind, we now put forward a PRISM model of how the preattentive system can determine three-dimensional orientation from lines in the image. In this model, both the interpretation of lines as edges and of regions as surfaces enter into the process, the two being used in a coordinated fashion. Although neither computational theory nor empirical results are complete enough yet to allow all details to be filled in, we believe that this process can be described in rough outline. The detailed mechanisms we propose should not be viewed as assertions that particular processes are necessary, but rather as demonstrations that processing can be carried out at the requisite speed.

The PRISM model can be separated conceptually into two distinct phases: (a) the generation of one candidate interpretation at each junction, followed by (b) a limited checking of the consistency between these local interpretations. Although the two phases are necessarily applied in sequence (the second phase operating on values determined by the first), each phase is carried out in parallel across the visual field. The results of each of these phases of the model are illustrated for five of the search items in Figure 9.

Phase 1: Local Estimates

The first stage of the interpretation process assigns candidate interpretations to the lines and regions about each junction in the image. To optimize the effectiveness of a rapid recovery system, interpretation should be based on quantities that are both useful and have a high likelihood of being estimated correctly from local measurements. Two such quantities are proposed here: (a) orientation estimates from arrow- and Y-junctions and (b) occlusion estimates from T-junctions. Because these two kinds of assignments are based on mutually exclusive sets of junctions, they can in principle be carried out simultaneously across the image.

Orientation estimates. Given a convex trihedral corner formed from orthogonal surfaces, the three-dimensional orientation of its constituent edges and surfaces can be recovered from the two-dimensional orientations of the lines about the corresponding junction in the image plane (see Appendix; Perkins, 1968). Thus, if corner convexity and orthogonality of surfaces are assumed, initial estimates of three-dimensional orientation can be made in parallel on all arrow- and Y-junctions in the image. This assignment is a purely local operation—it cannot take into account the extent


Figure 9. Schematic illustration of the outcome of the PRISM model for five search items (A–E) used in the experiments. (In Phase 1, local estimates are obtained for orientation [based on Y- and arrow-junctions] and occlusion [based on T-junctions]. In Phase 2, the junctions in each segmented region are checked for consistency. Check marks represent the successful completion of a stage of processing, Xs refer to a failure in the process, and dashes indicate that a given stage has no data to consider. Search rates are determined by the difference in interpreted orientation between target and distractor items.)

of these edges and surfaces or determine whether the estimates are consistent with those made for the rest of the scene.

The assumption of convexity follows naturally from the observation that objects tend to have more corners that are convex than concave. Convex corners determine the overall three-dimensional shape of an object, whereas concave corners correspond to indentations in and deformations of the global structure (Pentland, 1986). Concave corners are also unreliable in that they often result simply from contact between two or more objects (Biederman, 1985). Therefore, corner convexity is a reasonable default assumption for the rapid determination of object structure.

The assumption of orthogonality, on the other hand, is more difficult to justify on ecological grounds. Corners are rarely formed from perfectly orthogonal surfaces in the natural world. However, if there is no other way to determine three-dimensional orientation, the visual system may well assume orthogonality in order to get a "quick and dirty" first approximation. There is a great deal of psychophysical evidence that humans assume orthogonality in line drawings of both familiar and unfamiliar objects (Butler & Kring, 1987; Perkins, 1972; Shepard, 1981). They even "see" rectangular corners when they know orthogonality has been violated (Kubovy, 1986). In addition to these reasons, orthogonal angles are also natural defaults simply because they lie midway on the range of all possible angles between two surfaces.

If the orthogonality assumption does not hold for some junction, the resulting estimates will be at odds with those for the rest of the scene. The extent of this disagreement will depend on how closely the corner meets the orthogonality condition. If the disagreement between local estimates is severe enough, it will be detected at the stage of consistency checking and the consistency check will fail (discussed later).

Occlusion estimates. When one surface occludes another, their projections onto the image plane necessarily contact each other. To interpret a line drawing correctly, then, the lines must be split into groups, each corresponding to a separate collection of contiguous faces. As shown by work on line-labeling algorithms, this segmentation can be done using only three kinds of labels: convex, concave, and boundary edge (Horn, 1986). For segmentation based solely on local measurements, we propose a simplified variant of this scheme, namely, the use of T-junctions to mark particular lines as corresponding to boundary edges formed by occlusion.

To see how this comes about, consider the interpretation of the lines in a T-junction. The stem of the T can correspond to either a convex edge, a concave edge, or a boundary edge. Therefore, its status cannot be assigned unequivocally on the basis of purely local considerations. The situation, however, is quite different for the crossbar. Apart from extremely rare cases of accidental alignment, this line must correspond to a boundary edge that occludes the surface(S) associated with the stem. Consequently, it must belong to a different group of lines than the stem, a fact that can be signaled by marking the crossbar as an occluding boundary edge.

Similar reasoning shows that arrow- and Y-junctions cannot be used this way, unless an accidental alignment is assumed. Thus, the major determinants of segmentation are likely to be line interpretations arising from T-junctions alone. As in the case for the orientation estimates, these estimates can be computed locally and in parallel across the image. The validity of these interpretations is then tested by the subsequent phase of consistency checking.

Phase 2: Consistency Checking.

The local estimates of three-dimensional orientation and occlusion must be consistent with each other if the lines in the image correspond to orthographic projections of solid objects. One way this can be done rapidly and in parallel is through the pairwise comparison of estimates from neighboring junctions in each segmented group (i.e., junctions corresponding to corners deemed to be connected in the scene). If these estimates are compatible with each other, this will reinforce the reliability of the interpretation for that segment. If an inconsistency is detected, however, the interpretation of the segment will fail.

Orientation consistency. For trihedral junctions, the three-dimensional orientations of the edges are determined completely by the orientations of the surfaces, and vice versa (see Appendix). Strictly speaking, then, it is immaterial whether consistency checking is applied to orientation estimates for edges or for surfaces.

Consistency of surface orientation can be checked by testing whether the estimates of orientation assigned to a planar face are consistent with each other. This checking can be carried out in parallel for each of the candidate faces. Because this test involves the transmission of information across a region, the time required will increase with region size (Ullman, 1984). The speed of this transmission is difficult to ascertain, but it is reasonable to assume that it is comparable to the speed at which other kinds of spatial information are integrated across the visual field. Independent estimates based on contrast discrimination

(Jamar & Koenderink, 1983) and line drawing discrimination tasks (Enns & Girgus, 1986; Enns & King, 1990) suggest speeds of 20-30 ms per degree of visual angle. Because the size of the regions considered here is relatively small (1.5°) and the items relatively simple, this operation would add a small constant time factor to the interpretation process.

For the most part, similar considerations apply if consistency checking is based on edges rather than faces (see Mulder & Dawson, 1990). However, here L-junctions may also assist such a process. If the corner is assumed to be orthogonal, assigning a three-dimensional orientation to one of the edges will automatically determine the orientation of the other. Orientation information can then propagate around the boundaries of a face in tandem with orientation checking.

It is interesting to note here that if a line junction has a free end (as in the isolated junctions of Experiment 2), there will only be one estimate to consider. Because this estimate is not contradicted by any other, the consistency checking process will allow the interpretation to stand, even though the corresponding face cannot be completely delimited.

Occlusion consistency. To help ensure that orientation consistency is checked only over regions that correspond to actual faces or boundaries in the scene, it is useful to segment the lines into groups that correspond to separate objects. The local assignment of occlusion boundaries is a first step in this process (discussed earlier). As in the case of orientation estimates, these assignments must be checked for consistency with other estimates, and this may be done by propagating information along lines or across regions.

One way this testing could be done is by propagating the assignment of the occlusion boundary interpretation along lines connected by L-junctions. Such junctions generally correspond to corners of an object; if one line is marked as an occlusion boundary, so must the other.

Thus, the front plates of the items corresponding to the brackets in Figure 9C can be successfully segmented away from the remainder of the lines. For the truncated pyramids of Figure 9E, however, there is no consistent interpretation of the boundary edges. Note that it is possible for these edges to be seen as forming a "window" through which another object is viewed. However, because this "window" is not an acceptable boundary for an object, the surrounding lines will not be successfully segmented out of the item, and the interpretation process will fail.

Comparison With Empirical Results

We now examine the extent to which our PRISM model accounts for the results of the experiments described in this article. As stated in the introduction, we are interested mainly in those spatial relations that permit search to be more rapid than predicted by conventional models of attentive vision. However, in addition to finding examples of such rapid search, we also found other conditions in which search rates were intermediate between "rapid" and very slow. It is clearly more difficult to interpret these latter findings without making a commitment to a specific model of attention.

In what follows, we will take the generally accepted position that search rates reflect the signal-to-noise ratio of the target amid the distractors (Duncan & Humphreys, 1989; Treisman & Gormican, 1988). How serial or parallel pro-

cesses enter into all this is a somewhat independent question and does not directly concern us here. For present purposes, it is sufficient to show that relative rates of search can be predicted on the basis of the signal-to-noise ratios obtained from the PRISM model.

To begin with, the rapid search found for the drawings of blocks in Experiment 1 (Figures 2A and 9A) is a natural consequence of the PRISM model. All items are assigned orientation estimates and these items pass the consistency tests. Targets and distractors are therefore interpreted as blocks with different three-dimensional orientations, the differences being large enough for targets to be detected preattentively.

The items of Condition B in Experiment 1 (Figures 2B and 9E), on the other hand, contain T-junctions. Because the items cannot readily be interpreted as convex objects, no consistent segmentation is possible for the lines and the interpretation process fails. Search is consequently slow, in the range conventionally considered to be the result of attentive processes.

The slow search for the isolated T-junctions of Experiment 2 (Figures 3C) is also to be expected, because these junctions cannot give rise to estimates of three-dimensional orientation. According to the PRISM model, however, this quantity can be recovered for arrow- and Y-junctions. Targets are therefore distinguishable from distractors on the basis of this quantity, and search is consequently facilitated for these junctions. It is worth noting here that if the arrow-junctions (Figure 3A) are interpreted as two visible surfaces, then the orientations of these surfaces will differ considerably between target and distractor. This will give rise to very rapid search, as borne out by our data. In contrast, the region bounded by the right angle in the Y-junctions (Figure 3B) corresponds to a planar face that has a similar orientation in target and distractor. Because overlapping sets of features can cause a slowdown in search (Duncan & Humphreys, 1989), this may account for why search for arrow-junctions was faster than for Y-junctions.

The various search rates found in Experiment 3 (Figure 4) can also be accounted for by the model. The relatively fast search in Condition A reflects the differences in three-dimensional orientation assigned to Y-junctions. The items of Condition C are also expected to be detected rapidly—their line structure is similar to the blocks of Experiment 1, and so recovery of a distinctive three-dimensional orientation difference is expected. It is interesting to note that search for chevrons in Condition B was as rapid as that for the blocks in Condition C, despite the fact that chevrons contain dihedral rather than trihedral corners. Here the orientation estimate stage gives rise to local estimates of three-dimensional orientation that, in the absence of contradiction, are still available to form the basis for rapid search. In contrast, the T-junctions in Conditions D and E could not be interpreted consistently in terms of object boundaries, and so these junctions actively disrupted the interpretation process.

According to the PRISM model, the T-junctions in Experiment 4 (Figures 5 and 9C) give rise to a consistent segmentation of the lines. The two targets and the distractor from this experiment have been redrawn in Figure 10 in order to illustrate the interpretation process. One group of junctions corresponds to an occluding plate, the other to an occluded
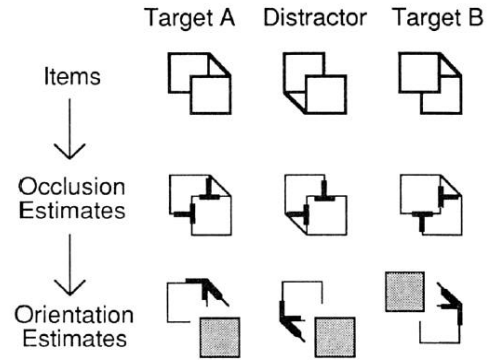

Figure 10. Illustration of the PRISM model applied to the items in Experiment 4 (Figure 5). (In Phase 1, T-junctions are used as the basis for segmenting the occluding plate from the remaining lines in each item. Arrow-junctions are used in this phase to estimate the three-dimensional orientation of the corresponding corners. The estimated corner in Target A differs from that of the distractor by 90°, whereas the corner in Target B differs by 180°. Search rates were accordingly faster for Target B.)

object that consists of a plate with an attached face (see Figure 10). Because the occluding plates cannot be assigned distinctive orientations, their orientations cannot be the basis of rapid search. Because the depth ordering of the two plates also cannot be used for this purpose (Enns & Rensink, 1990b), rapid search must be based on some other part of the item.

Disregarding the occluding plates, then, the target and distractor items differ only in the orientation of the remaining arrow-junctions. In target A (Figure 10) this junction signals an attached face that differs from the attached face in the distractor by 90°. This moderate difference in three-dimensional orientation is reflected in the intermediate search rates found for this condition (Figure 5A).

Target B in Figure 10 can be analyzed similarly. However, here the relevant arrow-junction in the target corresponds to a corner oriented at a considerably different angle from that of the distractor—the faces attached to the occluded plates now differ by 180° rather than the 90°. In keeping with this increased distinctiveness, search rates were considerably faster than in the previous condition. Interestingly, they were now in the range of search rates found for the Y-junctions alone (Figure 3).

In Experiment 5 (Figures 6 and 9D) the items contained arrow- and Y-junctions that violated the orthogonality constraint. As such, the local estimates made for these junctions will not be correct, and the associated orientation consistency check will fail. Targets will therefore be indistinguishable from distractors at the preattentive level, and search will consequently be quite slow.

The necessity of explicit junctions in the line drawings (Condition A of Experiment 6; see Figure 7A) can simply be attributed to the fact that junctions form the basis of the local estimates of three-dimensional orientation. Without these, the final phase of the interpretation process has nothing to go on. The insufficiency of junctions, on the other hand (Condition B of Experiment 6; see Figure 7B), stems from the failure at the second stage to join the local estimates into an interpretation of a coherent object. There are two ways

this failure may have occurred. First, each of the junctions in an item might have been considered separately, effectively increasing the number of items to be searched. Second, the junctions in these items have fairly short lines. If this caused scatter in the three-dimensional orientations assigned to the junctions of an item, there would no longer be any single value by which the target could be identified. Such heterogeneity is known to reduce search rates for image features (Duncan & Humphreys, 1989; Treisman, 1988) and the results of Experiment 6B suggest that heterogeneity can do the same for scene-based features.

Implications for Theories of Early Vision

Preattentive vision . Our results have important   implications for conventional theories of preattentive vision. In this section, we will begin by briefly describing three representative theories. We will then spell out several implications that apply to these theories generally, before focusing on implications that are specific to each.

According to feature integration theory (Treisman, 1986, 1988), the preattentive visual system maintains a set of spatiotopic maps that record the presence of elementary features at the corresponding locations in the image. The coded features are specific values along dimensions such as orientation, length, width, and color. Visual search for an item defined by activity in a single map can be conducted without any need for attention. In contrast, search for a conjunction of features between two sets of maps (e.g., orientation and color), or for the relative locations of features within the same set of maps (e.g., the two lines that define a T-shape), requires an attentive system to scan groups of items in a serial fashion.

The preattentive processes proposed by texton theory (Julesz, 1984, 1986) are similar to feature integration theory in many ways. Here the elementary features are geometric elements (i.e., textons), which are elongated blobs that are characterized by intrinsic properties of orientation, length, color, and so forth. These are thought to form the basis set of features in preattentive vision; rapid search is possible only for items containing distinctive textons. For most items defined by the spatial relations between textons, search will require attention and therefore be slow and serial. The only exception to this rule is a small set of compound textons such as intersections (i.e., two blobs of different orientation superimposed over each other). Spatial relations like these are thought to facilitate subsequent processes of form perception.

Resemblance theory (Duncan & Humphreys, 1989) differs from the other theories in that it postulates a continuum of search efficiency, corresponding to the ease with which items in a display can be selected.   Large items are differentiated more readily than small ones, and a linking operation is postulated to group together items with similar properties. Nonetheless, it is similar to the other theories in that only a small set of simple image features constitutes the basis for form and object perception. Visual similarity, and therefore the efficiency of selection, is defined in terms of the transformations (e.g., translation and rotation) required to turn the target item into the distractor item in the image plane.

There are two general implications of our results that apply equally well to each of these theories. First, our results show that preattentive processes do more than simply register and group together elementary properties of the two-dimensional image—they are also capable of determining properties of the corresponding three-dimensional scene. The features now known to be registered at the preattentive stage include three-dimensional orientation (this study; Enns & Rensink, 1990b; Epstein & Babler, 1990) and the direction of light in the scene (Enns & Rensink, 1990a; Ramachandran, 1988).

A second implication is that preattentive vision does not only perform local measurements—it also appears to employ cooperative processes to construct consistent local interpretations of the image. This sensitivity to the system of relations among features was found to occur for the preattentive detection of lighting direction (Aks & Enns, 1991; Enns & Rensink, 1990a) and was also seen clearly in the present study.

These findings have specific implications for feature integration theory. In particular, they show that the line elements in at least some preattentive maps must have locations and orientations that are represented to a fairly high degree of precision. These elements are therefore not free-floating (Treisman, 1986), nor are their positions coded only coarsely (Cohen & Ivry, 1990).

An implication for texton theory is that not all blob intersections are afforded equal weight in preattentive vision. T-junctions do not receive the same preattentive processing as arrow and Y-junctions. Our model presents a compelling reason why this should be: T-junctions simply do not contain the same kind of information about surface orientation as do arrow- and Y-junctions.

Finally, although some of our results are consistent with the heterogeneity effects predicted by resemblance theory, others are not. For example, the influence of spatial relations found in our study occurs at spatial scales much smaller than those predicted by resemblance theory. Furthermore, resemblance theory does not explain why arrow- and Y-junctions should be preferred over T-junctions. It is apparent that resemblances cannot be measured simply in terms of image properties—scene-based properties must be considered as well.

Recovery of three-dimensional structure. The experiments described here provide strong evidence that three-dimensional orientation can be determined by processes at preattentive levels. In addition, a PRISM model has been proposed that shows how this may be done. How do our data and model relate to other theories concerned with the problem of recovering three-dimensional orientation from two-dimensional images?

Several theories of human image understanding start with the premise that the elements of object perception are three-dimensional or volumetric solids (Biederman, 1985; Leeuwenberg, 1988; Pentland, 1986). There are at least two motivations for this. The first is an evolutionary argument: A reproductive advantage should accrue to organisms that can extract three-dimensional information rapidly from an image (e.g., Gibson, 1966; Ramachandran, 1988). The second is computational: Object descriptions based on volumetric primitives achieve an attractive balance in the inherent trade-off between the complexity of the primitive elements and the

complexity of the description of a scene using those elements (e.g., Biederman, 1985; Pentland, 1986).

We will restrict our discussion of volumetric theories to Biederman's (1985) recognition-by-components theory, both because it is the most thoroughly developed of these accounts and because it was specifically designed for the domain of line drawings. According to this theory, the recognition process begins by segmenting a line drawing into regions bounded by deep concavities. The lines within a region are then assigned to one of 36 possible volumetric primitives (or geons). Each geon is defined by a hierarchical catalogue of features. The first division in this hierarchy is between the principal axis of the object depicted in a region and the object's cross section, which is swept out along the axis. Lower-order divisions in the hierarchy provide a more detailed description of the properties of both the principal axis and the cross section. Empirical support for this theory comes largely from speeded object-naming tasks that examine the effects of geon number and line deletion on naming times.

At a general level, our results are consistent with this theory, because they demonstrate that some aspects of three-dimensional structure can be recovered early in the visual stream. However, our results also differ from its predictions in several ways. First, we have shown that isolated junctions can themselves be detected rapidly (Experiment 2). Unless there is a reliable way to assign a geon to a small fragment of a line drawing, volumetric primitives cannot account for this finding. Second, the failure to detect nonorthogonal objects rapidly in Experiment 5 shows that the preattentive system does not readily interpret all convex three-dimensional objects, including such apparently well-formed geons as the object with a diamond-shaped cross section (Condition B in Experiment 5). This is difficult to explain if geons are used at preattentive levels. Third, the inability of preattentive processes to interpret drawings containing line deletions suggests that our task is tapping a lower level of visual processing than the naming task on which Biederman's (1985) theory is based. Perhaps the three-dimensional orientations determined at preattentive levels provide the basis for geon-like representations at higher levels.

Other theories deliberately avoid the use of volumetric primitives, relying instead on fairly detailed knowledge of the three-dimensional shapes of objects in the scene (e.g., Brooks, 1981; Lowe, 1987). Recognition of objects is carried out by matching features in the image against predictions obtained by projecting the model onto the image plane. We will consider Lowe's SCERPO model as representative of model-based theories.

In SCERPO, the only information about the scene that is directly available concerns the three-dimensional shapes of the objects that may be present. Both three-dimensional orientation and viewing direction are then recovered by determining which values of these properties allow the given image to be considered as a scene viewed from a single viewpoint.

To allow initial estimates of these properties to be made, SCERPO uses a set of image features whose relations to scene-based properties remain invariant over a wide range of orientations and viewing directions. For instance, parallel edges in the scene will generally map to parallel lines in the image. This and several other invariant features are used to reduce the search among the space of possible interpretations. A final interpretation is then selected from the remaining candidates on the basis of the goodness of fit between the hypothesized feature and the feature actually present in the image.

Interestingly, SCERPO avoids the use of depth estimates obtained from model-independent operations on the image. As Lowe (1987) noted, information about depth is often incomplete, and even if it were available, it would be difficult to recover in a reasonable amount of time. Our results suggest that model-based systems such as SCERPO could be extended to use local estimates of three-dimensional orientation in at least some domains (i.e., where a PRISM model could be applied). A more general point, however, is that these systems could use scene-based features in addition to image-based ones to increase their speed and effectiveness.

Future Directions

An important issue that deserves further work concerns the nature of the constraints used in early vision to make possible the rapid interpretation of images. The present results suggest that preattentive vision takes advantage of the orthogonality assumption to interpret line drawings rapidly. Other constraints remaining to be investigated include assumptions about parallel and collinear edges (Kubovy, 1986), constraints on junctions formed by four or more lines (Lee, Haralick, & Zhang, 1985; Waltz, 1972), and the properties of curved surfaces bounded by piecewise smooth lines (Lee et al., 1985; Malik, 1987).

A second issue concerns the extent to which the rapid interpretation of line drawings is learned or innate. Studies comparing the ability of younger and older children to interpret line drawings under attention-demanding conditions have shown that there is considerable improvement through the early school years (Enns & Girgus, 1986; Enns & King, 1990). Additional tests will be needed to determine if this is also true for the preattentive interpretation of line drawings.

A third issue, and one that we have only yet touched on, concerns the degree to which image and scene features are represented abstractly in early vision. Experiments 1 and 7 suggest that three-dimensional orientation can be recovered equally well from drawings based on lines or luminance edges. Does this result generalize to other media? Do the possible media for representing edges all contribute their results to a common representation, or are several representations constructed? Preliminary work in our lab suggests that two-dimensional orientation is based on an abstract representation that does not distinguish among lines, edges, or even some texture boundaries (Enns & Wig, 1989). Further tests will be needed to extend these findings to three-dimensional orientation.

Conclusions

This article has demonstrated that preattentive vision is capable of more sophisticated processing than has generally been assumed. In particular, it has presented psychophysical data suggesting that three-dimensional orientation can be recovered rapidly from an image consisting of simple line drawings. A PRISM model has been proposed, showing how this can be done by processes operating rapidly and in parallel across the visual field.

Taken together, the results of this article have important implications for a revised view of early vision, at both a theoretical and a methodological level. We will briefly discuss four of the more important implications.

1. It is unnecessarily restrictive to assume that the parallel processes of early vision operate only on simple geometric elements. Although there must indeed be an initial stage that analyzes the retinal input in this way, our findings show that there must also be subsequent stages based on more complex properties. These properties are obtained neither by taking purely local measurements at each point in the image, nor by the operation of global processes such as regularization (Horn, 1986). Rather, they are calculated by rapidly acting processes that use information contained in a neighborhood of limited extent.

2. The elements of early vision may be characterized by environmental relevance. How then can the elements of early vision be characterized if geometrical simplicity alone is too restrictive a criterion? Our results show that these elements describe at least some properties of the three-dimensional scene. As several researchers have pointed out (e.g.,Walters, 1987; Weisstein & Maquire, 1978), the early determination of scene properties, even if incomplete, would facilitate processes further along the visual stream. Our results have shown that such a strategy is indeed used at pre-attentive levels, at least for the recovery of three-dimensional orientation. Other work has indicated that such a strategy might also be used to recover lighting direction (Enns & Rensink, 1990a; Ramachandran, 1988). It will be interesting to see which other properties can be recovered. Preliminary reports suggest that length may be registered only after size constancy mechanisms have operated on the image (Ramachandran, 1989). Brightness constancy and color constancy may also operate at these levels.

3. The elements of early vision must be rapidly computable. As we have argued, preattentive processes cannot afford the time required for complete interpretations such as those given by line labeling. How then is time managed for these "quick and dirty" processes? Are these processes simply allowed to run to completion on a given input, or are they given some fixed span of time in which to "do their best. " Our results suggest the latter. In all experiments, the intercepts of RT functions remained essentially the same, no matter how complex the items used or how steep the RT slope. If recovery processes are carried out in parallel, this implies that a fixed amount of time is allotted for their operation. Because information across the visual field is transmitted at a finite speed, this time constraint also provides an upper limit on the size of the neighborhood over which information is integrated. In this context, it is interesting to note that Walters (1987) found that junction type could affect the perceived brightness of a line, with apparent brightness increasing with line length to a maximum of 1.5°. It may well be that a similar spatial limit exists for recovery processes at preattentive levels.

4. The criteria of feature complexity, environmental relevance, and processing speed should be used to test other modules of early vision. The existence of features at pre-attentive levels for the interpretation of depth from single images suggests that other modules of early vision be studied from this new perspective. To what extent do other modules use scene-based features? It would be interesting to determine, for example, whether motion perception or stereopsis is able to use the spatial relations we have studied here (Cavanagh, 1987). A comparison of the features used by various modules may help shed light on how they operate, and how they are related to one another.

References

Aks, D. J. & Enns, J. T. (1991). Influence of apparent curvature and contrast polarity on visual search. Ms. submitted for publication.

Beck, J. (1982). Textural segmentation.In J. Beck (Ed.), *Organization and representation in perception*(pp. 285-317). Hillsdale, NJ: Erlbaum.

Beck, J., Rosenfeld, A. & Ivry, R. (June, 1989). Line segregation. Paper presented at the Workshop on Human and Machine Vision, North Falmouth, MA.

Biederman, I. (1985). Human image understanding: Recent research and a theory. *CVGIP*, 32, 29-73.

Brooks, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artif. Intell.*, 17, 285-348.

Butler, D. L. & Kring, A. M. (1987). Integration of features in de-pictions as a function of size. *Percept Psychophys*, 41, 159-164.

Cavanagh, P. (1987). Reconstructing the third dimension Inter-actions between color, texture, motion, binocular disparity, and shape. *CVGIP*, 37, 171-195.

Cavanagh, P., Arguin, M. & Treisman, A. (1990). Effect of stimulus domain on visual search for orientation and size. Ms. submitted or publication.

Chen, L. (1982). Topological structure in visual perception. *Science*, 218, 699-700.

Chen, L. (1990). Holes and wholes: A reply to Rubin and Kanwisher. *Percept & Psychophys*, 47, 47-53.

Clowes, M. B. (1971). On seeing things. *Artif; Intell*, 2, 79-116.

Cohen, A. & Ivry, R. (1990). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 650-663.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458.

Enns, J. T. (1990). Three dimensional features that pop out in visual search.In D. Brogan (Ed.),Visual search(pp. 37-45). London: Taylor & Francis.

Enns, J. T. & Girgus, J. S. (1986). A developmental study of shape integration over space and time. *Develop Psych*, 22, 491-499.

Enns, J. T. & King, K. A. (1990). Components of line-drawing interpretation. *Develop Psych*, 26, 469-479.

Enns, J. T., Ochs, E. P. & Rensink, R. A. (1990). VSearch: Macintosh software for experiments in visual search. *Behavior Research Methods, Instruments, & Computers,* 22, 118-122.

Enns, J. T. & Rensink, R. A. (1990a). Influence of scene-based properties on visual search. *Science*, 247, 721-723.

Enns, J. T. & Rensink, R. A. (1990b). Sensitivity to three-dimensional orientation in visual search. *Psychol Science*, 1, 323-326.

Enns, J. T. & Wig, P. S. (1989,November). The medium and the message in early vision. Paper presented at the meeting of the Psychonomic Society, Atlanta, GA.

Epstein, W. & Babler, T. (1990). In search of depth. *Percept & Psychophys*, 48, 68-76.

Fogel, I. & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61, 103-113.

Garey, M. R. & Johnson, D. S. (1979). *Computers and Intrac-tability: A Guide to the theory of NP-completeness*. New York: Freeman.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gurnsey, R. & Browse, R. A. (1989). Asymmetries in visual texture discrimination. *Spatial Vision*, 4, 31-44.

Guzman, A. (1968). Decomposition of a visual scene into three-dimensional bodies. AFIPS Fall Joint Conferences, 33, 291-304.

Holliday, I.E.& Braddick, O. J. (August, 1989). *Search for stereoscopic slant direction is parallel*. Paper presented at the 12th European Congress on Visual Perception, Brussels, Belgium.

Horn, B. K. P. (1986). Robot vision. Cambridge, MA: MIT Press.

Huffman, D. A. (1971). Impossible objects as nonsense sentences.In R. Meltzer & D. Michie (Eds.),Machine Intelligence 6(pp. 295-323). New York: Elsevier.

Humphreys, G. W., Quinlan, P. T. & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. Journal of Experimental Psychology: General, 118, 258-279.

Jamar, J. H. T. & Koenderink, J. J. (1983). Sine-wave gratings: Scale invariance and spatial integration at suprathreshold contrast. *Vision Research*, 23, 805-810.

Jolicoeur, P., Ullman, S. & MacKay, M. (1986). Curve tracing: A possible basic operation in the perception of spatial relations. *Memory & Cognition*, 14, 129-140.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience*, 7, 41-45.

Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, 54, 245-261.

Kirousis, L. & Papadimitriou, C. (1988). The complexity of recognizing polyhedral scenes. *J. Computer and System Sciences*, 37, 14-38.

Klein, R. & Farrell, M. (1989). Search performance without eye movements. *Percept & Psychophys*, 46, 476-482.

Kubovy, M. (1986). *The psychology of perspective and renaissance art.* Cambridge, England: Cambridge University Press.

Lee, S. J., Haralick, R. M. & Zhang, M. C. (1985). Understanding objects with curved surfaces from a single perspective view of boundaries. *Artif Intell*, 26, 145-169.

Leeuwenberg, E. (1988,November). On geon and global precedence in form perception. Paper presented at the meeting of the Psychonomic Society, Chicago.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artif Intell*, 31, 355-395.

Mackworth, A. K. (1973). Interpreting pictures of polyhedral scenes. *Artif Intell*, 4, 121-137.

Mackworth, A. K. (1976). Model-driven interpretation in intelligent vision systems. *Perception*, 5, 349-370.

Malik, J. (1987). Interpreting line drawings of curved objects. *International Journal of Computer Vision*, 1, 73-103.

McLeod, P., Driver, J. & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, 332, 154-155.

Mulder, J. A. & Dawson, R. J. M. (1990). Reconstructing polyhedral scenes from single two-dimensional images: The orthogonality hypothesis.In P. K. Patel-Schneider (Ed.), Proceedings of the 8th Biennial Conference of the CSCSI(pp. 238-244). Palo Alto, CA: Morgan-Kaufmann.

Nakayama, K. & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320, 264-265.

Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Percept& Psychophys*, 41, 191-201.

Pentland, A. P. (1986). Perceptual organization and the representation of natural form.*Artif Intell*, 28, 293-331.

Perkins, D. N. (1968). Cubic corners. *M. I. T. Research Laboratory of Electronics Quarterly Progress Report*, 89, 207-214.

Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular parallelopipeds. *Percep & Psychophys*, 12, 396-400.

Ramachandran, V. S. (1988). Perceiving shape from shading *Scientific American*, 259, 76-83.

Ramachandran, V. S. (1989,November). Is perceived size computed before or after visual search?Paper presented at the meeting of the Psychonomic Society, Atlanta, GA.

Ramachandran, V. S. & Plummer, D. J. (May, 1989). Preattentive perception of 3-D versus 2-D image features. Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Roberts, L. (1965). Machine perception of three-dimensional solids.In J. Tippett (Ed.),Optical and electro-optical information processing(pp. 150-197). Cambridge, MA: MIT Press.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.

Shepard, R. N. (1981). Psychophysical complementarity.In M. Kubovy & J. R. Pomerantz (Eds.),Perceptual organization(pp. 279-342). Hillsdale, NJ: Erlbaum.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologia*, 30, 276-315.

Stevens, K. A. (1978). Computation of locally parallel structure. *Biological Cybernetics*, 29, 19-28.

Sutter, A., Beck, J. & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Percept & Psychophys*, 46, 312-332.

Townsend, J. T. & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge, University Press.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, 255, 106-115.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201-237.

Treisman, A., Cavanagh, P., Fischer, B., Ramachandran, V. S. & von der Heydt, R. (1990). Form perception and attention: Striate cortex and beyond.In L. Spillman & J. S. Werner (Eds.),*Visual perception* (pp. 273-316). San Diego, CA: Academic Press.

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15-48.

Treisman, A. & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. Journal of Experimental Psychology: General, 114, 285-310.

Tsotsos, J. K. (1988). A "complexity level" analysis of immediate vision. International Journal of Computer Vision, 1, 303-320.

Ullman, S. (1984). Visual routines. Cognition, 18, 97-159.

Walters, D. (1987). Selection of image primitives for general purpose visual processing. Computer Vision, Graphics, and Image Processing, 37, 261-298.

Waltz, D. L. (1972). Generating semantic descriptions from drawings of scenes with shadows (AI-TR-271, Project MAC, MIT).(Reprinted in P. H. Winston, Ed. , 1975,The psychology of computer vision,pp. 19-92, New York: McGraw-Hill)

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America*, 4, 2006-2021.

Weisstein, N. & Harris, C. S. (1974). Visual detection of line segments: An object superiority effect. Science, 186, 752-755.

Weisstein, N. & Maquire, W. (1978). Computing the next step: Psychophysical measures of representation and interpretation.In A. R. Hansen & E. M. Riseman (Eds.),Computer vision systems(pp. 243-260). San Diego, CA: Academic Press.

Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419-433.

Zucker, S. W. (1987). Early vision. In S. C. Shapiro (Ed.),*The encyclopedia of artificial intelligence* (pp. 1131-1152). New York: Wiley.

Appendix

Mathematical Proof for the Recovery of Three-Dimensional Orientation From Line Drawings of Polyhedral Corners

The model of early vision proposed in this article relies heavily on the assertion that the three-dimensional orientations of the edges and surfaces of a convex polyhedral corner can be recovered from the corresponding line junctions in the image whenever these edges and surfaces are orthogonal. The following is a self-contained mathematical proof of this statement, first proved by Perkins (1968). It should be kept in mind that this treatment proves only that the information in a junction is sufficient to recover three-dimensional orientations under these assumptions. No claim is made about the particular representations and algorithms used by the recovery process itself.

Before proceeding with the proof, we would like to point out that surfaces at a corner are orthogonal if and only if the edges formed by their intersections are orthogonal as well. This implies that the normal to a surface defined by two of the edges will always be parallel to the remaining edge. Thus, the three-dimensional orientations of the surfaces can be taken directly from those of the edges, and vice versa. In what follows, only the orientations of the edges will be discussed.

Consider a Y-junction corresponding to a convex corner, such as is shown in Figure A1. The vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are formed by following each of the lines of the junction out to some fixed arbitrary distance from the point of intersection. These vectors are simply projections of the edges of the corner onto the image plane—they have no component oriented toward the viewer. Thus, they can be defined as:

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + 0\mathbf{k}$$
$$\mathbf{B} = B_x\mathbf{i} + B_y\mathbf{j} + 0\mathbf{k}$$
$$\mathbf{C} = C_x\mathbf{i} + C_y\mathbf{j} + 0\mathbf{k}$$

where subscripts refer to the components of these vectors in the x, y, and z dimensions of some coordinate space, and $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ refer to unit vectors along the axes of this. For present purposes it is convenient to take x and y to be the horizontal and vertical dimensions in the image plane. The z direction will be taken as the line of sight, assumed to be normal to the image plane, and in the direction toward the viewer.

Imagine now a set of vectors, $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, along the corresponding edges of the corner in the scene (Figure A1). These vectors are defined to be of unit length, so that they form an orthonormal set. If the x and y axes in the scene-based coordinate system are taken to be the same as those in the image, with the z-direction toward the viewer, the scene-based and image-based vectors can be related as follows:

$$\mathbf{a} = a\mathbf{A} + a_z\mathbf{k} = aA_x\mathbf{i} + aA_y\mathbf{j} + a_z\mathbf{k} \quad (A1a)$$
$$\mathbf{b} = b\mathbf{B} + b_z\mathbf{k} = bB_x\mathbf{i} + bB_y\mathbf{j} + b_z\mathbf{k} \quad (A1b)$$
$$\mathbf{c} = c\mathbf{C} + c_z\mathbf{k} = cC_x\mathbf{i} + cC_y\mathbf{j} + c_z\mathbf{k} \quad (A1c)$$

where a, b, and c are positive scalars. Because accidental alignments are disregarded here, these scalars must always be nonzero. The scalars a_z, b_z, and c_z, on the other hand, must have negative values, because the assumption of convexity means that a, b, and c must all be pointing away from the viewer.
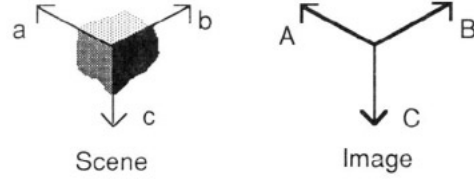

Figure A1. Vectors a, b, and c refer to edges in the scene; vectors A, B, and C refer to corresponding lines in the image.

As represented in Equation A1, the z-components of the edge vectors (a_z, b_z, and c_z) appear to be free parameters. However, they are constrained to yield vectors of unit length, and so:

$$a_z = -\sqrt{1 - a^2(A_x^2 + A_y^2)} = -\sqrt{1 - a^2(\mathbf{A}\cdot\mathbf{A})} \quad (A2a)$$
$$b_z = -\sqrt{1 - b^2(B_x^2 + B_y^2)} = -\sqrt{1 - b^2(\mathbf{B}\cdot\mathbf{B})} \quad (A2b)$$
$$c_z = -\sqrt{1 - c^2(C_x^2 + C_y^2)} = -\sqrt{1 - c^2(\mathbf{C}\cdot\mathbf{C})} \quad (A2c)$$

the negative signs indicating that all edge vectors are directed away from the viewer. Taken together, Equations A1 and A2 show that the orientations of $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ in three-dimensional space can be completely determined once a, b, and c are known. These three scalars thus describe completely the three degrees of freedom available for the orientation of an orthogonal corner in three-dimensional space.

To find a, b, and c, begin by observing that the edge vectors form an orthonormal set, and that the corner they define is convex. This leads to the relations:

$$\mathbf{a} = \mathbf{b}\times\mathbf{c} \quad (A3a)$$
$$\mathbf{b} = \mathbf{c}\times\mathbf{a} \quad (A3b)$$
$$\mathbf{c} = \mathbf{a}\times\mathbf{b} \quad (A3c)$$

Note that these equations are cyclic permutations of each other (i. e. any equation can be obtained from any other by the repeated substitution of $\mathbf{a} \to \mathbf{b}$, $\mathbf{b} \to \mathbf{c}$, and $\mathbf{c} \to \mathbf{a}$).

Using the relations of Equation A1 in the right-hand side of Equation A3, the z-components of the edge vectors can be written:

$$a_z = bc(B_xC_y - B_yC_x) = bc(\mathbf{B}\times\mathbf{C})\cdot\mathbf{k} \quad (A4a)$$
$$b_z = ca(C_xA_y - C_yA_x) = ca(\mathbf{C}\times\mathbf{A})\cdot\mathbf{k} \quad (A4b)$$
$$c_z = ab(A_xB_y - A_yB_x) = ca(\mathbf{A}\times\mathbf{B})\cdot\mathbf{k} \quad (A4c)$$

Because accidental alignments are assumed not to occur, z-components must be nonzero. As Equation A4 makes clear, this is equivalent to disregarding T-junctions, for which two of the three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are parallel.

Consider now the other two components of the edge vectors described by Equation A3, starting with Equation A3a. Substitution using Equation A1 yields:

$$a_x = aA_x = bB_yc_z - cC_yb_z \quad (A5a)$$
$$a_y = aA_y = -bB_xc_z - cC_xb_z \quad (A5b)$$

Multiplying Equation A5a by A_y and Equation A5b by A_x leads to:

$$b(B_yA_y + B_xA_x)c_x = c(C_yA_y + C_xA_x)b_x \qquad (A6)$$

Substituting the values of b_z and c_z in Equations A4b and A4c, and dividing through by a leads to:

$$\frac{b}{c} = \sqrt{\frac{(\mathbf{C}\cdot\mathbf{A})(\mathbf{C}\times\mathbf{A})\cdot\mathbf{k}}{(\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\times\mathbf{B})\cdot\mathbf{k}}} \qquad (A7a)$$

Because we are considering Y-junctions here, all angles between the constituent lines must be between 90° and 180°, and so all terms on the right-hand side must be negative. As such, their signs cancel out to give a positive quantity under the square root. Thus, the ratio b/c is a positive real number. Similarly, by cyclic permutation, the other ratios can be written:

$$\frac{c}{a} = \sqrt{\frac{(\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\times\mathbf{B})\cdot\mathbf{k}}{(\mathbf{B}\cdot\mathbf{C})(\mathbf{B}\times\mathbf{C})\cdot\mathbf{k}}} \qquad (A7b)$$

$$\frac{a}{b} = \sqrt{\frac{(\mathbf{B}\cdot\mathbf{C})(\mathbf{B}\times\mathbf{C})\cdot\mathbf{k}}{(\mathbf{C}\cdot\mathbf{A})(\mathbf{C}\times\mathbf{A})\cdot\mathbf{k}}} \qquad (A7c)$$

Finally, to establish the absolute value of one of the scaling factors, equate Equations A2a and A4a to get:

$$a_z = -\sqrt{1 - a^2(\mathbf{A}\cdot\mathbf{A})} = bc(\mathbf{B}\times\mathbf{C})\cdot\mathbf{k}$$

Squaring and using Equations A7b and A7c leads to (A8):

$$1 - a^2(\mathbf{A}\cdot\mathbf{A}) = [\frac{a^4(\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\times\mathbf{B})\cdot\mathbf{k}}{(\mathbf{B}\cdot\mathbf{C})}][\frac{(\mathbf{C}\cdot\mathbf{A})(\mathbf{C}\times\mathbf{A})\cdot\mathbf{k}}{(\mathbf{B}\cdot\mathbf{C})}]$$

This can be rewritten as:

$$Da^4 + A^2a^2 - 1 = 0 \qquad (A9)$$

where

$$A^2 = \mathbf{A}\cdot\mathbf{A} \quad \text{and}$$

$$D = \frac{[(\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\times\mathbf{B})\cdot\mathbf{k}][(\mathbf{C}\cdot\mathbf{A})(\mathbf{C}\times\mathbf{A})\cdot\mathbf{k}]}{(\mathbf{B}\cdot\mathbf{C})^2}$$

Note that D is always a positive quantity, because each of the four factors in its numerator is always negative.

Treating A9 as a quadratic equation in a^2 yields the solution:

$$a^2 = \frac{-A^2 \pm \sqrt{A^4 + D}}{2D} \qquad (A10)$$

with only the positive solution being used, because $a^2 > 0$.

Taking the positive square root of Equation A10 then yields the value of a. The values of b and c can then be determined using Equations A7b and A7c, respectively. Note that Equation A10 shows a to be inversely proportional to the length of $\mathbf{A}$. Similarly, the values of b and c are inversely proportional to $\mathbf{B}$ and $\mathbf{C}$, respectively. Thus, when they are substituted into Equation A1, the estimates of a, b, and c will be independent of the lengths initially used for $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. As such, all the relevant information is contained in the angles between the lines of the junction.

A similar treatment can be developed for the arrow junction formed by a convex corner (see Figure 1 for examples). For this case, a_x, $(\mathbf{A}\cdot\mathbf{C})$, and $(\mathbf{A}\cdot\mathbf{B})$ are all positive quantities, but apart from such details, the proof proceeds in the same way. Note that similar proofs can also be developed for junctions corresponding to concave corners, simply by changing the sizes of the z-components of the edge vectors. However, because the PRISM model is based on junctions assumed to correspond to convex corners, proofs regarding concave corners are not required here.