Visual Search II, D. Brogan, ed.
London: Taylor & Francis. 1992.

Chapter 4

A model for the rapid interpretation of line drawings in early vision

J.T. Enns and R.A. Rensink

Abstract

According to the prevailing view, the elements of early visual processing are characterized by simple geometric properties such as length, orientation, and curvature. We demonstrate in this chapter that this view must be revised—the elements of early vision need not be geometrically simple. Instead, they can be characterized in terms of environmental relevance, computational architecture, and processing speed. We begin by summarizing the conventional view of early vision and point to several problems it encounters. We then re-examine the role played by the elements of early vision, arguing that it is advantageous for them to describe environmentally relevant properties, even if these quantities are not always valid. As an illustration, we develop a computational model for the rapid recovery of one important scene property from line drawings—the three-dimensional orientation of objects. Data from recent visual search experiments in humans are presented in support of the model.

Early vision

Since the days of von Helmholtz (1867-1962), vision researchers have distinguished between *early* and *later* stages of visual processing. Researchers studying the physiology of vision use ‘early’ to refer to processes up to and including the striate cortex, and ‘later’ to refer to all subsequent stages (Zucker, 1987). For psychophysical researchers, ‘early’ refers to preattentive and ‘later’ to attentive processes (Treisman, 1986). Although the constructs in these two areas are not yet linked explicitly, (Treisman et al., 1990), researchers in both areas believe that early vision involves specialized processes that are rapid (i.e., they take place within 50-100 s), spatially parallel (i.e., they operate simultaneously across the visual field) and automatic (i.e., are relatively uninfluenced by moment-to-moment changes in the goal of the organism). The visual features thought to be computable by such processes are generally characterized as

properties of very simple geometric elements, including the orientation, length, curvature, and motion direction of elongated blobs (Beck, 1982; Julesz, 1984; Treisman et al., 1990).

Questioning the conventional view

We have been questioning this conventional view of early vision in our laboratory at the University of British Columbia for several years. In this section we will spell out some of these questions, focusing first on the processes and then on the representations of early vision.

Processes

Recent empirical findings have challenged the conventional dichotomy of 'early' versus 'later' vision. One of the primary diagnostics of this distinction, the visual search task, consistently yields search rates that vary smoothly from very fast (i.e., less than 10 ms per item) to very slow (i.e., more than 100 ms per item) (Treisman and Souther, 1985; Duncan and Hurnphreys, 1989). Another diagnostic, the texture segmentation task, indicates that texture segregation ranges continuously from being almost immediate (i.e., less than 50 ms exposure is required) to being very effortful (i.e., more than 200 ms is required). (Nothdurft, 1985; Callaghan , 1986; Enns, 1986; Taylor and Badcock, 1988; Callaghan, 1989). These findings are now leading to proposals that the processes of early vision may blend smoothly into those of later vision, with the point of intersection depending primarily on the similarity between the elements involved (Julesz, 1986; Treisman and Gormican, 1988; Duncan and Humphreys, 1989; Humphreys *et al*., 1989). If similarity is indeed being assessed at the earliest stages of vision, then processes more complex than simple feature registration must be occurring.

Representations

The view that the features of early vision are geometrically simple has been challenged by reports that rapid search is possible for targets defined only by conjunctions of their features. Such features have included binocular disparity and motion (Nakayama and Silverman, 1986), motion and form (McLeod *et al.*, 1988), saturated colours, large forms and distinctive orientations (Treisman, 1988; Wolfe et al., 1989) and spatial relations among line elements that are sufficiently long (Duncan and Hurnphreys, 1989; Humphreys *et al.*, 1989).

Another challenge has come from the discovery that early vision is sensitive to aspects of the three-dimensional scene that gives rise to the two-dimensional image (Ramachandran, 1988; Hollidayand Braddick, 1989; Ramachandran and Plummer, 1989; Enns, 1990; Enns and Rensink, 1990a; Enns and Rensink, 1990b; Epstein and Babler, 1990). Although these features are complex conjunc-

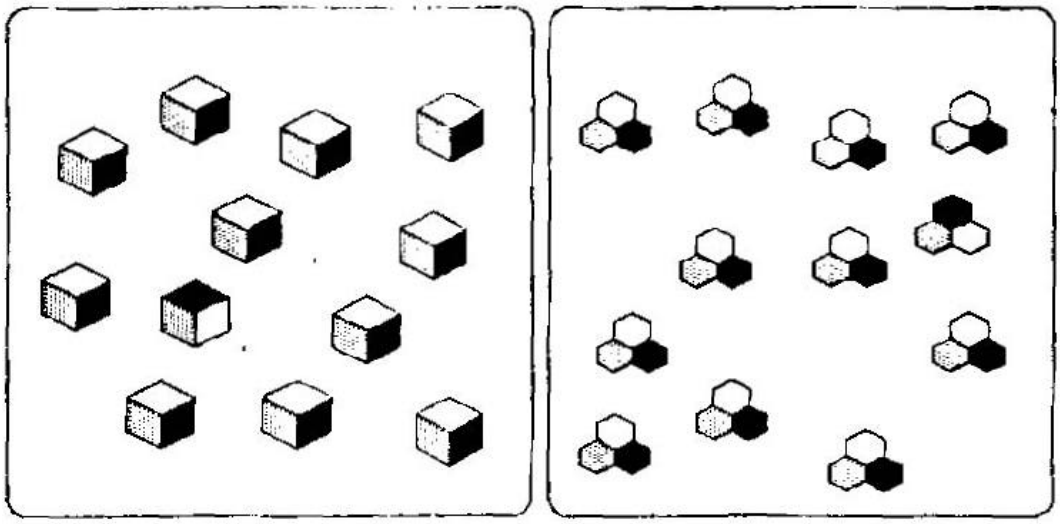

Figure 4.1(a) Rapid search is possible for items defined by spatial relations between polygons of *different luminance, provided that the relations correspond to the property of lighting direction in the corresponding scene; (b) similar conjunctions that do not readily correspond to three-dimensional objects result in slow search (from Enns and Rensink, 1990a).*

tions when defined as image elements, they can be shown to be simple features when defined with regard to the corresponding scene. For example, rapid search is possible for items defined by the spatial relations between polygons of different luminance (see Figure 4.1). However, it can also be shown that the visual system does not treat these items as arbitrary collections of features. Rapid search is possible only when the 'feature conjunctions' correspond to objects with different lighting direction in the scene (Figure 4. 1(a). Similar conjunctions that do not readily correspond to three-dimensional objects result in slow search (Figure 4.1(b)).

A second example shows that rapid search can be based on the three-dimensional orientation of objects (see Figure 4.2). Search is rapid when items can be interpreted as blocks with different three-dimensional orientations in the scene (Figure 4.2(b)), but much slower when items do not lend themselves readily to such an interpretation (Figure 4.2(b)). Empirical findings such as these, therefore, provide convincing evidence that complex, environmentally relevant features are represented in early vision. But how is this accomplished?

General characteristics of rapid recovery in early vision

To understand a visual process, it is essential to determine not only the representations used, but also the function of the process (Marr, 1982). Thus, as a first step towards establishing a framework for the rapid recovery of environmentally relevant information we will discuss our views on the function of early vision. Having done so, we then reconsider the processes and representations that can be used to carry it out.

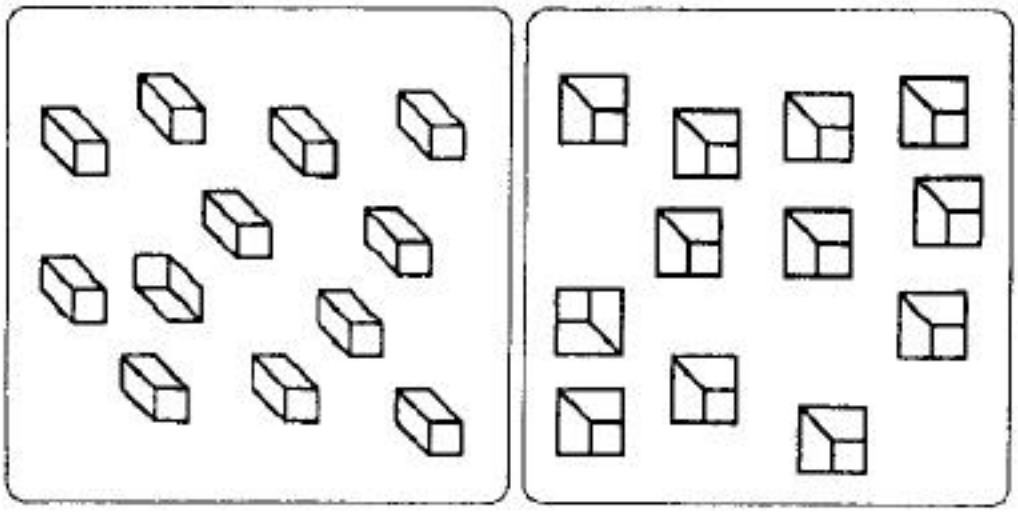

Figure 4.2(a) Rapid search is possible for items defined by spatial relations between lines, provided that the relations correspond to the property of three-dimensional orientation in the corresponding scene; (b) items that do not lend themselves to such an interpretation (from Enns & Rensink, 1990b).

Function

The goal of every biological vision system is to provide information to the organism about its surrounding environment. This information can be said to be 'useful' if it can be used to influence the actions taken by the organism. We take it as axiomatic that the early vision system must contribute to this goal in some way.

In the conventional view, early vision simply transforms retinal patterns of light into sets of simple geometric elements organized in topographic maps. This leaves the later visual processes with considerable work to do if such tasks as object recognition and scene perception are to be carried out within several hundred milliseconds from display onset. How could these tasks be accomplished in such a limited amount of time? We decided to consider an alternative function for early vision—that early vision had evolved as a high-speed system for the delineation of objects in a three-dimensional world (Enns, 1992). If early vision could recover even a limited number of properties of the three-dimensional scene, it would be able to guide the operation of more flexible processes further down the visual stream (Weisstein and Maquire 1978; Walters, 1987).

In thinking about this possibility, we found it helpful to distinguish between the information content of a visual feature and its validity. Information content can be thought of, loosely speaking, as the 'amount' of information contained in the feature. This can vary from pixel-by-pixel intensity information, to a complete description of the intrinsic scene properties. Information validity refers to the extent to which the properties measured by the system are the ones actually in the image or scene. We noted that there appeared to be an inherent trade-off between these two quantities in any visual system. For example, the conventional view of early vision assumes the validity of the information to be of

great importance as it will be the basis for subsequent processing. Because of this emphasis on validity, only features of relatively low information content can be extracted from the image. For example, all edges in the image of some particular length may be represented. If increased information content is desired at this stage (e.g. information about surface orientation), early vision would require more global processes to preserve the same degree of validity, thereby losing its great speed.

But what if the requirement of validity were relaxed somewhat in early vision? This would permit a substantially higher information content to be recovered rapidly, automatically and in parallel. For instance, it might be possible to obtain 'quick and dirty' estimates of surface orientation all over the image. These estimates would not always be valid, but they might on average help to guide immediate actions. Furthermore, relaxing validity in this way does not mean that the ultimate results of visual processing are necessarily less reliable as validity can still be established via constraints at later stages of processing. In essence, this means that more effective processing can be done early on by lifting the demand for validity from the shoulders of early vision and distributing it more evenly over a number of processing levels.

Processes

What kind of processes are consistent with the function of early vision as sketched above? We believe that neither the analysis of function nor the theories of processing are detailed enough at this time to allow a choice to be made between the dichotomous and continuous views of processing. However, we can outline a few considerations that must apply in either case.

Consider first the issue of processing speed. Recall that eye and hand movements each take approximately 200 ms to initiate, and that even covert movements of attention often take at least 50 ms. If the results of early vision are to be useful for guiding motor actions and for guiding later processing, they must be available within the first 50-100 ms of processing. It is therefore essential that processing be rapid.

A similar argument holds for parallel processing. If a system has a limited number of discrete processors, and speed is important, it is generally more efficient to have the processors do their work simultaneously than sequentially. Note that this assumption is in keeping with the comments made earlier about the content versus the validity of the recovered information. Parallel processes allow for higher information throughput than do serial processes, but the cost associated with this increase in information is a decrease in its validity. Later serial processes are required to determine whether the quantities recovered in parallel are consistent with one another and with the scene as a whole.

Given that the processes of early vision are both rapid and parallel, and that later processes are generally neither, it follows that top-down influences must not affect the actual operation of the low-level processes. The processes of early vision must therefore be largely automatic: once initiated, they will run to

completion in an all-or-none fashion. Note that this does not rule out all top-down influence. Higher level considerations may still determine which low-level processes will be run.

Representations

Given these assumptions about the function and structure of early vision, what can be said about the representations it employs? We believe that the elements of these representations can be described in terms of the following three characteristics:

1. *Local measurements*. Elements must be based on local measurements, since these are the only kinds that can be computed in parallel across the image. However, they do not need to be as 'dense' or as 'local' as every point in the image. For instance, they could conceivably be computed over spatial regions of limited extent at a relatively sparse set of locations in the visual field. We will call these limited regions 'neighbourhoods'.

2. *Relaxed validity*. To be computed rapidly the descriptions must run the risk of being invalid some of the time. There are two ways in which in which the validity requirement can be relaxed to maintain high information content In general, the time to complete a computation increases both with the number of candidate interpretations and with the degree of consistency checking between neighbourhoods (Garey and Johnson, 1979). Rapid processing can therefore be maintained by limiting computations to only a few 'useful' candidates and by keeping consistency checking to a minimum.

3. *Environmentally relevant properties*. The small number of candidate interpretations considered in each neighbourhood should ideally be relevant to the larger task of the visual system. As a first approximation, therefore, they should at least be relevant to the task of determining the surface shape of objects and the layout of objects in the larger scene, This assumption prevents the system from having to respond to stimuli of arbitrary complexity, Presumably, evolution has endowed the visual system with the ability to compute quantities that correspond to environmentally relevant properties most of the time.

A model with these three characteristics will be referred to here as a PRISM model of early vision, since it is based on the parallel and rapid interpretation of scene magnitudes. Such a model will provide a rapid 'first pass' of a visual image, picking out a 'best' interpretation at each location and passing the rest of the two-dimensional descriptions on to higher-level processes. The interpretations formed in this way would be not form a valid reconstruction of the scene every lime, since the small number of interpretations considered at each location would often fail to match the physical world. What could be expected, however, is that these matches would occur at least some of the time, so that environmentally relevant properties would be recovered at a several locations in the visual field. These descriptions should be useful for guiding immediate actions (e.g. eye movements, grasping, locomotion) and for guiding attentive processes further along the visual stream.

Blocks world interpretation through line-labelling

To illustrate how this framework can be applied to early visual processing, we shall develop a specific model for the rapid recovery of three-dimensional orientation from line drawings. We chose to start with the domain of line drawings, both because it has been studied extensively in the field of computational vision, and because humans interpret line drawings very rapidly (Biederman, 1985). As background to this model, then, we will first discuss the general problem of line drawing interpretation.

Given an image composed of several orthographically projected line drawings, how might the corresponding three-dimensional scene be recovered? Computational work on this question is based on the blocks world, a scene domain of polyhedral objects consisting only of trihedral corners (i.e. corners formed from three polygonal faces) (Clowes, 1971; Huffman, I971; Waltz, 1972; Mackworth, 1973). The objects are assumed to have uniform reflectances on all surfaces, so that no information about their structure is available from shading. Furthermore, viewing direction and the direction of lighting are held constant, with the two directions being made coincident to avoid shadows. This results in an image domain consisting of straight-line segments connected by dilinear or trilinear junctions. The problem is then to recover from these images the remaining scene properties of surface orientation and location.

Blocks world interpretation is based on 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, while the third is formed from the boundary of a face that occludes a second, noncontiguous surface or background. To interpret a line drawing correctly, each line must be labelled as corresponding to a particular kind of edge, with the labelling being consistent for all lines in the image.

Several algorithms to carry out the line- labelling process have been developed (e.g. Waltz, 1972; Horn, 1986; Mulder and Dawson, 1990). 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 4.3). There also exists a fourth class of dilinear junctions, L-junctions, which correspond to corners of single visible faces.

As is evident in Figure 4.3, trilinear junctions 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 immediate neighbours. This process is iterated until the interpretation at each junction in the drawing is consistent with those at all other junctions in the image.

Note that the three trilinear junctions differ in the kind of quantitative information they carry about the scene. T-junctions most often correspond to

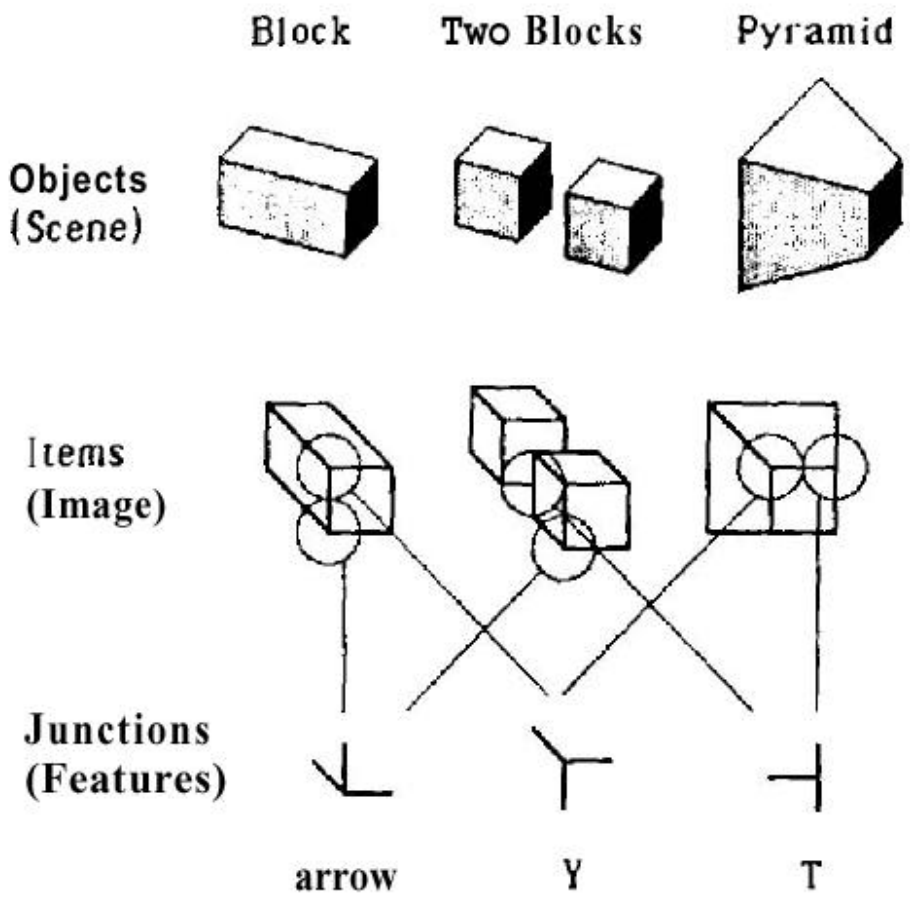

Figure 4.3 In line drawings of polyhedral objects, junctions involving three fines fall into three classes: arrow-, Y- and T-junctions. However, there is not a unique correspondence between a given junction and its three-dimensional interpretation. This can only be determined by considering the entire system of junctions.

recover the orientations of the surfaces at the corresponding corner, provided that the surfaces are mutually orthogonal to one another. The law of Perkins (1968) states that for an arrow-junction corresponding to an orthogonal corner, the sum of the two smallest angles must be at least 90°; for Y-junctions each of the two angles must be at least 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 Mackworth, 1976). Mulder and Dawson (1990) have extended these ideas recently, showing that this information can be used to recover the three-dimensional orientations of all the surfaces and edges of a large class of polyhedral objects.

It is important to point out 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 (see Figure 4.3). This junction is consistent with the laws of Perkins (1968), but does not actually 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.

A PRISM model for the recovery of three-dimensional orientation from line drawings

We now consider how the process of line drawing interpretation might be carried with a PRISM model. First, the requirement of local measurement can be met by assuming that trihedral junctions form the basis for the computations in each neighbourhood.

Second, the requirement of relaxed validity can be met by considering only the most likely interpretations for each of the junctions. For instance, T-junctions can be used in the computation of object segmentation; arrow- and Y-junctions can be used to estimate three-dimensional orientation at each of their locations. Limited consistency checking of these estimates can be ensured by having each neighbourhood communicate with only its nearest neighbours.

Third, the demand for environmental relevance is already met by choosing object segmentation and three-dimensional orientation as the quantities of interest. Object segmentation is essential for the interaction of the organism with discrete entities in its environment; the three-dimensional orientation of surfaces is crucial for positioning limbs on the basis of visual information (e.g. grasping, locomotion over uneven terrain).

The model itself can be separated into two distinct phases: (i) the generation of one candidate interpretation for each trilinear junction, followed by (ii) a nearest-neighbour check of the consistency between estimates. Although the two phases are necessarily applied in sequence (the second phase operating on values determined by the first) each phase itself is carried out in parallel across the visual field. The results of each phase of the model are illustrated in Figure 4.4 for some of the visual search items used in the experiments to be reported.

Phase 1: local estimates of orientation and occlusion

The three-dimensional orientation of a convex and orthogonal trihedral corner can be recovered from the projected orientations of the corresponding lines in the image plane (Perkins, 1968). Therefore, if comer convexity and mutual orthogonality can be assumed, estimates of three-dimensional orientation can be The assumption of convexity follows naturally from the observation that convex corners determine the overall three-dimensional shape of an object. Concave corners correspond to indentations in and deformations of the global structure (Pentland, 1986). Thus, corner convexity appears to be a reasonable default assumption. The assumption of orthogonality is more difficult to justify. Corners are rarely formed from perfectly orthogonal surface in the natural world. However, if there is no other way to determine three-dimensional

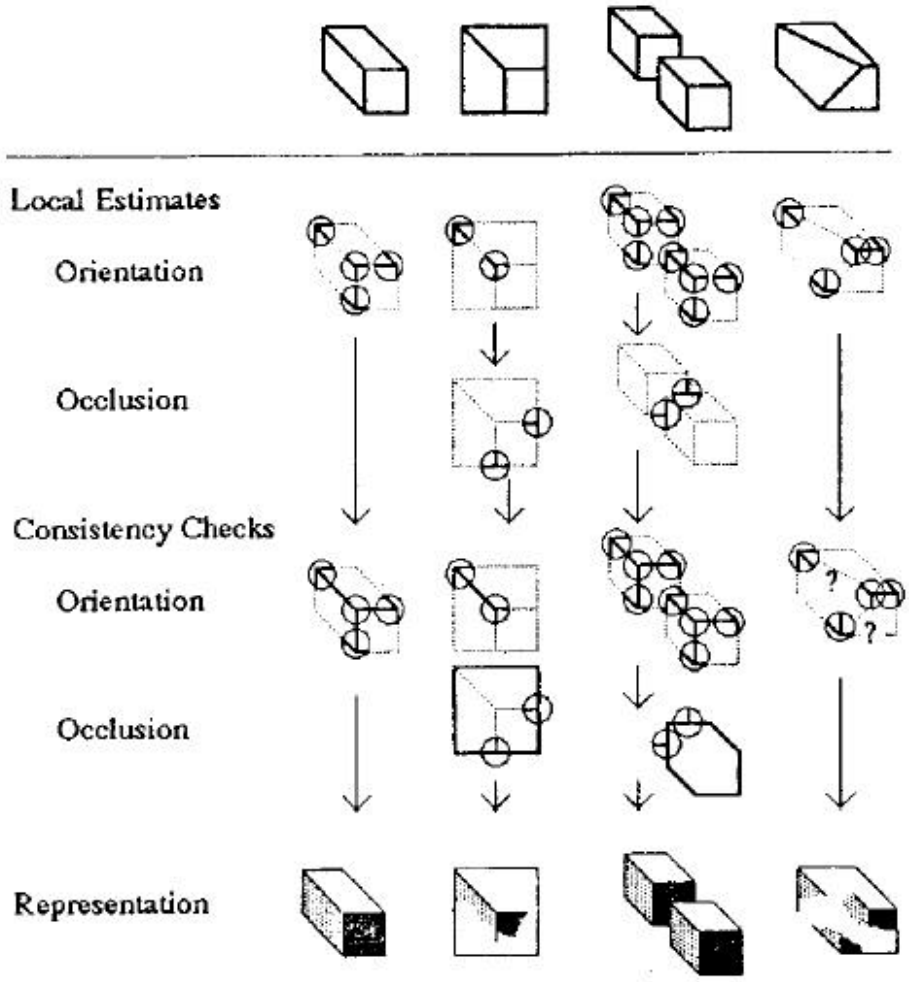

Figure 4.4 Illustration of the PRISM model applied to several items in the visual search experiments. In the first phase (local estimation), estimates of three-dimensional orientation are obtained from arrow-and Y-junctions; estimates of occlusion are based on T-junctions. In the second phase (consistency checking), neighbouring three-dimensional orientation estimates and occlusion estimates are checked for consistency.

orientation, the visual system may well assume mutual 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 (Perkins, 1972; Shepard, 1981; Butler and Kring, 1987). They even 'see' rectangular corners when they know orthogonality has been violated (Kubovy, 1986). In addition to these reasons, orthogonal angles may also be natural defaults simply because they lie midway on the range of all possible angles between two surfaces.

When one surfaces occludes another in the scene, 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. To segment the image, we propose a simplified variant of the scheme used for the blocks world, 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 interpretation of the stem of the T cannot be determined on the basis of purely local considerations—it could be a convex, concave or boundary edge. However, the situation is quite different for the crossbar. Apart from cases of accidental alignment, this line corresponds to a boundary edge that occludes the surface(s) associated with the stem. Consequently, it must belong to a different group of lines, which can be signalled by marking the crossbar as an occluding boundary edge.

Phase 2: consistency checking of orientation occlusion

The local estimates of orientation and occlusion must be consistent with each other if the lines in the image correspond to orthographic projections of solid objects. This can be done rapidly and in parallel by only comparing estimates from immediately neighbouring junctions. If these estimates are compatible, they will reinforce the validity of the interpretation. If an inconsistency is detected, the interpretation will fail.

Consistency of orientation estimates can be carried out in parallel for each segmented group of neighbourhoods. Since this test involves the transmission of information across neighbourhoods, 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 is integrated across the visual field. Independent estimates based on contrast discrimination (Jamar and Koenderink, 1983) and line drawing discrimination tasks (Enns and Girgus, 1986; Enns and King, 1990) suggest speeds of 20-30 ms per degree of visual angle. Since the size of the regions considered here are relatively small (1.5°), this operation would add a small constant time factor to the interpretation process.

To ensure that orientation consistency is checked only over regions that correspond to actual faces or boundaries in the scene, it is useful to first segment neighbourhoods into groups that correspond to separate objects. As in the case of orientation estimates, these assignments must be checked for consistency with their neighbours. One way this can 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 be the other.

Comparison of the PRISM model with empirical results

To compare the PRISM model with the results of visual search experiments, we took the generally accepted position that search rates reflect the signal-to-noise

ratio of the target amidst the distractors (Duncan and Humphreys, 1989; Treisman and Gormican, 1988). How serial or parallel processes 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.

We only summarize the methods used, since they are available elsewhere (Enns, 1990; Enns and Rensink, 1990a, 1990b). A standard visual search task was used in which observers searched for one target item among a total of 1,6 or 12 items (e.g. Treisman, 1988; Treisman and Gormican, 1988; Wolfe et al., 1989). Target and distractor items were composed of identical line segments that differed only in their spatial arrangement (see Figure 4.5). The target was present on a random one-half of the trials. The dependent variable of interest was the slope of the reaction time (RT) functions over display size, or search rate. As pointed out earlier, there is no sharp boundary between fast and slow rates of search. We use 'rapid search' to refer to target-present search rates (RT slopes) of less than 10- 15 ms per item. This speed is well above accepted estimates of attentional movement across the visual field (Julesz, 1984; Jolicoeur et al., 1986; Treisman and Gormican, 1988).

Early vision is sensitive to quantitative relations among lines. The rapid search found for the drawings of blocks in Figure 4.2(a) is a natural consequence of the PRlSM model. As shown in Figure 4.4, all junctions in these items can be assigned orientation estimates and these junctions pass the consistency tests. Targets and distractors are therefore interpreted as blocks with different three-dimensional orientations.

The items in Figure 4.2(b), on the other hand, contain T-junctions. Since these junctions do not lead to a consistent segmentation of the outlining contour, as shown in Figure 4.4, the items cannot be interpreted as convex objects. Search is consequently slow, in the range conventionally considered to be the result of attentive processes.

Early vision is not equally sensitive to all trihedral junctions. The slow search found for isolated T-junctions (Figure 4.5(a)) is also to be expected, since these junctions cannot give rise to estimates of three-dimensional orientation. According to the PRISM model, however, this quantity can be recovered for Y- and arrow-junctions (Figure 4,5(b), 5,5(c)) and search is consequently more rapid for these junctions. It is worth noting that if the arrow-junctions are interpreted as portions of 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 the data. In contrast, the region of the right-angle in the Y-junctions corresponds to a planar face that has a similar orientation in target and distractor. Since overlapping sets of features slow down search (Duncan and Humphreys, 1989). these arrow-junctions should lead to faster search than the Y-junctions.

Early vision is sensitive to the context in which junctions appear. The range of search rates found for the items in Figure 4.5(d)-(f) are also explained by the

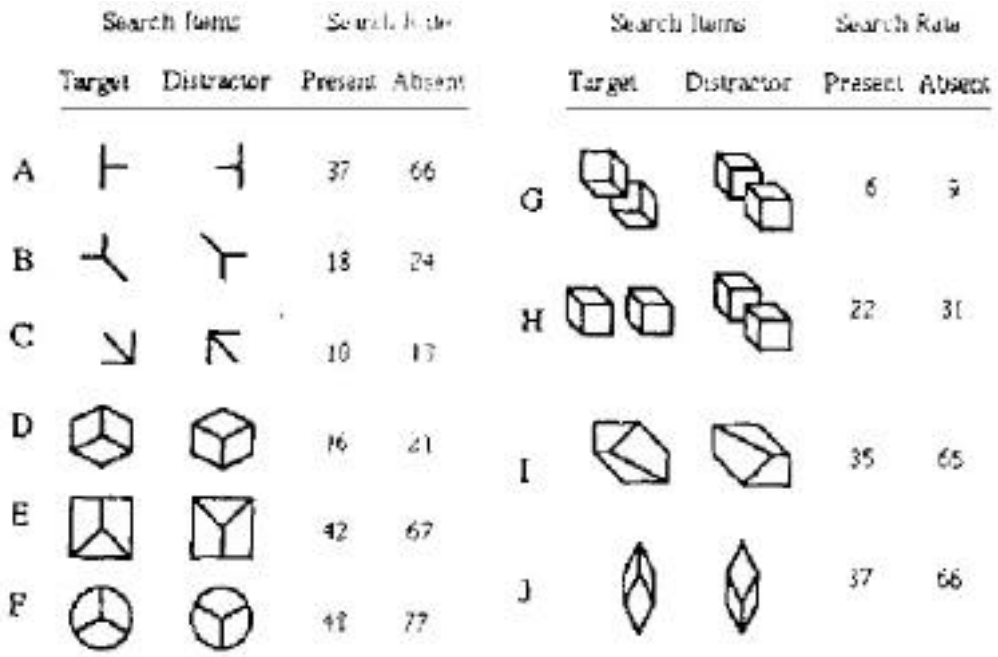

Figure 4.5 Selected search items and corresponding search rates from visual search experiments designed to test the PRISM model.

model. The rapid search for items in Figure 4.5(d) is expected since their line structure is similar to the blocks of Figure 4.2. In contrast, items in Figure 4.5(e)-(f) could not be interpreted consistently because their T-junctions did not result in object segmentation. As in the case of the pyramids in Figure 4.2, the failure to find consistency among these junctions actively disrupts the interpretation process.

Figure 4.5(d) shows that if arrow- and Y-junctions are connected together by lines, search is no slower than for any of the individual junctions. In contrast, the presence of a single T-junction in Figure 4.5(e) 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. The items in Figures 4.5(f) contain three T-junctions, and yet they are only subtly different from Figure 4.5(e). None the less, these small changes are sufficient to slow down search dramatically.

This context dependency also extends to T-junctions. Unlike those in previous items (Figure 4.5(e)-(f)), the T-junctions in Figure 4.5(g) give rise to a consistent segmentation of the lines. Consequently, search is rapid for these items. However, although early vision can apparently use T-junctions to segment objects from one another, it cannot use the T-junction itself as the basis for rapid detection. Items in Figure 4.5(h) differ in the presence] absence of T-junctions and yet search remains slow for these items. Taken together, these results show that early vision is very sensitive to the entire system of line relations in an item.

Early vision is also sensitive to the orthogonality constraint. In Figure 4.5(i), the items contain arrow- and Y-junctions that violate the orthogonality constraint.

These items had the same outline as those in Figure 4.2, 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 non-parallel orientations of the resulting lines, Figure 4.5(j) used drawings with similar Y-junctions, but in which parallel line orientations were maintained. Since both of these sets of items violate the orthogonality constraints, the local estimates made for their junctions will not be correct, and the associated orientation consistency check will fail. Targets are therefore indistinguishable from distractors in early vision and search is quite slow.

Conclusion

The visual search experiments described here have demonstrated that early human vision is much more sophisticated than has generally been assumed. In particular, they show that early vision can recover three-dimensional orientation from the information contained in line drawings alone. Our proposed PRISM model shows how this can be done by processes operating automatically, rapidly and in parallel across the visual field.

These results have three important implications for a revised view of early vision. First, it is unnecessarily restrictive to assume that early vision operates only on simple geometric elements. Although there must indeed be an initial stage which analyses the retinal input in this way, our findings show that there must also be subsequent stages of parallel processing based on more complex properties.

Second, the elements of early vision may be characterized by environmental relevance. Our results show that early vision is sensitive to at least the three-dimensional orientation of surfaces in the scene and the relation of occlusion. Other work has indicated that it is also sensitive to lighting direction (Ramachandran, 1988; Enns and Rensink, 1990a). It will be interesting to see which other properties can be recovered in the early stages of processing.

Finally, the elements of early vision must be rapidly computable. As we have argued, early visual processes cannot afford the time required for complete and valid interpretations of the scene. Instead, they appear to have a limited amount of time in which they 'do their best' to recover a limited set of environmentally relevant properties. These informationally complex, albeit sometimes invalid, representations can then be used to guide reflexive actions. Given sufficient time, they can also be used by the later, attentive processes to recover representations of the scene that are more valid.

References

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

Biederman, I., 1985, Human image understanding: recent research and a theory, C*omputer Vision, Graphics and Image Processing,* 32, 29-73.

Butler, D. L. and Kring, A. M., 1987, Integration of features in depictions as a function of size, *Perception and Psychophysics, 41,* 159-65.

Callaghan, T. C., 1989, Interference and dominance in texture segregation: hue, geometric form, and line orientation, *Perception and Psychophysics,* 299-311.

Callaghan, T. C., Lasaga, M. I. and Garner, W. R., 1986, Visual texture segregation based on orientation and hue, *Perception and Psychophysics,* 39, 32-8.

Clowes, M. B., 1971, On seeing things, *Artificial Intelligence, 2,* 79- 116.

Duncan, J. and Humphreys, *G.* W., 1989, Visual search and stimulus similarity, *Psychological Review, 96,* 433 *-58.*

Enns, J, T., 1986, Seeing textons in context, *Perception* and *Psychophysics,* 39, 143-7.

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

Enns, J. T., 1992, The nature of selectivity in early human vision, in Burns, B. (Ed.) *Percepts, Concepts, and Categories: The Representation and Processing of Information,* pp. 39-74, Amsterdam: Elsevier Science.

Enns, J. T., and Girgus, J. S., 1986, A developmental study of shape integration over space and time, *Developmental Psychology,* 22, 491 -9.

Enns, J. T., and King, K. A., 1990, Components of line drawing interpretation. *Developmental Psychology,* 26, 469 -79.

Enns, J. T., and Rensink, R. A., 1990a, Influence of scene-based properties on visual search, Science, 247, 72 1 -3.

Enns, J. T., and Rensink, R. A., 1990b, Sensitivity to three-dimensional orientation in visual search, *Psychological Science,* 1, 323-6.

Epstein, W., and Babler, T., 1990, In search of depth, *Perception and Psychophysics,48,* 68-76.

Garey, M. R. and Johnson, D.S., 1979, *Computers and intractability: A Guide to the Theory of NP-completeness,* New York; W. H. Freeman.

Holliday, I. E. and Braddick, O. J., 1989, 'Search for stereoscopic slant direction is parallel', presentation at the 12th European Congress on Visual Perception.

Horn, B. K. P., 1986, *Robot Vision,* Cambridge: MIT Press.

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

Humphreys, G. W., Quinlan, P. T. and Riddoch, M. J., 1989, Grouping processes in Visual search: effects with single- and combined-feature targets, *Journal of Experimental Psychology: General,* 118, *258-79.*

Jamar, J. H. T. and Koenderink, J. J., 1983, Sine-wave gratings: scale invariance and spatial integration at suprathreshold contrast, *Vision Research,* 23, *805- 10.*

Jolicoeur, P., Ullrnan, S. and MacKay, M., 1986, Curve tracing: a possible basic operation in the perception of spatial relations, *Memory and Cognition,* 14, 129-40.

Julesz, B., 1984, A brief outline of the texton theory of human vision, *Trends in Neuroscience,* 7, 41-5.

Julesz, B., 1986, Texton gradients: the texton theory revisited, *Biological Cybernetics,* 54, 245-61.

Kubovy, M., 1986, *7he Psychology of Perspective and Renaissance Art,* Cambridge. UK: Cambridge University Press.

Mackworth, A. K., 1973, Interpreting pictures of polyhedral scenes, *Artificial Intelligence,* 4, 121-37.

Mackworth, A. K., 1976, Model-driven interpretation in intelligent vision systems, *Perception, 5,* 349-70.

Marr, D., 1982, *Vision,* San Francisco: W. H. Freeman.

McLeod, P., Driver, J. and Crisp, J., 1988, Visual search for a conjunction of movement and form is parallel, *Nature,* 332, 154-5.

Mulder, J. A. and Dawson, R. J. M., May 1990, Reconstructing polyhedral scenes from single two-dimensional images: The orthogonality hypothesis, in Patel-Schneider, P. K. (Ed .) *Proceedings of the 8th Biennial Conference of the CSCSI,* pp. 238-44, Palo AIto, CA: Morgan-Kaufmann.

Nakayarna, K. and Silverman, *G.* H., 1986, Serial and parallel processing of visual feature conjunctions, *Nature,* 320, 264-5.

Nothdurft, H. C., 1985, Sensitivity for structure gradient in texture discrimination tasks, *Vision* Research, 25, *957-68.*

Penrland, A. P., 1986, Perceptual organization and the representation of natural form, *Artificial Intelligence,* 28, 293-331.

Perkins, D. N., 1968, Cubic corners, *MIT Research Laboratory of Electronics Quarterly Progress Report,* 89, 207- 14.

Ptrkins, D. N., 1972, Visual discrimination between rectangular and nonrectangular parallelopipeds, *Perception and Psychophysics, 12,* 396-400.

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

Ramachandran, V, S. and Plummer, D. J., 1989, 'Preattentive perception of 3-D versus 2-D image features', presentation at the Association for Research in Vision and Ophthalmology, Sarasota, Florida.

Shepard, R. N., 1981, Psychophysical complementarity, in Kubovy, M. and Pomerantz, J. R. (Eds) *Perceptual Organization,* pp. 279-342, Hillsdale, NJ: Erlbaum.

Taylor, S, and Badcock, D., 1988, Processing feature density in preattentive perception, *Perception and Psychophysics,* 44, 55-42 .

Treisman, A,, 1986, Features and objects in visual processing, *Scientific American,* 255, 106-15.

Treisman, A., 1988, Features and objects; the fourteenth Bartlett memorial lecture, *Quarterly Journal of Experimental Psychology,* 40A, 201-37.

Tieisman, A, and Gormican, S., *1988,* Feature analysis in early *vision:* evidence from search asymmetries, Psychological Review, 95, 15-48.

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

Treisman, A., Cavanagh, P., Fischer, B., Ramachandran, V. S. and von der Heydt, R., 1990, Form perception and artention: striate cortex and beyond, in Spillman, L. and Werner, J. S. (Eds), *Visual Perception,* pp. 273-316, New York: Academic Press.

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

von Helmholtz, H ., 1867-1967, *Treatise on Physiological Optics,* VoI. 3, in Southall, J. P. C. (Ed. and Trans.) NY: Dover.

Walters, D., 1987, Selection of image primitives for general purpose visual processing, *Computer Vision, Graphics, and huge Processing,* 37, 26 1-98.

Waltz, D. t., 1972, Generating semantic descriptions from drawings of scenes with shadows, A1-TR-271, Project MAC, MIT (Reprinted in Winston, P. H. (Ed.), 1975, *The psychology Computer Vision,* pp. 19-92, New York: McGraw-Hill.

Weissrein , N. and Maquire, W., 1978, Computing the next step: psychophysical measures of representation and interpretation, in Hansen, A. R. and Riseman, E. M. (Eds), *Computer Vision Systems,* pp. 243-60, New York: Academic Press.

Wolfe, J. M., Cave, K. R, and Franzel, S. L., 1988, Guided search: an alternative to the feature integration model for visual search, *Journal* of *Experimental Psychology: Human Perception and Performance,* 15, 4 19-33.

Zucker, S. W., 1987, Early vision, in Shapiro, S. C. (Ed.) *The Encyclopedia of Artificial*

References

The research was supported by grants from NSERC (J. E.; R. R. via R. J. Woodham) and UBC CICSR (R. R.). Correspondence concerning this chapter may be addressed to the first author, Department of Psychology, University of British Columbia, Vancouver, Canada V6T 1Y7.