
Research Report

**SENSITIVITY TO THREE-DIMENSIONAL ORIENTATION IN
VISUAL SEARCH**
James T. Enns¹ and Ronald A. Rensink²¹Department of Psychology and ²Department of Computer Science, University of British Columbia

Abstract—*Previous theories of early vision have assumed that visual search is based on simple two-dimensional aspects of an image, such as the orientation of edges and lines. It is shown here that search can also be based on three-dimensional orientation of objects in the corresponding scene, provided that these objects are simple convex blocks. Direct comparison shows that image-based and scene-based orientation are similar in their ability to facilitate search. These findings support the hypothesis that scene-based properties are represented at preattentive levels in early vision.*

Visual search is a powerful tool for investigating the representations and processes at the earliest stages of human vision. In this task, observers try to determine as rapidly as possible whether a given target item is present or absent in a display. If the time to detect the target is relatively independent of the number of other items present, the display is considered to contain a distinctive visual feature. Features found in this way (e.g. orientation, color, motion) are taken to be the primitive elements of the visual systems.

The most comprehensive theories of visual search (Beck, 1982; Julesz, 1984; Treisman, 1986) hypothesize the existence of two visual subsystems. A *preattentive* system detects features in parallel across the visual field. Spatial relations between features are not registered at this stage. These can only be determined by an *attentive* system that inspects serially each collection of features in the image.

Recent findings, however, have argued for more sophisticated preattentive processes. For example, numerous reports show features to be context-sensitive (Callaghan, 1989; Enns, 1986; Nothdurft, 1985). Others show that spatial conjunctions of features permit rapid search under some conditions (McLeod, Driver, & Crisp, 1988; Treisman, 1988; Wolfe, Franzel, & Cave, 1988). These findings suggest that spatial information can be used at the preattentive stage.

Recent studies also suggest that the features are more complex than previously thought. For example, rapid search is possible for items defined by differences in binocular disparity (Nakayama & Silverman, 1986), raising the possibility that stereoscopic depth may be determined preattentively. Indeed, it appears that the features do not simply describe two-dimensional aspects of the image, but also describe attributes of the three-dimensional *scene* that gave rise to the image. Ramachandran (1988) has shown that the convexity/concavity of surfaces permits spontaneous texture segregation, and Enns and Rensink (1990) have found that search for shaded polygons is rapid when these items can be interpreted as three-dimensional blocks. Although the relevant scene-based properties present at preattentive levels have not yet been completely mapped out, likely candidates include lighting direction, surface reflectance, and three-dimensional orientation.

Correspondence should be addressed to Ronald A. Rensink or James T. Enns, Department of Psychology, University of British Columbia., 2136 West Mall, Vancouver BC V6T 1Z4, Canada. Email: rensink@psych.ubc.ca; jenns@psych.ubc.ca.

In this paper we systematically examine one of these candidates: three-dimensional orientation. Rapid search can be based on the orientation of items in the image plane (Beck, 1982; Julesz, 1984; Treisman, 1986), and it is natural to ask whether the same holds for orientation in the three-dimensional scene. To answer this question, we conducted a series of visual search experiments using the simple black-and-white items shown in Figures 1 to 3.

Many of the target and distractor items were projections of rectangular objects differing in the three-dimensional orientation of their principal axes. These items always contained the same set of image-based lines and polygons. Therefore, if rapid search were possible, it would depend necessarily on the spatial relations that capture the three-dimensional orientation of the corresponding objects.

GENERAL METHOD

A Macintosh computer was used to generate the displays, control the experiments, and collect the data (Enns, Ochs, & Rensink, 1990). In each experiment, observers searched for a single target item among a total of 1, 6, or 12 items. The target was present in half the trials, randomly distributed throughout the trial sequence. Items were distributed 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 location by ±0.5° to prevent influences of item collinearity.

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.

Five to ten observers completed 4 to 6 sets of 60 test trials in each condition of a given experiment. Target presence or absence was reported by pressing one of two response keys. Observers were instructed to maintain fixation and to keep errors below 10%.

DATA ANALYSES

Observers were quite accurate overall, each making fewer than 10% errors in Experiments 1 and 3, and fewer than 20% in Experiment 2. However, there were systematic differences in accuracy between conditions (see Figures 1 and 2). Target present trials led to more errors than target absent trials, consistent with previous reports for visual search (Klein & Farrell, 1989; Humphreys, Quinlan, & Riddoch, 1989). More important for present purposes was the observation that errors tended to increase with response time, indicating that observers were not simply trading accuracy for speed.

The response time (RT) data were analyzed the same way for 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 $r = .71$ to 1.00 across conditions and experiments). Second, the estimated slope and intercept parameters were submitted to analyses of variance. T-tests (Fisher's *least-significant difference tests*) were used to examine the significant main effects involving these parameters; the reported *p*-values therefore refer to these tests.

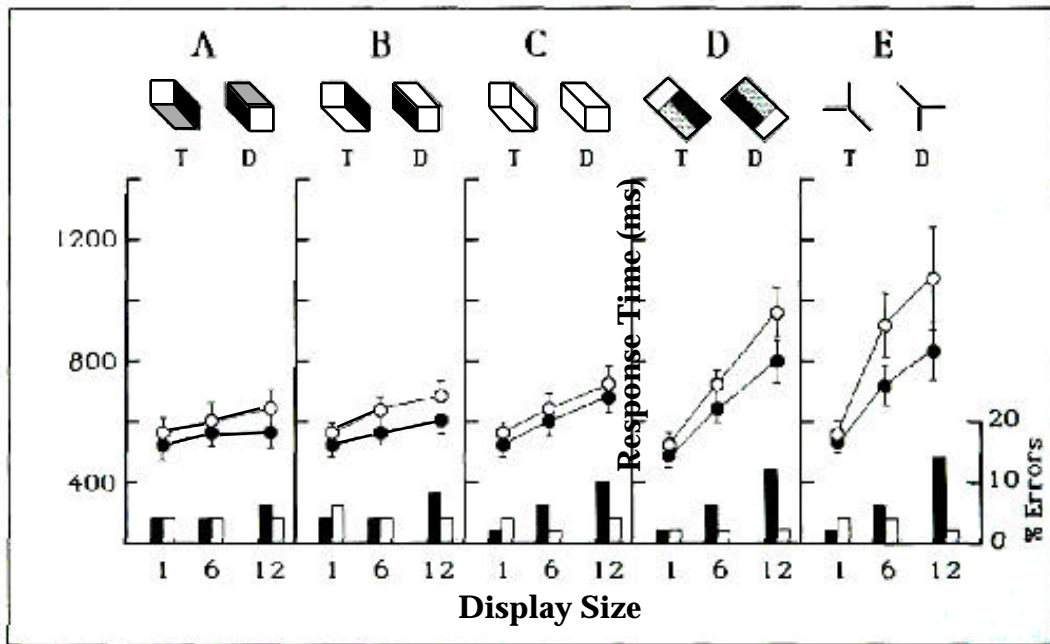


Fig. 1. Experiment 1: The target (T) and distractor (D) items in the five conditions (A to E). Filled circles and bars represent data from target present trials; open circles and bars represent target absent trials. (A) Search is very rapid when shading and line relations correspond to convex blocks differing in three-dimensional orientation. (B and C) search is slowed slightly when shading is removed, but is (D and E) very slow when based on spatial relations that do not correspond to three-dimensional objects.

EXPERIMENT 1

We first examined the relative contributions of shading and line relations to visual search (see Figure 1). The shaded polygons in Condition A could be detected very rapidly, regardless of how many distractor items were in the display (RT slopes of 6 ms per item for target present; 9 ms per item for target absent). Omitting the shading from one of the polygons in Condition B increased the search rate slightly (RT slopes of 9 and 12 ms per item). When all shading was removed in Condition C, a similar small increase occurred (RT slopes of 12 and 14 ms per item). These search rates did not differ reliably at each step, although Conditions A and C were reliably different ($p < .05$). These results show clearly that items corresponding to blocks of distinctive three-dimensional orientations can be rapidly detected.

Search items with the same spatial relations between shaded polygons, but without a three-dimensional interpretation, yielded much slower search in Condition D ($p < .05$; RT slopes of 26 and 29 ms per item). Search was also slowed down when the common outlines of the items were removed in Condition E ($p < .05$; RT slopes of 28 and 47 ms per item). These results show clearly that it is the *system* of spatial relations that is important for rapid search, rather than the spatial relations at any particular point.

EXPERIMENT 2

What kinds of spatial relations allow items to be rapidly detected when they correspond to objects of different three-dimensional orientation? The line junctions in the simple convex blocks of Experiment 1 were either (a) simple L-junctions corresponding to corners of single faces, (b) arrow-junctions corresponding to corners formed from two visible faces, or (c) Y-junctions corresponding to corners formed from three visible faces. To test the necessity of these conditions, Experiment 2 used projections of U-shaped brackets

formed from three rectangular plates (see Figure 2). For these objects, only two surfaces meet at each corner. Consequently, arrow-junctions must correspond to two-plate corners, and T-junctions to local surface occlusions.

The first three conditions in Experiment 2 repeated Conditions A-C in Experiment 1 with brackets in place of the blocks. Conditions D and E examined the influence of simple occlusion, signaled by shading relations and by line relations, respectively (see Figure 2).

For all conditions tested, search was far slower for the brackets than for the blocks ($p < .05$; RT slopes for Condition A: 28 and 42 ms; Condition B: 52 and 78 ms; Condition C: 52 and 86 ms; Condition D: 40 and 58 ms; and Condition E: 59 and 97 ms per item). Furthermore, removal of shading now caused search to slow down considerably (Condition A was reliably faster than Conditions B and C, $p < .05$). These results indicate that this particular system of line relations does not permit rapid search. Indeed, the similar slow search rates in Conditions B, C, and E ($p > .05$) show that search was little affected by the particular relations between lines in these items.

A comparison of Experiments 1 and 2 shows that the preattentive system is very sensitive to the system of line relations present in an item. For example, the items in Condition C of Experiment 1 differed from their counterparts in Experiment 2 only by the placement of a few small lines, yet the difference in search rates was considerable. This sensitivity is difficult to explain in terms of local image operations such as spatial filtering or simple feature detection. Furthermore, if only two-dimensional aspects of the image are relevant, why would there be such sensitivity to spatial relations?

A simpler explanation is that three-dimensional orientation was recovered preattentively for the items in Conditions A-C of Experiment 1, allowing them to be rapidly detected. Conversely, the set of spatial relations present in the other items did not allow this quantity to be recovered efficiently, entailing a slower rate of search.

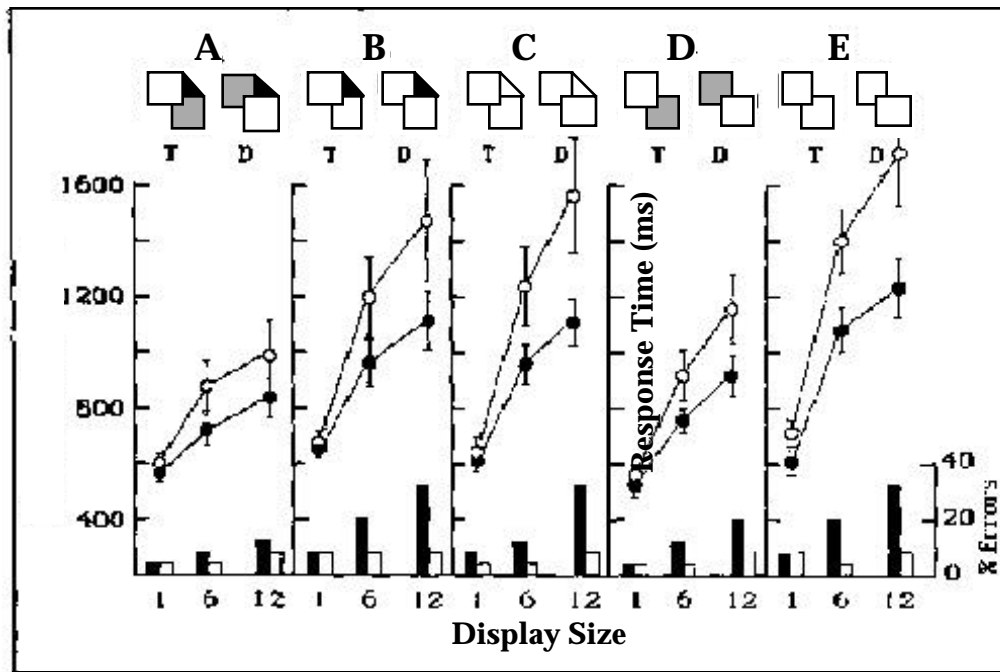


Fig. 2. Experiment 2: Search is slow when three-dimensional orientation is signalled by feature relations corresponding to U-shaped brackets. Shading clearly influences this effortful search task (A versus B and C; D versus E).

EXPERIMENT 3

To test directly the status of three-dimensional orientation as a pre-attentive feature, we next compared the influence of image-based and scene-based orientation on visual search. The target item in each condition corresponded to a block oriented upward and to the left of the line of sight (see Figure 3). Rotating and reflecting the target image generated seven different sets of distractors (Conditions A-G), with each condition having different diagnostic orientations. These conditions tested all possible ways in which the target orientations could differ from the distractor orientations. The distractors within a condition were presented with equal probability in each display and over each set of trials.

Multiple regression models based on the three orientation features were used to predict mean search rates in each condition (averaged over the 10 subjects). These models fit the *present* and *absent* data very well ($r^2 = .90$ and $.88$, respectively). When none of the orientation features were diagnostic, the models predicted search rates of 19 ms per item for target present trials and 74 ms per items for target absent trials. Image-based orientation was estimated to reduce search by 15 and 46 ms per item ($p < .05$), vertical axis orientation by an additional 9 and 42 ms per item ($p < .05$), and horizontal axis orientation by another 4 and 42 ms per item ($p < .05$). The, the two scene-based features together were at least as useful as the image-based one for improving search.

A comparison of Conditions D and G helps to illustrate the influence of orientation at both the image and scene level (see Figure 3). These conditions differ only in that a third distractor is present in Condition G. If search is determined solely by image orientation, this distractor should increase the distinctiveness of the target, and thereby speed up search. In fact search is slowed down considerable ($p < .05$), owing to the increased heterogeneity of scene-based orientation that is contributed by this distractor. Comparisons between Conditions B and F and between C and F are similarly instructive.

GENERAL DISCUSSION

These findings support the hypothesis that scene-based features are represented at preattentive levels (Enns & Rensink, 1990). They are also consistent with previous reports of effortless texture segregation based on shading patterns corresponding to convex/concave surfaces (Ramachandran, 1988). However, the present results contribute two unique findings. The first is that rapid search can be based entirely on the spatial relations between lines—shading is not required. The second is that rapid search can be based on the three-dimensional orientation of simple blocks and not simply on the convexity/concavity of surfaces.

These results have several important implications for the study of early vision. To begin with, they prompt speculation that three-dimensional orientation is signaled by neurons early in the visual pathway. In principle, the slant and tilt of a surface can be determined using the $2G$ filters assumed to exist in the visual cortex (Pentland, 1984), and it may be that three-dimensional orientation is represented at this level. However, the results of Experiment 1 (conditions A vs D and C vs E) show that a complete set of line relations is needed. This finding cannot be explained purely in terms of the spatially localized processing believed to be done by cells at this level. Either there exist striate cells sensitive to these systems of relations, or else visual search must access areas much higher in the cortical hierarchy.

These findings also have clear implications for computational models of early vision. In particular, they show that the parallel processes at this level must be sophisticated enough to extract three-dimensional orientation from line-drawings of convex blocks. Possibly there is early recognition of three-dimensional volumetric primitives such as *geons* (Biederman, 1985). One of the defining features of a geon is its principal axis, and it may be that the three-dimensional orientation of this axis can be used directly as a basis for search. Alternatively, the three-dimensional structure of the object may be recovered via a line-labelling technique applied to the lines of each

	Target	Distractor(s)	Diagnostic Orientation		Search Rate	
			Scene		Image	
			Horiz.	Vert.	Present	Absent
A			+	+	6	7
B				+	+1	0
C			+		1	0
D				+	7	17
E			+		15	47
F					+1	17
G					22	87

Fig. 3. Experiment 3: The target and distractor items, the diagnostic orientations, and corresponding search rates (me per item) in Conditions A to G. Rapid search is possible when either scene-based (A) or image-based (F) orientation is diagnostic of the target. Multiple regression models based on all seven conditions show that scene- and image-based orientation are comparable in their ability to direct search.

item in the image (Mackworth, 1976). It has been shown mathematically that if the corners of an object are formed from surfaces that are mutually orthogonal, the orientation of these surfaces can be recovered from the image. In this view, the failure of the preattentive system to extract these orientations from the U-shaped objects in Experiment 2 would reflect an interesting limitation of its underlying processes. In any event, the results described here show clearly that more sophisticated processing is being carried out in parallel by the human visual system than has generally been assumed.

Acknowledgements—This research was supported by NSERC (J.E.; RR through R.J. Woodham) and UBC CICSAR (R.R.). We thank Eric Ochs for programming assistance and Andrew MacQuistan and Surjit Jagpal for collecting data. We are also grateful to an anonymous reviewer for suggesting condition D in Experiment 1.

REFERENCES

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

Biederman, I. (1985). Human image understanding: Recent results and a theory. *Computer Vision, Graphics, and Image Processing*, **32**, 29-73.

Callaghan, T.C. (1989). Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & Psychophysics*, **46**, 299-311.

Enns, J.T. (1986). Seeing textons in context. *Perception & Psychophysics*, **39**, 143-147.

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

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

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

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

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

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**, 258-279.

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

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

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

Nothdurft, H.C. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, **25**, 195-1968.

Pentland, A.P. (1984). Local shading analysis. *IEEE Transactions: PAMI*, **6**, 170-187.

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

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.

Wolfe, J. M., Franzel, S. L., & Cave, K. R. (1988). Parallel visual search for conjunctions of color and form. *Journal of the Optical Society of America, A*, **4**, 95.

(Received 12/13/89; Accepted 2/12/90)