Surfaces versus features in visual search

Zijiang J. He & Ken Nakayama

Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA

OFTEN implicit in the interpretation of visual search tasks is the assumption that the detection of targets is determined by the feature-coding properties of low-level visual processing[1,2]. But higher level processes have also been implicated as visual search ability is enhanced in a depth plane[3] or when two-dimensional shapes are interpreted as three-dimensional forms[4,5]. Here we manipulate binocular disparity to degrade visual search, so that otherwise identical features become parts of surfaces through perceptual completion, rendering them less clearly distinguishable as targets and distractors. Our results indicate that visual search has little or no access to the processing level of feature extraction but must have as an input a higher level process of surface representation.

Stereoscopic depth can have a profound effect on the interpretation of surface continuation behind occluders[6] and can control perceived transparency[7]. For example, consider what happens to an L-shaped target when binocular disparity is manipulated (see stereograms, Fig. 1). In Fig. 1*a*, the L shape is seen in front, standing alone and appearing to cover the adjacent square. In Fig. 1*b*, the L-shape is seen behind, but no longer appears as an L-shape, instead taking on the appearance of a square continuing behind the other[6]. The strong perceptual continuation of a surface behind an occluder, referred to as 'amodal' completion[8] has been shown to influence a wide variety of seemingly low-level visual functions, in particular motion perception[9,10] and image segmentation[6].

A similar question may be asked regarding the underlying mechanisms of visual search, where an observer must find an odd target among distracting elements. If visual search is dictated by the properties of features, as has been generally assumed, then subtle manipulations that leave these features intact should not influence performance. Alternatively, if visual search has little or no access to visual features, then such manipulations as seen in Fig. 1 should have a great effect. When targets and distractors are perceived as surfaces behind an occluder and thus no longer seen as Ls and reversed Ls, the search for the odd target should become much more difficult.

The various stimulus elements used in our first experiment are illustrated in Fig. 1*a* and *b*. The search stimulus consisted of one target (L-shape or mirror-reversed L-shape) and several distractors (L or reversed L) accompanied by squares. The target and distractor were always in a same depth plane, but appeared to be in a different depth plane relative to the squares. Seven observers (three naive) were used. Observers viewed the stereograms using a phase haploscope device, consisting of shuttered glasses to synchronize transmission in each eye with alternating TV monitor frames. The task was to search for a target among randomly distributed distractors and to release a button when the target was found. Responses in naive observers

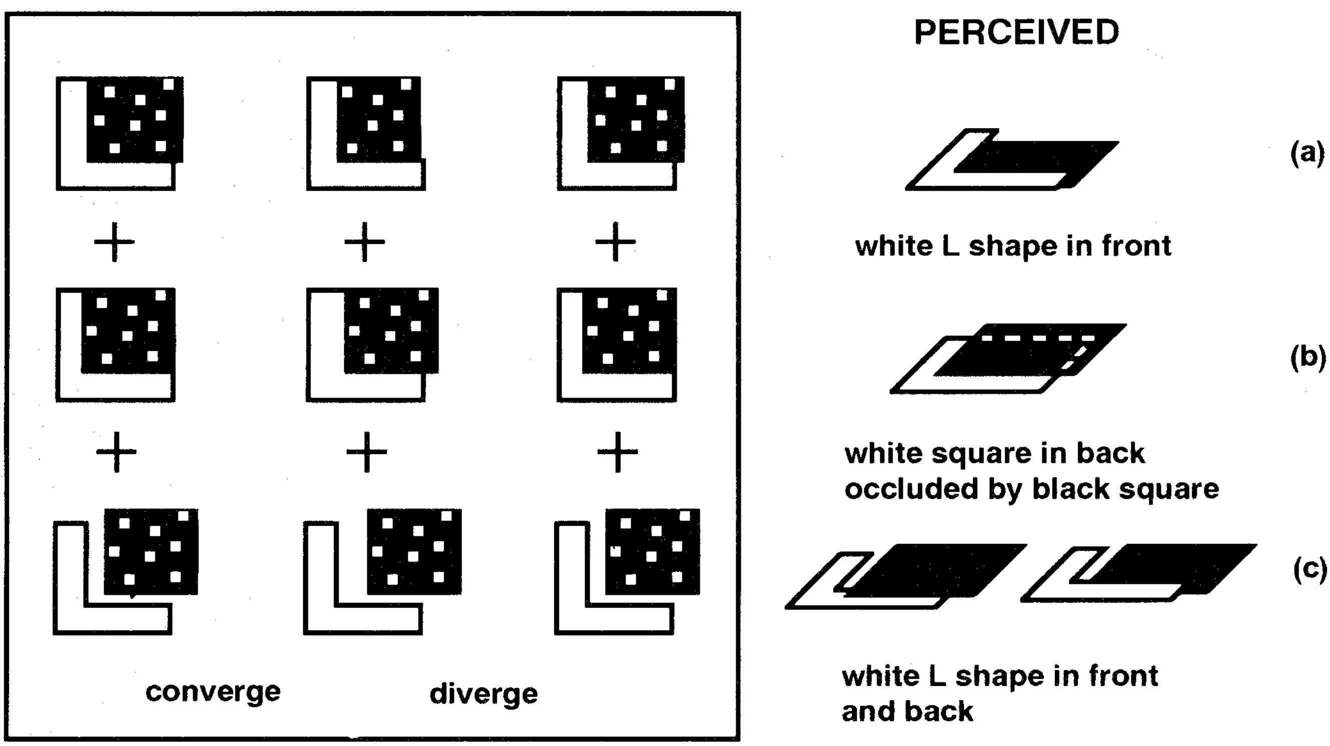

FIG. 1 Three image stereograms (in box) designed for convergent or divergent fusion. For convergent fusion, left and centre images should be used, ignoring the right image. For divergent fusion, use centre and right images. Perceived shapes are illustrated to the right. An L-shape is seen in front for row a, and behind for row b. In row b, the L becomes extended to form a larger surface, 'amodally completing behind the square occluder. Row c, stimuli used in the second experiment, which have the same disparity differences but with a 9.2′ separation betwen L and square. In both experiments, target and distractors were distributed randomly within a rectangular area (13.4°×9.6°). The luminance of the grey background was 1.7 cd m^{-2}; the white L, 4.8 cd m^{-2} and the black squares 0.6 cd m^{-2}. The size of the L was 1°×1°, with the horizontal limb subtending 11.5′. Crosses are provided to aid in free fusion.

TABLE 1 Mean reaction time differences between presentation of L-shape behind and L-shape in front of square (ms)

Observer	Experiment 1*	Experiment 2
B.A.	1,154	−90
P.V.	1,083	63†
Q.W.‡	1,434	79
S.S.‡	752	−35
T.K.‡	1,253	48
T.W.	1,270	189†
Z.H.	2,279	107
Average	1,316 ± 474	52 ± 92

* The reaction time differences of each observer were significantly different (t-test; $P < 0.01$).

† Reaction time differences significantly different (t-test; $P < 0.05$).

‡ Naive observer.

were verified by requiring them to indicate whether the target appeared on the left or right half of the screen. In all experiments, location errors never exceeded 10% and had a mean of about 5%. Reaction times corresponding to incorrect position reports were not counted in final results.

Reaction time histograms for observer P.V. are shown in Fig. 2*a*. There is a significant shift in the reaction time distribution between the presentation of the target behind the square (upper histogram) and in front of the square (lower histogram). The other observers gave very similar results (Table 1, first column). For seven observers, the mean reaction time differences between target behind and target in front ranged from 752 ms to 2,279 ms, with an average of 1,316 ms.

Our results support the view that the increased reaction time when the L-shape is behind is caused by the perceptual or amodal[8] completion of a putative surface behind the square, because perceptual completion is possible only in this case. But it is important to rule out the possibility that detection of targets might simply be more effective if they are in a front disparity plane[11].

To test this possiblity we conducted a second experiment, comparing the detection of L-shapes in front of and behind the square, as in the previous experiments, but where there was no opportunity for perceptual completion behind the square. This was accomplished by creating a small gap between the square and L-shape, such that they did not share a common border (Fig. 1*c*). The manipulation of binocular disparity no longer provides the opportunity for amodal surface completion[10]. In both cases, the L-shapes appear as distinct entities. If our hypothesis regarding the role of surface completion is correct, then reversing binocular disparity should not lead to an appreciable difference in search reaction times for this experiment.

Indeed we found that there was very little difference between the presentation of L-shape in front and behind, when the Ls did not share a common border with the squares. Observer P.V.'s reaction time histograms for this experiment are shown in Fig. 2*b*. Table 1 (second column) shows the difference between mean reaction times (behind − in front) for all seven observers. The mean difference for the group is 52 ± 92 ms, which is significantly smaller than reaction time difference without separation (1,316 ± 474 ms). This control experiment strongly supports our view that the difficulty arising from the L-shape-behind case is caused by surface completion behind the square, rendering target and distractor much less distinguishable at a surface representation level.

In these experiments, one concern is whether the stimuli used are appropriate to activate simple feature mechanisms differentially. There is no neurophysiological data to indicate the presence of 'filters' or units selectively tuned for L- and reversed L-shapes. However, simple orientation mechanisms would have different outputs for each pattern, which are likely to be sufficient for activation of different populations of orientation units. For example, Ls and Xs in many search tasks are easily distinguishable and a number of studies have indicated that a combination of low-level visual processing mechanisms could easily distinguish between these simple shapes[12-14]. This assumption is supported by our third experiment, where we used a behavioural criterion which has been used to define a simple feature search, namely the existence of unvarying reaction times as distractor number is increased.

We used the same elements as the first experiment (Fig. 1*a*, *b*), varying the distractor number to see how this influenced reaction times in the two depth conditions. The relation between mean reaction times and number of distractors for two observers is shown in Fig. 3. There is a distinct qualitative difference in performance between the different depth conditions. When the L-shape is behind, reaction times increase dramatically as distractor number is increased. When the L-shape was seen in front, the slope is essentially flat with little or no increase in reaction time with increasing distractor number. Such flat functions are believed to indicate that search depends on discerning differences based on simple feature representations[15].

Our results indicate that binocular disparity plays a large part in the slowing down of visual search but only in cases where it facilitates surface completion behind an adjoining occluder. When completion occurs, the target and distractor become perceptually more alike, in this case each appearing as a square, and thus become harder to distinguish than an L and a mirror-reversed L. This indicates that visual search is applied at a much higher level of visual representation than feature detection, most probably at the level of perceived shapes or surfaces. The result, that subtle changes which did not alter the features themselves had a profound effect in making visual search much more difficult, supports our original supposition that features cannot be accessed directly. The results also provide a possible explanation of the recent experiments showing that different polyhedral[4] and shaded spherical forms[5] are much easier to find in visual search experiments. Instead of assuming that such forms are

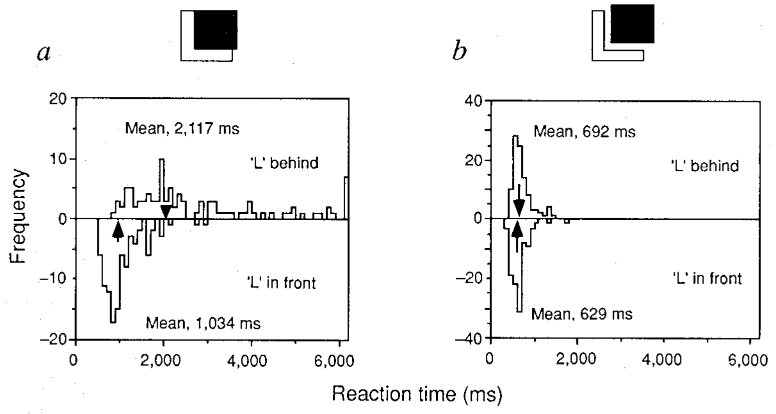

FIG. 2 Reaction time histograms (100 trials) of observer P.V. in the first experiment (*a*) where amodal completion of the L behind the square was possible and the second experiment (*b*) where it was not. The histograms above and below the horizontal axis represent reaction times obtained when Ls were behind and in front of the squares, respectively (disparity, 12.4′). Arrows indicate the median reaction times. In both experiments, 20 distractors were used.

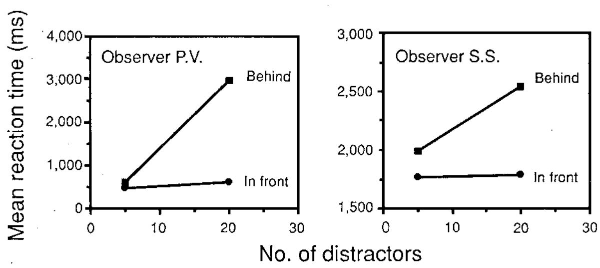

FIG. 3 Relation between mean reaction time (100 trials per point) and the number of distractors, with Ls placed either in front or behind.

simple features, we think it more appropriate to regard them as perceived three-dimensional shapes, distinguished at an object or surface level.

We attach particular significance to the fact that visual search is impaired rather than improved by changing a surface representation. It could be argued that the improvement in performance with particular three-dimensional shapes[4,5] is not surprising, because information more meaningfully configured for higher-level vision is also available to be searched. Here we draw an additional conclusion, that the feature-detection level of representation cannot be directly accessed by visual search mechanisms, and that in order to perform a visual search task, the system must operate at a higher level. □

Received 13 April; accepted 27 July 1992.

1. Julesz, B. *Nature* **290,** 91–97 (1981).
2. Treisman, A. & Gelade, G. *Cognitive Psychol.* **12,** 97–136 (1980).
3. Nakayama, K. & Silverman, G. H. *Nature* **320,** 264–265 (1986).
4. Enns, J. T. & Rensink, R. *Psychol. Sci.* **1,** 323–326 (1990).
5. Kleffner, D. & Ramachandran, V. *Percept. Psychophys.* **52,** 18–36 (1992).
6. Nakayama, K., Shimojo, S. & Silverman, G. H. *Perception* **18,** 55–68 (1989).
7. Nakayama, K. & Shimojo, S. *Cold Spring Harb. Symp. quant. Biol.* **40,** 911-924 (1990).
8. Kanisza, G. *Organization in Vision, Essays in Gestalt Perception* (Prager, New York, 1979).
9. Shimojo, S., Silverman, G. H. & Nakayama, K. *Vision Res.* **29,** 619–626 (1989).
10. Shimojo, S. & Nakayama, K. *Perception* **19,** 285–299 (1990).
11. Fox, R. & Patterson, R. *Percept. Psychophys.* **30,** 513–520 (1981).
12. Beck, J. *Percept. Psychophys.* **1,** 300–302 (1966).
13. Bergen, J. R. & Julesz, B. *Nature* **303,** 696–698 (1983).
14. Malik, J. & Perona, P. *J. opt. Soc. Am.* **A7,** 923–932 (1990).
15. Treisman, A. & Souther, J. *J. exp. Psychol: Gen.* **114,** 285–310 (1985).

ACKNOWLEDGEMENTS. This work was supported in part by a grant from Air Force Office of Scientific Research.