1993-01

A Neural Theory of Visual Search: Recursive Attention to Segmentations and Surfaces

A NEURAL THEORY OF VISUAL SEARCH: RECURSIVE ATTENTION TO SEGMENTATIONS AND SURFACES

Stephen Grossberg, Ennio Mingolla, and William D. Ross

January 1993

Technical Report CAS/CNS–93–023

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

A Neural Theory of Visual Search: Recursive Attention to Segmentations and Surfaces

Stephen Grossberg,[1], Ennio Mingolla[2] and William D. Ross[3]
Center for Adaptive Systems and Department of Cognitive and Neural Systems
Boston University, 111 Cummington Street, Boston, Massachusetts 02215 USA

Abstract

A neural theory is proposed in which visual search is accomplished by perceptual grouping and segregation, which occurs simultaneous across the visual field, and object recognition, which is restricted to a selected region of the field. The theory offers an alternative hypothesis to recently developed variations on Feature Integration Theory (Treisman, and Sato, 1991) and Guided Search Model (Wolfe, Cave, and Franzel, 1989). A neural architecture and search algorithm is specified that quantitatively explains a wide range of psychophysical search data (Wolfe, Cave, and Franzel, 1989; Cohen, and Ivry, 1991; Mordkoff, Yantis, and Egeth, 1990; Treisman, and Sato, 1991).

Introduction

Recently, a psychophysical paradigm has been used to discover the mechanisms of visual search (Treisman, and Gelade, 1980; Nakayama, and Silverman, 1986; Pashler 1987; Wolfe, Cave, and Franzel, 1989). In this paradigm a target item is pre-defined, either semantically or by visual exposure, and the observer is required to determine whether it is present in a scene that includes distractor items. Response times for scenes with various numbers of items can then be compared to determine whether search is accomplished by parallel processing of the entire visual field or serial investigation of each item (Figure 1).

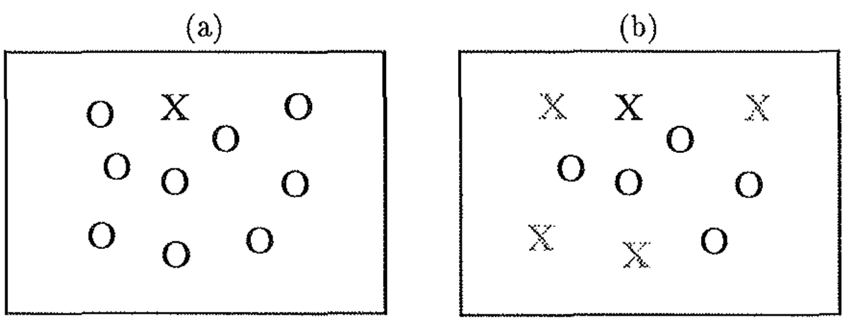


Figure 1: *Search scenes. (a) The target (black X) is distinguishable by a single feature (form). Response time is fast and doesn't vary with the number of distractor items. (b) The target is only distinguishable by the conjunction of two features (form and color).*

[1]Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175), DARPA (AFOSR 90-0083, and ONR N00014-92-J-4015) and the Office of Naval Research(ONR N00014-91-J-4100).

[2]Supported in part by British Petroleum (89-A-1204), DARPA (AFOSR 90-0083, and ONR N00014-92-J-4015), the National Science Foundation (NSF IRI 90-00530) and the Office of Naval Research (ONR N00014-91-J-4100),

[3]Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175), British Petroleum (89-A-1204), the National Science Foundation (NSF IRI 90-00530), and the Office of Naval Research (ONR N00014-91-J-4100)

The alternatives of parallel processing and serial search are not, however, exhaustive. While target and distractor items can be arbitrarily defined in the construction of test scenes, it does not follow that they will each immediately give rise to distinct candidates for further analysis. Accurate and complete segmentation of a visual scene containing multiple objects is arguably more difficult than rapid visual search for a single target.

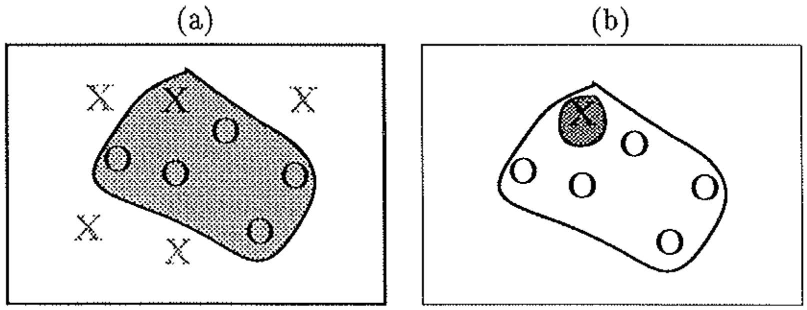


Figure 2: *Illustration of search by recursive segmentation. (a) Initially a region of target color which includes multiple items is separated as a candidate. (b) Next this region is searched for target form. Thus, fast search times can be achieved without parallel processing of feature conjunctions.*

For targets defined by a conjunction of color and form (Figure 1b), search for a conjunctive target could proceed as a two-step recursive process (Figure 2). Initially, a multi-item region defined by a single target feature, such as color, would be separated in parallel from the rest of the scene. Next, spatial registration of the other target feature within that multi-item candidate region would guide target search. This two-step recursive process would yield fast search times that within the "item= object" paradigm could be misinterpreted as evidence for simultaneous or parallel processing of feature conjunctions. Such a recursive search may be based upon the visual representation generated within a preattentive theory of visual perception (Grossberg, Mingolla, and Todorovic, 1989; Grossberg, 1987; Grossberg, 1992) in which a Boundary Contour System (BCS) generates emergent boundary segmentations, and a Feature Contour System (FCS) fills-in surface properties of brightness, color and depth. These surface representations are generated on separable slabs that segregate combinations of color and depth from one another. The BCS and FCS preattentive visual representations reciprocally interact with an attentive Object Recognition System (ORS), which is modeled as an Adaptive Resonance Theory (ART) network (Carpenter, and Grossberg, 1988). The search process summarized below sheds new light on how these reciprocal interactions may be organized.

A Recursive Search Algorithm

Our model of visual search can be summarized as a four step process (see Figure. 3). In step 1, preattentive processing of the visual scene results in retinotopic registration of stimulus features. In step 2, these retinotopic featural arrays support boundary segregation and surface grouping, which segment the scene into separate candidate regions, say via an FCS color-depth slab. This step has been assumed by others to immediately and correctly define scenic objects, which in test scenes are target or distractor items. In step 3, a candidate region is selected for further analysis. This step could be influenced by top-down bias by the ORS of target features in a directed search. Finally in step 4, feature groupings within the candidate region must be compared to the stored target representation. A mismatch between all these feature groupings and the stored

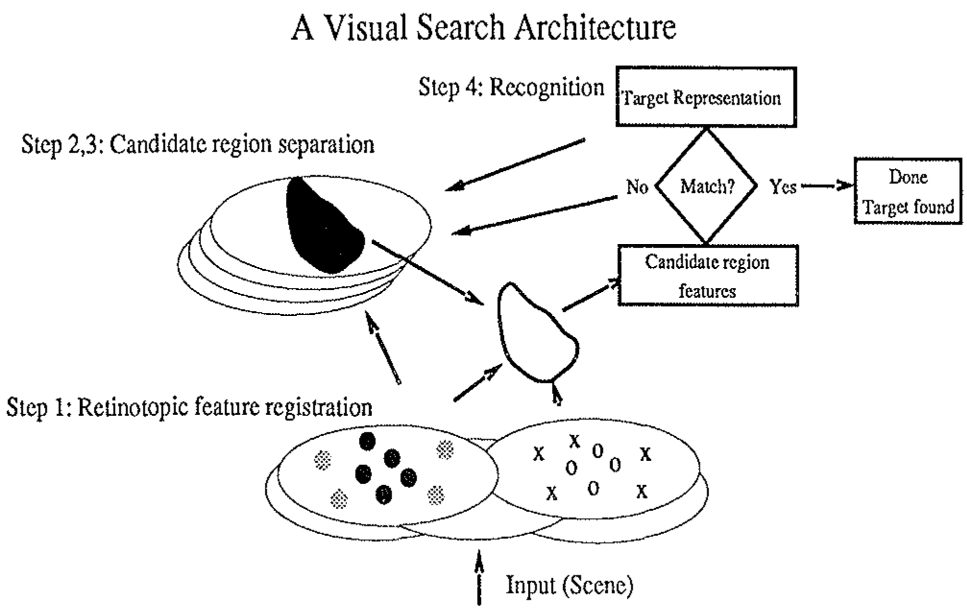


Figure 3: *A visual search architecture. The scene is analysed in parallel through the retinotopic registration of stimulus features and surface properties. Next candidate target regions are serially separated for recognition by the interaction of top-down priming and bottom-up segregation and grouping processes.*

target representation causes a return to step 3 for reset of the old region and selection of a new candidate region. A partial mismatch between the features in a multi-item candidate region and the stored target features may trigger a more vigilant search within the candidate region. This would cause a return to step 2 in order to segment the candidate region on the basis of a new featural dimension. If this recursive process does not yield a target match, the entire candidate region could then be reset. Search self-terminates when a match is found.

There may be no sharp distinction between segmentaion and selection. The separation of candidate targets may in fact combine act scene segmentation and selection. Segmentation of the entire scene would emerge gradually over during such a search. In computer simulations, segmentation and selection are lumped into a constant duration, which is added to search time for each candidate target.

Our simulations instantiate three segmentation principles; (1) Segmentation occurs on the basis of a single featural dimension at a time. (2) Featurally similar items can be grouped into the same candidate region if they can be connected by uninterrupted spatial paths whose width corresponds roughly to the diameter of items. (3) The probability that item groupings become candidate targets is a function of stimulus saliency. A fourth principle governs candidate object selection: (4) Top-down priming restricts target search to candidate regions that match the target in the featural dimension by which the scene is segmented.

The prediction of the mean search or response time (RT) for scenes of a given number of items requires assigning a duration for each of the four steps as well as the algorithmic computation of scene segmentation and search based on the four principles stated above. The average result of the algorithm for "target-present" conjunctive scenes can be approximated by:

$$RT = K + N \times (S + M), \tag{1}$$

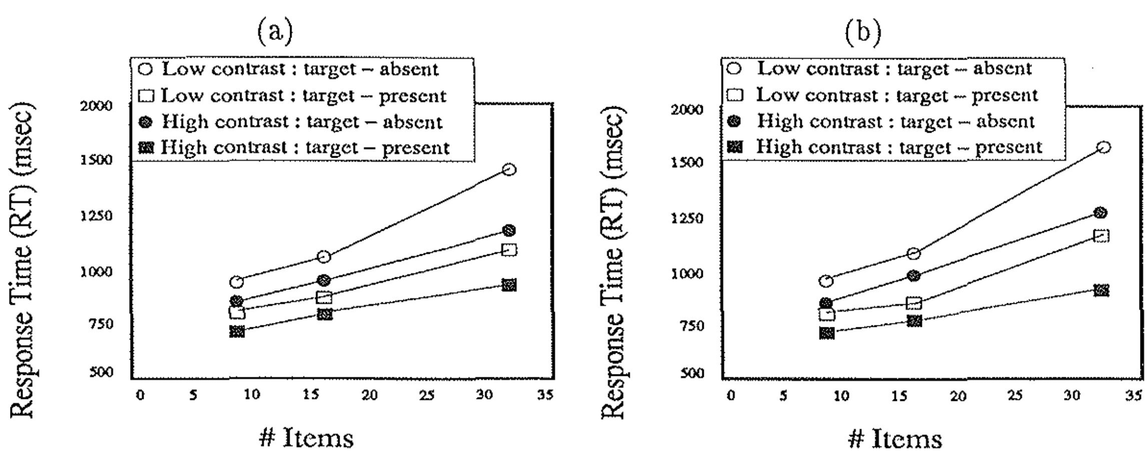
response time for target absent scenes is given by:

$$RT = K + N \times 2 \times (S + M), \tag{2}$$

where, K is the duration necessary to complete step 1: retinotopic feature registration, S is the duration necessary to seperate a candidate region by steps 2, and 3: segmentation and selection, M is the duration necessary to compare the candidate region and the target representation in step 4, and N is the average number of candidate regions into which the scenes containing a certain number of items are initally segmented. Each multi-item candidate region must be recursively searched resulting in a recursion factor of 2 in each equation for conjunctive scenes. In equation 1 this factor is cancelled, since in target-present scenes, on average only half of the candidate regions (N/2) have to be evaluated before the target is found.

Simulation of psychophysical data

A program which creates psychophysical scenes was written. The program generates scenes with various numbers, sizes, types and featural saliency of items randomly distributed across predefined potential locations. It converts the experimental specifications listed in the methods sections of psychophysical reports into model inputs. The algorithm was also programmed in order to compute RT on the simulated scenes. This program consists of procedures for scene segmentation, candidate selection and candidate evaluation. All simulations were run with a single choice of step durations and segmentation parameters. Response time (RT) curves were generated by averaging the result of algorithm results over 20 trials for each scene size.



Experiment
Redrawn from
Wolfe *et al.* (1989, Expt. 7)

Simulation

Figure 4: *Increasing stimulus saliency yields faster conjunctive search.*

Recursive conjunctive search

Treisman, and Gelade (1980) conducted tachistoscope studies indicating serial conjunctive search. Wolfe, Cave, and Franzel (1989) repeated the experiments using high-contrast scenes on a Cathode Ray Tube (CRT) and found fast conjunctive searches that could be explained by a parallel search process. Both these results are simulated by our algorithm. The scenes used to test the algorithm were generated using experimental description given in experiment 7 of Wolfe *et al.* (1989). In order to fit the curves, only 30% of the spatially allowable multi-item candidate objects were allowed to form. In the high contrast case, they were all allowed to form. Figure 4 shows the close match between simulation and experimental data.

Clumped vs. Spread-out conjunctive search

Cohen and Ivry (1991, Experiment 3) found that in scenes in which conjunctive items are clumped together, search apparently proceeds by serial processing, while in spread-out scenes search is much faster. This result is also explained by the algorithm, in which multi-item candidates are only formed by items that can be connected by uninterrupted paths of item width. Figure 5 shows the close match between simulation and experimental data.

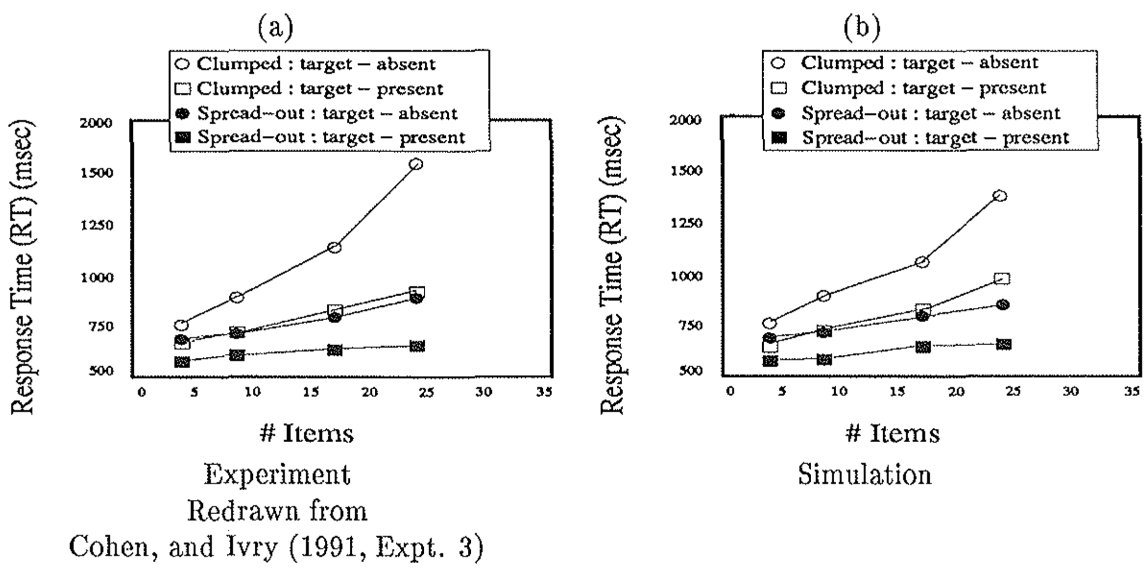


Figure 5: *Spreading out items yields faster conjunctive search.*

Additional experimental findings

Data on fast response to six-item scenes with two targets (Mordkoff, Yantis, and Egeth, 1990), the additivity of feature effects (Treisman, and Sato, 1991), and some findings on color-color cojunctive search are also quantitatively explained by the algorithm. The theory agrees with data indicating that item grouping processes influence search (Humphreys, Quinlan, and Riddoch, 1989), and recent findings suggesting that visual search mechanisms cannot directly access feature representations but instead must operate at a higher level (Zijiang, and Nakayama, 1992).

Referrences

Carpenter, and Grossberg, (1988) The ART of adaptive pattern recognition by a sel-organizing neural network. *Computer*, **21**, 77-88.

Cohen, and Ivry, (1991) Density effects in conjunctive search: Evidence for a coarse location mechanism of feature integration. *Journal Experimental Psychology.*, **17**(4), 891-901.

Grossberg, (1987) Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory. *Perception and Psychophysics*, **41**, 117-158.

Grossberg, (1992) 3-D Vision and Figure-Ground Separation by Visual Cortex, Tech. Rept. CAS/CNS-TR-92-019, Boston University.

Humphreys, Quinlan, and Riddoch, (1989) Grouping processes in visual search: Effects with single and combined feature targets. *Journal Experimental Psychology: General*, **118**, 258-279.

Grossberg, Mingolla, and Todorovic (1989) A neural network architecture for preattentive vision. *IEEE Transactions on Biomedical Engineering*, **36**, 65-84.

Mordkoff, Yantis, and Egeth (1990) Detecting conjunctions of color and form in parallel. *Perception and Psychophysics*, **5**, 157-168.

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

Pashler, (1987) Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics*, **41**, 191-201.

Treisman, and Gelade, (1980) A feature-integration theory of attention. *Cognitive Psychology*, **16**(3), 97-136.

Treisman, and Sato, (1991) Conjunction Search Revisited. *Journal Experimental Psychology*, **16**(3), 459-478.

Wolfe, Cave, and Franzel, (1989) Guided Search: An alternative to the feature integration model of visual search. *Journal Experimental Psychology*,**15**(3), 419-433.

Zijiang, and Nakayama, (1992) Surface features in visual search. *Nature*, **359**, 231-233.