Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 45 (2005) 2987–2992

Vision Research

www.elsevier.com/locate/visres

Serial visual search from a parallel model ☆

Seth A. Herd *, Randall C. O'Reilly

Department of Psychology, University of Colorado Boulder, 345 UCB, Boulder, CO 80309, USA

Received 26 September 2003; received in revised form 19 July 2005

Abstract

We tested a parallel neural network model of visual search, and found that it located targets more quickly when allowed to take several fast guesses. We suggest that this serially iterated parallel search may be the mode used by the visual system, in accord with theories such as the Guided Search model. Furthermore, in our model the most efficient mode of processing varied with the type of search. If the nature of visual search varies with task demands, seemingly contradictory findings can be reconciled.

Keywords: Neural network; Attention; Visual search; Parallel

1. Introduction

There is a longstanding debate regarding the nature of visual search: do we look for an object by moving attention across a scene serially, one object at a time, or by processing everything in that scene at once, in parallel? (see Wolfe, 1998 for a review). In a standard visual search (VS) experiment, the participant is asked to search a display for a particular target object among many distractors, and quickly decide whether the target is present. For search tasks such as finding a single red object among many green ones ("feature search"), the search time does not depend on the number of distractors. For most searches, however, extra distractors slow search by an amount roughly proportional to their number.

This led to a Feature Integration Theory (FIT), which states that people shift attention serially from one object to the next, deciding for each whether it is the target (Treisman & Gelade, 1980). This process was said to be necessary when conjunctions of object features (color, shape, size, orientation, etc.) differentiate targets from distractors, e.g., searching for a red X among green X's and red O's (conjunction search).

However, these results could also be the result of inefficient parallel search processes. Theories of this type are supported by a variety of evidence (Chelazzi, 1999; Duncan & Humphreys, 1989). Deco and Zihl (2001) presented a simple parallel model that reproduced the finding of feature search times independent of number of objects in the search display, and conjunction searches times linearly dependent on number of objects. That model embodied a theory with no serial aspects.

We constructed and further explored a computational model of this type, and discovered a relevant and probably general feature of its behavior: it worked faster if allowed to operate in a partly serial manner. We therefore offer a reinterpretation of this class of model in which it supports the Guided Search model of Wolfe and colleagues. Our interpretation supports the idea that visual search is often partly serial—a parallel process may guide attentional fixations, so that easy "pop-out" searches require only one fixation, very difficult searches may require individual inspection of each item, while intermediate difficulty searches like standard conjunction searches require only a few fixations on

☆ This work was supported by ONR Grants N00014-00-1-0246 and N00014-03-1-0428.

* Corresponding author.
E-mail addresses: sethherd@psych.colorado.edu (S.A. Herd), oreilly@psych.colorado.edu (R.C. O'Reilly).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.07.017

average. This work suggests that the degree to which search is serial varies across both task conditions, and with individual strategies.

2. Methods

The core of our model is similar to that of Deco and Zihl (2001) in structure and basic function (Fig. 1), but we interpret its performance quite differently (see Sections 3 and 4). It includes a retinotopic feature layer, in which each unit represents a specific feature in a specific location, and location layer that represents any features at a given location. These functions match those known to exist in early ventral visual stream areas, and late dorsal stream areas, respectively. In addition, a template layer holds on line the features of the target. This function is probably performed by prefrontal areas.

As a first step, we replicated the modeling results of Deco and Zihl (2001) using a different modeling framework. We used the Leabra modeling framework, previously used to model a wide range of psychological phenomena (O'Reilly, 1998; O'Reilly & Munakata, 2000). The Leabra framework is designed to mimic principles of cortical processing. Units are based on the dynamics of single pyramidal neurons, and use the point neuron approximation (including ion currents and membrane potential).

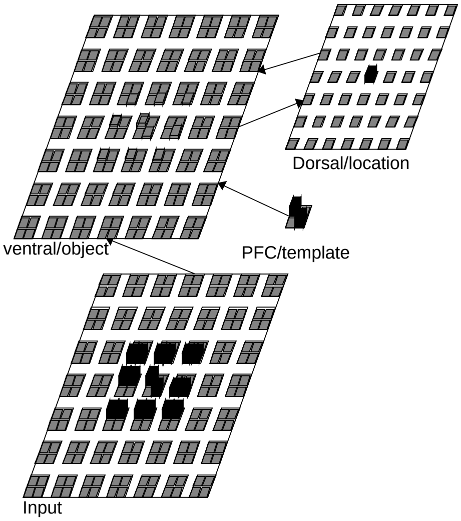

Fig. 1. Input layer is externally set to represent a nine-object conjunction search with the target in the center. In the input and object layers, two units of the four-unit group in each location represent different colors, while the other two represent different shapes. The four units in the PFC/template layer share this representation. The connection from input to object layer is one-to-one, with uniform weights. All four units at each location in the ventral/object layer project to the one unit in the corresponding location in the dorsal/location layer, and these connections are reciprocal. Each of the four units in the PFC/template layer connects to the one matching unit in every location in the object/ventral layer. The response criteria is the activation of any location unit above a threshold of .5; we interpret this response as completing the focus of spatial attention upon a certain location.

The principles of the model's function can be understood in terms of spreading activation. Each trial begins with an input pattern clamped onto the input layer, and a template pattern clamped onto the PFC/template layer. Activation then spreads from these units to those they are connected to in the ventral/object layer. Those units receiving activation from both the input and template will quickly become more active.

This activity in turn spreads to the location layer, and when one location unit reaches an activity of .5, the trial is terminated. We interpret this as a commitment of spatial attention to that location. This happens only when units at one ventral/object location have become more active than those at any other competing location. The winning location is most likely to be the location containing the target, although this likelihood varies with how quickly the model is allowed to settle, as explained in the results section.

In more depth, the Leabra framework functions as follows. The membrane potential V_m is updated as a function of ionic conductances, g, with reversal (driving) potentials, E, as follows:

$$\frac{\mathrm{d}V_\mathrm{m}(t)}{\mathrm{d}t} = \tau \sum_c g_c(t)\overline{g_c}(E_c - V_\mathrm{m}(t)). \tag{1}$$

There are three channels (c): e is the excitatory input; l the leak current; and i is the inhibitory input. The overall conductance is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the network, and a constant $\overline{g_c}$ that controls the relative influence of the different conductances.

The excitatory net input/conductance $g_\mathrm{e}(t)$ or η_j is computed as the proportion of open excitatory channels as a function of sending activations times the weight values:

$$\eta_j = g_\mathrm{e}(t) = \langle x_i w_{ij} \rangle = \frac{1}{n}\sum_i x_i w_{ij}. \tag{2}$$

The inhibitory conductance is computed via the kWTA function described in the next section, and leak is a constant.

Activation communicated to other cells (y_j) is a thresholded (Θ) sigmoidal function of the membrane potential with gain parameter γ:

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma[V_\mathrm{m}(t) - \Theta]_+}\right)}, \tag{3}$$

where $[x]_+$ is a threshold function that returns 0 if $x < 0$ and x if $X > 0$. Note that if it returns 0, we assume $y_j(t) = 0$, to avoid dividing by 0. To produce a less discontinuous deterministic function with a softer threshold, the function is convolved with a Gaussian noise kernel.

2.1. k-Winners-take-all inhibition

Leabra uses a kWTA function to achieve sparse distributed representations, with two different versions having different levels of flexibility around the k out of n active units constraint. Both versions compute a uniform level of inhibitory current for all units in the layer as follows:

$$g_i = g^{\Theta}_{k+1} + q(g^{\Theta}_k - g^{\Theta}_{k+1}), \tag{4}$$

where $0 < q < 1$ is a parameter for setting the inhibition between the upper bound of g^{Θ}_k and the lower bound of g^{Θ}_{k+1}. These boundary inhibition values are computed as a function of the level of inhibition necessary to keep a unit right at threshold:

$$g^{\Theta}_i = \frac{g^*_{\mathrm{e}}\bar{g}_{\mathrm{e}}(E_{\mathrm{e}} - \Theta) + g_l\bar{g}_l(E_l - \Theta)}{\Theta - E_i}, \tag{5}$$

where g^*_{e} is the excitatory net input.

In the *average-based* kWTA version (used for this model), g^{Θ}_k is the average g^{Θ}_i value for the top k most excited units, and g^{Θ}_{k+1} is the average of g^{Θ}_i for the remaining $n-k$ units. This version allows for more flexibility in the actual number of units active depending on the nature of the activation distribution in the layer and the value of the q parameter (which is typically between .5 and .7 depending on the level of sparseness in the layer, with a standard default value of .6). Activation dynamics similar to those produced by the kWTA function have been shown to result from simulated inhibitory interneurons that project both feedforward and feedback inhibition (O'Reilly & Munakata, 2000).

3. Results

Our model initially produced results quantitatively similar to those of the previous model, and to behavioral results. We obtained nearly flat search slopes in the feature search condition, and a linear increase in time to settle with additional distractors in conjunction search (Fig. 2).

This linear increase was driven by noise: the search cost per distractor varied with the amount of gaussian noise applied to the net input current on each time step (Fig. 2). According to this type of model, varying behavioral search slopes result from a larger signal/noise ratio for more easily discriminated stimuli.

The model's performance stems from the fact that only feature units that enjoy both bottom up (input) and top down (target template) inputs become active enough to influence the competition among location units. There is only one such unit in the feature search condition, the target feature in the target location. In the conjunction search condition, one target feature is present at each location, but both target features are present at the target location, allowing that location to dominate if enough evidence is accumulated to minimize the effects of noise.

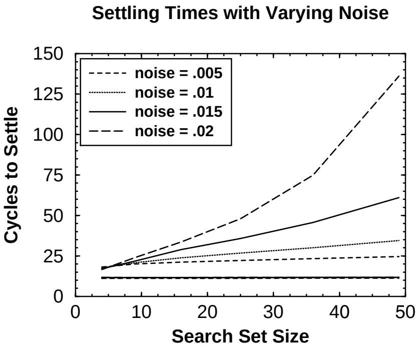

Fig. 2. Settling times for our model. Sloped lines are conjunction search; lower flat lines are feature searches. The amount of noise affects the settling slope for conjunction search, but does not affect the feature search settling time. We assume that human reaction times are proportional to these settling times, plus constant times for motor responses and object identification.

Units needed two sources of input to become active because the leabra algorithm uses a thresholded activation function Eq. (3). Without this threshold, we would expect to see a contribution from inputs with no support from the PFC/template layer, and therefore a search cost even in the feature search condition, as is often observed experimentally. However, this cost would be very small, since only very large contributions from noise could overcome the lack of top-down support.

This model can be understood as a diffusion process model in which information is accumulated over time in a noisy environment, with more noise present for each distractor that shares a target feature. It is thus possible to speed the settling process at the cost of accuracy. Many variables could affect the system in this way. We chose to vary the starting value of the membrane potential. This has the effect of placing the system closer to settling, so that less evidence is needed to produce an attentional fixation. It also seems that this is a likely variable for online adjustment by the cognitive system; providing extra diffuse input before a trial will provide a baseline activity level, and put the system closer to its response threshold.

Raising the system's baseline activity level produced a dramatic speedup of settling, at the cost of an equally dramatic reduction in accuracy (Fig. 3).

Is this reduction in accuracy disastrous for the performance of the system? It is if we assume that every missed location is a missed trial; behavioral performance usually shows a less than 10% error rate. However, if we instead assume that the system checks the accuracy of its response with an object identification process, then

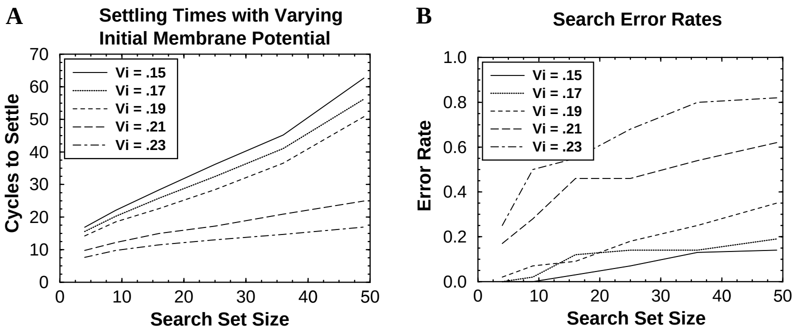

Fig. 3. (A) Location process times for varying starting states. Locating a potential target is dramatically speeded by larger starting membrane potentials, corresponding to a lowered threshold. (B) Error rates rise rapidly as the location process becomes faster.

chooses a new location if that object does not match the target template, then risking wrong location guesses could be a good strategy.

For simplicity we assume, rather than explicitly model, this object identification process. This identification may happen by virtue of the dorsal visual stream providing extra activation to that location in early ventral stream areas, so that higher areas respond predominantly to that information versus information from surrounding distractors. This account is in general accord with the biased competition model of Desimone and Duncan (1995), but our model does not depend on these details. We assume only that this process takes some amount of time to identify the object at the location selected by the model, gives a response if the object is the target, and triggers a new iteration of the whole process if the object is not the target.

If every missed location process results in a repeat of that process, the total search time will be given by (location time + identification time)/(1 − $P(error)$), since the series $1 + x + x^2 + x^3 + x^4 \cdots$ converges to $1/(1 - x)$ for $x < 1$. That series corresponds to the total number of location processes that will be completed on average when $x = P(error)$, or alternately, one plus the average number of errors per trial.

The speedup of search proved so dramatic that the system can afford one or even more missed attentional fixations, depending on assumptions about how long the identification process takes, and the signal strength and noise level. Fig. 4A gives the total search times under the assumption that an identification process takes 10 extra cycles. Even though that process is fairly costly, it can be seen that less conservative location processes are competitive with those that locate the target on the first try.

This assumption is probably still too conservative; it seems unlikely that no information is retained from the location process after the first settling process. If we assume that later location processes take 1/2 the time of the first, due to retained information, search efficiency is biased even further toward processes that make some mistakes in the interest of a faster location process (Fig. 4B). In this case, an intermediate parameter setting is the most efficient over the whole range of display sizes, while the most efficient search parameters vary with changing display size. Of course each missed location

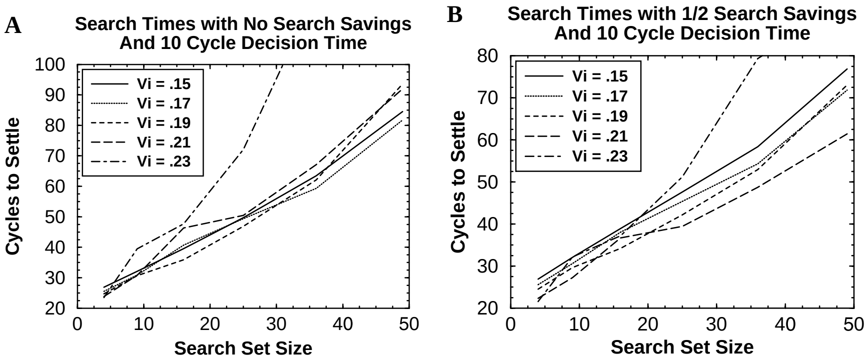

Fig. 4. (A) Total conjunction search times under the consideration that there is an object identification process that takes the equivalent of 10 processing cycles of the model, about the same amount of time the system takes to settle on the location of the target in a feature search. (B) Total conjunction search times under the same considerations for identification times, but with the more reasonable assumption that some location information is retained so that additional location processes take 1/2 the time of the first. Note that which line is lowest, and therefore the most efficient search parameters, changes between these two sets of assumptions.

process results in a wasted object identification process, so if object identification is very slow relative to the location process, a conservative (and therefore parallel) process will be most efficient.

4. Discussion

We used a model in which target localization is parallel and capacity-unlimited. We replicated earlier work indicating that such a process can produce the linearly increasing reaction times with set sizes. We went on to test the speed/accuracy tradeoff within the model, and discovered that the model gained so much in speed that in some situations it was faster to obtain a correct answer by running it several times at low accuracy rather than once with high accuracy. This finding suggests that human visual search may be performed serially by default because it is faster than performing search in parallel.

This conception of search processes, based on an entirely parallel target location process, has converged with the Guided Search model (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989), in which a serial search is guided by a parallel "saliency map" operation. If we assumed that the process retained all of its information instead of enough to cut settling time in half, as in Fig. 4B, we would have exactly reproduced Wolfe's guided search model. We do not make this assumption because Wolfe's own work has shown that location information retention is not nearly perfect, (e.g., Horowitz & Wolfe, 2001).

Our model therefore differs from Wolfe's in assuming that the time consuming parallel process must be run again for each unsuccessful attentional fixation (although some information from the previous parallel process may be retained). This follows from the following train of logic: observers generally prefer eye movements in standard conjunction tasks (Shen, Reingold, & Pomplun, 2003); eye movements massively disrupt representations in the early ventral stream areas; and those areas are widely identified with the feature maps that guide search (reviewed in Shipp, 2004). The implication is that the time taken by versus the accuracy of the parallel stage becomes an important tradeoff under parametric control of the observer. Thus, we predict different search patterns for different strategies on the same search task, as well as among different search tasks as predicted by Guided Search.

Like the Guided Search model, our model does not specify the conditions under which search is terminated. An effective strategy should assume that no target is present after a number of unsuccessful guesses, or after a conservative settling process does not settle in a given time. The criteria for a "no" response will vary with the internal parameters (strategy) used for the search, and the physical parameters of the search. Therefore, we have dealt only with target present responses, leaving this issue to be addressed by future work.

Although we have wound up in nearly the same theoretical position as Guided Search, we have reached this position from a very different route. Guided Search assumes that a large amount of noise is inevitable in the guidance process; we have assumed that the effective amount of noise varies with the amount of time flexibly allowed to that process. Thus, guidance is not inaccurate because it must be, but because it may be faster to quickly guess at and check a few locations rather than waiting for a more certain guess at the target location.

In our model, the parameters that lead to the fastest search depend on how long an identification process will take, the amount of noise in the system, and the search display size. The first two parameters can be expected to vary with the perceptual discriminability of target vs distractors, while the participants knowledge of the display size can be varied experimentally. Our analysis predicts that subjects should be measurably more efficient for searches in which they know the display size before the trial.

According to this analysis, parallel neural network models can support the conclusion that, under many conditions, search will have a small number of serial fixations. This conclusion corresponds well to the finding from eye tracking experiments that participants in visual search tasks that allow eye movements show a small number of fixations in searching relatively large displays (Brown & Gilchrist, 2000; Williams & Reingold, 2001).

This type of model can potentially account for the full efficiency range of visual search findings. The discriminability of targets from distractors can be modeled by changing noise amounts, by reducing the difference in input values for different stimuli, or both. Informal experimentation suggest that these changes can produce a range of search efficiencies. However, findings of slow feature search, and cases in which little or no information seems to be guiding search, require a more complex explanation. We are working on models that explain these effects as results of the increasing size of receptive fields outside the fovea, and in neurons receptive to more subtle visual features (Herd & O'Reilly, in preparation).

As a final note, the current work has an interesting link to theories of visual search inspired by Signal Detection Theory. A major criticism of visual search theories is that they are "high threshold", that is, they do not allow for a distractor to be misidentified as a target. High threshold theories have been convincingly rejected in the domain of simple detection (Palmer, Verghee, & Pavel, 2000). The current model avoids this criticism in that the parallel stage of the model is 'low threshold'; it often misidentifies a distractor location as a target location.

However, studies of search overwhelmingly show many more misses than false alarms, implying that a low threshold model is not the only factor. We therefore theorized a second identification process, which checks the identity of the object at the selected location, and restarts search if it is not the target. In the current model this process is truly 'high threshold'; we assume it never mistakes a distractor for a target. Because false alarms certainly do occur in most search tasks, a more realistic model would include a decision process that uses a relatively high threshold for the identification process, but that does sometimes mistake a distractor for a target. This two stage arrangement may be more efficient than simply using a high decision criteria for the parallel location process, because it directs the (likely) time consuming work of more certain identification process only to locations that are likely to contain a target.

References

Brown, J. M. F. V., & Gilchrist, I. D. (2000). Saccade target selection in visual search: The effect of information from the previous fixation. *Vision Research, 41*(1), 87–95.

Chelazzi, L. (1999). Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research, 62*, 195–219.

Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition, 8*, 119–140.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuro science, 18*, 193.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3).

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics, 63*(2), 272–285.

O'Reilly, R. C. (1998). Six principles for biologically-based computational models of cortical cognition. *Trends in Cognitive Sciences, 2*(11), 455–462.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge MA: MIT Press.

Palmer, I., Verghee, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*, 1227–1268.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive viusal search: The distractor-ratio effect. *Canadian Journal of Experimental Psychology, 57*(2), 76–96.

Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences, 8*(5), 223–230.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review, 8*(3), 476–488.

Wolfe, J. M. (1994). Guided search 2.0—a revised model of visual search. *Psychonomic Bulletin and Review, 1*(2), 202–238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Philadelphia: Psychology 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 & Performance, 15*(3), 419–433.