UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Dynamic Gating in Vision

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 14(0)

Authors

Postma, Eric O.
van den Herik, H. Jaap
Hudson, Patrick T.W.

Publication Date

1992

Peer reviewed

Dynamic Gating in Vision*

Eric O. Postma, H. Jaap van den Herik and Patrick T. W. Hudson

Department of Computer Science
University of Limburg
P.O. Box 616, 6200 MD Maastricht/The Netherlands
postma@cs.rulimburg.nl

Abstract

Visual attention requires the selection of salient regions and their remapping into a position-invariant format. We propose the *dynamic-gating* model capable of *autonomous* remapping. It combines the localization network of Koch and Ullman (1985) with a modified shifter-circuit network (Anderson & Van Essen, 1987). Autonomous selection and remapping of salient regions result from local gating dynamics and local connectivity, implying that scaling to large problem sizes is straightforward.

1. Introduction

The visual system is equipped with a highly flexible process that enables the allocation of computational resources to a restricted part of the retinal image. When operating independently of eye movements, this process is called *covert attention* (Posner and Presti, 1987). This contribution proposes a neural model of covert attention. In order to allow scaling to large problem sizes (e.g., vision), we impose the implementational restrictions of *local processing* and *local connectivity*. By distributing task load and executing computations in parallel we aim at achieving a performance matching that of the human visual system (Nelson & Bower, 1990). Section 2 discusses behavioral findings on covert attention. In Section 3, dynamic remapping is suggested to play an important role in the visual system. Section 4 presents the dynamic-gating model and Section 5 gives some simulation results. Finally Section 6 evaluates the model.

2. Covert attention

A common metaphor for covert attention is of a spotlight illuminating part of the retinal image. The processing of stimuli captured by the spotlights' beam is enhanced at the cost of the processing of stimuli lying outside the beam. Behavioral studies revealed that covert attention indeed behaves much like a spotlight. Sagi and Julesz (1986), for instance, found performance on the detection of a test flash to be enhanced when attention was directed on a to-be-identified stimulus in its vicinity. The enhancement was strongest at the stimulus location and dropped gradually at increasing eccentricity. Other findings indicate that the attentional beam can be expanded to cover larger retinal regions (Eriksen, 1990), coincided by an enhancement that is inversely proportional to the size of the attended area. Apparently, then, the computational resources invoked in covert attention are limited causing a trade-off between resolution and viewing angle.

The speed of moving the spotlight to new locations has been reported to exceed the speed of eye movements (even 4 to 5 times, Saarinen & Julesz, 1991). Other findings (Remington & Pierce, 1984; Kwak, Dagenbach, & Egeth, 1991) show that attention jumps in a *time invariant* fashion (i.e., with a speed proportional to distance) suggesting an underlying mechanism reminiscent of the saccades associated with overt attention (Posner, 1980).

Covert attention is here assumed to be based on two processes: *localization* and *identification*. In localization, a conspicuous area is selected very rapidly in order to align the attentional beam with it. In identification, the contents at the attended region are matched against internal object representations. Such a matching requires the computation of an object-centered frame of reference. We will focus on how localization and the construction of an object-centered frame of reference can be modeled given our implementational restrictions of local processing and local connectivity.

3. Dynamic remapping

Detailed visual analysis of complex patterns requires the allocation of large computational resources. It is not feasible to allocate multiple recognizers for all positions in the visual field. Rather a single dedicated recognition module, dynamically linked to a (spatially contiguous) part of the retinal input (cf. the attentional spotlight), is a much more parsimonious solution (cf. Van Essen, Anderson, & Felleman, 1992). Recent neurophysiological findings suggest such a process to operate in concert with visual attention. Duhamel, Colby and Goldberg (1992) found receptive fields of neurons in the parietal cortex to shift in anticipation of eye movements. As a result, the internal representation of the visual scene is remapped to match the retinal pattern after saccade completion. Desimone, Moran and Spitzer (1989) found attentionally modulated receptive-field shifts in inferior temporal cortex. Considering the common functional distinction of a "where" (occipitoparietal) pathway dealing with the task of encoding spatial relations and a "what" (occipitotemporal) pathway involved in the task of object recognition (e.g., Mishkin & Appenzeller, 1987; Goodale & Milner, 1992), a tentative view holds that dynamic remapping accounts for maintaining constancies appropriate for the task at hand.

Construction of perceptual reference frames (e.g., object-centered or viewer-centered) may proceed by appropriate sampling of retinally contiguous spatial (and feature) patterns. Such selective sampling can effectively be realized by dynamic remapping. Two complementary neural models using dynamic remapping have been proposed in the literature. One model deals with the localization of conspicuous patterns, the other accounts for the remapping of patterns into an appropriate reference frame. Below, we briefly discuss these models integrated in the dynamic-gating model presented in Section 4.

The shifter-circuit network

The remapping of patterns has been hypothesized to occur through dynamic routing by Anderson and Van Essen (1987). The shifter-circuit network shown in Figure 1a consists of a hierarchy of three concatenated *shifter circuits* accommodating the routing of a contiguous pattern in the input towards the output. Each shifter circuit shifts incoming patterns to the left or to the right by selectively enabling the transmission lines pointing in one direction while disabling the lines pointing in

the other direction. Patterns presented at the bottom of a shifter circuit are, therefore, remapped without distorting the internal pattern structure. Critical for proper functioning of a shifter circuit is the requirement to align enabled (disabled) transmission lines. To fulfil this requirement Anderson and Van Essen (1987) introduced one *shift-control* module for each shifter circuit. A shift-control module contains two neurons, one contacting all transmission lines pointing leftwards and one contacting all transmission lines pointing rightwards. At any time only one of the shift-control neurons is active, effectively disabling signal flow through all the lines it contacts. A shifter-circuit architecture with L layers (i.e., shifter circuits) and k shifting directions ($k = 2$ in Figure 1a) is capable of remapping a subpattern of length i from any position in an input layer of length $i + k^L - 1$ towards the output layer.

The localization network

Koch and Ullman (1985) proposed a hierarchical network that enables localization and detection of conspicuous features. The structure of their model is depicted in Figure 1b. It consists of L layers with k^l links in layer l ($l = 1$ is the top layer, $k = 2$ in Figure 1b). A Winner-Take-All (WTA, Feldman & Ballard, 1982) competition among k adjacent transmission lines (neurons) within a subtree results in the selection of a local maximum input value. After selection, the "winning" value is propagated (remapped) to the next layer upwards. The maximum value is available at the top layer. Figure 1b (bottom) shows the localization of the maximum by a concatenated sequence of "winning" transmission lines (arrows). A feedback network of auxiliary units (each associated with a single upward transmission line) determines the position of the maximum. After the maximum has reached the top layer, auxiliary elements compare the state of their associate transmission line with the state of the one directly above. A unit becomes active if both lines are winners, otherwise it remains inactive. Consequently, activated auxiliary units trace the concatenated winning-line sequence in the reverse direction.

4. The dynamic-gating model

The localization and shifter-circuit networks can be combined to form an autonomous "spotlight" that orients towards conspicuous regions. A salient pattern detected by the localization network can steer
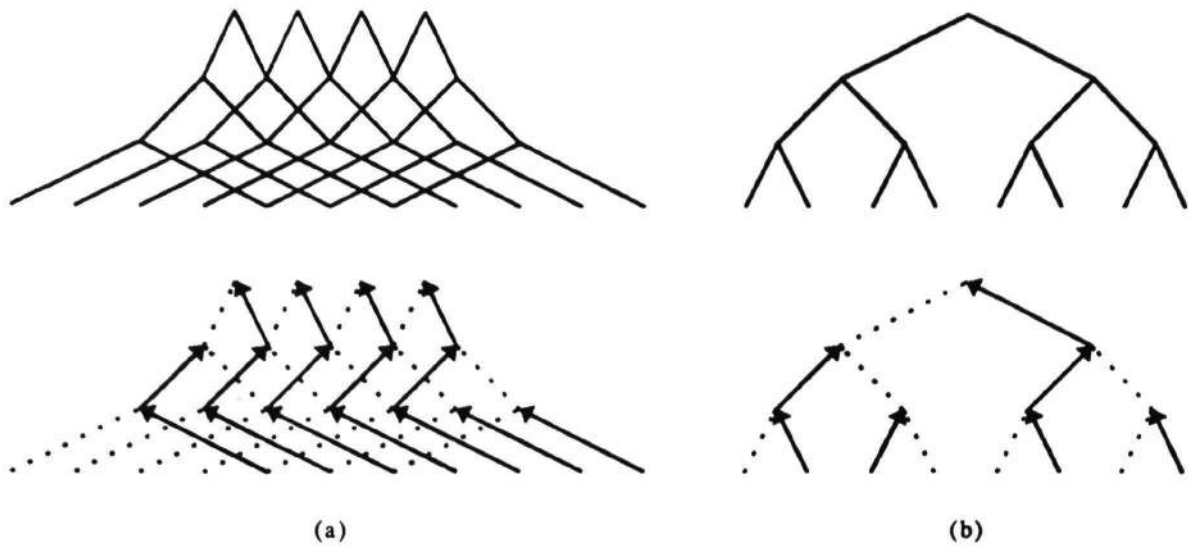
Figure 1: *(a) The shifter-circuit network proposed by Anderson and Van Essen (1987, top) and its selection/remapping of a pattern of length 4 (bottom). (b) The localization network proposed by Koch and Ullman (1985. top) and its localization of the maximum input value.*

the shifter-circuit network so that the pattern is routed towards its top layer where it can be matched against prototypical object representations. Our implementational constraints of locality of processing and connectivity guide the successive modifications necessary for full integration. This approach manages to retain the neurophysiological plausibility of its component networks. In particular the combination provides a model for dynamic remapping within the occipitotemporal pathway.

Network structure

Although the network structures shown in Figure 1 have similar characteristics they are still incompatible since the magnitude of shifts in localization networks increase (when going upward) whereas in shifter-circuit networks they decrease. To maintain local connectivity, similar functions (i.e., localizing and remapping) must be executed in adjacent areas (e.g., Nelson & Bower, 1990). Therefore we have modified the shifter-circuit network into a structure that matches the localization network by reversing the order of shifter circuits (the *reversed shifter-circuit network*). As shown in Figure 2a and b, the reversed shifter-circuit network matches the structure of the localization network. Although the reversed network employs a larger number of transmission lines in comparison to the original, it achieves the same remapping capacity given an equal number of layers. (Both the standard and

reversed network shown in Figure 1a and 2a can remap a pattern of length 4 out of an input of length 11.) Both networks sample the same input, but the localization network samples at a lower resolution (see below).

Gating dynamics

The elements of both networks have a combined gating and competition (selection) function. The gating element (or *gate*, represented by a circle in Figure 2) proposed herein performs both functions simultaneously. It competes with the links in its local neighborhood and, if it wins the WTA-competition, gates a local input value to its output. The dynamics of the attentional spotlight requires considerable and rapid flexibility. Many WTA-schemes react relatively slowly to any change in the input. For this reason we proposed local stochastic gating dynamics based on the neuron model of Little (e.g., Little & Shaw, 1975; Postma, van den Herik, & Hudson, 1992). The intrinsic noise of a gate is exploited to enable it to respond rapidly to changing input. At the same time there is a limit; too much flexibility causes inherent instability (see below).

Horizontal connectivity

The nature of a gate's horizontal interactions (i.e., its local neighborhood) differs for the two networks. In the localization network winners are determined
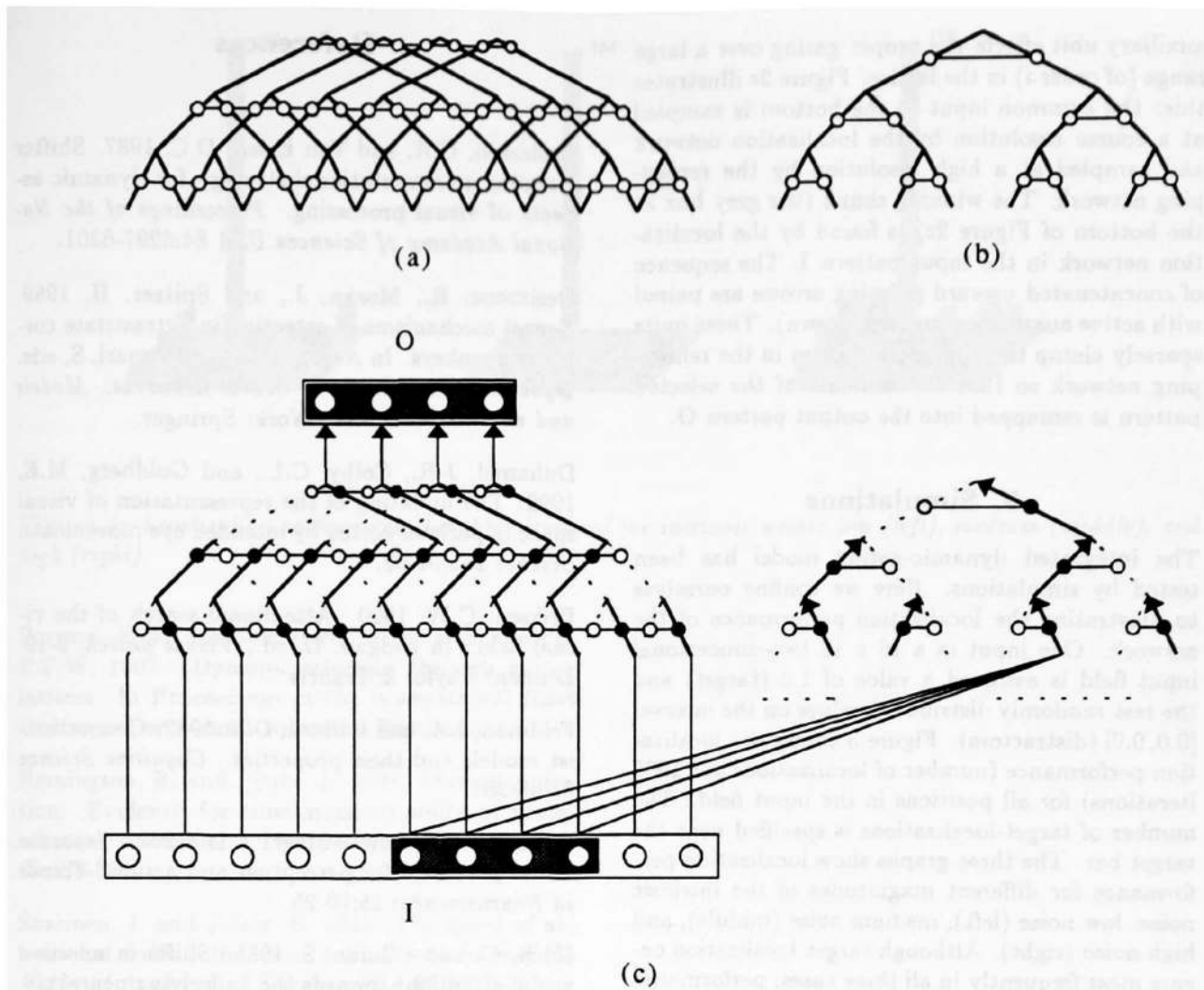
1040

Figure 2: *(a) Reversed shifter-circuit network. (b) Localization network. (c) Integrated dynamic-gating network. The circles represent open (black) or closed (white) gates.*

locally, whereas in the remapping network winners are determined over the length of the output pattern. These opposing characteristics are directly related to the distinct functions of both networks and should be reflected in the definition of the neighborhood of the gates. The horizontal lines in Figure 2 illustrate the appropriate connectivity pattern of both networks. In the localization network, connectivity is restricted to non-overlapping local clusters of 2 gates (for the one-dimensional network and $k = 2$), i.e., each gate is inhibited by a single neighbor (cf. Koch & Ullman, 1985). In the remapping network each gate is inhibited by 2 neighbors, i.e., neighborhoods overlap. In two dimensions, horizontal interactions within a layer form a field of independent clusters in the localization network and a

lattice structure in the remapping network. Within each cluster, there is a single open gate. In the lattice, a regular spatial pattern of open gates (each surrounded by a circular neighborhood of closed gates) reflects aligned gating within a single layer. The latter represents a special case of the Ising Lattice of statistical mechanics and exhibits global gating behavior on the basis of local interactions (see Postma, van den Herik, & Hudson, 1992, for a more detailed treatment of these issues).

Integration

In the integrated dynamic-gating model, the auxiliary units *clamp* a small subset of gates within a layer of the remapping network. A single active

auxiliary unit effects the proper gating over a large range (of order i) in the lattice. Figure 2c illustrates this: the common input at the bottom is sampled at a course resolution by the localization network and sampled at a high resolution by the remapping network. The winning chunk (the grey box at the bottom of Figure 2c) is found by the localization network in the input pattern I. The sequence of concatenated upward pointing arrows are paired with active auxiliary units (not shown). These units sparsely clamp the appropriate gates in the remapping network so that the contents of the selected pattern is remapped into the output pattern O.

5. Simulations

The integrated dynamic-gating model has been tested by simulations. Here we confine ourselves to illustrating the localization performance of the network. One input in a 16 x 16 two-dimensional input field is assigned a value of 1.0 (target) and the rest randomly distributed values on the interval $[0.0, 0.9]$ (distractors). Figure 3 shows the localization performance (number of localizations per 1000 iterations) for all positions in the input field. The number of target-localizations is specified near the target bar. The three graphs show localization performance for different magnitudes of the intrinsic noise: low noise (left), medium noise (middle), and high noise (right). Although target localization occurs most frequently in all three cases, performance is best at a medium noise level. At low-noise levels, localization tends to "stick" to local maxima (distractors) whereas at high-noise levels it becomes unstable. The optimal (intermediate) noise level combines stability against input noise with vigilance for input change.

6. Evaluation

In conclusion we may state that we have succeeded in formulating a neural model of the attentional spotlight. By obeying the implementational requirements of locality of processing and connectivity, we arrived at an architecture that can be scaled to large visual inputs. As a direct continuation we are currently studying the performance of the dynamic-gating model on a range of covert-attention tasks.

References

Anderson, C.H. and Van Essen, D.C. 1987. Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences USA* 84:6297-6301.

Desimone, R., Moran, J., and Spitzer, H. 1989. Neural mechanisms of attention in extrastriate cortex of monkeys. In Arbib, M.A. and Amari, S. eds. *Dynamic interactions in neural networks: Models and data*, 169-182. New York: Springer.

Duhamel, J-R., Colby, C.L., and Goldberg, M.E. 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255:90-92.

Eriksen, C.W. 1990. Attentional search of the visual field. In Brogan, D. ed., *Visual Search*, 3-19. London: Taylor & Francis

Feldman, J.A. and Ballard, D.H. 1982. Connectionist models and their properties. *Cognitive Science* 6:205-254.

Goodale, M.A. and Milner, A.D. 1992. Separate visual pathways for perception and action. *Trends in Neurosciences* 15:20-25.

Koch, C. and Ullman, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* 4:219-277.

Kwak, H-W., Dagenbach, D., and Egeth, H. 1991. Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics* 49:473-480.

Little, W.A. and Shaw, G.L. 1975. A statistical theory of short and long term memory. *Behavioral Biology* 14:115-133.

Mishkin, M. and Appenzeller, T. 1987. The anatomy of memory. *Scientific American* 256:62-71.

Nelson, M.E. & Bower, J.M. 1990. Brain maps and parallel computers. *Trends in Neurosciences*, 13:403-408.

Posner, M.I. 1980. Orienting of attention. *Quarterly Journal of Experimental Psychology* 32:3-25.

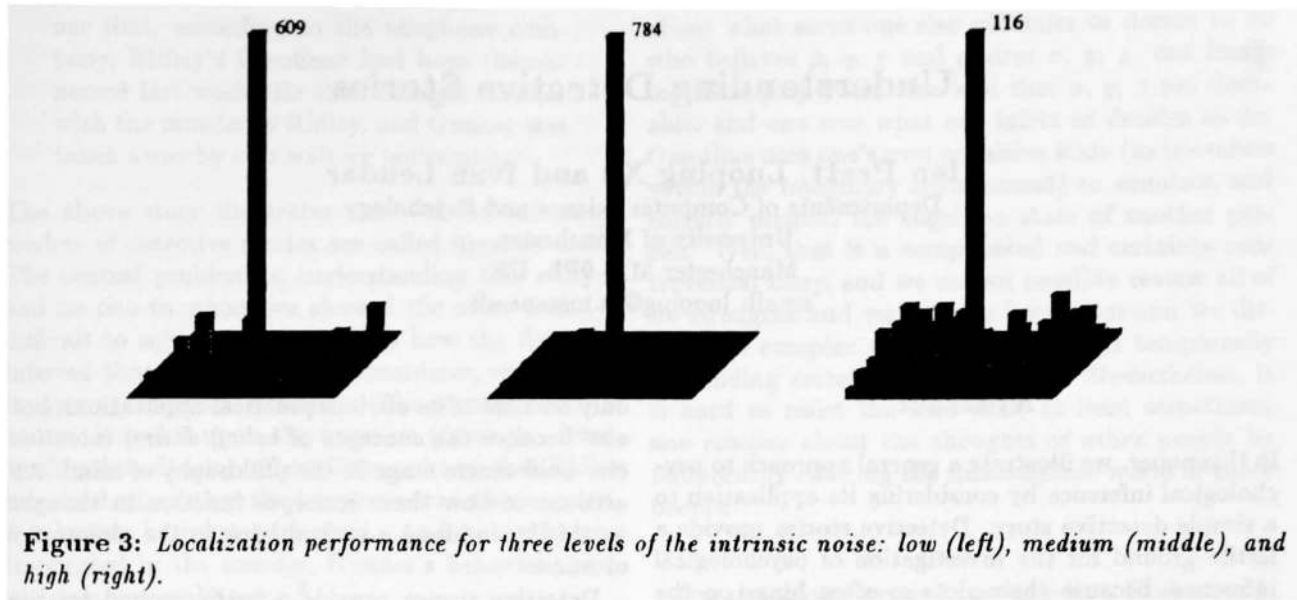Posner, M.I. and Presti, D.E. 1987. Selective attention and cognitive control. *Trends in Neurosciences* 10:13-17.

Figure 3: *Localization performance for three levels of the intrinsic noise: low (left), medium (middle), and high (right).*

Postma, E.O., van den Herik, H.J., and Hudson, P.T.W. 1992. Dynamic selection through gating lattices. In Proceedings of the International Joint Conference on Neural Networks, Baltimore.

Remington, R. and Pierce, L. 1984. Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics* 35:393-399.

Saarinen, J. and Julesz, B. 1991. The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences USA* 88:1812-1814.

Sagi, D. and Julesz, B. (1986). Enhanced detection in the aperture of focal attention during simple detection tasks. *Nature* 321:693-695.

Van Essen, D.C. and Anderson, C.H. 1990. Information processing strategies and pathways in the primate retina and visual cortex. In Zornetzer, S.F., Davis, J.L., and Lau, C., eds., *An introduction to neural and electronic networks*. Academic Press, San Diego, CA.

Van Essen, D.C., Anderson. C.H., and Felleman, D.J. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science* 255:419-423.