A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information

Bruno A. Olshausen,[1,3] Charles H. Anderson,[1,2,3] and David C. Van Essen[1,3]

[1]Computation and Neural Systems Program, California Institute of Technology, Pasadena, California 91125, [2]Jet Propulsion Laboratory, Pasadena, California 91109, and [3] Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

We present a biologically plausible model of an attentional mechanism for forming position- and scale-invariant representations of objects in the visual world. The model relies on a set of *control neurons* to dynamically modify the synaptic strengths of intracortical connections so that information from a windowed region of primary visual cortex (V1) is selectively routed to higher cortical areas. Local spatial relationships (i.e., topography) within the attentional window are preserved as information is routed through the cortex. This enables attended objects to be represented in higher cortical areas within an object-centered reference frame that is position and scale invariant. We hypothesize that the pulvinar may provide the control signals for routing information through the cortex. The dynamics of the control neurons are governed by simple differential equations that could be realized by neurobiologically plausible circuits. In preattentive mode, the control neurons receive their input from a low-level "saliency map" representing potentially interesting regions of a scene. During the pattern recognition phase, control neurons are driven by the interaction between top-down (memory) and bottom-up (retinal input) sources. The model respects key neurophysiological, neuroanatomical, and psychophysical data relating to attention, and it makes a variety of experimentally testable predictions.

[Key words: visual attention, recognition, model, gating, visual cortex, pulvinar, control]

Of all the visual tasks humans can perform, pattern recognition is arguably the most computationally difficult. This can be attributed primarily to two major factors. The first is that in order to recognize a particular object, the brain must go through a matching process to determine which of the countless objects it has seen before best matches a particular object under scrutiny. The second factor is that any particular object can appear at different positions, sizes, and orientations on the retina, thus giving rise to very different neural representations at early stages of the visual system.

Research on associative memories has provided some insight as to how the problem of pattern matching can be solved by neural networks (e.g., Hopfield, 1982; Kanerva, 1988). However, it is far less clear how the brain solves the second problem to produce object representations that are *invariant* with respect to the dramatic fluctuations that occur on the sensory inputs. Our goal here is to propose a neurobiological solution to this problem that is detailed enough in its structure to generate useful experimental predictions.

Our basic proposal is similar to a psychological theory put forth by Palmer (1983), in which it was proposed that the process of attending to an object places it into a canonical, or object-based, reference frame. It was suggested that the position and size of the reference frame could be set by the position and size of the object in the scene (assuming it was roughly segmented), and that the orientation of the reference frame could be estimated from relatively low-level cues, such as elongation or axis of symmetry (see also Marr, 1982). The computational advantage of such a system is obvious: only one or a few versions of an object need to be stored in order for the object to be recognized later under different viewing conditions. The disadvantage, of course, is that a scene containing multiple objects requires a serial process to attend to one object at a time. However, psychophysical evidence suggests that the brain indeed employs such a sequential strategy for pattern recognition (Bergen and Julesz, 1983; Treisman, 1988).

Palmer made no attempt to describe a neural mechanism for transforming an object's representation from one reference frame to another, because his was primarily a psychological model. Various other models have been proposed for transforming reference frames using neural circuitry (Pitts and McCulloch, 1947; Hinton, 1981a; Hinton and Lang, 1985; von der Malsburg and Bienenstock, 1986). Of these, only the proposal of Pitts and McCulloch can be viewed truly as a neurobiological model. However, their proposal—that the brain averages over all possible transformations of an object via a scanning process—cannot be reconciled with our current understanding of the visual cortex.

In this article we propose a neurobiological mechanism for routing retinal information so that an object becomes represented within an object-based reference frame in higher cortical areas. The mechanism is modified and expanded from an earlier proposal (Anderson and Van Essen, 1987) for dynamically shifting the alignment of neural input and output arrays without loss of spatial relationships. The model presented here allows both shifting and scaling between input and output arrays, and it also provides a solution for controlling the shift and scale in an

Received Dec. 3, 1992; revised Apr. 29, 1993; accepted May 6, 1993.

We thank Christof Koch, John Tsotsos, Andreas Herz, Harold Hawkins, Ed Connor, and Bill Press for critical commentary and helpful suggestions. This work was supported by ONR Grant N00014-89-J-1192 and U.S. PHS Grant MH19138-02 (T32). Portions of this work were described in CNS Memo 18, California Institute of Technology, August 1992.

Present address of all authors, Department of Anatomy and Neurobiology, Box 8108, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110.

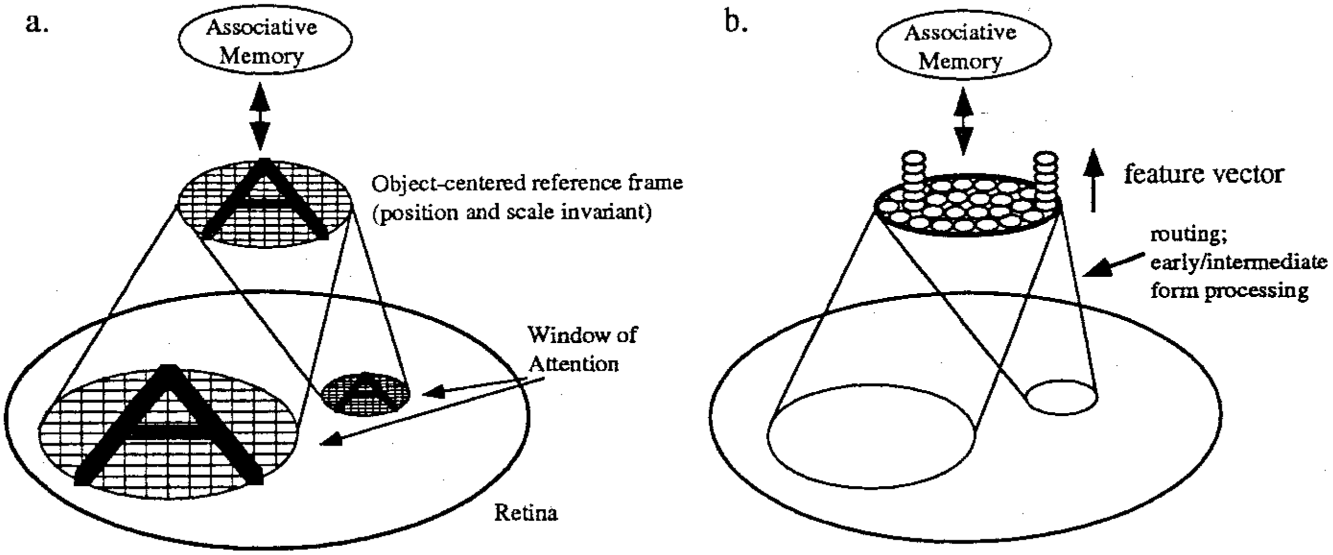
a.
Associative Memory
Object-centered reference frame (position and scale invariant)
Window of Attention
Retina
b.
Associative Memory
feature vector
routing; early/intermediate form processing

Figure 1. Shifting and rescaling the window of attention. The image within the window of attention in the retina is remapped onto an array of sample nodes in an object-centered reference frame. *a,* In the simplest scheme, each "pixel" in the object-centered reference frame represents image luminance. *b,* More realistically, each pixel should presumably correspond to a feature vector that integrates over a somewhat larger spatial region and represents orientation, depth, texture, and so on.

autonomous fashion. While the model is clearly an oversimplification in some respects, it respects key neuroanatomical constraints and is consistent with neurophysiological and psychophysical data relating to directed visual attention.

We begin with a description of the basic model—the *dynamic routing circuit*—and its autonomous control. Subsequent sections describe the proposed neurobiological substrates and mechanisms, predictions of the model, and a comparison with other models that have been proposed for visual attention and recognition.

The Model

The goal of our model is to provide a neurobiologically plausible mechanism for shifting and rescaling the representation of an object from its retinal reference frame into an object-centered reference frame. Information in the retinal reference frame is represented on a neural map (the topographic representation in V1), and we hypothesize that information in the object-based reference frame is also represented on a neural map, as illustrated in Figure 1. This does not necessarily imply that only "pixels" can be routed into the high level areas, as drawn in Figure 1*a*; each sample node in the high level map could be expanded into a feature vector representing various local image properties, such as orientation, texture, and depth, that are made explicit along the way (Fig. 1*b*).

In order to map topographically an arbitrary section of the input onto the output, the neurons in the output stage need to have dynamic access to neurons in the input stage. In the brain, this access must necessarily be obtained via the physical hardware of axons and dendrites. Since these pathways are physically fixed for the time scale of interest to us (< 1 sec), there needs to be a way of dynamically modifying their strengths. We propose that the efficacy of transmission along these pathways is modulated by the activity of *control neurons* whose primary responsibility is to dynamically route information through successive stages of the cortical hierarchy.

A dynamic routing circuit

Figure 2*a* shows a simplified, one-dimensional dynamic routing circuit (the next section discusses how this circuit can be scaled up as a model of the visual cortex). It consists of an input layer of 33 nodes, an output layer of five nodes, and two layers in between. Additionally, a set of control units make multiplicative contacts onto the feedforward pathways in order to change connection strengths. This network has been constructed so that

1. the fan-in (number of inputs) on any node is the same—in this case 5,
2. the spacing between inputs doubles at each successive stage, and
3. the number of nodes within a layer is such that the spread of its total input field just covers the layer below.

This connection scheme has the attractive property of keeping the fan-in on any node fixed to a relatively low number while allowing the nodes in the output layer access to any part of the input layer. This property will be important in scaling up the model.

An example of how the weights might be set for different positions and sizes of the window of attention is shown in Figure 2, *b* and *c*. When the window is at its smallest size (same resolution as the input stage, Fig. 2*b*), the weights are set so as to establish a one-to-one correspondence between nodes in the output and the attended nodes in the input. When the window is at a larger size, the weights must be set so that multiple inputs converge onto a single output node, resulting in a lower-resolution representation of the contents of the window of attention on the output nodes. If the input representation were to contain nodes tuned for different spatial frequencies, then the low-frequency nodes would be primarily used when the window of attention is large, whereas the high-frequency nodes would be

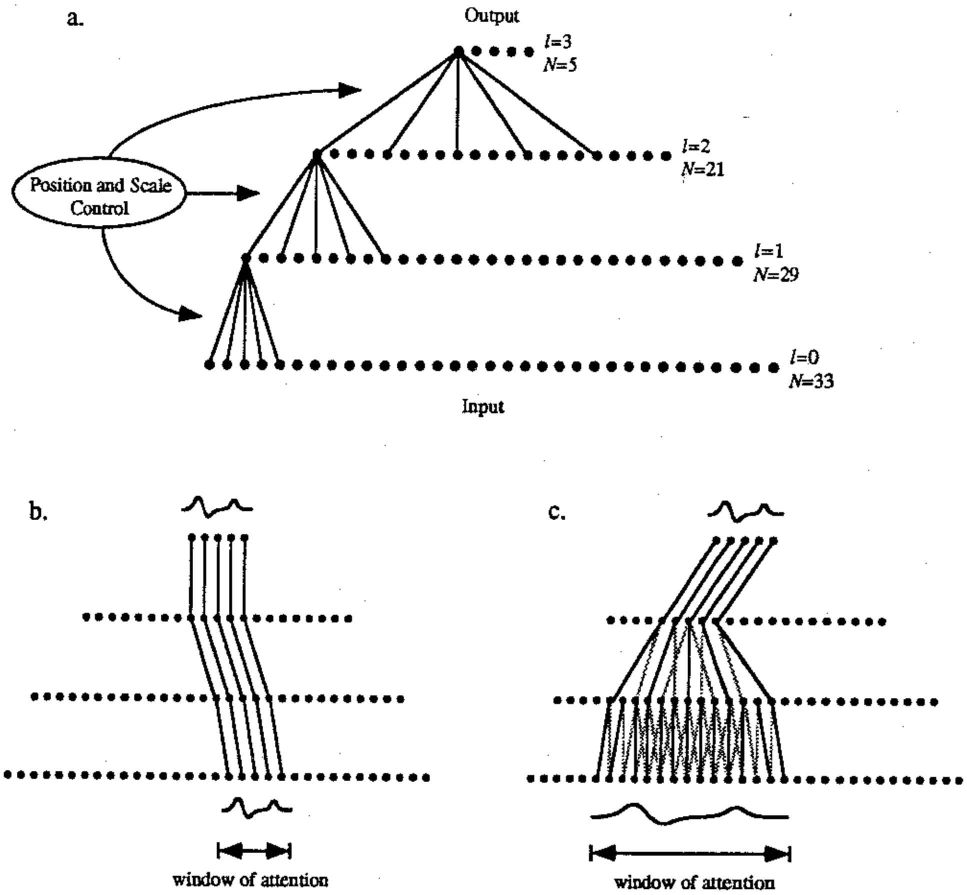

Figure 2. A simple, one-dimensional dynamic routing circuit. *a,* Connections are shown for the leftmost node in each layer. The connections for the other nodes are the same, but merely shifted. N denotes the number of nodes within each layer, and l denotes the layer number. A set of control units (not explicitly shown) provide the necessary signals for modulating connection strengths so that the image within the window of attention in the input is mapped onto the output nodes. *b* and *c,* Some examples of how connection strengths would be set for different positions and sizes of the window of attention. The gray level of each connection denotes its strength. Each node, I_i^l, essentially interpolates from the nodes below by forming a linear weighted sum of its inputs:

$$I_i^l = \sum_j w_{ij}^{l-1} I_j^{l-1},$$

where w_{ij}^l denotes the strength of the connection from node j in level l to node i in level $l + 1$. If a gaussian is used as the interpolation function, then w_{ij}^l is given by

$$w_{ij}^l = \exp\left[-\frac{(j - \alpha_l i - d_l)^2}{2\sigma_l^2}\right],$$

where the parameters d_l, α_l, and σ_l denote the amount of translation, scaling, and blurring, respectively, in the transformation from level l to level $l + 1$. The overall translation, scaling, and blurring of the entire circuit (d, α, and σ) is then given by $d = d_0 + \alpha_0(d_1 + \alpha_1 d_2)$, $\alpha = \alpha_0\alpha_1\alpha_2$, $\sigma^2 = \sigma_0^2 + \alpha_0^2(\sigma_1^2 + \alpha_1^2\sigma_2^2)$. Note that the lowest layers are best suited for small, fine-scale adjustments to the position and size of the attentional window, while the upper layers are better suited for large, coarse-scale adjustments.

used when the window is small. Thus, much of the image smoothing could be accomplished by using a set of hardwired filters, and then switching between these filters depending on the size of the attentional window.

The challenge in controlling the routing circuit lies in properly setting the synaptic weights to yield the desired position and size of the window of attention. Low levels of the circuit are well suited for making fine adjustments in the position and scale of the window of attention, whereas higher levels are best suited for coarse control. In general, though, there are an infinite number of possible solutions in terms of the combinations of weights that could achieve any particular input–output transformation.

Control

Our analysis of how information flow can be controlled is aided by visualizing the routing circuit in "connection space," as shown in Figure 3*a*. This diagram shows the connection matrix for a simple one-dimensional routing circuit composed of two layers—an input layer and an output layer. The horizontal axis represents the nodes constituting the input layer of the network; the vertical axis represents the nodes constituting the output layer. An "×" at coordinate (j, i) in connection space denotes that a physical connection exists from node j in the input to node i in the output; the lack of an "×" at (j, i) implies that

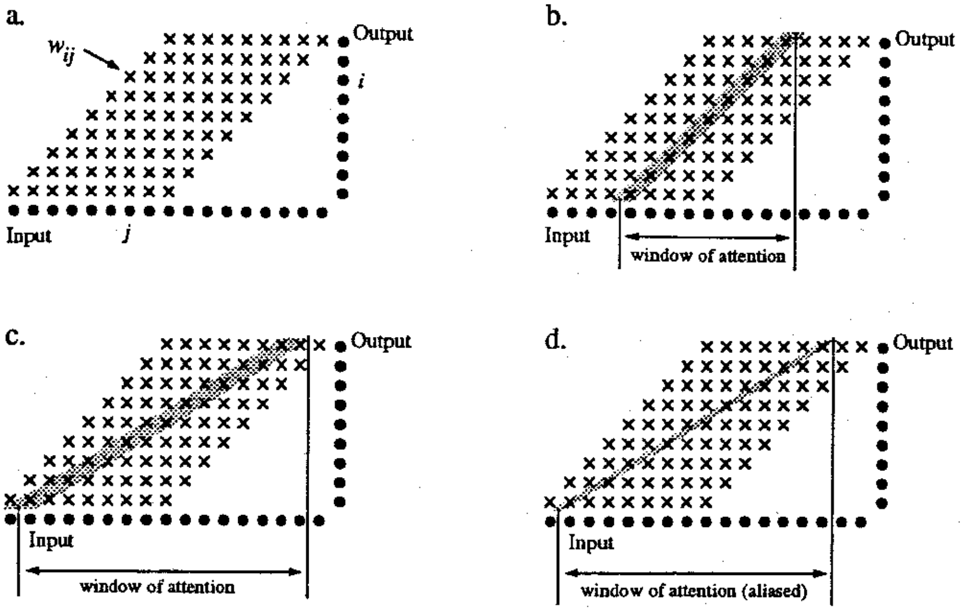

Figure 3. An illustration of "connection space." The input contains 17 sample nodes and the output contains nine sample nodes. *a*, Each × denotes a physical connection from an input node to an output node. We shall denote the effective strength of the connection from node j in the input to node i in the output as w_{ij}. *b* and *c*, The *stippled region* indicates those connections that need to be enabled ($w_{ij} > 0$) in order to map the region within the window of attention onto the output nodes. *d*, If the width of the enabled region is too small, then aliasing will result; an exaggerated case is illustrated here (i.e., some output nodes will be lacking any input, leading to spurious patterns in the output).

no connection pathway exists between those nodes. We denote the strength of the connection at (j, i) as w_{ij}. Note that for a two-dimensional routing circuit the connection matrix would require four dimensions to display. We will use the one-dimensional routing circuit for ease of illustration, but the concepts developed here are readily extendible to two dimensions.

If the window of attention is to be of a certain position and size, then the strength of each connection, w_{ij}, needs to be set appropriately. Figure 3*b* shows how this would look in connection space for an attentional window centered within the input array with a scale factor of one (i.e., no magnification). The stippled area represents those connections that are enabled; the remaining connections are effectively disabled by mechanisms discussed below. If the window of attention is to shift to the left or right, then the band of enabled connections must translate across the connection matrix. Changing the size of the window of attention corresponds to tilting the band of open connections, as shown in Figure 3*c*. Note that the band of open connections must also be widened as it is tilted (corresponding to blur); otherwise, aliasing would occur, leading to spurious patterns in the output representation (Fig. 3*d*).

By viewing the routing circuit in this way, it can be seen that the problem of setting the position, size, and blur of the window of attention amounts to one of generating the proper patterns of active synapses in connection space. How this might be accomplished by the control units depends on how they are connected to the feedforward synapses of the routing circuit. One possible scenario is for each control unit to modulate the strength of a single physical connection (j, i), as illustrated in Figure 4*a*. If a given control unit were "on," then its corresponding connection would be enabled, and if it were off then the connection would be disabled. Nearly any remapping could then be accomplished by simply activating the control units corresponding to the connections we wish to enable. However, this scheme would require an enormous number of control units for a scaled-up system. Since the set of remappings we wish to accomplish (translations and scalings) is but a minute fraction of all possible remappings, this scheme would arguably constitute a waste of computational resources. Another possibility would be for the control units to gate connections globally so that each unit is responsible for effecting a single position and scale of the window of attention, as shown in Figure 4*b*. However, this scheme would require a large fan-out for each control unit in a scaled-up system. This could cause implementation difficulties and render the circuit neurobiologically implausible.

Our proposed solution to the control problem minimizes both the number of control units and the fan-out required by having each control unit modulate a local group of synapses—or a *control block* in connection space (Fig. 4*c*). The problem of forming the desired patterns in connection space then becomes an approximation problem, in which the control blocks form the basis functions and the activations of the corresponding control units form the coefficients. That is, the connection strengths w_{ij} would be determined according to

$$w_{ij} = \sum_k c_k \Psi_k(j, i), \qquad (1)$$

where c_k denotes the activity of the kth control unit, and the function $\Psi_k(j, i)$ specifies the shape of the kth control block in connection space. In order to facilitate their ability to approximate patterns in connection space, the control blocks should not have sharp boundaries; rather, they should have a gaussian-like taper and overlap one another somewhat. Shaping the control blocks as in Figure 4*c* would be most optimal for realizing translations, but could also be used to approximate scalings as well, as shown in Figure 4*d*. It may well be possible to optimize the shape of the control blocks using appropriate learning algorithms, but the strategy illustrated here will suffice for our immediate purposes.

An alternative way of expressing Equation 1 that will be useful later is

$$w_{ij} = \sum_k c_k \Gamma_{ijk}, \qquad (2)$$

where $\Gamma_{ijk} = \Psi_k(j, i)$. In this sense, Γ_{ijk} denotes the weight with which c_k modulates the strength of synapse (j, i). Note that $\Gamma_{ijk} = 0$ for most combinations of i, j, and k, since each control neuron modulates only a small fraction of the many possible synapses (j, i).

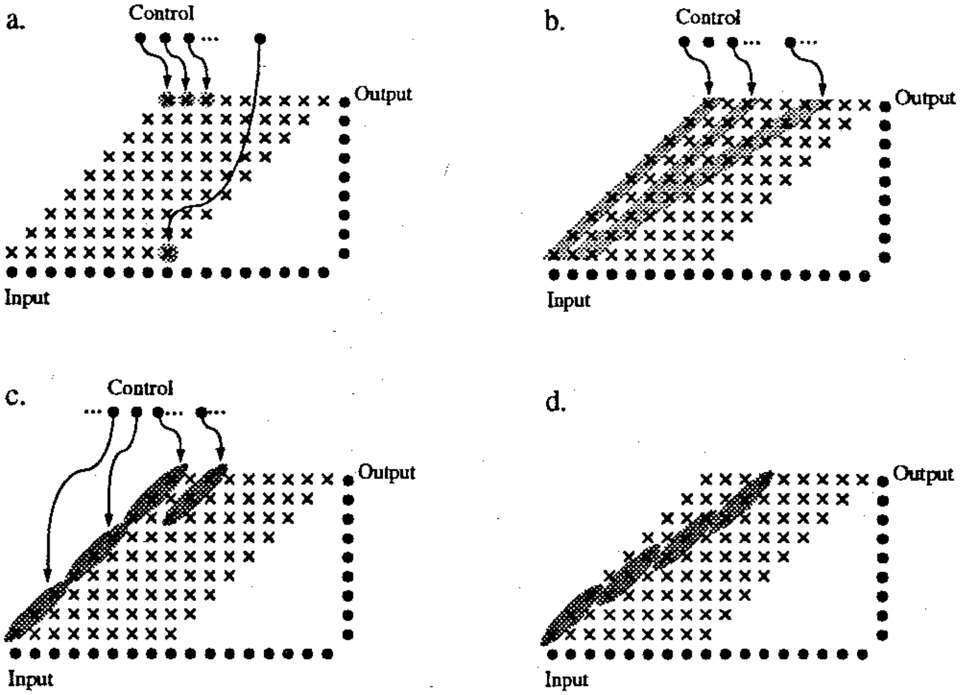

Figure 4. Some possible control scenarios. *a*, Each control unit modulates the strength of a single connection. *b*, Each control unit modulates the strength of a large number of connections in order to effect a global position and scale of the window of attention. *c*, Each control unit modulates a local group of connections, or a "control block." *d*, Approximating a desired position and scale of the window of attention using control blocks.

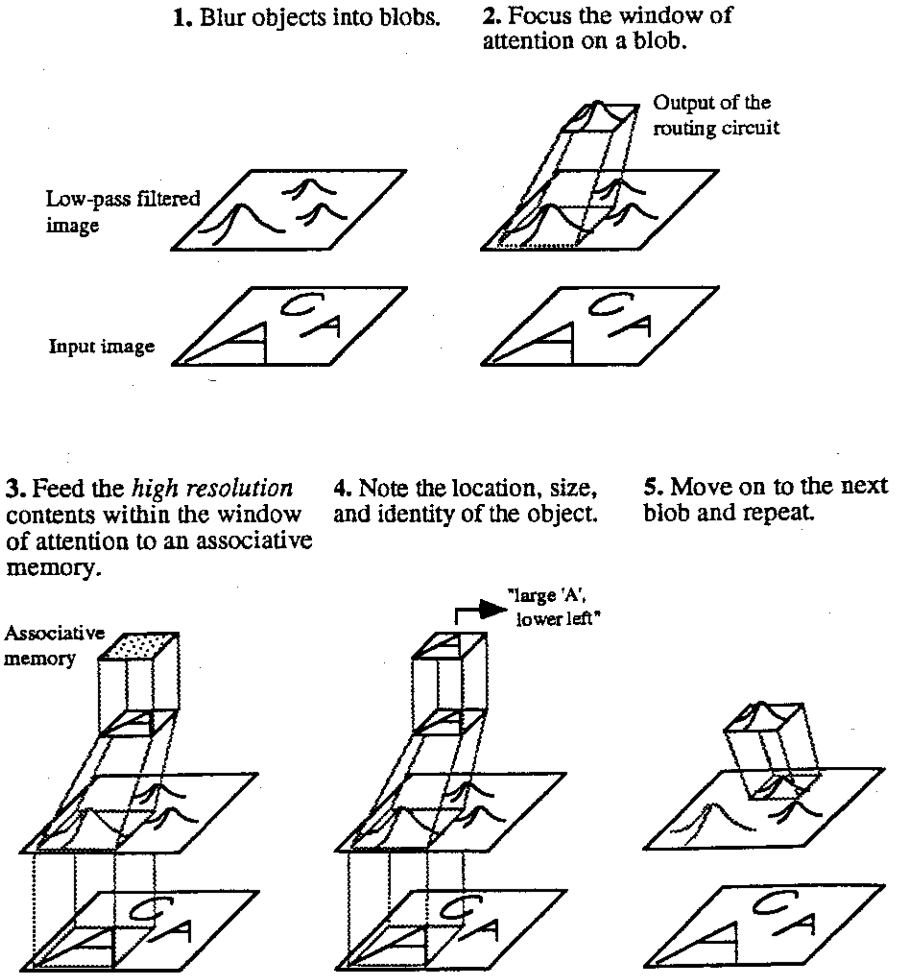
1. Blur objects into blobs.
2. Focus the window of attention on a blob.
Output of the routing circuit
Low-pass filtered image
Input image
3. Feed the *high resolution* contents within the window of attention to an associative memory.
4. Note the location, size, and identity of the object.
5. Move on to the next blob and repeat.
Associative memory
"large 'A', lower left"

Figure 5. A simple attentional strategy for an autonomous visual system. Objects are preattentively segmented via low-pass filtering. Once an object has been localized, the contents of the window of attention are fed to an associative memory for recognition. This process is then repeated ad infinitum, or until all interesting locations have been attended.

Autonomous control

Up to now we have described an essentially "open loop" model of visual attention. That is, given a desired position and size for the window of attention, one could manually set the activity of the control units of the network so that the image within the window is remapped onto the output units of the network. We now describe how the network may be autonomously controlled when provided only with visual input and no external commands beyond the initial task specification.

System objective. The purpose of attention is to focus the neural resources for recognition on a specific region within a scene. Thus, it would make sense for the attentional window to be automatically guided to salient, or potentially informative, areas of the visual input. Salient areas can often be defined on the basis of relatively low-level cues—such as pop-out due to motion, depth, texture, or color (e.g., Koch and Ullman, 1985; Anderson et al., 1985). Here, we utilize a very simple measure of salience based on luminance pop-out in which attention is attracted to "blobs" in a low-pass-filtered version of a scene. (A blob may be defined simply as a contiguous cluster of activity within an image.) In reality, attention can also be directed via voluntary or cognitive influences, but these are not incorporated into our present model.

We propose the following simple but useful strategy for an autonomous visual system (see Fig. 5).

1. Form a low-pass-filtered version of the scene so that objects are blurred into blobs.
2. Select one of the blobs from the low-pass image—whichever is brightest or largest—and set the position and size of the window of attention to match the position and size of the blob.
3. Feed the high-resolution contents of the window of attention to an associative memory for recognition.
4. If a match with one of the memories is close enough (by some as yet unspecified criterion), then consider the object to have been recognized; note its identity, location, and size in the scene. If there is not a good match, then consider the object to be unknown; either learn it or disregard it.
5. Now inhibit this part of the scene and go to step 2 (find the next most salient blob).

The following three subsections describe the details for carrying out steps 2, 3, and 5. Step 1 is trivial, whereas step 4 is a high-level problem beyond the scope of this article (cf. Carpenter and Grossberg, 1987; Mumford, 1992).

Focusing attention on a blob. We begin by formulating a solution for a simple one-dimensional routing circuit with one or more gaussian blobs presented to the input units, as shown in Figure 6*a*. The values on the output units, I_i^{out}, are computed from the input units, I_j^{in}, via

$$I_i^{\text{out}} = \sum_j w_{ij} I_j^{\text{in}} \tag{3}$$

$$= \sum_j \sum_k c_k \Gamma_{ijk} I_j^{\text{in}}. \tag{4}$$

Note that Equation 4 is obtained by substituting Equation 2 into Equation 3. In this simple circuit the Γ_{ijk} are set so that each control unit c_k corresponds to a global position of the window of attention, but in general this need not be the case.

In order to focus the window of attention on a blob in the input, the network's "goal" is to fill the output units with a blob while maintaining a topographic correspondence between the input and output (Fig. 5, step 2). Since the dynamic variables in this network are the c_k, we need to formulate an equation governing the dynamics of c_k that accomplishes this objective. We can accomplish the first part of the objective by letting c_k follow the gradient of an objective function, E_{blob}, that provides a measure of how well a blob is focused on the output units. One possible choice for E_{blob} is the correlation between the actual values on the output units, I_i^{out}, and the desired blob shape, G. That is,

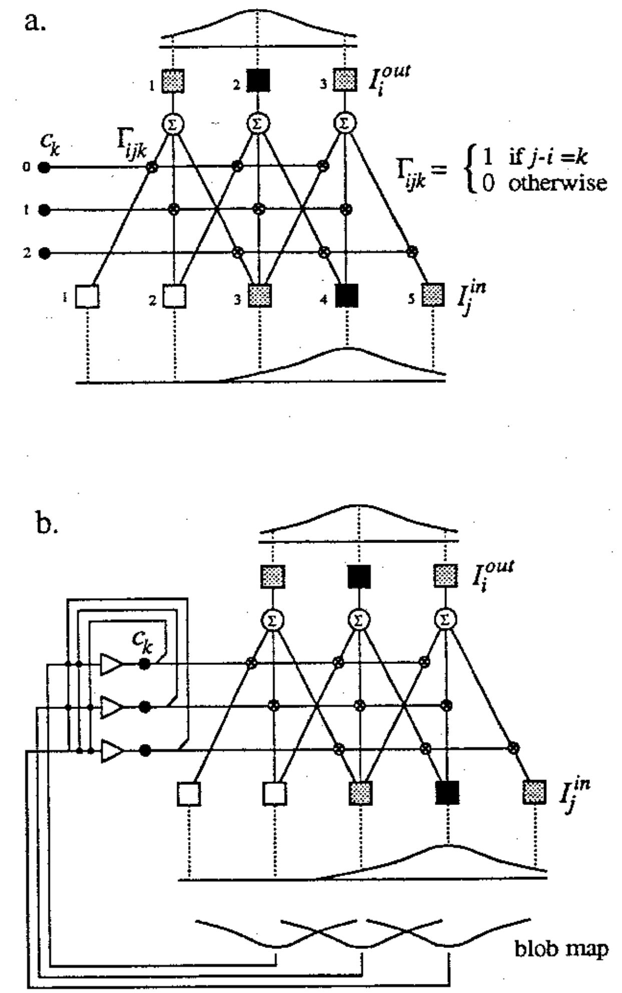
a.
1
2
3
I_i^{out}
c_k
Γ_{ijk}
0
1
2
$\Gamma_{ijk} = \begin{cases} 1 & \text{if } j\text{-}i = k \\ 0 & \text{otherwise} \end{cases}$
1
2
3
4
5
I_j^{in}
b.
I_i^{out}
c_k
I_j^{in}
blob map

Figure 6. *a*, A simple one-dimensional routing circuit with a gaussian blob presented to the input units. Each control unit corresponds to a different position of the window of attention: left (c_0), center (c_1), or right (c_2). For example, in order to accomplish the remapping shown, the values on the control units should be $c_2 = 1$ and $c_0 = c_1 = 0$. *b*, The same circuit with control circuitry added to autonomously focus the window of attention on a blob in the input. Each control unit essentially has a gaussian receptive field in the input layer. The control units then compete among each other, via negatively weighted interconnections, such that only the control unit corresponding to the strongest blob in the input prevails. The combined leaky integrator and squashing function (Eqs. 7, 8) are denoted by the *amplifier symbol.*

$$E_{\text{blob}} = -\sum_i I_i^{\text{out}} G_i,$$

$$G_i = \exp[-(i - \mu)^2/\sigma^2]. \tag{5}$$

The second part of the objective (maintaining topography) can be accomplished by letting c_k follow the gradient of a constraint function, $E_{\text{constraint}}$, that favors valid control states—that is, those corresponding to translations or scalings of the input–output transformation. One possible choice for $E_{\text{constraint}}$ is

$$E_{\text{constraint}} = -\frac{1}{2} \sum_{k,l} c_k T^c_{kl} c_l, \tag{6}$$

where the constraint matrix T^c is chosen so as to couple the control neurons appropriately. For the simple circuit of Figure 6*a*, each control neuron corresponds to a different position of the window of attention, so we could define T^c as

$$T^c_{kl} = \begin{cases} -1 & k \neq l \\ 0 & k = l. \end{cases}$$

This has the effect of punishing any state in which two or more control units are active simultaneously, and thus forces a winner-take-all solution. (The more general case using control blocks is described below.)

A dynamical equation for c_k that simultaneously minimizes both E_{blob} and $E_{\text{constraint}}$ is given by

$$c_k = \sigma(u_k), \tag{7}$$

$$\frac{du_k}{dt} + \tau^{-1} u_k = \eta \sum_i \sum_j G_i \Gamma_{ijk} I^{\text{in}}_j + \eta\beta \sum_l T^c_{kl} c_l, \tag{8}$$

where the constants τ and η determine the rate of convergence of the system, and the constant β determines the contribution of $E_{\text{constraint}}$ relative to E_{blob}. A sigmoidal squashing function (σ) is used to limit c_k to the interval [0, 1]. (A derivation of Eq. 8 is given in the Appendix.)

A neural circuit for computing Equations 7 and 8 is shown in Figure 6*b*. The first term on the right of Equation 8 is computed by correlating the gaussian, G, with a shifted version of the input (the amount of shift depends on the index k). The second term is computed by forming a weighted sum of the activities on the other control units. These two results are then summed together and passed through a leaky integrator and squashing function to form the output of the control unit, c_k. Thus, the c_k essentially derive their inputs directly from a "blob map," and then compete among each other so that the c_k corresponding to the strongest blob prevails.

The circuit of Figure 6 could easily be modified to allow for different sizes of the window of attention by adding another set of control units for each desired size of the window of attention. The control units corresponding to a large window of attention would then derive their inputs from a coarse-grained (low-resolution) blob map, while control units corresponding to a small window of attention would derive their inputs from a fine-grained (high-resolution) blob map. All of these units would then compete with one another so that the window of attention is constrained to a single position and scale. (See example in the next section.)

In a more biologically plausible scenario, the control units would be configured into control blocks, like those shown in Figure 4*c*. In this case, Equation 8 states that the input to each c_k would be computed by correlating the gaussian values, G_i, and the input values, I^{in}_j, that are "connected" via that control unit (specified by Γ_{ijk}). Note that since the G_i are fixed, the term $\Sigma_i\, G_i \Gamma_{ijk}$ (Eq. 8) can essentially be considered a fixed weight. Also, the constraint matrix, T^c, would need to be modified in this case so that those control units corresponding to a common translation or scale reinforce each other ($T^c_{kl} > 0$), while control units that are not part of the same transformation inhibit each other ($T^c_{kl} < 0$), as illustrated in Figure 7. This scheme has the effect of introducing many local minima, however, and so the control neurons need to be more tightly constrained in order to converge on states that preserve local spatial relationships. We have accomplished this by utilizing a coarse-to-fine control architecture (B. Olshausen, unpublished observations). In this scheme, routing is at first performed by a small number of control neurons on a low-pass-filtered version of the image, and this smaller set of control neurons is then used to constrain the activities of the fine-grained control neurons routing the high-resolution information.

Recognition. Once the window of attention has been focused on a blob, the underlying high-resolution information can also be fed through the routing circuit and into the input of an associative memory for recognition. However, it is likely that the initial estimation of position and size made by routing the blob would be only approximately correct, and this may cause problems for matching the high-resolution information. Thus, it would be desirable to have the associative memory help adjust the position and scale of the attentional window while it converges. How, then, shall the associative memory be incorporated into the control of the routing circuit?

If a Hopfield associative memory (Hopfield, 1984) is used for recognition, then we can replace E_{blob} with the associative memory's "energy" function, E_{mem}, which is defined as

$$E_{\text{mem}} = -\frac{1}{2} \sum_i \sum_j T_{ij} \text{V}_i V_j + \sum_i \frac{1}{R_i} \int_0^{V_i} g_i^{-1}(V)\,dV - \sum_i V_i I^{\text{mem}}_i. \tag{9}$$

In this equation the V_i denote the output voltages on the associative memory neurons, T_{ij} denotes the connection strength between neurons i and j, I^{mem}_i denotes the inputs to the memory, and g_i is a squashing function such as tanh(x). Normally, the only dynamic variables are the V_i, which evolve by following a monotonically increasing function, g_i, of the gradient of the energy. That is,

$$V_i = g_i(u^m_i) \tag{10}$$

$$C_i \frac{du^m_i}{dt} = \frac{-\partial E_{\text{mem}}}{\partial V_i} = \sum_j T_{ij} V_j - \frac{u^m_i}{R_i} + I^{\text{mem}}_i, \tag{11}$$

where C_i and R_i are constants that determine the integration time constant of each neuron. The dynamics of Equations 10 and 11 can be implemented in simple, neural-like circuitry. Note that the effect of minimizing E_{mem} is to simultaneously maximize (1) the similarity between the neuron voltages, V_i, and one of the stored patterns superimposed in the T_{ij} matrix (first term of E_{mem}), and (2) the similarity between the V_i and the inputs I^{mem}_i (last term of E_{mem}). (The second term of E_{mem} is the "leaky integrator term," which is unimportant for now. See Appendix.)

Since the inputs of the associative memory are to be obtained directly from the outputs of the routing circuit ($I^{\text{mem}}_i = I^{\text{out}}_i$), the control neurons, c_k, become additional dynamic variables hidden in the last term of E_{mem}. By letting the c_k follow the gradient of E_{mem}, along with the V_i, the combined associative memory/routing circuit should relax to the closest stored pattern and to the correct position and size of the window of attention simultaneously.

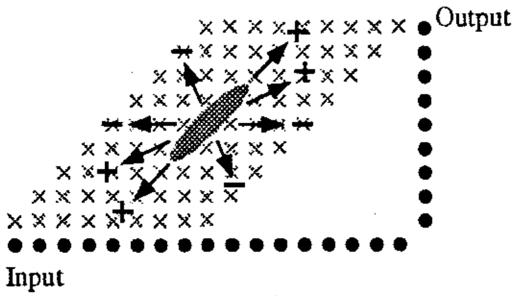

Figure 7. Control unit interactions when configured into control blocks. The control unit corresponding to the block shown (*stippled region*) should have excitatory connections ($T^c_{kl} > 0$) to other control units whose blocks form a consistent position and size of the window of attention—that is, those blocks lying along the "+" directions. Inhibitory connections ($T^c_{kl} < 0$) should be formed with control units whose blocks are inconsistent with this one—that is, those along the "−" directions. This scheme is somewhat analogous to the way constraints are imposed in the Marr/Poggio stereo algorithm (Marr and Poggio, 1976).

A dynamical equation for c_k that simultaneously minimizes both E_{mem} and $E_{\text{constraint}}$ is given by

$$c_k = \sigma(u_k), \tag{12}$$

$$\frac{du_k}{dt} + \tau^{-1}u_k = \eta \sum_i \sum_j V_i \Gamma_{ijk} I_j^{\text{in}} + \eta\beta \sum_l T^c_{kl} c_l. \tag{13}$$

(A derivation is given in the Appendix.)

A neural circuit for computing Equations 12 and 13 is shown in Figure 8. The first term on the right of Equation 13 is computed by correlating the inputs, I_j^{in}, and outputs, V_i, whose connection pathways are influenced by control unit c_k (specified by Γ_{ijk}). The other terms are computed as before. Thus, the main qualitative difference between this circuit and the "blob finder" (Fig. 6) is that the control is guided by the interaction between top-down and bottom-up signals rather than purely bottom-up sources.

In order to avoid local minima, it would be advantageous to perform the combined process of pattern matching, shifting, and scaling in a coarse-to-fine manner by utilizing information at multiple scales (e.g., Witkin et al., 1987; Buhmann et al., 1990). In this way, the low-pass information can be used initially to send the memory into the right part of its search space; the initial output of the associative memory can then be used to better refine the position and scale of the window of attention before allowing in higher-resolution information. A crude form of such a coarse-to-fine strategy has been utilized in the computer simulation below.

Shifting attention. Once an object has been recognized, the window of attention should move on to another interesting part of the scene. One way this could be accomplished would be for the control units to be self-inhibited through a delay. Thus, when a group of control units are active for some time (long enough for recognition to take place) they should begin to shut off. This will then allow other blobs or interesting items to compete successfully for control of the window of attention (see also Koch and Ullman, 1985).

Computer simulation

Figure 9 shows the results of a computer simulation of a simple attentional system for recognizing objects, based on the principles elucidated above. The network begins in blob search mode, attempting to fill the output of the routing circuit with something interesting. In Figure 9*a*, the network has settled on the "A," since it has the greatest overall brightness in the input. (Since the shapes used in this example are so compact and simple, we have bypassed the step of prefiltering them into blobs. Thus, during blob search, an object is low-pass filtered by the routing circuit itself.) After settling on a potentially interesting object, the network is switched into recognition mode and the output of the routing circuit is fed to an associative memory. Two patterns—"A" and "C"—have been previously stored in the associative memory. The blurred version of the object initially drives the inputs of the associative memory to begin the pattern search. If the position of the window of attention is slightly off, the blurred version of the object is not affected much and still sends the memory searching in the correct direction. As the associative memory converges, control units compute the correlation between memory outputs and retinal inputs and set their activation correspondingly. This tends to maximize the similarity between the outputs of the memory and the outputs of the routing circuit, which will also refine the position of the attentional window so that the high-resolution components can be properly matched (Fig. 9*b*). After allowing a fixed amount of time for the associative memory to converge (another time constant or two), the simulation states the position and presumed identity of the object. The current control state is then self-inhibited and the network switches back into blob search mode. This then puts the next interesting object at a competitive advantage in attracting the window of attention so that it may also be recognized (Fig. 9*c*,*d*).

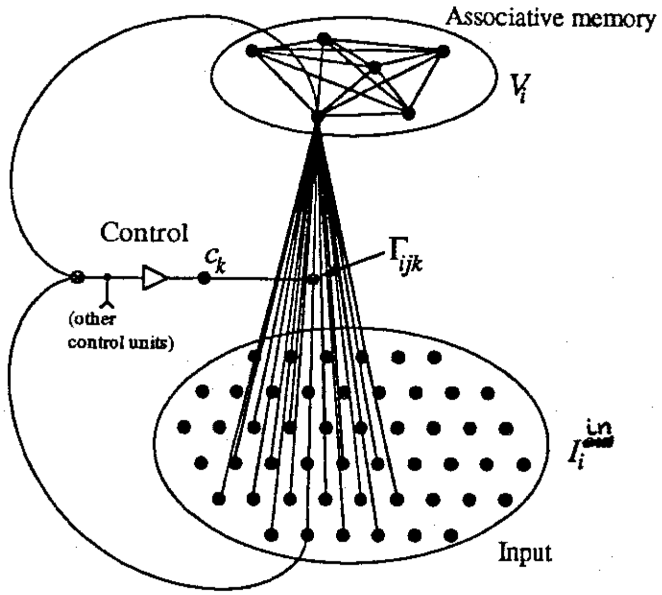

Figure 8. An autonomous routing circuit for recognition. Each node of the associative memory receives its external input from an output node of the routing circuit. Hence, each node of the associative memory has dynamic connections to many input nodes. The outputs of the associative memory are then fed back and correlated with the inputs to drive the control units.

Summary of the model

By using control neurons to modulate connection strengths dynamically, we have derived simple, neural-like circuits for shifting and rescaling the information from an input array into a higher level, object-centered reference frame. We assumed that

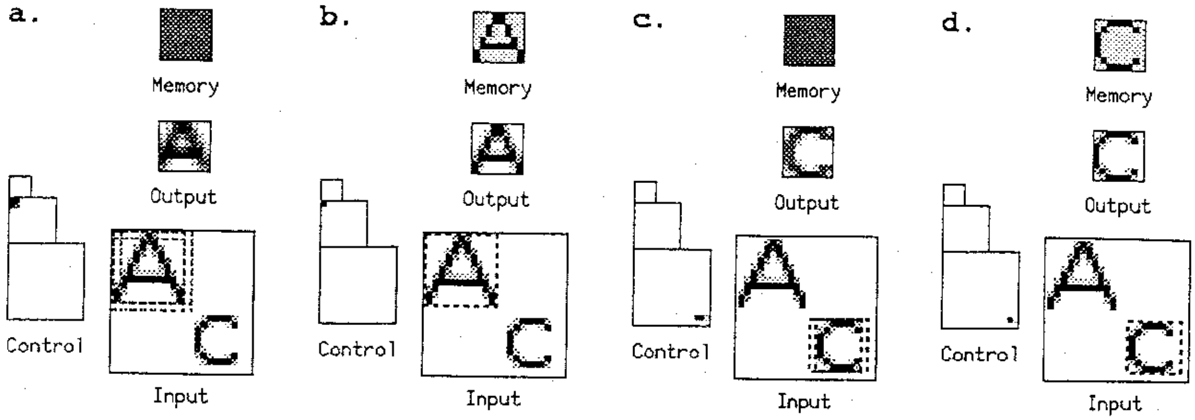

Figure 9. Computer simulation of a simple attentional system for recognizing objects. The input to the routing circuit consists of a 22 × 22 array of sample nodes and the output of the routing circuit is an 8 × 8 array of sample nodes. There are three sets of control units, each one corresponding to a different size of the window of attention [small (8×8), medium (11×11), and large (16×16)]. Each control neuron within a set corresponds to a particular position of the window of attention. The Hopfield associative memory network ("Mem output"; see Fig. 8) is composed of 64 units, fully interconnected and arranged into an 8×8 grid (i.e., one node for each output of the routing circuit). The *dashed outline* within the input array denotes the position and size of the window of attention. *a*, The network begins in blob search mode, attempting to fill the output of the routing circuit with something interesting. The blurring function of the routing circuit has been facilitated in this case by setting the constraint matrix so that neighboring positions of the window of attention only weakly inhibit each other. The network has settled on the *A* since it has the greatest overall brightness. *b*, The network is then switched into recognition mode and settles on the identification of the object. The position and size of the object are encoded in the activities of the control neurons. After a fixed amount of time, the current control state is self-inhibited and the network is switched back into blob search mode. *c* and *d*, The *C* is now at a competitive advantage in attracting the window of attention (*c*) and is subsequently recognized by the associative memory (*d*).

a useful strategy for an autonomous visual system would be to focus its attention on interesting regions within a scene and attempt to recognize whatever is there. From this basic assumption, we derived equations for governing the dynamics of the control neurons in both "preattentive" (blob search) and "attentive" (recognition) modes. Although these circuits have been greatly oversimplified for the purpose of illustration, the basic principles can be extended to larger, scaled-up routing circuits composed of multiple stages. We now turn to the issue of how such circuits may possibly be implemented in the brain.

Neurobiological Substrates and Mechanisms

Figure 10*a* shows the major visual processing centers of the primate brain. Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds through a hierarchy of visual areas that can be subdivided into two major functional streams (Ungerleider and Mishkin, 1982). The so-called "form" pathway leads ventrally through V4 and inferotemporal cortex (IT) and is mainly concerned with object identification, regardless of position or size. The so-called "where" pathway leads dorsally into the posterior parietal complex (PP), and seems to be concerned with the locations and spatial relationships among objects, regardless of their identity. The pulvinar, a subcortical nucleus of the thalamus, makes reciprocal connections with all of these cortical areas (cf. Robinson and Petersen, 1992). The following sections describe how we envision the dynamic routing circuit mapping onto this collection of neural hardware.

Cortical areas

The "form" pathway. Figure 10*b* shows the scaled-up routing circuit that we propose as a model of attentional processing in visual cortex. The different stages of the network correspond to the major cortical areas in the "form" pathway. There are two stages for V1: V1a corresponding to layer 4C, and V1b corresponding to superficial layers, since V1 has about twice the density of neurons per unit surface area as the rest of neocortex (O'Kusky and Colonnier, 1982). The remaining areas—V2, V4, and IT—occupy one stage apiece. Each node within a stage represents, in the simplest sense, a sample of image luminance. More realistically, each node would correspond to a feature vector that is represented by the activity profile on a large group (hundreds or thousands) of neurons in each visual area. For example, in V1, each group would include cells selective for various orientations, and spatial frequencies, in a small region of visual space. It is impractical at this stage to include these characteristics explicitly in our model, but we contend that these details can safely be neglected for now without losing the predictive value of the model.

The input layer of the network (V1, layer 4C) contains approximately 300,000 samples of the retinal image (~550 nodes across in one dimension). This corresponds roughly with the number of complete spatial samples delivered by the 10^6 optic nerve fibers when one takes into account the fact that information is divided into on- and off-channels, magno and parvo streams, and different spectral bands (Van Essen and Anderson, 1990). The number of nodes in the other layers is dictated by the rules specified in the previous section, given a fan-in of 1000 inputs per node (~30 inputs in one dimension). The sizes of the first four layers scale roughly with the relative sizes of each corresponding cortical area (V1 = 1120 mm^2, V2 = 1190 mm^2, V4 = 540 mm^2; Felleman and Van Essen, 1991). IT is disproportionately large, perhaps because it includes a complex of multiple areas, some of which may be devoted to specialized aspects of pattern recognition. Only a relatively small portion of IT would be required to represent the actual contents of the window of attention.

The fan-in for each node is about 1000 inputs, which is rea-

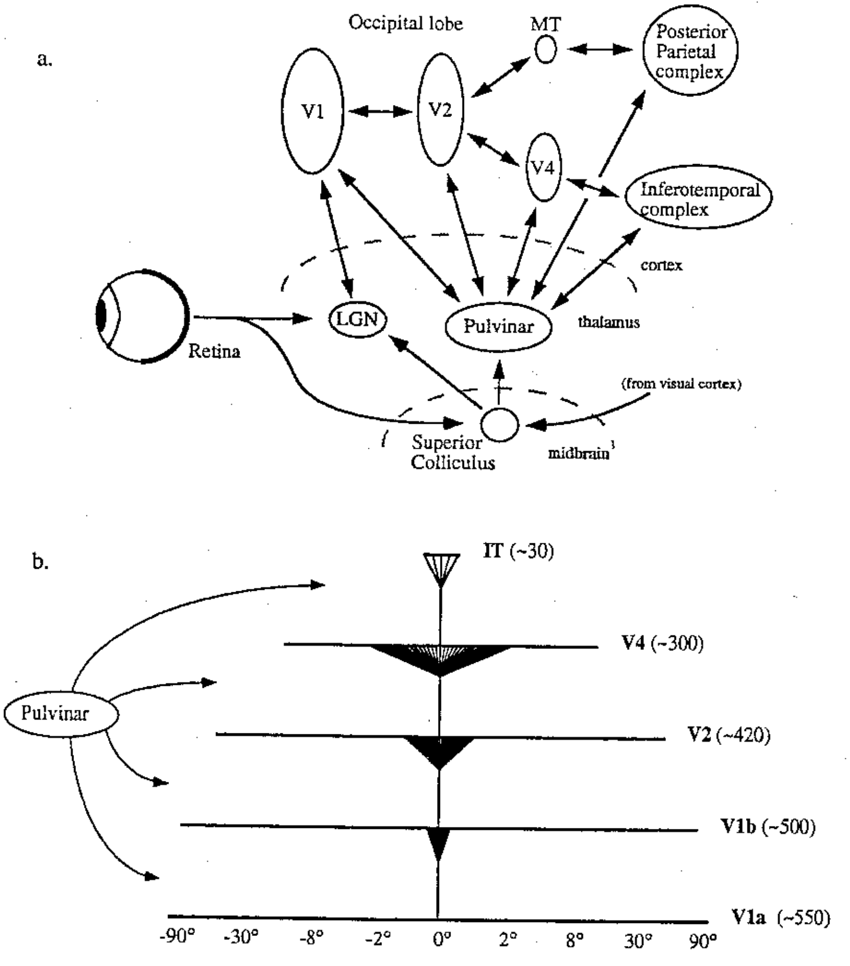

Figure 10. *a,* Major visual processing pathways of the primate brain. To avoid clutter, many known connection pathways (e.g., V4–PP) are not shown. *b,* Proposed neuroanatomical substrates for dynamic routing. The label beside each layer indicates the corresponding cortical area and the number of sample nodes in one dimension. The number of sample nodes in two dimensions is approximately the square of this number. At the *bottom* is shown a scale of the approximate eccentricity of the input nodes to the circuit. Connections are shown for the center node in each layer. (Individual nodes are indistinguishable here because of their density.) Control signals originate from the pulvinar to effectively gate the feedforward synapses.

sonable for cortical neurons (Cherniak, 1990; Douglas and Martin, 1990a). Note that without the multistage hierarchy, a fan-in of nearly 10^6 would be required for the neurons in IT, which is several orders of magnitude beyond what is neuroanatomically plausible. Also, the resulting receptive field sizes (in the all-connections-open state) are consistent with the observed increase in the size of classical receptive fields as one proceeds upward through the form pathway (Gattass et al., 1985).

The output of the network, which represents the contents of the window of attention, contains approximately 1000 sample nodes, or a window size of about 30 × 30 nodes. This then corresponds to the spatial resolution of the window of attention in our model. This estimate is roughly consistent with several lines of psychophysical evidence, including studies of spatial acuity, contrast sensitivity to gratings, and recognition (Campbell, 1985; Van Essen et al., 1991). While we certainly allow for some give and take on all of these numbers, we believe this circuit contains the essential components to explain how information can be routed from a shiftable and scalable window of attention in V1 into IT while preserving spatial relationships.

In order to better visualize the operation of this circuit, we have created a computer simulation of an "open loop" version of the model (i.e., manually controlled). Given a user-specified position and size for the window of attention, the program appropriately gates the feedforward connections at each stage in the routing circuit so that only the contents of the window of attention are routed to IT. Figure 11 shows some example outputs of the simulation when attention is focused on different items within a scene. Note that regions outside the window of attention in each cortical area are blurred, because there is no need to gate the inputs selectively to a neuron if it is not being attended to. The specific predictions generated by this circuit will be discussed in the next section.

The "where" pathway. The posterior parietal cortex (PP) is known to play an important role in attentional processes. Some studies have reported that neurons in this area show an enhanced response to attended targets within their receptive fields, even when no eye movements are made (Bushnell et al., 1981). Others have reported a threefold enhancement for *unattended* targets when the animal is in an attentive state (Mountcastle et al., 1981), or even a relative suppression for attended targets as opposed to unattended targets (Robinson et al., 1991; Steinmetz et al., 1992). These latter results suggest that PP may be representing the locations of potential attentional targets, as opposed to targets already being attended. This is also supported by lesion studies that show that damage to the parietal lobe in humans hinders the ability of other objects in the field of view to attract the attentional window away from the currently attended location (Posner et al., 1984). Thus, we propose that PP may act as a "saliency map" (e.g., Koch and Ullman, 1985), analogous to the blob map utilized in the simple attentional system described previously. These neurons would then drive the control neurons that compete for control of the window of attention.

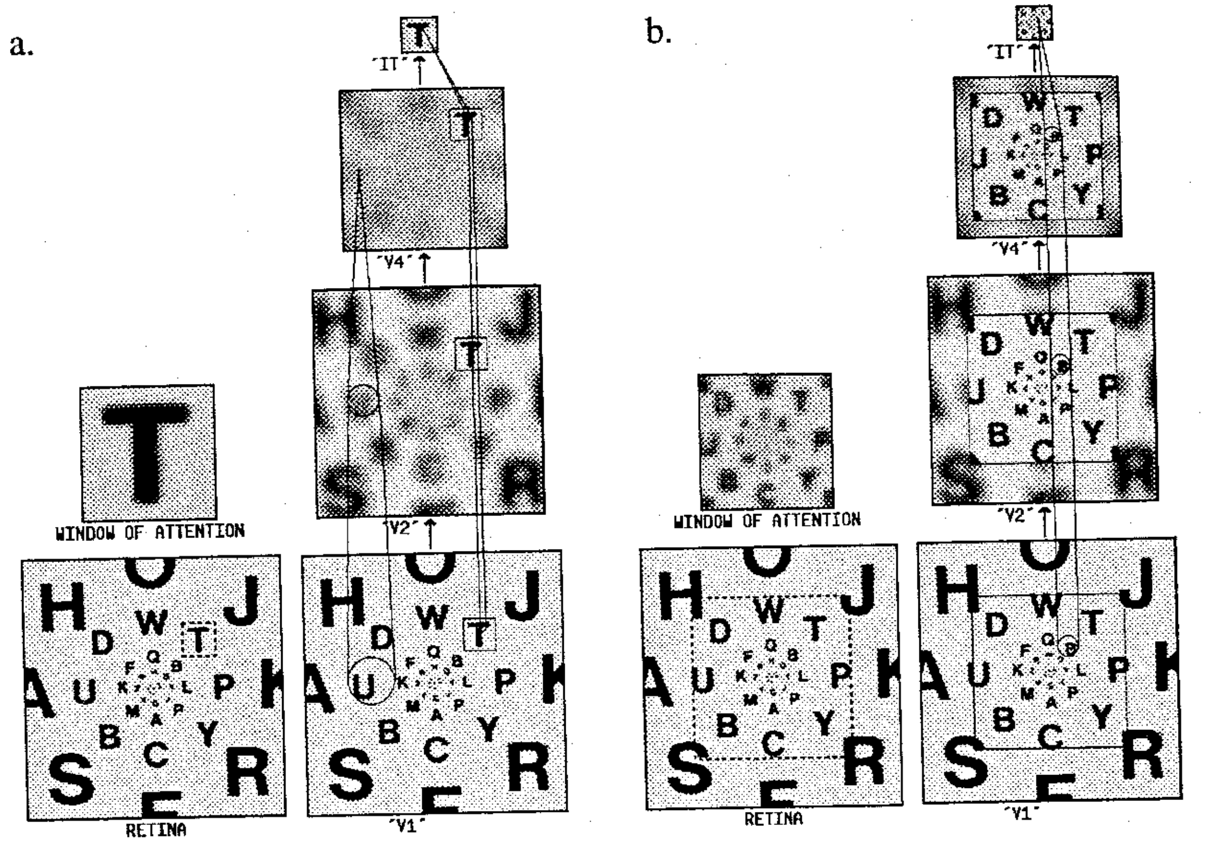

Figure 11. Computer simulation of a scaled-up, cortical dynamic routing circuit (no autonomous control). In both *a* and *b*, the *bottom left image* shows a hypothetical retinal image, and the *dashed outline* within this image indicates the position and size of the window of attention. The *image above* this shows the output of the routing circuit—the contents of the 30×30 window of attention. The *four images to the right* show four stages of the routing circuit: *V1* (essentially a copy of the retina), *V2*, *V4*, and *IT* (the output). *a*, Attention is focused on the letter *T* at highest resolution (i.e., connections between input and output are 1:1). *b*, Attention is focused on a larger region of the scene, and so resolution is sacrificed within the window of attention. In each case, the receptive field of a hypothetical IT cell is shown (small in *a* and large in *b*); in *a*, the receptive field of a V4 cell outside the window of attention is also shown. A more realistic simulation utilizing a log-polar lattice has also been constructed, but the essential predictions of the model are more easily conveyed with this simpler version of the circuit. [The image used in this example was obtained from Anstis (1974).]

This proposal contains at least two potential weaknesses, however. One possible drawback is that PP neurons typically have relatively long latencies—~100 msec (Robinson et al., 1978; Duhamel et al., 1992)—which is hard to reconcile with psychophysical data that imply that attention takes ~50 msec to move to a new location in the visual field (Nakayama and Mackeben, 1989; Saarinen and Julesz, 1991). A possible solution to this dilemma is that the superior colliculus may supplement PP by acting as a crude saliency map, but with a quicker response time due to its direct retinal input (the latency of neurons in the superficial layers of the superior colliculus is in the range of 40–50 msec; Goldberg and Wurtz, 1972). The other drawback of this proposal is that currently available anatomical data seem to offer relatively few direct pathways by which PP could influence the control neurons for modulating information flow in the "form" pathway. However, there do exist indirect pathways, such as through the superior colliculus, that may provide viable alternatives (see below).

Subcortical areas

We hypothesize that the pulvinar complex plays an important role in providing the control signals required for the routing circuit. The pulvinar is reciprocally connected to all areas in the form pathway, thus making it a plausible candidate for modulating information flow from V1 to IT. The pulvinar also receives a massive projection from the superior colliculus, which is known to encode the direction of saccade targets and may also be involved in setting up attentional targets (Posner and Petersen, 1990; Gattass and Desimone, 1991, 1992). In addition, neurophysiological studies (Petersen et al., 1985, 1987), lesion studies (Rafal and Posner, 1987; Bender, 1988; Desimone et al., 1990), and positron emission tomography studies (LaBerge and Buchsbaum, 1990; Corbetta et al., 1991) of the pulvinar suggest that it plays a role in engaging visual attention, or filtering out unattended stimuli.

A subcortical nucleus such as the pulvinar also has the important property of being spatially localized while at the same time being able to communicate with vast areas of the visual cortex. The relative proximity of pulvinar neurons to each other would facilitate the competitive and cooperative interactions among the control neurons, which are necessary to enforce the constraint of maintaining spatial relationships within the attentional window. Although it is not known whether such interactions exist among pulvinar neurons, Ogren and Hendrickson (1979) have reported the existence of interneurons with elaborate dendritic trees approaching 600 μm in diameter, which could mediate communication among pulvinar neurons. In addition, neuropharmacological experiments by Petersen et al. (1987) have shown that enhancing or depressing inhibition within the pulvinar can respectively slow down or speed up attentional shifts, which is suggestive of lateral inhibitory connections within the pulvinar. An analogous function might also be served by the reticular nucleus of the thalamus, which is an inhibitory structure through which pulvinar neurons project on their way to the cortex. One study in *Galago* (Conley and Diamond, 1990) has shown that the pulvinar projects quite diffusely into the reticular nucleus, which would be desirable for a winner-take-all type circuit.

To first order, it would make sense for each stage of the routing circuit to have its own set of control neurons. The anatomical subdivisions of the pulvinar correspond roughly with this scheme, insofar as the inferior pulvinar projects mainly to lower areas (V1, V2) and the lateral and medial pulvinar to higher areas (V4, IT). The control neurons for the lower stages would need to compete only locally, since these stages would be more concerned with making local adjustments in the position and scale of the window of attention. Control neurons at the highest stage would need to compete globally, since these stages are setting the position and scale of the window of attention for the entire scene.

The number of control neurons that would be required for the routing circuit depends on how many cortical synapses are modified by each control neuron. Theoretically, the minimal number of control neurons is given by

$$\text{\# of control neurons} = \frac{(\text{\# of output nodes}) \times (\text{fan-in per node})}{(\text{\# of synapses per control block})}.$$

Assuming that the control blocks comprise approximately 1000 synapses each, then the number of control neurons required for each stage of the routing circuit would be about the same as the number of output nodes of each stage (since the fan-in per node is also about 1000). Thus, ~250,000 control neurons would be required for the first stage, ~175,000 for the second stage, and so on, which is well within the estimated number of neurons in the pulvinar. (The pulvinar has somewhat lower neuronal density than the LGN, but also is several times larger. Since the LGN contains $\sim 10^6$ projection neurons, this would constitute a reasonable lower bound for the number of neurons in the pulvinar.) However, each output node in the circuit actually corresponds to a multitude of neurons representing various features, such as orientation, spatial frequency, and so on. Thus, each pulvinar control neuron would require an additional fan-out for controlling the inputs to all the neurons corresponding to an output node. Since there may be hundreds of neurons for each node, the pulvinar neurons would need to amplify their fan-out via other neurons (a fan-out of 100,000 for pulvinar neurons is probably too large to be plausible). This could possibly be subserved by neurons residing in the deeper layers (5 and 6) of the cortex (see Van Essen and Anderson, 1990). Control would then be implemented in a hierarchical fashion, with each pulvinar neuron specifying how information is routed between nodes, and cortical control neurons specifying how information is routed between the neurons belonging to each node.

The simple autonomous routing circuits of Figures 6 and 8 suggest an interesting role for the projections to the pulvinar from the parietal and temporal lobes and the superior colliculus. During "blob search," the pulvinar might be influenced primarily from a saliency map of targets in the parietal lobe or superior colliculus. During recognition, top-down influences from IT might then take over to refine the position and size of the attentional window for object matching. The pulvinar would then alternate between these two modes of input as attention moves from one object to the next. A potential weakness of this proposal, however, is that the anatomical evidence suggests that PP and IT project mostly to segregated portions of the pulvinar (Baleydier and Morel, 1992). On the other hand, there is some overlap near the border between the lateral and medial portions of the pulvinar where these two streams intermingle. As noted already, parietal cortex may also communicate with IT-recipient pulvinar indirectly through the superior colliculus.

An alternative means by which IT could supply top-down guidance to the control neurons would be via corticocortical feedback pathways. Under this scenario, control neurons within the cortex would be driven by feedback signals emanating from IT once the pulvinar neurons have roughly set the position and size of the window of attention. The pulvinar's role would thus be analogous to that of a general in an army—coarsely specifying a plan of action, which the cortical control neurons refine into a concise remapping under top-down, or object-based, guidance from IT.

Gating mechanisms

Neural gating mechanisms are believed to play an important role in many aspects of nervous system function. For example, the extent to which a noxious stimulus is perceived as painful varies greatly as a function of one's emotional state and other external factors. This is subserved at least in part by gating mechanisms in the spinal cord, where descending fibers from the raphe nuclei form part of a control system that modulates pain transmission via presynaptic inhibition in the dorsal horn (Fields and Basbaum, 1978). Gating mechanisms are also thought to play an important role in sensorimotor coordination; for example, there are many instances in which spinal cord central pattern generators gate sensory inputs according to the phase of the movement cycle in which the input occurs (Sillar, 1991). A somewhat different form of gating seems to take place in the LGN, where thalamic relay cells exhibit two distinct response modes: a *relay* mode, in which cells tend to replicate retinal input more or less faithfully, and a non-relay *burst* mode, in which cells burst in a rhythmic pattern that bears little resemblance to the retinal input (Sherman and Koch, 1986). In this instance, the reticular nucleus of the thalamus is thought to be the source of the signal that switches the LGN into the nonrelay burst mode.

Although there is as yet no explicit evidence for gating mechanisms in the visual cortex, there are several possible biophysical mechanisms that would allow control neurons to gate synapses along the V1–IT pathway. Presynaptic inhibition, as in the spinal cord, would probably provide the most localized gating effect. However, to date there exists no morphological evidence for this type of synapse in the visual cortex (Berman et al., 1992). Postsynaptically, a control neuron could decrease or possibly nullify the efficacy of a corticocortical synapse via shunting inhibition. Evidence for this type of mechanism playing a role in orientation or direction tuning is mixed, with some for (Pei et al., 1992; Volgushev et al., 1992) and some against (Douglas et al., 1988). Another possible postsynaptic gating mechanism could be realized via the combined voltage- and ligand-gated NMDA receptor channel, which has been shown to play an important role in normal visual function (Miller et al., 1989; Nelson and Sur, 1992). In this case, a control neuron could effectively boost the gain of a corticocortical synapse by locally depolarizing the membrane in the vicinity of the synapse. Also, there exist voltage-gated Ca^{2+} channels in dendrites (Llinas, 1988) that could provide nonlinear coupling between inputs. Evidence for nonlinear interactions of this type have been reported for synaptic inputs into layer 1 of neocortex (Cauller and Connors, 1992). All of these mechanisms, and possibly others, offer a multiplicative-type effect that is suitable for gating information flow through the cortex (see also Koch and Poggio, 1992).

Under an inhibitory gating scheme scheme, such as shunting or presynaptic inhibition, the control neurons would need to become active only when attention is actively engaged on an object. The finer the resolution desired within the window of attention, the more the control neurons would need to be engaged. The absence of any activity on the control neurons would correspond to all connections being open (the inattentive state), in which case neurons in IT would exhibit the very large receptive fields observed in anesthetized or inattentive animals (Gross et al., 1972; Desimone et al., 1984).

Under an excitatory gating scheme, such as via NMDA receptors, one would need to hypothesize the existence of a gain control mechanism working in concert with the control neurons. When no control signals are provided, cortical input would be rather weak, and the firing threshold of pyramidal cells should be lowered to let all information through. When control signals are present to boost the gain of individual synapses, however, the threshold should be raised. This way, the unboosted synapses will be essentially suppressed to a relatively low strength. Threshold adjustment could perhaps be subserved by chandelier cells, which make strong inhibitory connections exclusively onto the axon initial segment of pyramidal cells (Douglas and Martin, 1990b). Evidence that gain control mechanisms indeed exist in visual cortex has been established in previous physiological studies (Ohzawa et al., 1982; Pettet and Gilbert, 1992).

From a computational viewpoint, gating of inputs within individual dendrites provides a much higher degree of flexibility than would merely gating the outputs of pyramidal cells. Since the output of a pyramidal cell may branch to several cortical areas and make synaptic connections to a multitude of neurons, any modulation of the cell's output will simply be duplicated at all these subsequent input points. Gating inputs within the dendrites, on the other hand, allows the nonlinear computation of many intermediate results ($\Sigma_k c_k \Gamma_{ijk} I_j^{in}$) within the postsynaptic membrane, which can then be summed together within a single cell. This results in a computational structure that is orders of magnitude richer (Mel, 1992), and provides a higher degree of flexibility in sculpting patterns in connection space (see Fig. 4). We believe the demonstrable computational advantage of dendritic gating mechanisms for visual processing motivates the need to specifically look for such mechanisms experimentally. (See also Desimone, 1992, for a discussion of output vs input gating mechanisms.)

Discussion

Because of its detailed neurobiological correlates, the routing circuit model makes a number of interesting predictions that can be tested experimentally. In this section we discuss these predictions, as well as the differences between our model and other network models that have been proposed for visual attention and invariant pattern recognition. We also describe some generalizations of the model, and briefly outline the unresolved issues that remain as topics for future research.

Predictions

Neurophysiology. The most obvious prediction of the dynamic routing circuit model is that the receptive fields of cortical neurons should change their position or size as attention is shifted or rescaled. This effect should be especially pronounced in higher cortical areas. Some support for this prediction comes from the neurophysiological findings of Moran and Desimone (1985) in areas V4 and IT of primate visual cortex. As schematized in Figure 12, they found that if two bar-shaped stimuli were placed within the classical receptive field (CRF) of a V4 cell, and the animal was trained to attend to only one of them, then the cell's response to the unattended stimulus was substantially attenuated. This is what one would expect from our routing circuit, since the pathways between the cell and the unattended stimulus would be effectively disabled in this case (Fig. 12*c*). They also found that the V4 cell responded to an unattended stimulus anywhere within its CRF when the animal attended a stimulus outside the CRF. This effect is also predicted by the model, because once a V4 cell lies outside the region of interest in V4 it no longer needs to restrict its inputs (Fig. 12*d*). Indeed, other targets of V4, such as those in PP, would presumably be interested in the information from regions lying outside of the attentional beam.

While Moran and Desimone's findings offer some support for attentional modulation effects predicted by the model, they did not attempt to map receptive fields under different attentional conditions with any precision; thus, their results do not address the more specific effects predicted by the model. One would expect a cortical receptive field to shift as the attentional window is translated, and to expand or shrink as the attentional window is made larger or smaller, respectively. We predict that the optimal spatial frequency for the cell should change as well, shifting to high spatial frequency for a small window of attention, and to low spatial frequency for a large window of attention. These predictions can be tested by giving the animal a task that forces it to attend to a region of a specific size and location, and then probing the receptive field with a neutral (behaviorally irrelevant) stimulus to measure its extent. Preliminary results using such a paradigm suggest that the receptive fields of V4 cells do indeed translate toward attentional foci in or near the classical receptive field (Connor et al., 1993). In its present simple form, our model predicts that V4 receptive fields could become up to 100-fold smaller than the CRF (in one dimension) when attention is at highest resolution. While this extreme is unlikely, given the evidence for complex receptive fields in V4 (Desimone and Schein, 1987; Gallant et al., 1993), there remains a pressing need to resolve empirically the extent to which cortical receptive fields can dynamically change position and size.

Another physiological prediction of the model is that lesions to the pulvinar, the hypothesized control center, should dramatically degrade attention and pattern recognition abilities. While there is substantial evidence linking pulvinar lesions to attentional defects (Rafal and Posner, 1987; Bender, 1988; Desimone et al., 1990), some pattern recognition abilities appear to be relatively unimpaired by pulvinar lesions (Mishkin, 1972; Chalupa et al., 1976; Nagel-Leiby et al., 1984; Bender and Butter, 1987). One possible reason for the apparent sparing of pattern recognition is that the tasks used in these studies generally

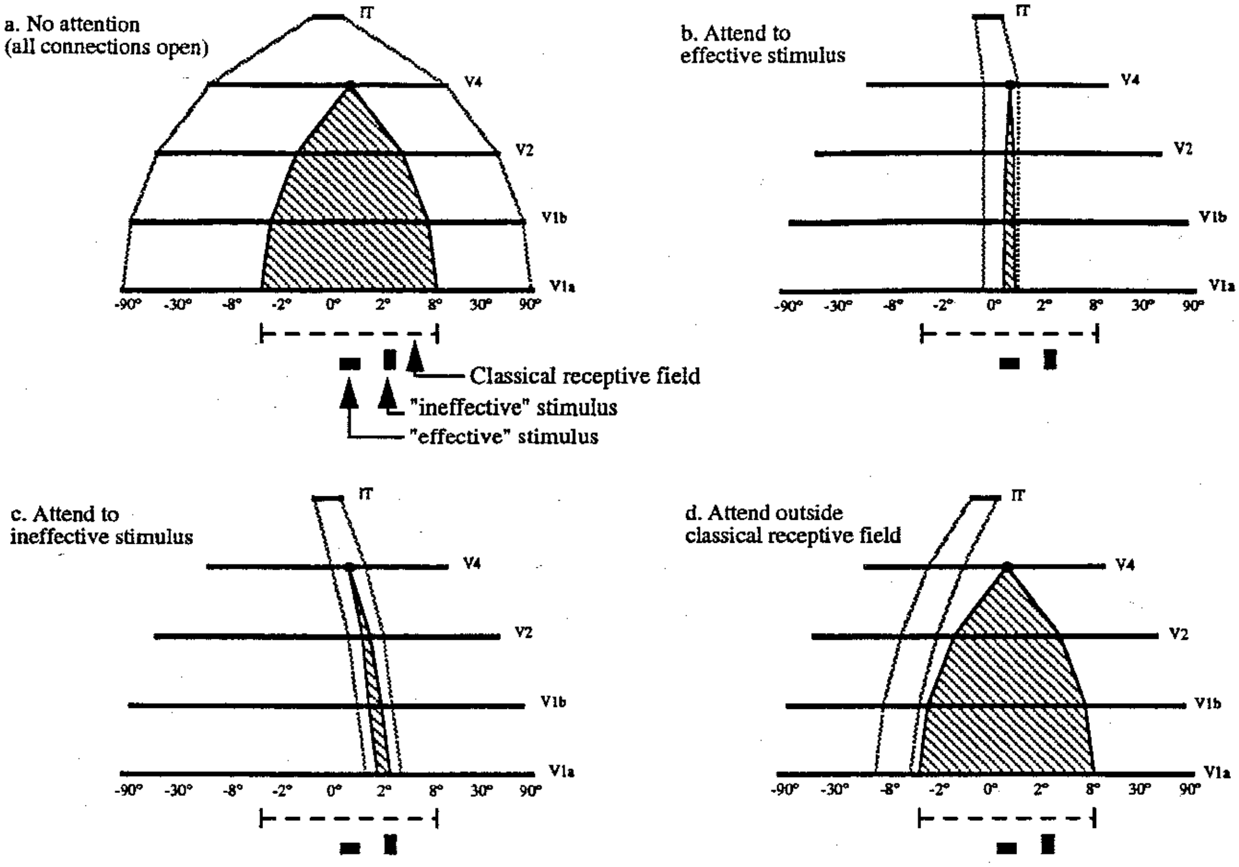

Figure 12. The dynamic routing circuit interpretation of the Moran and Desimone (1985) experiment. The node in layer V4 indicates the cell under scrutiny. The *hatched region* indicates those connections to the cell that are enabled; the others are disabled. The bounds of the window of attention in each area are shown by the *stippled lines. a,* In the nonattentive state, all connections will be open and the effective stimulus can excite the cell anywhere within its CRF. *b,* When attending to the effective stimulus, the cell's response should be unaltered since the neural pathways to the stimulus are still open. *c,* When attending to the ineffective stimulus, the cell's response should decrease substantially since the neural pathways to the effective stimulus are gated out. *d,* When attending outside the cell's CRF, there is no need to gate the cell's inputs since it is no longer taking part in the process of routing information within the window of attention.

were very simple, such as distinguishing a large "N" from a "Z" (Chalupa et al., 1976). It is conceivable that such a task could be carried out even when the fidelity of the remapping process has been compromised. A more rigorous test using stimuli that demand the full spatial resolution capacity of the window of attention would be better suited to test the effect of pulvinar lesions on recognition abilities. Pulvinar lesions would also be expected to diminish the result found by Moran and Desimone (1985), and it would be interesting to repeat this experiment while reversibly deactivating the pulvinar.

The physiological responses to be expected from pulvinar neurons depend on how they are configured to gate information flow in the cortex. In an inhibitory gating scheme, one would expect enhanced responses from pulvinar neurons projecting to areas of the cortex within and immediately surrounding the attentional beam, and little or no response from pulvinar neurons projecting to those areas of the cortex substantially outside the attentional beam. In an excitatory gating scheme, one would expect to find enhanced responses from pulvinar neurons projecting to areas of the cortex within the attentional beam only. Petersen et al. (1985) have reported such an enhancement effect for neurons in the dorsomedial portion of the pulvinar (which is connected with PP), but not in the inferior or lateral portion (which is connected to V1–IT). The lack of enhancement in these latter areas may be due to the fact that the task used in this experiment was very simple (detecting the dimming of a spot of light). Again, a more appropriate task would be one that fully taxes the capacity of the attentional window, as this would require the greatest participation from the control neurons in gating out irrelevant information.

Neuroanatomy. The routing circuit model predicts that the size of the cortical region from which a cell receives its input should increase by about a factor of 2 at each stage in the hierarchy of visual areas in the form pathway. While there is some evidence in support of this prediction—for example, connections between V4 and IT are more diffuse than connections between V1 and V2 (Van Essen et al., 1986, 1990; DeYoe and Sisola, 1991)—more accurate and higher resolution data are needed in order to confirm or contradict this prediction. Also, since the distribution of connections in the routing circuit becomes more patchy at higher levels (see Fig. 10*b*), one would expect a retrograde injection in V4 or IT to result in a patchy distribution in the lower level, which indeed has been reported (Felleman and McClendon, 1991; Felleman et al., 1992).

Another anatomical prediction of the model is that the terminations of pulvinar–cortical projections should be suitably positioned for effective modulation of intercortical synaptic strengths. The pulvinar is known to project to the output layers (2, 3) of V1 and to both the input and output layers (3, 4) of extrastriate areas V2, V4, and IT (Benevento and Rezak, 1976; Ogren and Hendrickson, 1977; Rezak and Benevento, 1979). These synapses are suspected to be excitatory since they are of the asymmetric type (in layers 1 and 2; Rezak and Benevento, 1979). However, it is not known whether the pulvinar afferents make synapses with inhibitory interneurons or directly onto the dendrites of pyramidal cells.

Finally, the model predicts that there should exist lateral inhibitory and excitatory connections within the pulvinar in order to enforce the constraint of preserving spatial relationships within the window of attention. This prediction is partially supported by the existence of interneurons within the pulvinar (Ogren and Hendrickson, 1979), but it remains to be seen if the axons of projection neurons have collaterals that spread horizontally within the pulvinar, or to what extent the reticular nucleus of the thalamus might subserve this role.

Psychophysics. The number of sample nodes in the top layer of our routing circuit is predicated on the notion that the spatial resolution of the window of attention is limited to the equivalent of about 30 × 30 pixels. This prediction shares a basic similarity to Nakayama's (1991) "iconic bottleneck" theory, although his estimate (~100 pixels total) is somewhat lower than ours. The 30 × 30 estimate is roughly consistent with several lines of psychophysical evidence, including studies of spatial acuity, contrast sensitivity to gratings, and pattern recognition (Campbell, 1985; Van Essen et al., 1991). However, one problem with this analysis is that the critical data were derived from experiments in which visual attention was not explicitly controlled. In particular, most of the experiments had display times long enough to permit multiple shifts of attention (although we doubt that this would have been a major contaminating factor in most cases).

On the other hand, those experiments that have been directed at studying the amount of "resources" allocated during visual attention have largely ignored the issue of spatial resolution. For example, various studies have reported evidence for a "zoom lens" model of attention in which the density of processing resources decreases as the size of the attentional window increases (Eriksen and St. James, 1986; Shulman and Wilson, 1987). However, these experiments were not designed to measure spatial resolution explicitly. Also, Verghese and Pelli (1992) have attempted to measure the information capacity of the window of attention, which they conclude to have an upper bound of about 50 bits. However, they studied only two tasks—detecting a nonmoving target among moving distractors, or detecting a nonflashing square among flashing squares—neither of which is well suited for measuring spatial resolution. A more appropriate experiment might be one that tested pattern discrimination ability as a function of the position, size, and resolution of an object. In this case, our present model predicts that performance would drop off sharply once the spatial frequency content of the stimulus exceeded approximately 15 × 15 cycles per object.

The model also makes some interesting predictions with regard to the dynamics of visual attention. For example, once a location has been attended to in the visual field it should be difficult to stay there or immediately revisit the site, because the control neurons corresponding to that part of the visual field would be transiently inhibited from firing. There is some evidence for such a mechanism, in that involuntary attentional fixations tend to be transient (Nakayama and Mackeben, 1989) and appear to be inhibited from return (Posner and Cohen, 1984). The amount of time that it takes the attentional window to shift from one location to another would be expected to be roughly independent of the distance between locations. Unlike eye saccades, there is no obvious reason why the control neurons should sequence through all intervening positions of the attentional window. Rather, moving the locus of attention would require merely inhibiting the current control state and activating a new one. This prediction is most consistent with Remington and Pierce's (1984) study showing time-invariant shifts of visual attention, although other studies (e.g., Tsal, 1983) are in disagreement (but see Eriksen and Murphy, 1987, for a critical commentary on these and other studies). On the other hand, if attention were actually to track a stimulus, then one would indeed expect a smooth transition of activity across the control neurons. It is interesting to note that Cavanagh (1992) has discovered some forms of visual stimuli that produce a motion percept only when tracked with attention. We speculate that the progression of activity across the control neurons is what underlies one's perception of motion in such cases.

Comparison with other models

Control versus synchronicity. A number of other models of visual attention and pattern recognition have been proposed that rely on the synchronous firing of neurons in order to change connection strengths (e.g., Crick, 1984; von der Malsburg and Bienenstock, 1986; Crick and Koch, 1990). We contend that a key disadvantage of such approaches is that information about the effective connection state at any one point in time is not explicitly encoded anywhere in the system. In our model, this information is encoded explicitly in the activities of the control neurons, which then allows it to be utilized advantageously in a number of ways.

One way that information about connectivity can be utilized is in constraining the active connections between retinal- and object-based reference frames to be in accordance with a global shift and scale transformation. This constraint is incorporated in our model via the competitive and cooperative interactions among the control neurons (Eq. 6). During object recognition, this constraint drastically reduces the number of degrees of freedom in matching points between the retinal and object-centered reference frames, because once a few point-to-point correspondences have been established, the number of potential matches between other pairs of points is greatly reduced. In machine vision, this is known as the *viewpoint consistency constraint*, and it has proved to be a powerful computational strategy for object recognition systems (Hinton, 1981b; Lowe, 1987).

Another advantage of having knowledge of the active connection state readily available is that the ensemble of control neurons together form a neural code for the current position and size of the window of attention. Therefore, information about the position and size of an object can be obtained by simply reading out the state of the control neurons. In addition, it would also be possible for the control neurons to warp the reference frame transformation in order to form object representations that are invariant to distortion (e.g., handwritten digits), in which case information about the particular shape of the object (e.g., its slant or style) could also be preserved. Note that such information is typically lost in networks that utilize feature hierarchies of complex cells (Fukushima, 1980, 1987; LeCun et al., 1990) or Fourier transforms (e.g., Pollen et al., 1971; Cavanagh, 1978, 1985) for forming position-, scale-, and/ or distortion-invariant representations.

Our model can also explain how attention may be directed "at will," or by other modalities, to the extent that those areas of the brain having access to the control neurons (such as parietal cortex) can directly influence where attention is directed. This also provides a convenient format for mediating the access to control among various competing demands. While such forms of top-down control are not impossible to incorporate in models based on synchronicity-gated connections, its implementation would seem to be less straightforward.

Control-based network models. A number of other network models of attention and recognition have also utilized the concept of control neurons for directing information flow. Niebur et al. (1993), Desimone (1992), LaBerge (1990, 1992), Ahmad (1992), and Posner et al. (1988), among others, have proposed models that involve the pulvinar as a control site for routing information from a select portion of the visual scene. In addition, Tsotsos (1991) and Mozer and Behrmann (1992) have proposed somewhat more abstract connectionist models that utilize gating units to control attention. However, none of these models preserve spatial relationships within the window of attention, which we consider to be a critical component of the routing process.

Hinton and Lang (1985) and Sandon (1990, 1988) have proposed control-based models that do preserve spatial relationships within the window of attention and share the same basic principle as the model presented here—that is, remapping object representations from retinal into object-centered reference frames via a third set of units (equivalent to control in our model). Although these models attempt to explain various psychophysical data, they do not contain the necessary level of neurobiological detail to give them strongly predictive value in biology.

Postma et al. (1992) have proposed a neural model based upon the original shifter circuit proposal (Anderson and Van Essen, 1987) to account for translational invariance in visual object priming (Biederman and Cooper, 1992). This model shares many similarities to the model presented here, including top-down (template-driven) control, but it differs in the specifics of the control structure. Most notably, Postma et al. have proposed an interesting solution to controlling a hierarchical shifter circuit based on a series of stages of local, winner-take-all circuits.

Control as a general computational strategy

Besides being advantageous for the control of visual attention, we believe that the strategy of utilizing explicit control neurons may be a useful computational principle employed by the brain in other domains as well. A different perspective of dynamic control is illustrated in Figure 13. In most neural network models, the output of a neuron is computed by forming the inner product of a weight vector, $\vec{w}$, with the inputs to the neuron, and then passing the result through a nonlinearity. The weight vector may change on a slow time scale in order to optimize the network for performing a certain task, but typically $\vec{w}$ remains fixed over the relatively short time in which the task is actually performed (e.g., <1 sec). By having control neurons available to modify $\vec{w}$ on a short time scale, the computation being carried out by the network can be dynamically reconfigured and optimized for the particular task at hand. This added degree of flexibility reduces the neural resources required for solving a complicated task, since it is no longer necessary to have dedicated, specialized networks with fixed connections to deal with each variation of a task (Van Essen et al., 1993).

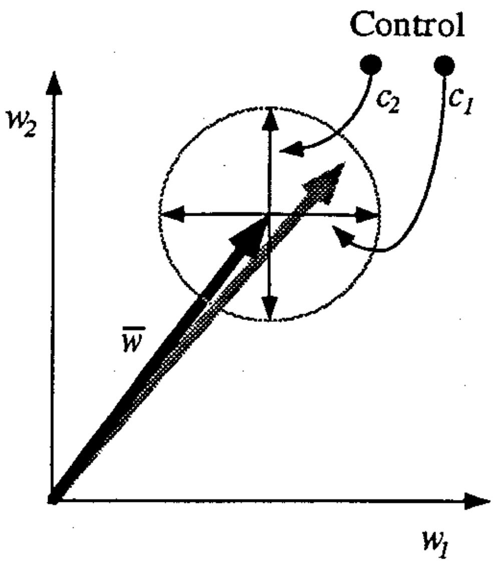

Figure 13. A more general way of viewing control. A weight vector with two components, w_1 and w_2, is shown. Control neurons c_1 and c_2 modulate each of these components, respectively, to change the weight vector dynamically. Thus, the weight vector may be able to occupy any region within the *circular outline* in order to optimize the network for the particular input and task at hand.

Unresolved issues

The dynamic routing circuit as described in this article is intended as a "zero-th order" model, and as such many details have been neglected or oversimplified. Here we outline some of the more important unresolved issues that remain as topics for future research.

Features instead of pixels. As already noted, one key neurobiological characteristic neglected in the present model is the known preponderance of feature-selective cells in the visual cortex. V1, for example, is known to contain cells tuned for various orientations and spatial frequencies, and V2 and V4 contain cells that seem to be tuned for more complex stimuli (von der Heydt and Peterhans, 1989; Gallant et al., 1993). How does this affect the routing process? One possible strategy, as mentioned earlier, would be to route information primarily from low-spatial-frequency cells when the window of attention is large, and from high-spatial-frequency cells when the window of attention is small. More generally, dynamic routing need not necessarily be restricted to the space domain, but could work across feature domains as well.

Feedback pathways. We have described how information can be routed in the feedforward pathways, but we have more or less ignored the feedback pathways that are known to exist in abundance in the visual cortex. Mumford (1992) has sketched a theory proposing that the role of these feedback pathways is to relay the interpretations of higher cortical areas to lower cortical areas in order to verify the high-level interpretation of a scene. Such a mechanism would obviously be of use for step 4 of our proposed strategy for an autonomous visual system. Under this scenario, it would be necessary to route information flow within the feedback pathways as well to ensure that the high-level interpretation is matched against the appropriate region within the cortical area below (i.e., within the window of attention). Another possible role for information flow in the feedback pathways may be to refine the tuning characteristics of lower-level cortical cells based upon the interpretations made in higher cortical areas (see, e.g., Tsotsos, 1991).

Pop-out in multiple dimensions. In the simple autonomous visual system we have proposed, "blobs" were the only salient features used to attract the window of attention. How might other salient features—such as pop-out due to motion or texture gradients—be incorporated into the preattentive system? How would the demands among these different saliencies be mediated?

Integration across multiple attentional shifts. How are the various "snapshots" obtained by the window of attention incorporated to form an overall percept of a scene? One possibility, as outlined by Hinton (1981b), is that a compact representation of each object is maintained in the form of the activities on a set of neurons within a "scene buffer." Each attentional fixation would then write its contents into a different part of the buffer, depending on the position and size of the attentional window as well as the orientation of the eyes, head, and body with respect to the environment (see also Baron, 1987).

Rotation and warp. Our model accounts for how reference frames can be shifted and rescaled, but it does not address rotation and other distortions (e.g., handwritten characters). The ability to rotate or warp reference frames could probably be included in the model without much difficulty, since this would just involve another form of routing. Moreover, for foveated objects the log-polar representation in V1 would convert rotations into approximate linear shifts on the cortex (Schwartz, 1980), which may facilitate the routing.

Three-dimensional objects. How are three-dimensional objects represented neurally, and how is information in the retinal reference frame transformed to match this representation? One possibility, as advanced by Poggio and Edelman (1990), is that three-dimensional objects are actually represented by a few characteristic two-dimensional views, and that a match to the retinal representation is achieved by interpolating among these views. In this case, the routing circuit would be required to reposition and rescale the object properly so that the interpolation could take place.

Learning. Although the model we have presented here is neurobiologically plausible in terms of the number of neurons, connectivity, and computational mechanisms required, it remains to be seen whether such a system can self-organized or fine tune itself with experience, beginning with only roughly appropriate connections. A hint as to how this may be accomplished has been described by Foldiak (1991), who has demonstrated how a complex cell can learn translation invariance using the objective function of "perceptual stability." In our model, perceptual stability would be desired in IT, and the control neurons would need to learn how to configure themselves to maintain a stable percept as an attended object moves or changes size on the retina. More generally, there is a clear need to devise learning rules for networks with control-like structures, or three-way interactions, rather than simple perceptron-type networks with two-way interactions only.

Concluding remarks

In order for us to make sense of the visual world, the brain must be capable of forming object representations that are invariant with respect to the dramatic fluctuations occurring on the retina. We have demonstrated how this feat may be accomplished by model neural circuits that are largely consistent with our current knowledge of neurophysiology and neuroanatomy. The model suggests several experiments—such as measuring attentional modulation of receptive field position and size, or measuring the spatial resolution of the window of attention—that may not have been obvious otherwise. As these experiments are carried out, the results will either help to increase our confidence in the model, or will suggest where it is wrong and how it might be revised. It is through this *combined* process of computational modeling and experimentation that we hope to understand how visual attention and recognition are actually implemented in the brain.

Appendix: Derivation of Autonomous Control Dynamics

Blob search

The total energy functional we wish to minimize is

$$E_{\text{total}} = E_{\text{blob}} + \beta E_{\text{constraint}}, \tag{A1}$$

where E_{blob} and $E_{\text{constraint}}$ are defined in Equations 5 and 6, and β is a constant determining the relative contribution of the constraint term. Letting c_k follow the gradient of this functional, we obtain

$$\begin{aligned}\frac{dc_k}{dt} &= -\eta\frac{\partial E_{\text{total}}}{\partial c_k}\\ &= -\eta\frac{\partial E_{\text{blob}}}{\partial c_k} - \eta\beta\frac{\partial E_{\text{constraint}}}{\partial c_k},\end{aligned} \tag{A2}$$

where η is a constant determining the rate of gradient descent.

As it stands, c_k is unbounded; hence E_{blob} and $E_{\text{constraint}}$ will also be unbounded and the network will not be guaranteed to converge. We can ameliorate this problem by letting c_k be a monotonically increasing function of another analog variable, u_k, that actually follows the gradient. That is,

$$c_k = \sigma(u_k), \tag{A3}$$

$$\frac{du_k}{dt} = -\eta\frac{\partial E_{\text{total}}}{\partial c_k}, \tag{A4}$$

$$\sigma(x) = [1 + \exp(-\lambda x)]^{-1}. \tag{A5}$$

This has the effect of limiting c_k to the interval [0, 1], but since we know a priori that the desired minimum of E_{blob} and $E_{\text{constraint}}$ lies in this range, the limitation does not present a problem.

Taking the derivative of E_{blob} and $E_{\text{constraint}}$ with respect to c_k yields

$$\frac{\partial E_{\text{blob}}}{\partial c_k} = -\sum_i \sum_j G_i \Gamma_{ijk} I_j^{\text{in}}, \tag{A6}$$

$$\frac{\partial E_{\text{constraint}}}{\partial c_k} = -\sum_l T_{kl}^c c_l, \tag{A7}$$

and so the dynamical equation for u_k is thus

$$\frac{du_k}{dt} = \eta \sum_i \sum_j G_i \Gamma_{ijk} I_j^{\text{in}} + \eta\beta \sum_l T_{kl}^c c_l. \tag{A8}$$

One remaining problem is that u_k must be computed via pure integration, which may cause implementation difficulties. We can convert the integrator to a more biologically plausible leaky integrator by adding to E_{total} the term

$$E_{\text{leak}} = \sum_k \int_{0.5}^{c_k} \sigma^{-1}(c)\, dc. \tag{A9}$$

The total energy functional is now defined as

$$E_{\text{total}} = E_{\text{blob}} + \beta E_{\text{constraint}} + \alpha E_{\text{leak}}, \tag{A10}$$

where the constant α determines the relative contribution of E_{leak}. [The effect of adding this term is discussed in Hopfield (1984). It essentially pushes c_k slightly away from 0 and 1.0, depending on the value of α and λ.]

Taking the derivative of E_{leak} with respect to c_k yields

$$\frac{\partial E_{\text{leak}}}{\partial c_k} = u_k, \tag{A11}$$

and so the final dynamical equation for c_k is now

$$c_k = \sigma(u_k),$$

$$\frac{du_k}{dt} + \tau^{-1} u_k = \eta \sum_i \sum_j G_i \Gamma_{ijk} I_j^{\text{in}} + \eta\beta \sum_l T_{kl}^{c} c_l, \tag{A12}$$

where the time constant τ is defined as $1/\eta\alpha$.

Recognition

Now the total energy functional is

$$E_{\text{total}} = E_{\text{mem}} + \beta E_{\text{constraint}} + \alpha E_{\text{leak}}, \tag{A13}$$

where E_{mem} is defined as in Equation 9. Note that Equation A13 is just the same as Equation A10, except with E_{blob} replaced by E_{mem}.

Taking the derivative of E_{mem} with respect to c_k yields

$$\frac{\partial E_{\text{mem}}}{\partial c_k} = -\sum_i \sum_j V_i \Gamma_{ijk} I_j^{\text{in}}, \tag{A14}$$

and so the new dynamical equation for c_k is thus

$$c_k = \sigma(u_k), \tag{A15}$$

$$\frac{du_k}{dt} + \tau^{-1} u_k = \eta \sum_i \sum_j V_i \Gamma_{ijk} I_j^{\text{in}} + \eta\beta \sum_l T_{kl}^{c} c_l. \tag{A16}$$

Note that this result is just the same as Equation A12, with the exception that G_i is replaced with V_i.

References

Ahmad S (1992) VISIT: a neural model of covert visual attention. In: Advances in neural information processing systems 4 (Moody JE, Hanson SJ, Lippman RP, eds), pp 420–427. San Mateo, CA: Kaufmann.

Anderson CH, Van Essen DC (1987) Shifter circuits: a computational strategy for dynamic aspects of visual processing. Proc Natl Acad Sci USA 84:6297–6301.

Anderson CH, Burt PJ, van der Wall GS (1985) Change detection and tracking. SPIE Vol 579–Intelligent Robots and Computer Vision, 72-78.

Anstis SM (1974) A chart demonstrating variations in acuity with retinal position. Vision Res 14:589–592.

Baleydier C, Morel A (1992) Segregated thalamocortical pathways to inferior parietal and inferotemporal cortex in macaque monkey. Vis Neurosci 8:391–405.

Baron RJ (1987) The cerebral computer. Hillsdale, NJ: Erlbaum.

Bender DB (1988) Electrophysiological and behavioral experiments on the primate pulvinar. In: Progress in brain research, Vol 75 (Hicks TP, Benedek G, eds), pp 55–65. New York: Elsevier.

Bender DB, Butter CM (1987) Comparison of the effects of superior colliculus and pulvinar lesions on visual search and tachistoscopic pattern discrimination in monkeys. Exp Brain Res 69:140–154.

Benevento LA, Rezak M (1976) The cortical projections of the inferior pulvinar and adjacent lateral pulvinar in the rhesus monkey: an autoradiographic study. Brain Res 108:1–24.

Bergen JR, Julesz B (1983) Parallel versus serial processing in rapid pattern discrimination. Nature 303:696–698.

Berman NJ, Douglas RJ, Martin KAC (1992) GABA-mediated inhibition in the neural networks of visual cortex. In: Progress in brain research, Vol 90 (Mize RR, Marc RE, Silito AM, eds), pp 443–476. New York: Elsevier.

Biederman I, Cooper EE (1992) Evidence for complete translational and reflectional invariance in visual object priming. Perception 20: 585–593.

Buhmann J, Lades M, von der Malsburg C (1990) Size and distortion invariant object recognition by hierarchical graph matching. Paper presented at the International Joint Conference on Neural Networks, San Diego, June.

Bushnell C, Goldberg ME, Robinson DL (1981) Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. J Neurophysiol 46:755–772.

Campbell FW (1985) How much of the information falling on the retina reaches the visual cortex and how much is stored in the visual memory? In: Pattern recognition mechanisms (Chagas C, Gattass R, Gross C, eds), pp 83–95. Berlin: Springer.

Carpenter G, Grossberg S (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vision Graphics Image Process 37:54–115.

Cauller LJ, Connors BW (1992) Functions of very distal dendrites: experimental and computational studies of layer I synapses on neocortical pyramidal cells. In: Single neuron computation (McKenna T, Davis JL, Zornetzer SF, eds), pp 199–229. Cambridge, MA: Academic.

Cavanagh P (1978) Size and position invariance in the visual system. Perception 7:167–177.

Cavanagh P (1985) Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In: Models of the visual cortex (Rose D, Dobson VG, eds), pp 85–95. New York: Wiley.

Cavanagh P (1992) Attention-based motion perception. Science 257: 1563–1565.

Chalupa LM, Coyle D, Lindsley DB (1976) Effect of pulvinar lesions on visual pattern discrimination in monkeys. J Neurophysiol 39:354–369.

Cherniak C (1990) The bounded brain: toward quantitative neuroanatomy. J Cognit Neurosci 2:58—68.

Conley M, Diamond IT (1990) Organization of the visual sector of the thalamic reticular nucleus in *Galago*. Eur J Neurosci 2:211–226.

Connor CE, Gallant JL, Van Essen DC (1993) Effects of focal attention on receptive field profiles in area V4. Soc Neurosci Abstr, in press.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci 11:2383–2402.

Crick F (1984) Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci USA 81:4586–4590.

Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. Semin Neurosci 2:263–275.

Desimone R (1992) Neural circuits for visual attention in the primate brain. In: Neural networks for vision and image processing (Carpenter GA, Grossberg S, eds), pp 343–364. Cambridge, MA: MIT Press.

Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J Neurophysiol 57: 835–868.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062.

Desimone R, Wessinger M, Thomas L, Schneider W (1990) Attentional control of visual perception: cortical and subcortical mechanisms. Cold Spring Harbor Symp Quant Biol 55:963–971.

DeYoe EA, Sisola LC (1991) Distinct pathways link anatomical subdivisions of V4 with V2 and temporal cortex in the macaque monkey. Soc Neurosci Abstr 17:1282.

Douglas RJ, Martin KAC (1990a) Neocortex. In: Synaptic organization of the brain (Shepard GM, ed), pp 389–438. New York: Oxford UP.

Douglas RJ, Martin KAC (1990b) Control of neuronal output by inhibition at the axon initial segment. Neural Comput 2:283–292.

Douglas RJ, Martin KAC, Whitteridge D (1988) Selective responses of visual cortical cells do not depend on shunting inhibition. Nature 332:642–644.

Duhamel J, Colby L, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. Science 255:90–92.

Eriksen CW, Murphy TD (1987) Movement of attention focus across the visual field: a critical look at the evidence. Percep Psychophys 42: 299–305.

Eriksen CW, St James JD (1986) Visual attention within and around the field of focal attention: a zoom lens model. Percept Psychophys 40:225–240.

Felleman DJ, McClendon E (1991) Modular connections between area V4 and temporal lobe area PITv in macaque monkeys. Soc Neurosci Abstr 17:1282.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Felleman DJ, McClendon E, Lin K (1992) Modular segregation of visual pathways in occipital and temporal lobe visual areas in the macaque monkey. Soc Neurosci Abstr 18:390.

Fields HL, Basbaum AI (1978) Brainstem control of spinal pain-transmission neurons. Annu Rev Physiol 40:217–248.

Foldiak P (1991) Learning invariance from transformation sequences. Neural Comput 3:194–200.

Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36:193–202.

Fukushima K (1987) Neural network model for selective attention in visual pattern recognition and associative recall. Appl Optics 26: 4985–4992.

Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. Science 259:100–103.

Gattass R, Desimone R (1991) Attention-related responses in the superior colliculus of the macaque. Soc Neurosci Abstr 17:545.

Gattass R, Desimone R (1992) Stimulation of the superior colliculus (SC) shifts the focus of attention in the macaque. Soc Neurosci Abstr 18:703.

Gattass R, Sousa APB, Covey E (1985) Cortical visual areas of the macaque: possible substrates for pattern recognition mechanisms. In: Pattern recognition mechanisms (Chagas C, Gattass R, Gross C, eds), pp 1–20. Berlin: Springer.

Goldberg ME, Wurtz RH (1972) Activity of superior colliculus in behaving monkey. I. Visual receptive fields of single neurons. J Neurophysiol 35:542–559.

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35:96–111.

Hinton GE (1981a) A parallel computation that assigns canonical object-based frames of reference. Paper presented at the Seventh International Joint Conference on Artificial Intelligence 2. Vancouver, B.C., Canada.

Hinton GE (1981b) Shape representation in parallel systems. Paper presented at the Seventh International Joint Conference on Artificial Intelligence 2. Vancouver, B.C., Canada.

Hinton GE, Lang KJ (1985) Shape recognition and illusory conjunctions. Paper presented at the Ninth International Joint Conference on Artificial Intelligence. Los Angeles.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79: 2554–2558.

Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proc Natl Acad Sci USA 81:3088–3092.

Kanerva P (1988) Sparse distributed memory. Cambridge, MA: MIT Press.

Koch C, Poggio T (1992) Multiplying with synapses and neurons. In: Single neuron computation (McKenna T, Davis JL, Zornetzer SF, eds), pp 315–345. Cambridge, MA: Academic.

Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227.

LaBerge D (1990) Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. J Cognit Neurosci 2:358–372.

LaBerge D, Buchsbaum MS (1990) Positron emission tomographic measurements of pulvinar activity during an attention task. J Neurosci 10:613–619.

LaBerge D, Carter M, Brown V (1992) A network simulation of thalamic circuit operations in selective attention. Neural Comput 4:318–331.

LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1990) Backpropagation applied to handwritten Zip code recognition. Neural Comput 1:541–551.

Llinas RR (1988) The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242:1654–1663.

Lowe DG (1987) The viewpoint consistency constraint. Int J Comput Vision 1:57–72.

Marr D (1982) Vision. New York: Freeman.

Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science 194:283–287.

Mel BW (1992) NMDA-based pattern discrimination in a modeled cortical neuron. Neural Comput 4:502–517.

Miller KD, Chapman B, Stryker MP (1989) Visual responses in adult cat visual cortex depend on *N*-methyl-D-aspartate receptors. Proc Natl Acad Sci USA 86:5183–5187.

Mishkin M (1972) Cortical visual areas and their interactions. In: Brain and human behavior (Karczmar AG, Eccles JC, eds), pp 187–208. New York: Springer.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782–784.

Mountcastle VB, Andersen RA, Motter BC (1981) The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J Neurosci 1:1218–1235.

Mozer MC, Behrmann M (1992) Reading with attentional impairments: a brain-damaged model of neglect and attentional dyslexias. In: Connectionist approaches to natural language processing (Reilley RG, Sharkey NE, eds), pp 409–460. Hillsdale, NJ: Erlbaum.

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251.

Nagel-Leiby S, Bender DB, Butter CM (1984) Effects of kainic acid and radiofrequency lesions of the pulvinar on visual discrimination in the monkey. Brain Res 300:295–303.

Nakayama K (1991) The iconic bottleneck and the tenuous link between early visual processing and perception. In: Vision coding and efficiency (Blakemore C, ed), pp 411–422. Cambridge: Cambridge UP.

Nakayama K, Mackeben M (1989) Sustained and transient compounds of focal visual attention. Vision Res 29:1631–1647.

Nelson SB, Sur M (1992) NMDA receptors in sensory information processing. Curr Opinion Neurobiol 2:484–488.

Niebur E, Koch C, Rosin C (1993) An oscillation-based model for the neuronal basis of attention. Vision Res, in press.

Ogren MP, Hendrickson AE (1977) The distribution of pulvinar terminals in visual areas 17 and 18 of the monkey. Brain Res 137:343–350.

Ogren MP, Hendrickson AE (1979) The structural organization of the inferior and lateral subdivisions of the *Macaca* monkey pulvinar. J Comp Neurol 188:147–178.

Ohzawa I, Sclar G, Freeman RD (1982) Contrast gain control in the cat visual cortex. Nature 298:266–268.

O'Kusky J, Colonnier M (1982) A laminar analysis of the number of neurons, glia, and synapses in the visual cortex (area 17) of adult macaque monkeys. J Comp Neurol 210:178–290.

Palmer SE (1983) The psychology of perceptual organization: a transformational approach. In: Human and machine vision (Beck J, Hope B, Rosenfeld A, eds), pp 269–339. Orlando: Academic.

Pei X, Volgushev M, Creutzfeldt O (1992) A comparison of directional sensitivity with the excitatory and inhibitory field structure in cat striate cortical simple cells. Perception 21 [Suppl 2]:26.

Petersen SE, Robinson DL, Keys W (1985) Pulvinar nuclei of the behaving rhesus monkey: visual responses and their modulation. J Neurophysiol 54:867–886.

Petersen SE, Robinson DL, Morris JD (1987) Contributions of the pulvinar to visual spatial attention. Neuropsychologia 25:97–105.

Pettet MW, Gilbert CD (1992) Dynamic changes in receptive-field size in cat primary visual cortex. Proc Natl Acad Sci USA 89:8366–8370.

Pitts W, McCulloch WS (1947) How we know universals: the perception of auditory and visual forms. Bull Math Biophys 9:127–147.

Pollen DA, Lee JR, Taylor JH (1971) How does the striate cortex begin the reconstruction of the visual world? Science 173:74–77.

Poggio T, Edelman S (1990) A network that learns to recognize three-dimensional objects. Science 343:263–266.

Posner MI, Cohen Y (1984) Components of visual orienting. In: Attention and performance, X, control of language processes (Bouma H, Bouwhuis DG, eds), pp 531–556. Hillsdale, NJ: Erlbaum.

Posner MI, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci 13:25–42.

Posner MI, Walker JA, Friedrich FJ, Rafal RD (1984) Effects of parietal injury on covert orienting of attention. J Neurosci 4:1863–1874.

Posner MI, Petersen SE, Fox PT, Raichle ME (1988) Localization of cognitive operations in the human brain. Science 240:1627–1631.

Postma EO, van den Herik HJ, Hudson PTW (1992) The gating lattice: a neural substrate for dynamic gating. Paper presented at CNS*92, San Francisco, CA, July.

Rafal RD, Posner MI (1987) Deficits in human visual spatial attention following thalamic lesions. Proc Natl Acad Sci USA 84:7349–7353.

Remington R, Pierce L (1984) Moving attention: evidence for time-invariant shifts of visual selective attention. Percept Psychophys 35: 393–399.

Rezak M, Benevento LA (1979) A comparison of the organization of the projections of the dorsal lateral geniculate nucleus, the inferior pulvinar and adjacent lateral pulvinar to primary visual cortex (area 17) in the macaque monkey. Brain Res 167:19–40.

Robinson DL, Petersen SE (1992) The pulvinar and visual salience, Trends Neurosci 15:127–132.

Robinson DL, Goldberg ME, Stanton GB (1978) Parietal association cortex in the primate: sensory mechanisms and behavioral modulations. J Neurophysiol 41:910–932.

Robinson DL, Bowman EM, Kertzman C (1991) Convert orienting of attention in Macaque. II. A signal in parietal cortex to disengage attention. Soc Neurosci Abstr 17:442.

Saarinen J, Julesz B (1991) The speed of attentional shifts in the visual field. Proc Natl Acad Sci USA 88:1812–1814.

Sandon PA (1990) Simulating visual attention. J Cognit Neurosci 2:213–231.

Sandon PA, Uhr LM (1988) An adaptive model for viewpoint-invariant object recognition. Paper presented at the 10th Annual Conference of the Cognitive Science Society, Montreal, Canada, August.

Schwartz EL (1980) Computational geometry and functional architecture of striate cortex. Vision Res 20:645–669.

Sherman SM, Koch C (1986) The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. Exp Brain Res 63:1–20.

Shulman GL, Wilson J (1987) Spatial frequency and selective attention to local and global information. Perception 16:89–101.

Sillar KT (1991) Spinal pattern generation and sensory gating mechanisms. Curr Opin Neurobiol 1:583–589.

Steinmetz MA, Connor CE, MacLeod KM (1992) Focal spatial attention suppresses responses of visual neurons in monkey posterior parietal cortex. Soc Neurosci Abstr 18:148.

Treisman A (1988) Features and objects: the fourteenth Bartlett memorial lecture. Q J Exp Psychol 40A:201–237.

Tsal Y (1983) Movements of attention across the visual field. J Exp Psychol [Hum Percept] 9:523–530.

Tsotsos JK (1991) Localizing stimuli in a sensory field using an inhibitory attention beam. Technical report RBCV-TR-91-37, Department of Computer Science, University of Toronto.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behavior (Ingle DJ, ed), pp 549–586. Cambridge, MA: MIT Press.

Van Essen DC, Anderson CH (1990) Information processing strategies and pathways in the primate retina and visual cortex. In: An introduction to neural and electronic networks (Zornetzer SF, Davis JL, Lau C, eds), pp 43–72. New York: Academic.

Van Essen DC, Newsome WT, Maunsell JHR, Bixby JL (1986) The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. J Comp Neurol 244:451–480.

Van Essen DC, Felleman DJ, DeYoe EA, Olavarria J, Knierim J (1990) Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey. Cold Spring Harbor Symp Quant Biol 55:679–696.

Van Essen DC, Olshausen B, Anderson CH, Gallant JL (1991) Pattern recognition, attention, and information bottlenecks in the primate visual system. In: Proceedings of the SPIE Conference on Visual Information Processing: from neurons to chips, Vol 1473 (Mathur BP, Koch C, eds), pp 17–28. Burlingame, WA: SPIE.

Van Essen DC, Anderson CH, Olshausen BA (1993) Dynamic routing strategies in sensory, motor, and cognitive processing. In: Large scale neuronal theories of the brain (Koch C, Davis J, eds), in press. Cambridge, MA: MIT Press.

Verghese P, Pelli DG (1992) The information capacity of visual attention. Vision Res 32:983–995.

Volgushev M, Pei X, Vidyasagar TR, Creutzfeldt OD (1992) Orientation-selective inhibition in cat visual cortex: an analysis of postsynaptic potentials. Perception 21 [Suppl 2]:26.

von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. J Neurosci 9:1731–1763.

von der Malsburg C, Bienenstock E (1986) Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain. In: NATO ASI series, Vol F20, Disordered systems and biological organization (Bienenstock E, Fogelman Solie F, Weisbach G, eds), pp 247–272. Berlin: Springer.

Witkin AP, Terzopoulos D, Kass M (1987) Signal matching through scale space. Int J Comput Vision 1:133–144.