An inhibitory beam for attentional selection

John K. Tsotsos

A scheme to solve the problem of selective visual attention using hierarchical winner-take-all circuits was proposed by Koch and Ullman (1985). Their idea has played a prominent role in many current theories of visual attention. However, their scheme imposes unreasonable computational restrictions and leads to an architecture that is not as biologically plausible as once believed. This presentation will propose a new version of winner-take-all networks that solves the problem of locating and localizing items in the visual field and shows how to implement the idea of an inhibitory attentional beam. The scheme is based on the foundation laid by Koch and Ullman, but incorporates several novel changes and additions which permit a proof of convergence with constant time convergence properties, address the issue of saliency maps and the binding across representations, and includes much tighter comparisons to the biology. The results are directly applicable to any sensory field and are not restricted to vision.

1. Introduction

The space of neural responses may be considered as a huge, multi-dimensional search space, where each point represents the strength of response of one hypothesis existing in space and in time. More than one hypothesis exists at each point in space and time, each one competing with the rest to describe the perceptual objects or events at that point. The goal of visual processing then could be to find the set of strongest hypotheses such that they are most consistent with the input and the reasons for considering this particular input in the first place. The identification of this maximal hypothesis set is the goal of the search task. The simplest, parallel method of maximum selection is perhaps the winner-take-all (WTA) scheme proposed by Feldman and Ballard (1982). WTA is an iterative, neurally-plausible, and very simple mechanism to find a unique maximum, which is based on mutual

inhibition among units that are all connected to one another[1]. However, it must be emphasized that the space is not small and the search problem is potentially intractable (Tsotsos, 1989). A WTA on its own does not offer a plausible solution simply because it would have to operate over too large a search space; connectivity requirements could not be realized nor could convergence be guaranteed if arbitrary weights are permitted. Further, since the maximal hypothesis set is the target of the search, a unique hypothesis does necessarily exist.

Koch and Ullman (1985) proposed a particular form of the winner-take-all process for shifts in selective visual attention. The key points of their proposal are: 1) a number of features are computed and represented in parallel in different topographical maps; 2) a selective mapping exists from these representations into a central non-topographic representation such that this central representation contains the properties of only a single location in the visual scene at any time; 3) conspicuity of locations in the topographic maps is the major determinant of selection and selection is realized by a winner-take-all network; 4) inhibition of the selected item will cause an automatic shift towards the next most conspicuous location. Additional rules for the computation in step 3 are proximity and similarity preferences; they are not detailed nor implemented however. The Koch and Ullman WTA idea has played a role in many views and theories of attention and explanations of attentional phenomena, including those of Anderson and Van Essen (1987), Nakayama and Mackeben (1989), and Fuster (1990).

Koch and Ullman reject a general WTA process based on mutual inhibition: they point out that the connectivity requirements make the scheme biologically implausible (each unit must be connected to all others, thus for an n-unit field, n^2 connections are required for the WTA process). Further, they argue, the scheme is not guaranteed to converge if arbitrary weights and initial configurations are permitted. These problems led them to a proposal for implementation of the method that involved a new inhibition rule and a tree of intermediate nodes. They are correct on both arguments; however, they may have tried to solve far too general a task. There are number of problems with the Koch and Ullman scheme:

(*a*) Winner-take-all networks possess a troublesome feature: because they are based on mutual inhibition, the winning unit has a final response strength that is attenuated. Thus, although the location of the winner may be known, its value is not meaningful. If the value is to be passed on to a central representation as they suggest, some additional mechanism is then required to recover that value. Further, the winning value is not representative of the items selected in visual space because

[1] It is a special case of general relaxation labeling procedures (Hummel and Zucker, 1983) as is the new version of the updating rule to be presented later in this paper.

it was initially computed using the non-attended receptive field. A re-computation is required for this and such a re-computation was not included.

(*b*) The proposal requires the WTA to act as a gating function on the saliency map; however, the saliency map combines features into a conspicuity measure and thus does not represent any feature explicitly. No other mechanism is proposed to achieve the proper gating action.

(*c*) Koch and Ullman separate selection from input processing, that is, the input is first processed in whatever manner is appropriate by individual units (such as line detectors) and then selection takes place. It seems that this separation may be artificial and inappropriate given recent results from single-cell recordings in awake and behaving primates (such as (Moran and Desimone, 1985)), where the behavior of a single cell makes it appear as if the receptive field changes properties as a result of attentive selection. Thus, a scheme that integrates attentive selection with interpretive processing may be more biologically plausible.

(*d*) Conspicuous locations need not be single locations, but may be regions. Koch and Ullman do not permit multiple winners in their algorithm.

(*e*) The intermediate tree of computations of very small branching factor which is necessitated by the connectivity and convergence properties of the WTA rule does not seem to have an immediate biological counterpart. There is no evidence for a tree in the strict computational sense; however, the computations do seem to be hierarchical in nature.

(*f*) A side-effect of using a tree of intermediate WTA computations over very small networks is that a shift of attention requires time proportional to the number of branches of the tree that must be traversed upwards from the previous focus and then downwards to the new focus of attention and this traversal distance is related to topographic distance. Koch and Ullman use the results of Shulman, Remington and McLean (1979) and of Tsal (1983) as part of the justification for their shifts of attention requiring time proportional to topographic distance. Those authors found that attention shifts with constant velocity passing through and processing intermediate locations as it moves. However, Remington and Pierce (1984) show that distance has no effect on attention shifts; there is no attentional gradient. They further point out a very important constraint: Efficient coordination with the saccadic eye movement system in reading or visual search tasks would dictate rapid, time-invariant movements to match saccade dynamics. To complicate the matter even more, Ericksen and Murphy (1987) claim that prior experimental evidence is conflicting, with assumptions dubious and data interpretation problematic. They say that attention shifts

pose an "open problem". More recently, Kröse and Julesz (1989) found no proximity effect and conclude that a parallel scheme is needed to find prospective locations which are then checked by a slow serial process.

Further analysis and a new approach seem to be required.

2. An inhibitory attentional beam

Complexity analysis of the problem of visual search has led to several constraints that are important for the design of a vision system (Tsotsos, 1990; Tsotsos, 1988; Tsotsos, 1989). Note that the collection of constraints below form a sufficient but not necessary set:

1 parallelism of sufficiently high degree;

2 hierarchical organization through the abstraction of prototypical visual knowledge in order to cut search time at least logarithmically;

3 localization of receptive fields, noting that the physical world is spatio-temporally localized and that objects and events, and their physical characteristics, are not arbitrarily spread over time and space;

4 using the observation that not all visual stimuli require all possible parameter types for interpretation, separable, logical maps permit selection of individual maps as required;

5 hierarchical abstraction of the input token arrays so as to maintain semantic content yet reduce the number of elements;

6 tokens of visual parameters at high resolution cannot be directly accessed, but must be obtained by the tuning of computing units and through the input abstraction hierarchy;

7 predictions for the overall configuration of the visual system in terms of lower bounds on the size and number of maps and upper bounds on the required degree of parallelism;

8 an inhibitory attentional beam.

Most of these constraints are consistent with the increasing literature on single-cell recordings in striate and extrastriate visual cortex from awake and behaving primates (Bushnell et al., 1991; Fuster and Jervey, 1981; Braitman, 1984; Moran and Desimone, 1985; Mountcastle et al., 1987; Haenny and Schiller, 1988; Haenny et al., 1988; Spitzer et al., 1988; Maunsell et al., 1988; Motter, 1988; Fuster, 1988, 1990; Andersen et al., 1990). Also, the constraints on structure are consistent with primate visual cortex neuroanatomy (Felleman and Van Essen, 1991). Most importantly however, complexity analysis quantitatively confirms what has been widely believed for a long

time: selective attention is a major contributor to reducing the amount of computation. The nature of attention is to inhibit the portions of a sensory field that are not selected.

Constraints (5) through (8) require some elaboration. The beam concept is a result of a connectivity argument: low spatial resolution at the highest levels of the visual hierarchy and fixed connectivity suggests attentive access of visual information through the input abstraction hierarchy. If it is true that the sizes of visual maps are small, then complete connectivity across a higher level map is not biologically infeasible as Koch and Ullman claimed. In humans, it is known that V1 has on average 2100 hypercolumns per hemisphere (Stensaas et al., 1974) and it is known from primate studies that the higher areas are smaller yet. If a hypercolumn represents processing units of all types and all constrained to the same visual space[2], and adjacent hypercolumns represent processing in physically adjacent visual space, then it is not unreasonable to consider the spatial resolution of a visual map in the cortex to be determined by the number of hypercolumns it contains. Thus, across a visual map, each column in the map would require only a small number of additional connections in order to implement a completely connected network for a winner-take-all process.

Finally, the common idea of an attentional spotlight does not permit any way for the message of selection to reach the input items that are in fact selected. This, and the previously described considerations led to the concept of an inhibitory attentional beam, introduced in (Tsotsos, 1990). This beam is rooted at the top of an abstraction pyramid so that at the top level there is a spotlight selecting a single item (or items). The beam illuminates the sub-pyramid whose apex is defined by the spotlight at the top and has a simple internal structure consisting of a "pass zone" and an "inhibit zone". The pass zone of the beam is required for obvious reasons: this is the pathway through the pyramid that is selected for further processing. The inhibit zone is needed because if a single unit is selected at the top level of the hierarchy, then it has a receptive field which is potentially very large, much larger than the selected item (it would correspond exactly to the sub-pyramid of the beam). The sensory stimuli within the receptive field but which are not selected may be regarded as noise interfering with the processing of the selected items. The "signal-to-noise" ratio is greatly enhanced if the unwanted stimuli are attenuated or eliminated. Selection leads to processing of the selected input as if it were the only stimulus in the sensory field. Desimone et al. (1990) have concluded that attention may not be necessary if there is only one item in the visual field. Within the beam explanation, the reason why they and others observe no attentional effect in their experiments with single items is clear. Although the beam mechanism

[2] This is known to be true at least for V1 and MT in primates (see Albright and Desimone, 1987).

may always be active, its effect on a field of only one stimulus cannot be observed since the beam seeks to achieve a single stimulus visual field.

It is important to distinguish the beam concept in this paper from that of Posner (1980) and from the feature integration theory of Treisman and colleagues (1988). Posner proposes an attentional system that enhances performance in a spatially restricted region. Treisman on the other hand, believes that attention is required to integrate features. The experimental paradigms within which each of these ideas was developed differ from one another (see Briand and Klein, 1987 for detailed discussion on this). More to the point, neither concept has been detailed sufficiently so as to offer a mathematical framework and an implementation or circuit model. The beam in this paper refers in principle to the fact that attention must operate in the three-dimensional structure of the brain and in practice to the circuits that provide downward flowing information within the visual processing hierarchy, and not simply the two-dimensional region of visual space on which the spotlight shines as Posner defines it[3] and as Treisman uses it.

3. Realization of the inhibitory beam

As described earlier, the basic idea proposed by Koch and Ullman is useful and the new scheme in this paper borrows much from it. It motivated the basic processing algorithm shown below as Algorithm 1. However, it has been modified so that multiple items in the input and time-varying input can be handled. Also, the final stage of the algorithm is not simply the routing of information as Koch and Ullman claim, but rather a recomputation using only the stimuli that were found as "winners" at the input level of the hierarchy.

```
1.  receive stimulus at input layer (continuously over time)
2.  receive task guidance at top layer (asynchronously, as it
       is available)
3.  do  4 through 8 forever
4.      compute output representation with task guidance if
            available or without if unavailable
5.      apply inhibitory beam using the WTA hierarchy and
            task guidance
6.      re-compute output representation
7.      extract selected output item
8.      inhibit item(s) at input that represented the
        selected item if unchanged
```

[3] "Attention can be likened to a spotlight that enhances the efficiency of detection of events within its beam"(Posner, 1980).

Algorithm 1.

The sequence of actions in the above algorithm leads to a timing of processing actions that is consistent with the observations and hypotheses of Fuster (1990) and of Desimone (1991, personal communication). The major steps of this algorithm will be described in more detail in several stages, beginning with the presentation of a new winner-take-all updating rule plus proof of convergence.

3.1. WTA updating rule

The updating rule presented by Koch and Ullman must be replaced so that the updating function does not suffer from the problems described earlier. The new winner-take-all process is non-destructive for the winner. That is, the winning units will maintain its actual response strength while the other units decay in strength. It is accomplished using a simple observation: if the inhibitory signal is based on the response differences, then an implicit but global ordering of response strengths is imposed on the entire network on the basis of pair-wise local information. The largest item will thus not be inhibited at all, but will participate in inhibiting all other units. The smallest unit will not inhibit any other units but will be inhibited by all. The new function which controls the WTA process for a given unit is:

$$s_i' = R\left[s_i - \frac{1}{\sum_{j\in M, j\neq i} w_{ij}}\left(\sum_{j\in M, j\neq i} w_{ij} R\left[s_j - s_i\right]\right)\right] \tag{1}$$

where R is a rectifying function ($R[x] = x$ if $x > 0$, otherwise $R[x] = 0$), $0 < w_{ij} \leq 1$, $0 \leq s_i \leq 1$, and M is the set of locations on the sensory map with non-zero strength. This can be enhanced to include a "variance" threshold and a scale parameter as follows:

$$s_i' = R\left[s_i - \frac{1}{\sum_{j\in M, j\neq i} w_{ij}}\left(\sum_{j\in M, j\neq i} w_{ij}\Delta_{ij}\right)\right] \tag{2}$$

if $0 < \Theta_i < s_j - s_i$ and $((x_i - x_j)^2 + (y_i - y_j)^2)^{\frac{1}{2}} < C$, then $\Delta_{ij} = s_j - s_i$

$$\text{else } \Delta_{ij} = 0$$

where C specifies a priori scale, Θ is the variance threshold, x_k, y_k is the location in the sensory field represented by unit s_k. The variance threshold could reflect the acceptable variation across an item. That is, responses could vary from one another within this threshold and still be considered as arising from the same physical stimulus. The scale parameter could place a bound on the sphere of influence of a given unit; units that are farther than C units from s_i would not affect s_i in any way.

Proposition: *The WTA updating rule of Equation (1) is guaranteed to converge for all inputs provided* $0 < w_{ij} \leq 1, 0 \leq s_i \leq 1$.

Proof: Let $\frac{1}{\sum_{j \in M, j \neq i} w_{ij}} \left(\sum_{j \in M, j \neq i} w_{ij} R[s_j - s_i] \right)$ be termed the contribution to a unit.

1 Since the contribution to unit i depends on a difference function, an ordering of units is implicitly imposed depending on their response magnitude.

2 Unit j will inhibit unit i only if s_j is larger than s_i. Thus, the largest units (a unique maximum is not required) will have a contribution of 0 and will remain unaffected by the iterative process.

3 All other units will have strictly positive contributions and thus will decay in magnitude or remain at zero.

4 The rectifying function guarantees that no unit receives an updated value that is negative and thus oscillations cannot occur.

5 The iterations are terminated when a stable state is reached (no units change in magnitude).

It is thus trivially shown that the process is guaranteed to converge and locate the largest items in the sensory field.

It is important that the convergence properties of this scheme be investigated. From the updating function, it is clear that the time to convergence depends only on three values: the magnitude of the largest unit, the magnitude of the second largest unit and the parameter Θ. The largest unit, recall, is not affected by the updating process at all. The largest unit, however, is the only unit to inhibit the second largest unit. The contribution term for all other units would be larger than for the second largest because those units would be inhibited by all larger units. This, along with the fact that they are smaller initially, means that they would reach the lower threshold faster than the second largest unit. Convergence is achieved when all units but one decay in value to Θ; therefore the time to convergence is determined by the time it takes the second largest element to reach this value. This makes the convergence time independent of the number of units in the WTA process. The amount of inhibition per iteration of the updating rule for the second largest unit s_2 where the largest unit is s_1 is given by equation 2 which simplifies for this situation to :

$$s_2' = 2s_2 - s_1$$

s_1 is constant. Convergence is achieved when $s_2' \leq \Theta$. At the kth iteration, $s_2^k = 2^k s_2^0 - (2^k - 1)s_1$. Convergence will thus require $\log_2((s_1 - \Theta)/(s_1 - s_2^0))$ iterations. A bound on this number of iterations is desirable. Arbitrarily

small differences between values are not allowed; the differences must be at least Θ, so the denominator of the logarithm can be no smaller than Θ. s_1 can be no larger than 1.0. Thus, the upper bound on the number of iterations is given by:

$$\log_2\left(\frac{1.0-\Theta}{\Theta}\right)$$

If convergence is required within K iterations, then the following expression gives the appropriate value of Θ that will guarantee the convergence:

$$\Theta = \frac{1}{2^K+1}$$

This tacitly assumes that $s_1 > \Theta$. If $K = 1, \Theta = 0.333$ and s_1 may not be large enough in some situations; moreover such a large theta may not be sensible given that it is a variance threshold for responses caused by the same physical stimulus. A "gain" parameter will solve this problem. Re-define the contribution to include a gain parameter $A, A \geq 1$, so that:

$$s_i' = R\left[s_i - \frac{A}{\sum_{j\in M, j\neq i} w_{ij}}\left(\sum_{j\in M, j\neq i} w_{ij}\Delta_{ij}\right)\right]$$

In this case, at the kth iteration, $s_2^k = (1+A)^k s_2^0 - ((1+A)^k - 1)s_1$ Convergence will require $\log_{1+A}((s_1+\Theta)/(s_1 - s_2^0))$ iterations. Using the same argument as above, if convergence is required within K iterations, then the following expression gives the appropriate value of Θ that will guarantee the convergence:

$$\Theta = \frac{1}{(1+A)^K+1}$$

Knowledge of the magnitudes of values with which the WTA process works will help determine appropriate values of A. Large gain amplifiers may be more expensive to realize in neural circuitry, as they are in silicon, and this may be an important constraint on the acceptable gain magnitudes. Of course, no gain is needed at all if the expense of more iterations is permissible. If 5 iterations are possible given the time constraints, Θ can be set to 0.03, for example, and gain control may not be necessary. In general, if the allowable variance is known and the maximum number of permissible iterations is given, then the gain may be set as:

$$A = \left(\frac{1-\Theta}{\Theta}\right)^{\frac{1}{K}} - 1$$

Constant time convergence of all WTA processes no matter their size can be guaranteed throughout the hierarchy. This is possible only because the iterative update is based on differences of units and thus only the largest and second largest values need be considered; a two-unit network is thus easy to characterize.

3.2. Pruning a hierarchy of winner-take-all processes

Next, assume a hierarchical representation where units are represented by their response strength. Connectivity from layer to layer need not be fixed and each layer (indeed, each unit) may have different connectivity patterns including overlap. WTA processes are set up for the top layer as a whole, and for each set of inputs for each unit everywhere else in the hierarchy as follows.

Suppose that the top layer of the visual processing hierarchy is completely connected in such a fashion so that a winner-take-all process operates across the entire visual field: it could compute the global winner. It could be autonomous and could accept guidance for areas or stimulus qualities to favor if that guidance were available but would operate independently of such guidance otherwise. This global winner would be represented by a receptive field that is necessarily very large since it is at the top of the visual processing hierarchy. In order to then localize the global winner in the sensory field, a hierarchy of WTA processes are activated as a result of the global winner. The global winner activates a WTA that operates only over its direct inputs. This would select the winner within the global receptive field. In this way, all of the branches of the hierarchy that do not contribute to the winner would be pruned from the search space. This pruning idea is then applied recursively to as many successively lower layers as is reasonable. The end result is that from a globally strongest response, the cause of that largest response is localized in the sensory field at the earliest levels.

Complete connectivity required for the top level is biologically feasible for representations of the size predicted by the complexity level analysis and for the visual areas where hypercolumn counts have been done as described earlier. Complete connectivity only within a receptive field for all other levels also does not violate connectivity constraints if each receptive field has a small number of thousands of inputs. This structure may be realized by inserting a layer of independent WTA gating networks in between each pair of layers as shown in Figure 1. There are two types of units in this hierarchy at each layer k (layers denoted by subscripts):

interpretive units (I_k) which receive visual input from below and perform processing related directly to the interpretation of that input (color, edges, motion for example); and,

gating units (G_k) which compute the winner-take-all result for a particular interpretive unit and gate input through to the next higher interpretive units.

It is hypothesized that these units exist as appropriate assemblies of neurons rather than as single neurons.

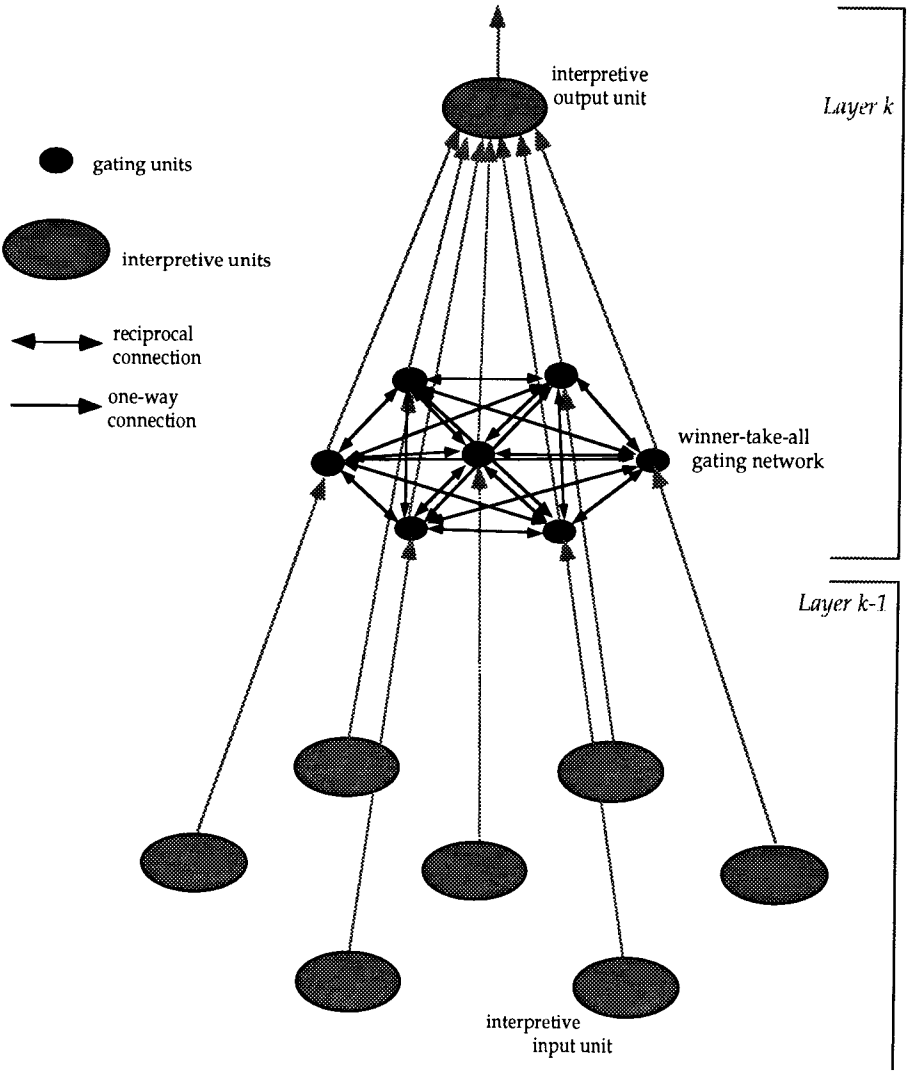

Fig. 1. *The interpretive input hierarchy with WTA processes across the inputs to each interpretive unit. In this example, the function of the gating units is to inhibit "non-winning" signals at each level, effectively pruning the hierarchy and implementing the inhibitory beam.*

An example of the implemented beam process in operation will be shown. Response strength of each node is computed in the same manner. Response equals the sum of all inputs: the examples shown thus code unweighted luminance. This simple demonstration generalizes easily to the case where response of a unit is given by any weighted-sum function of its inputs; these need not be uniform throughout the hierarchy but may be different at each

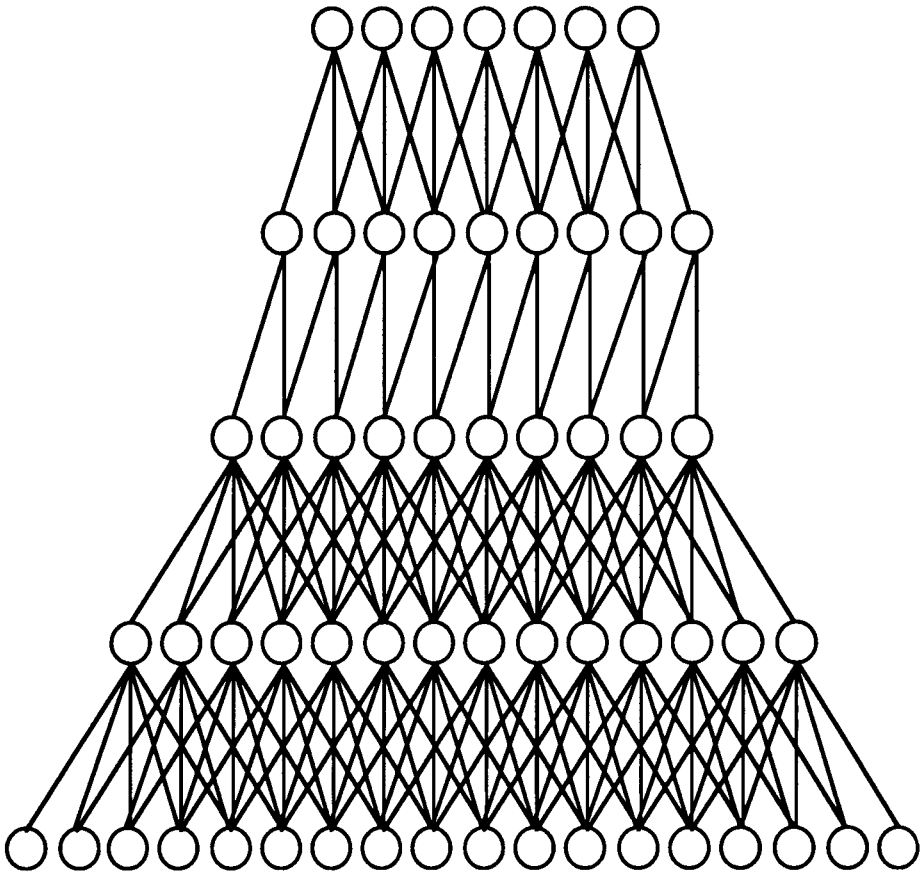

Fig. 2. *The sample hierarchy of interpretive units (the gating units are not shown) with an input layer, 3 intermediate layers and an output layer. This and following diagrams are coded as follows: solid circles are units which are active; solid lines are connections which are active; gray circles are inactive units and dashed lines are connections that are inhibited by the beam. Missing lines show pathways which are "don't cares".*

level. Although one dimensional sensory fields are shown, the extension to two or higher dimensional fields is trivial.

Figure 2 gives the basic hierarchy on which the process is demonstrated with only the interpretive units displayed (gating networks are omitted to clarify the figure). Connectivity may vary between levels and the number of levels may vary. The first figure shows the representation after the first application of step 4 of Algorithm 1 without task guidance, the next figure shows the representations after step 5 of Algorithm 1, and the last two figures show the representation after the second and third iterations through step 5.

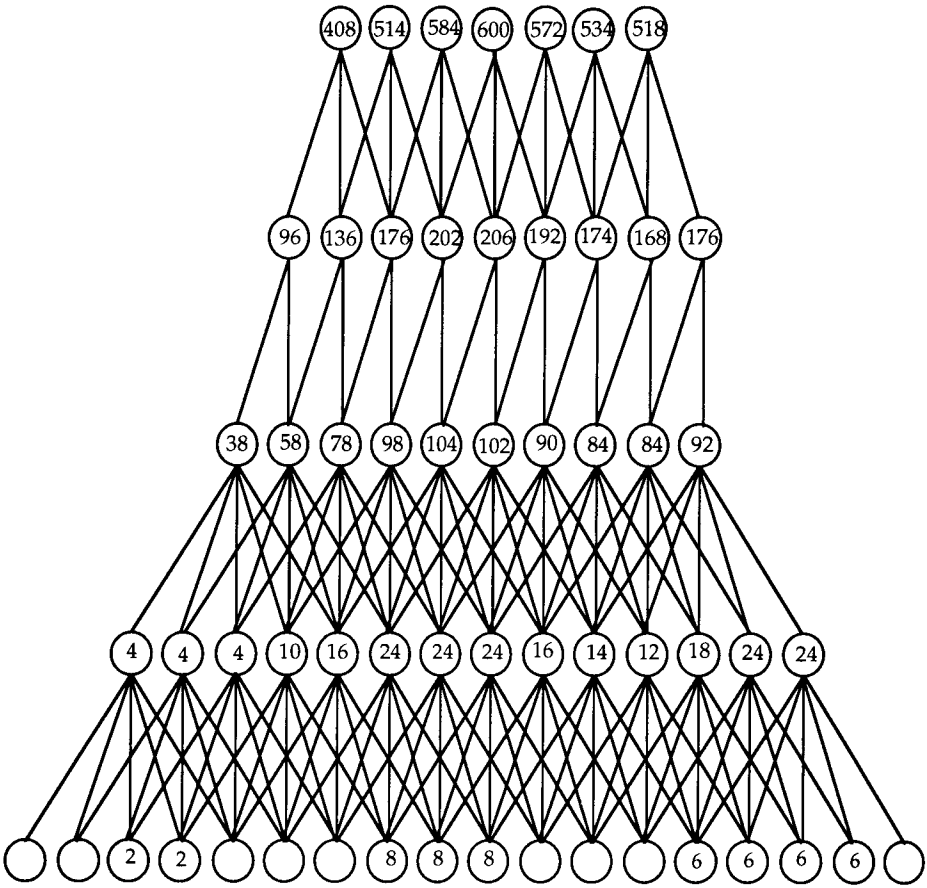

Fig. 3. *This is the initial configuration for the first example. If there is no attentional selectivity, the input, consisting of three items (a pair of '2's, a trio of '8's, and a quartet of '6's) would lead to the given output layer with the simple computation proposed for this example.*

If a simple stimulus pattern is applied to the input layer, as is shown in the input layer, the remaining nodes of the hierarchy will compute their responses based on a summation of their inputs resulting in the configuration of Figure 3. The first pass of the WTA scheme is shown in Figure 4. It is clear that the largest item is found and that the overall action is precisely the inhibitory beam that was proposed in (Tsotsos, 1990). It must be emphasized that this form of selection permits the response at the top of the hierarchy to be determined using only the data selected at the early stages of the hierarchy: it is as if only the selected data were present in the sensory field and any other conflicting stimuli did not exist. Once this first stimulus item is processed, it is inhibited as Koch and Ullman proposed and the next is selected (Figure 5 and then Figure 6 for the third item).

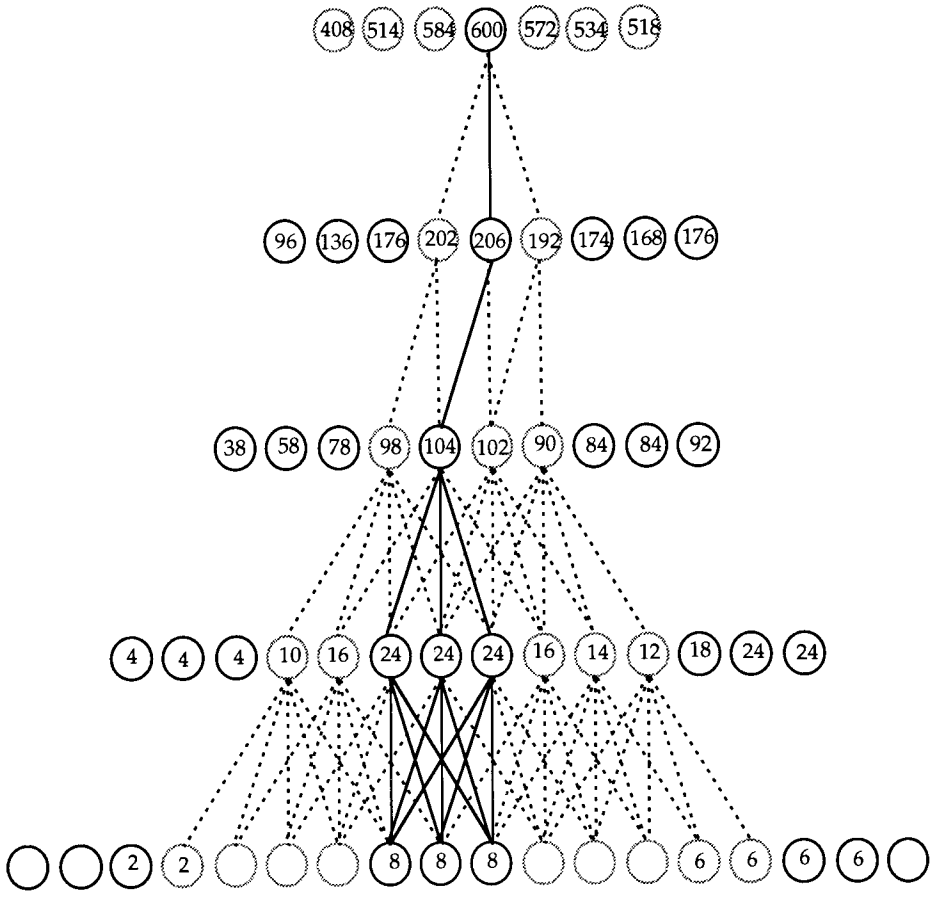

Fig. 4. *The first ("brightest") item is selected and any pathways leading to the same receptive fields in which the item is found are inhibited. The beam's pass and inhibit zones are clear. This, and successive examples showing the beam are snapshots taken during the computation at the point just after the beam is applied but before re-computations are performed on the selected items within the beam's pass zone.*

Culhane and Tsotsos (1992) show many more examples of the implemented process for real images, using both brightness and edge computations in the interpretive units.

3.3. Propagation of inhibitory effects and of task guidance

The hierarchy as shown in Figure 1 is now augmented by a parallel, interlaced hierarchy of beam units (B_k) which provide top-down guidance for visual selection, whether the selection be for regions in space or sub-ranges of visual feature to attend to (see Figure 7). As well, the beam hierarchy

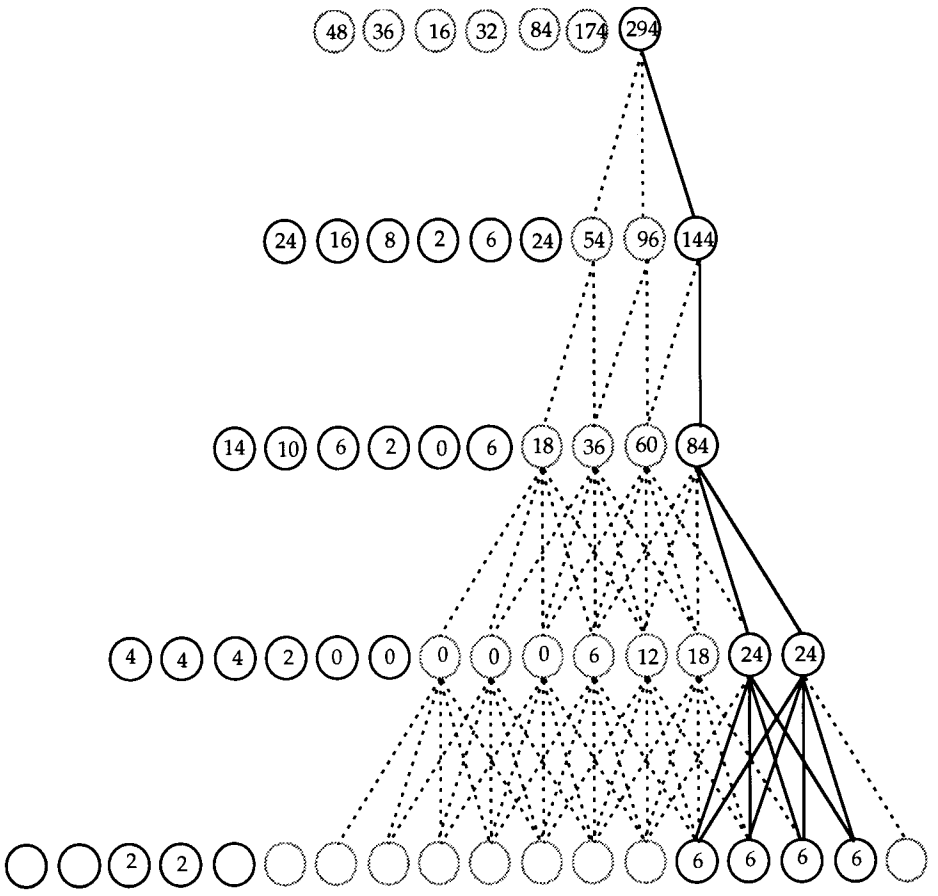

Fig. 5. *The first selected item is inhibited, the hierarchy is re-computed and the second brightest item is selected.*

provides an important conduit for upward-flowing information not directly associated with the interpretive computations, such as location information. Gating units may be directed by beam units and also provide output to the next layer of beam units downward. A brief overview of the function of the gating and beam units will be presented here and details can be found in (Tsotsos, 1991).

The top-down guidance can take several forms. It may simply select the type of computation that is permitted to pass through, for example, 45-degree oriented lines versus all orientations of lines. Or, it may both select and enhance responses from the selected feature type. In tests with the winner-take-all algorithm, it has been observed that the latter leads to enhancement of responses as observed in Spitzer, Desimone and Moran (1988) for attended units. Much more experimentation is required to discover the nature of such task-specific guidance signals.

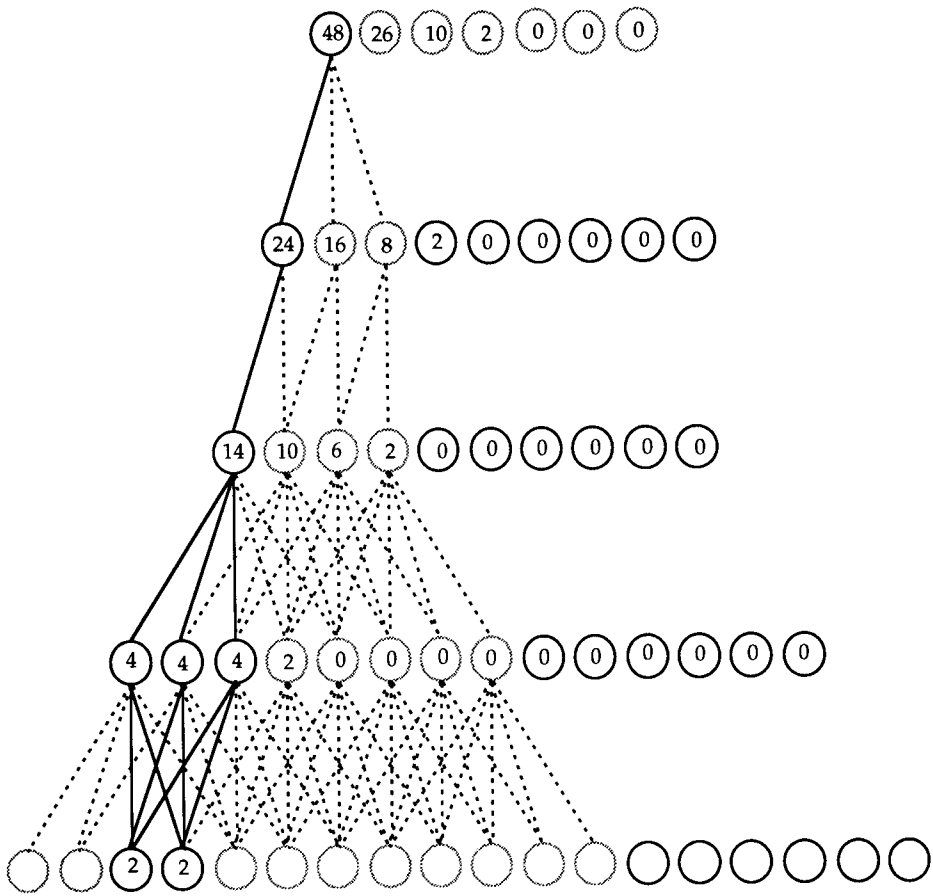

Fig. 6. *The second selected item is inhibited, the hierarchy is re-computed and the final item is selected.*

Each beam unit receives signals from many gating units as well as top-down task selection information. The algorithm which controls the selection process is presented in (Tsotsos, 1991). The important actions of the algorithm can be illustrated by considering what effect overlapping receptive fields have on the function of the beam hierarchy. Figure 8 shows an example of overlap.

Here, several problems must be considered. There is potential for ambiguity in three ways, each related to a different signal that is propagated along the hierarchy. Beam units receive three signals: task specific directions, inhibitory signals from the the gating units and location information from winning units below, and propagate these signals to higher or lower parts of the beam hierarchy as appropriate. The confluence of inhibitory magnitudes is handled by imposing a "min" function, that is the inhibition that is

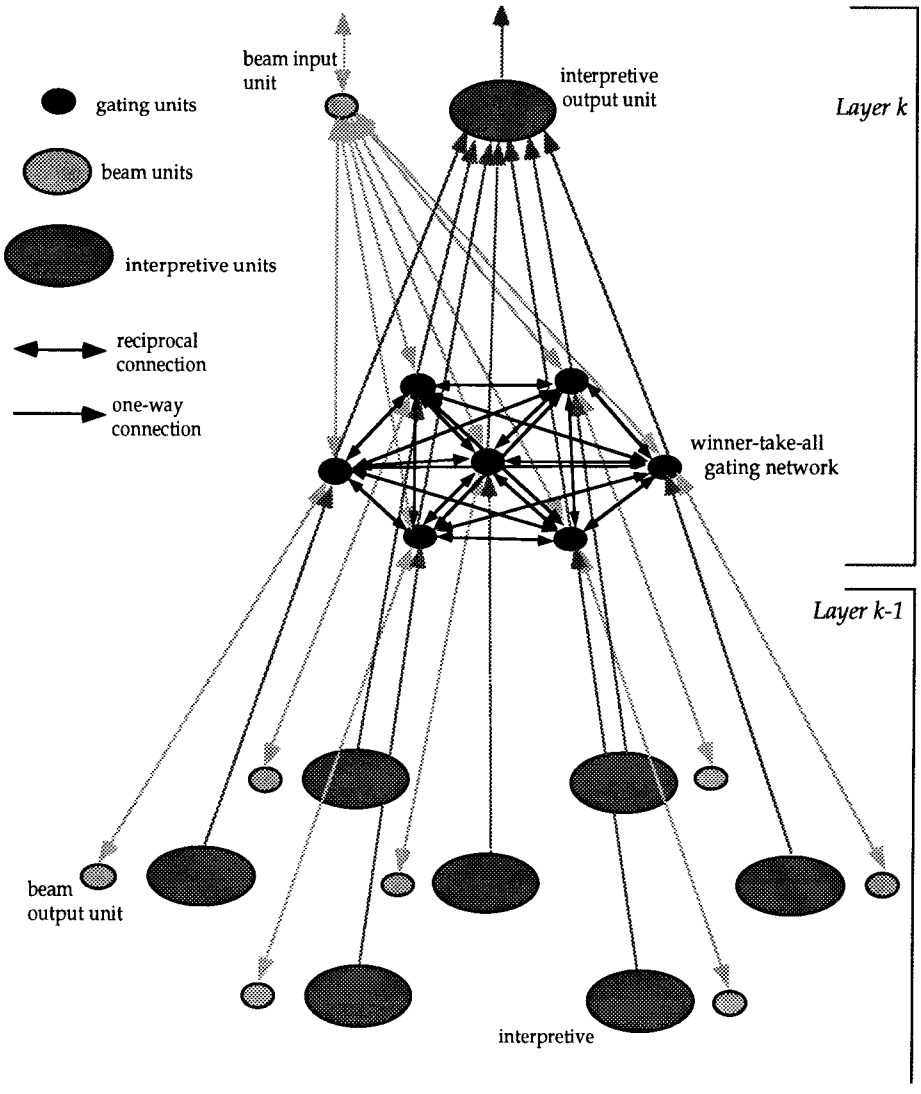

Fig. 7. *Two adjacent layers in the interpretive visual hierarchy. The inhibitory beam hierarchy is added to the structure in Figure 1.*

passed on downwards is the strongest of that received from all sources. This is perhaps the simplest resolution to the problem of confluence of inhibitory signals. Simple sums, weighted sums, products of differences would not lead to meaningful results or would not ensure that the inhibitory value is between 0 and 1.0. The top-down task guidance signals are passed through an "AND" function. That is, all task-specific signals must be satisfied by gating units below their point of confluence in order for the signal to pass. In this way, it is ensured that any pathways above the point of confluence

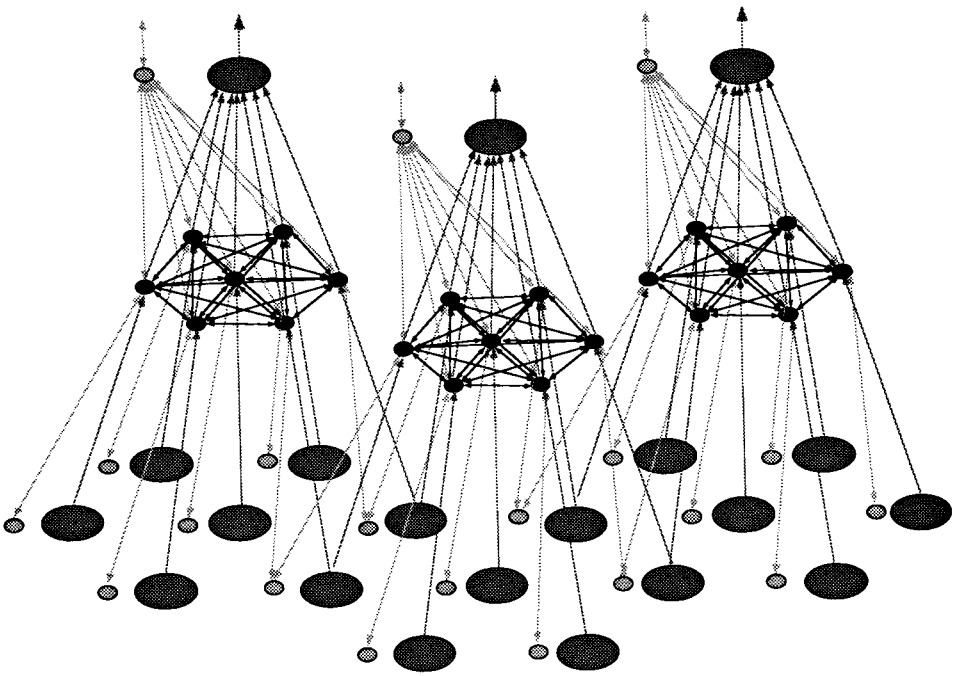

Fig. 8. *The downward flowing information from the beam units at the higher level have more than one target beam unit at the next lower level due to the fact that interpretive units have overlapping receptive fields.*

have their task requirements satisfied. Finally, the location signals are put together into a list creating a list of signals which can be decoded as described in (Tsotsos, 1991). This strategy for dealing with the confluence of downward flowing information is equally valid for confluence from multiple beams as for confluence within a beam.

4. Summary

The implementation of the inhibitory attentional beam has a number of important properties that make it preferable to the Koch and Ullman scheme both as a hypothesis of biological attentional mechanisms and also as a computational tool for attention in machine sensing systems:

- the scheme integrates attentive selection with interpretive processing resulting in behavior that is consistent with single-cell recordings in tasks requiring attention both with respect to changes in receptive field structure and temporal response changes;
- WTA process is provably convergent and requires constant time irrespective of the locations or numbers of the sensory items, consistent with current views on the scanpath of attentional fixations;

- the method permits multiple winners within a single representation;
- the intermediate layers of computations can be mapped directly onto the hierarchy of areas in the visual cortex;
- no single saliency map that combines features into a conspicuity measure is required;
- multiple beams are permitted across different representations of visual information;
- receptive fields of many sizes are permitted at each level with overlap and algorithms are provided to deal with the potential ambiguities;
- task knowledge can provide scale selection plus feature salience.

The result is a very fast, automatic, independent, continuous and reactive system.

In Felleman and Van Essen (1991) the pathways in the macaque visual cortex are described in great detail, including the fact that among the 32 visual areas there are 121 reciprocal-pair connections[4]. What could the functional significance for these paired connections be? The proposal in this paper requires such paired connections (see Figure 7). The layers of computation described in this proposal may be likened to visual areas. Thus, the question can be re-phrased as: could the implementation of an inhibitory attentional beam across visual areas be one of the functionalities of some of these reciprocal connections?

Acknowledgements

Allan Jepson and Sean Culhane provided useful discussion. The author is the CP-Unitel Fellow of the Canadian Institute for Advanced Research. This research was funded by the Information Technology Research Center, one of the Province of Ontario Centers of Excellence, the Institute for Robotics and Intelligent Systems, a Network of Centers of Excellence of the Government of Canada, and the Natural Sciences and Engineering Research Council of Canada.

[4] There are 32 areas with 305 known pathways connecting them. Of these pathways, 121 are known to be reciprocal pairs (one for each of top-down and bottom-up directions), 5 are known to be singlets (one direction only), and 58 pathways are not being critically tested.