

T Nathan Mundhenk · Laurent Itti

Computational modeling and exploration of contour integration for visual saliency

Received: 19 April 2004 / Accepted: 21 April 2005 / Published online: 26 August 2005
© Springer-Verlag 2005

Abstract We propose a computational model of contour integration for visual saliency. The model uses biologically plausible devices to simulate how the representations of elements aligned collinearly along a contour in an image are enhanced. Our model adds such devices as a dopamine-like fast plasticity, local GABAergic inhibition and multi-scale processing of images. The fast plasticity addresses the problem of how neurons in visual cortex seem to be able to influence neurons they are not directly connected to, for instance, as observed in contour closure effect. Local GABAergic inhibition is used to control gain in the system without using global mechanisms which may be non-plausible given the limited reach of axonal arbors in visual cortex. The model is then used to explore not only its validity in real and artificial images, but to discover some of the mechanisms involved in processing of complex visual features such as junctions and end-stops as well as contours. We present evidence for the validity of our model in several phases, starting with local enhancement of only a few collinear elements. We then test our model on more complex contour integration images with a large number of Gabor elements. Sections of the model are also extracted and used to discover how the model might relate contour integration neurons to neurons that process end-stops and junctions. Finally, we present results from real world images. Results from the model suggest that it is a good current approximation of contour integration in human vision. As well, it suggests that contour integration mechanisms may be strongly related to mechanisms for detecting end-stops and junction points. Additionally, a contour integration mechanism may be involved in finding features for objects such as faces. This suggests that visual cortex may be more information efficient and that neural regions may have multiple roles.

1 Introduction

In the visual world there are many things which we can see, but certain features, sets of features and other image properties tend to more strongly draw our visual attention toward them. A very simple example is a stop sign, in which combinations of red color and angular features of an octagon combine with a strong word “stop” to create something that hopefully we would not miss if we come upon it. Such propensity of some visual features to attract attention defines in part the phenomenon of visual saliency. Here we assert, as others (James 1890; Treisman and Gelade 1980; Koch and Ullman 1985; Itti and Koch 2001) that saliency is drawn from a variety of factors. At the lowest levels, color opponencies, unique orientations and luminance contrasts create the effect of visual pop-out (Treisman and Gelade 1980; Wolfe et al. 1998). Importantly, these studies have highlighted the role of competitive interactions in determining saliency – hence, a single stop sign on a natural scene backdrop is usually highly salient, but the saliency of that same stop sign and its ability to draw attention is strongly reduced as many similar signs surround it. At the highest levels it has been proposed that we can prime our visual processes to help guide what we wish to search for in a visual scene (Wolfe 1994; Miniussi et al. 2002; Navalpakkam and Itti 2002). Given the organization of visual cortex it has also been proposed that saliency is gathered into a topographic saliency map. This is a landscape of neurons in partnership and competition with each other. For instance, neurons that are most excited have the greatest ability to competitively suppress their neighbors. This creates a winner-take-all phenomenon, whereby the strongest and most unique features in an image dominate other features to become salient. However, in addition to direct uniform center-surround competition, it has been suggested by several studies that saliency is enhanced when a series of elements like the dashed lines on a road are aligned in a collinear fashion (Braun 1999; Li and Gilbert 2002; Peters et al. 2003). Such a phenomenon is part of what is known as contour integration. Here, instead of a global inhibition for surround, neurons can selectively enhance other neurons

T. Nathan Mundhenk (✉) · Laurent Itti
Computer Science Department
University of Southern California
Hedco Neuroscience Building, HNB-30A
3641 Watt Way Los Angeles,
CA 90089-2520 USA
E-mail: mundhenk@usc.edu, nathan@mundhenk.com

with a similar preference for image features. In this case, neurons will enhance if they have a preference for the same line orientation and are aligned by preference in a collinear or co-circular fashion. Neurons thus, compete with other neurons selectively, while enhancing the activity of others.

In contour integration, bar or Gabor elements (defined as the product of a Gaussian “bell-curve” and a sinusoidal grating) that are collinear, when observed, seem to enhance their ability to “Pop out” in an image that is also filled with other Gabors that are nonaligned noise elements (Field et al. 1993; Kovács and Julesz 1993; Braun 1999; Gilbert et al. 2000; Wu and Gilbert 2002). An example can be seen in Fig. 1, which shows Gabor elements of the same contrast, modulation, amplitude and size aligned into what seems to be an uneven circle. There is no direct physical link between the elements in this image that would give a direct cue as to their connectedness. Instead, the elements seem merely to point toward each other. The brain makes a functional gestalt leap and links these elements into a single unified contour (Wertheimer 1923; Koffka 1935). At the same time, the relative salience of the contour objects is elevated in the visual cortex. Thus, our brain reads between the lines as it were and creates the cognitive illusion of continuity even when objects along a contour are not physically connected. At the same time, our mind takes these contour elements and promotes their visual importance thus creating the effect of pop-out.

Several factors have been explored as being important to the phenomenon of contour integration. In particular, the properties of the elements in the contours can affect our ability to detect contours in a seemingly nonlinear fashion. For

instance, contours can be affected by continuity of colors, phase of Gabors and luminance of aligned foreground elements (Field et al. 2000; Mullen et al. 2000). Similarly, statistics of the background can also affect our perception of contours. For instance, if contour elements have a stronger collinear orientation compared with background elements, that is, they are more aligned, the contour is more visible (Polat and Sagi 1993a,b; Usher et al. 1999; Hess and Field 1999). Interestingly, when result data for enhancement of Gabor elements is plotted on a graph, enhancement for collinear elements is “U”-shaped. That is, a string of parallel Gabor elements, aligned like the steps on a ladder also have enhancement abilities, but diagonally oriented elements (elements which point in the same direction but are off-set like a staircase) have far less ability to enhance (Polat and Sagi 1993a,b; Yu and Levi 2000). Thus, as elements are rotated relative to each other, they have the strongest enhancement if the elements are aligned collinear or directly parallel to each other, but enhancement drops as elements are rotated between being collinear and parallel.

In addition to sameness of elements, contours also seem to become enhanced if the arrangement of the elements forms a closed loop (Kovács and Julesz 1993; Braun 1999). While there is some disagreement to the amount of pop-out from contour closure it is still nonetheless considered significant. This suggests that neurons sensitive to contour integration may perform some sort of linking to each other in a manner conceptually similar to a closed circuit like loop (Li 1998; Yen and Finkel 1998; Braun 1999; Prodhöhl et al. 2003). That is, neurons that do not directly touch may propagate effect to each other through their neighbors. Thus, ideally, if we imagine that contour integration is the result of neurons of preferred orientation linking to each other, we might conclude that contour integration may not just involve linking nearest neighbors to each other in a linear one-shot excitement, but may involve continuous reciprocation of neurons such that effects can propagate around a network. Such a notion is supported by current observations that all of the neurons on a contour that are thought to enhance each other in contour integration cannot be directly connected due to the limited reach of visual cortical axons. Thus, neurons in V1 and V2 are limited in the scope of their direct effect onto each other and should not cross the entire visual field. For contour closure effects to occur, especially over long contours, there should be some sort of network propagation (Li 1998; Yen and Finkel 1998; Braun 1999).

Contour integration can also be explored in both local and nonlocal ways. For instance, single Gabor element flankers and center-surround pedestals demonstrate that elements in a contour can enhance each other with only one flanker neighbor element to each side (Polat and Sagi 1993a,b; Zenger and Sagi 1996; Yu and Levi 2000). However, contours are further enhanced as elements are added (Braun 1999; Li and Gilbert 2002). This has become somewhat of a mystery for the reason that elements seem to enhance each other at distances that span beyond the size of the classical receptive field of neurons in the visual cortex (Braun 1999). Thus, adding

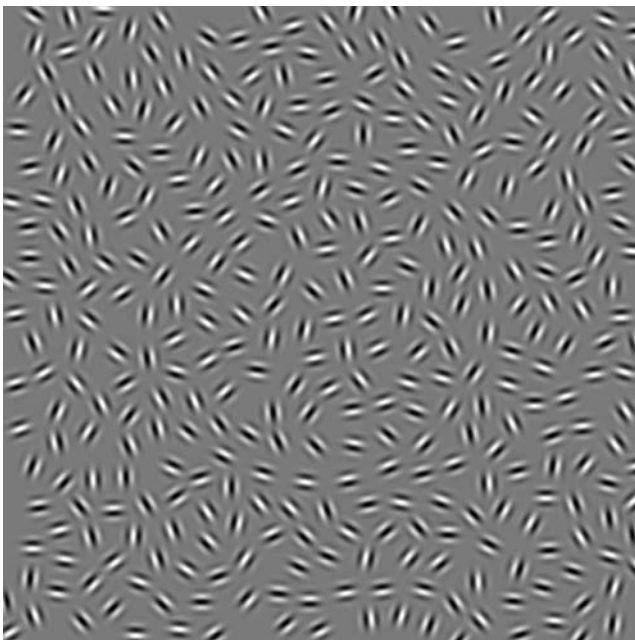


Fig. 1 This is an example of a contour created by Make Snake (Braun 1999). As can be seen, there appears to be a complete circle. However, the circle is created by unconnected Gabor wavelet elements. The mind connects these elements in a phenomenon known as contour integration

to the previous argument, there seems to be some ability for neurons in visual cortex to enhance a contour's perceptibility at locations represented by neurons that they are not *directly* connected to. Several theories have been advanced to explain how that can happen, for instance neural synchronization (Yen and Finkel 1998; Choe and Miikkulainen 2004), potential propagation (Li 1998) and fast plasticity (Braun 1999; Mundhenk and Itti 2003).

In addition to their saliency effects, it has also been suggested that contours play an important role in object identification. In particular, the ends of contours frequently referred to as end-stops and the junctions of contours may hold important data for the geometric interpretation of objects (Rubin 2001; Biederman et al. 1999). Thus, contour enhancement may not only be important for drawing our attention to the contours *qua* contours, but to the places at which those contours join with other contours and yield useful geometric information about objects for identification. Thus, it may be important for a mechanism that integrates contours for the sake of visual saliency to not only find contours, but to find the junctions at those contours even more salient. From this, we propose that a model of contour integration may do more than just enhance isolated contours. That is because more information is to be obtained from the junctions at contours. From an efficiency standpoint, junctions should also be detected if possible since this would reduce the number of neurons dedicated to the task of contour integration and end-stopping as well as speed up computation through parallel processing of information. This then could reduce redundancy and extra processing steps.

1.1 Computation

Traditionally, it has been a challenge to model contour integration. Two approaches are generally taken when trying to model contour integration. The first is the biological route (Yen and Finkel 1998; Li 1998; Grigorescu et al. 2003; Mundhenk and Itti 2003; Choe and Miikkulainen 2004; Ben-Shahar and Zucker 2004). In this method, the idea is to create a model of contour integration that explores how the brain may perform such activities. The other route is computational (Shashua and Ullman 1988; Guy and Medioni 1993), which is another important approach. However, these models tend to explore possibilities of contour integration computation or attempt to take a direct path to simulate contour integration for engineering applications. Here our approach is both. Our model attempts to explain saliency for contours in a manner that strives to illuminate the mechanisms that the brain uses, while attempting to optimize computation in order to be applied to visual saliency tasks in machine vision.

An important aspect of many contour integration algorithms has been the control of connectivity between computational elements. This is because, as has been mentioned, neurons seem to influence, beyond their own physical range, other neurons evaluating the same contour. This creates a situation where neural groups that process contour integration need to spread effect throughout the network while at

the same time controlling the network and preventing it from losing control. Some biological approaches have included a global normalization gain control and neural synchronization for this effect (Yen and Finkel 1998). We attempt to control our model by taking advantage of the properties of GABAergic interneurons to control local groups of neurons discretely. As we will describe later, the corresponding group that processes contours is broken into smaller local groups. Each local group is managed by its own single GABAergic interneuron, which controls gain by managing activity gradients for the local group it belongs to. Thus, each local group of neurons in the corresponding group has its own inhibitory bandleader to control its gain. The reason for taking this approach over global normalization is that we avoid direct influence between elements in the model that should not have direct interactions due to the limitations of the reach of neurons in visual cortex.

Our model will also attempt to explain how contour enhancement can extend beyond the typical receptive field of neurons by utilizing a fast plasticity (von der Malsberg 1981; von der Malsburg 1987) based on dopaminergic temporal difference like priming effects and pyramidal image size reduction. We will also show our model's abilities to perform similarly to humans in local enhancement tasks involving colinear aligned elements (Polat and Sagi 1993a,b) as well as in longer contour tasks with elements that enhance beyond the range of the neurons' receptive field.

In addition, our model will take into account physiological mechanism for contour integration by comparing our results to those of psychometric data. By fitting our algorithm to this data we will not only demonstrate the viability of our solution, but show we will have created a more complete solution in the process.

2 The model

2.0 Features

We have created a model, which we call carefully implemented neural network for integrating contours (*CINNIC*). Our model simulates the workings of a corresponding group of hyper-columns in visual cortex. We use the term "corresponding" to mean small proximate hyper-column groups, which correspond to the same basic task, for instance, integrating contours for saliency. In essence, it can be thought of as a cube of brain matter. Each neuron in a corresponding group connects to the many neighboring neurons within its reach. Each neuron in the corresponding group is sensitive to a distinct angle present in an image being observed by the model. That is, certain neurons activate more strongly when they are presented with a 45° line in their receptive field while others might be more sensitive to a 30° angle line. This means that each neuron in a hyper-column, and thus each neuron in the corresponding group has a preference to distinct angles (Hubel and Wiesel 1977). Contour integration is achieved in principle when neurons that are close and

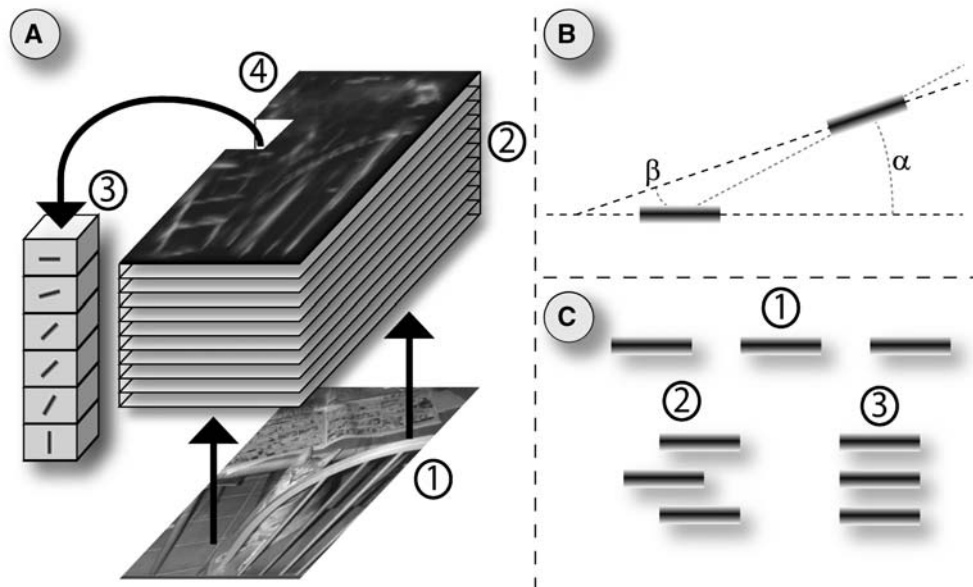


Fig. 2 **a** An image is taken (*I*) and is split into 12 orientation-filtered images (2), which are sent to their own layers in the corresponding group (3). Each of the 12 preferred orientations are rotated at 15 degrees (3). After interaction the output is collected at a top-level saliency map (4). **b** Interaction between layers is governed by collinearity. More collinear elements excite each other (α and β are small) while less collinear elements suppress each other (α and β are large). **c** Elements like (*I*) enhance, elements like (2) suppress, and highly parallel elements can enhance, like in (3)

have similar preferred orientations either enhance if they are collinear to each other, or suppress if they are parallel to each other. This is a method used widely (Yen and Finkel 1998; Li 1998; Grigorescu et al. 2003; Mundhenk and Itti 2003). Figure 2 shows an example of these simple rules for enhancement. It should be mentioned that the reason to suppress parallel flanking elements is to preserve the uniqueness of the visual item. For instance, a single line on a blank background should be more salient than a group of parallel lines (Treisman and Gelade 1980; Itti and Koch 2001). This can be intuitively imagined by thinking of one thin line drawn on a wall compared with a line on a pin stripe suit. It is easy to imagine that a single line on the wall is more salient and more likely to pop out than a single line amongst several others on the pin stripe suit.

An overview of the functioning of the network is as follows, as each neuron in the corresponding group fires, it transmits synaptic current to a neuron at the top of its hyper-column. This top-level neuron is a leaky integrator that stores charge received from neurons in its hyper-column. The way to imagine this is that the top level of leaky integrator neurons map one to one with an input image and creates a saliency map. Thus, an input pixel is connected to several neurons above it in a hyper-column and creates a one-to-one mapping for location between each hyper-column and an image pixel. That is, a hyper-column of neurons and its leaky integrator neuron on top maps spatially to exactly one pixel in an image, but then connects outwards to surrounding pixels in a center-surround architecture.

Each neuron has the ability to enhance its neighbor using dopamine-like priming connections. Thus, connectedness

among neurons in the corresponding group is enhanced by their ability to prime each other. The reason for this is that it allows activity of neurons to propagate. This gives neurons the ability to extend their influence beyond their own reach to neurons outside their receptive field. For instance, an active neuron primes its neighbor which causes its neighbor to become more active following that priming which in turn causes the neighbor to prime its neighbor and so on. Dopamine-like neurons are used in our model since they are fairly ubiquitous and can prime one another in 50–100 ms (Schultz 2002), which is well within the time span suggested for long-range contour integration of about 250 ms (Braun 1999). We state this because contour detection performance saturates at 12 Gabor elements. 50-ms priming may be the right amount of time for it to propagate in the network since depending on the exact speed of the network, a 10 or 12 cell networks effect will have met half way by this point in time. Additionally, this means that our model depends on a Hebbian-like associative priming where neurons that receive input in one epoch of our model enhance their neighbors firing in the next. Figure 3 shows a frame-by-frame example of this process. We reason for this method of propagation by observing that this process of priming has been observed and simulated in the brain, for instance in striatal neurons (Schultz 2002; Suri et al. 2001). Additionally, we should note that we emphasize the term *dopamine-like*. This is because other systems such as norepinephrine neurons in the locus coeruleus and Cholinergic neurons in basal forebrain also exhibit similar behavior (Schultz 2002), and while fast plasticity has been observed in higher cortical areas such as the prefrontal cortex (Hemple et al. 2000) and the rat visual cortex (Varela et al.

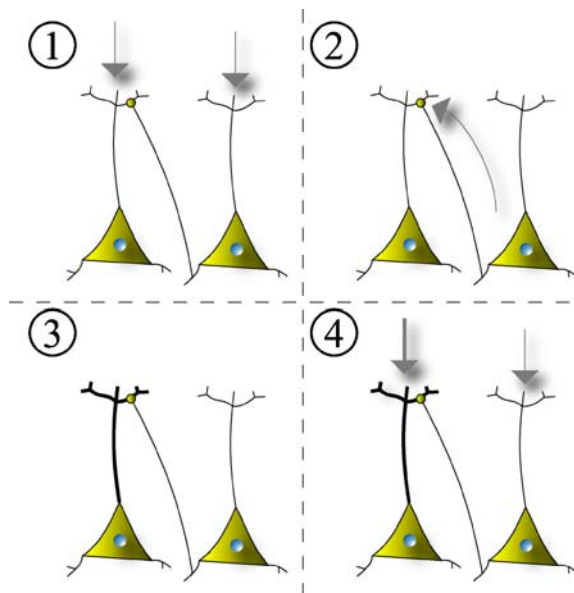


Fig. 3 An important element of the model is a fast plasticity term. In our model we follow the notion of priming via dopamine. (1) A neuron and its neighbor receive input. (2) The neuron on the right sends a signal to the neuron on the left. (3) The left neuron is now primed via dopamine. (4) When the neuron on the left receives another input, it is more likely to cross its firing threshold. This allows contour elements to propagate activity to other contour neurons that are not directly connected

1997) the time course and underlying mechanisms seem not to be understood well enough at the moment for our simulation. As such, we use the term dopamine-like since it seems that its mechanisms are generalizable enough for our purposes. Our model does not implement explicit temporal synchronization for propagation since it is our observation that evidence for its actions in V1 and V2 seem less certain, and that while some papers suggest explicit temporal synchronization based on their results (Lee and Blake 2001) as we mention in the discussion, they can also be accounted for by a fast plasticity mechanism. Our argument will then be for such a process based upon its feasibility as well as the fitness of such a mechanism to explain the processes which are observed in humans. As a last note we wish to point out that we do not object to explicit temporal synchronization at any theoretical level, it is to say, we believe that fast plasticity may better explain contour propagation.

Another feature of our model is that it controls runaway gain from over excitation of the corresponding group. It does this by using suppression of local groups of pyramidal neurons that are in subsections of the whole corresponding group. To accomplish this we hypothesize that medium sized basket type fast spiking (FS) interneurons are stimulated from one or few putative inputs from the top leaky integrator neuron and exhibit strong control over the neurons they efferently connect to. Such neurons have been observed in the brain in many areas, particularly in the pre-frontal cortex (Krimer and Goldman-Rakic 2001) and Striate Cortex (Shevelev et al. 1998; Pernberg et al. 1998). They need only one or few inputs and can give very strong inhibition. Here, these FS

parvalbumin-type interneurons are plausible since they require very few putative inputs in order to create inhibitory post-synaptic potentials (IPSP) (Krimer and Goldman-Rakic 2001). Further, they have been found to modulate pyramidal neuronal activity directly (Gao and Goldman-Rakic 2003), which are the type of neurons we have constructed our corresponding group from. A gradient-based suppression could be attained by having a second slow interneuron inhibit the first interneuron, this may be plausible since interneuron to interneuron connections are well known (Wang et al. 2004). If the activity of the first interneuron levels off, the second interneuron will catch up and suppress the first completely. Figure 4 shows a representation of this. Since interneurons can spike at a variety of rates (Bracci et al. 2003), the end result from this mechanism is that local groups of pyramidal neurons are inhibited proportionally to their local groups' sum excitation.

2.1 The process

In our computational model, before an image is sent to the corresponding group it must undergo some preprocessing. This takes several steps. The first is to take in a real world image. This can be a digital photograph, or an artificially-created stimulus such as an image of Gabors. The input image is filtered for orientation using Gabor wavelets. This creates several images, in our case 12, that have been filtered for orientation. In this model, 12 orientations are used since it is hypothesized that this is the number of the orientations the brain may use in V1 (Itti et al. 2000). The image is then reduced into three different scales of 64×64 , 32×32 and 16×16 pixels by using the pyramid method for image reduction (Burt and Adelson 1983). This yields 36 processed images, that is, 12 orientations by three scales. In the next stage, each scale is processed separately. As such, we have three independent sub-corresponding groups, one for each scale. Each orientation image is sent to a layer in the sub-corresponding group for its scale that is selective for that orientation. For instance, the 90° orientation image inputs directly only into the layer that is designated as selective for 90° orientations. This creates a sub-corresponding group with a stacked topology where each layer is comprised of neurons sensitive to only one orientation. To reiterate, the structure places neurons directly above each other, which receive direct input from the exact same location in the visual field. Thus, the result can be thought of as a cube of neurons where the i and j dimensions correspond to a specific location in the visual field and the α dimension corresponds to the preferred orientation of the neuron. To make this cube of brain matter a corresponding group, connections are established between the neurons.

Interaction between neurons is created using a hyper-kernel. Each hyper-kernel describes both the inhibitory and excitatory connections between neurons simultaneously rather than as two separate kernels where one is for inhibition and one is for excitation. This is done to *speed* up the computation operation and can be done since, if we neglect temporal

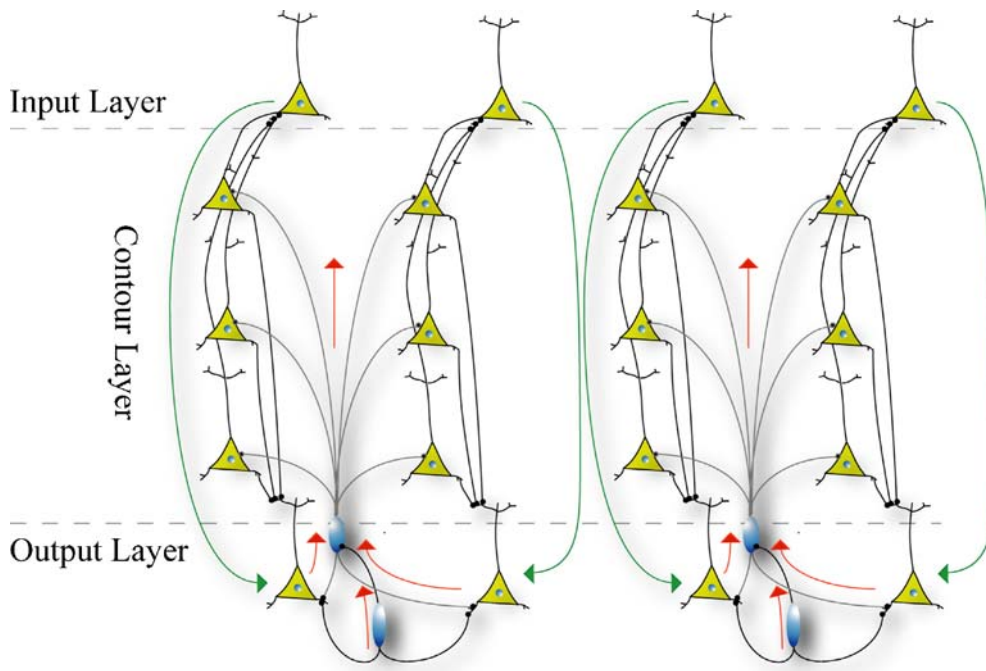


Fig. 4 Gain in the network is controlled by a Basket GABAergic interneuron-like connection scheme. This works by spatially grouping local neurons into groups that are all suppressed by a local interneuron for that group. This creates a gain control, but keeps such control local to within the theoretical spatial range of axonal arbors in V1 and V2

differences between excitation and inhibition at this level, the summation of inhibition and excitation to another neuron results in a mutually exclusive inhibition or excitation result. That is, the hyper-kernel is the summation of excitation and inhibition kernels. Figure 5 shows the “slices” of the kernel we used and how it is used to define how neurons interact with each other by defining the weights of excitation and inhibition. Each hyper-kernel slice has a reach of 12 pixels (reaching out to a span of 12 neurons) for excitation and ten for inhibition. It should be noted that this is the same across all scales. When the image is reduced, the kernel will reach across 1.4° of visual angle for 64×64 pixel scale image, 2.8° for the 32×32 scale image and 5.6° for the 16×16 scale image. Additionally, while the kernel at the 16×16 pixel scale is large in terms of visual angle, it has a relative lack of acuity since the image has been reduced dramatically. Thus, we still fall within size constraints for neuron reach since the kernel at 16×16 is still the same size. However, the image has shrunk.

In all, 144 slices are created for our hyper-kernel. These represent all the possible connections between two neurons in the corresponding group. That is, each neuron is selective for one of 12 orientations and can interact with another neuron, which can be selective for one of 12 orientations. This creates 12×12 possible interactions. The spatial relation for each hyper-kernel is handled within each slice. That is, each slice maps retinotopically. Orientation is thus handled between slices, while translation is handled within slices of the hyper-kernel. It can be seen then, that the hyper-kernel is stacked in the same way as the layers of a corresponding

group. Since it has the same topology, it can then *pass* over and through a corresponding group in much the same way a standard 2D kernel is passed over a standard 2D image. However, the process moves the hyper-kernel in 2 dimensions over the 3D corresponding group (with 4D connections), so in essence, the convolution adds an extra set of dimensions over 2D convolution. This can be thought of as moving a hypercube of 12 spatially overlapping cubes (one for each orientation) simultaneously in a Cartesian manner along 2D through a larger box of the same height (which can be thought of as the corresponding group).

Each orientation-selective neuron when stimulated by input from the image and by input from other neurons that excite it will send synaptic current to a top layer of leaky integrator neurons at the top of its hyper-column. The top layer of leaky integrator neurons is treated as a saliency map for these purposes. The top layer can reciprocate to control gain of local neurons using suppression from FS interneurons. That is, the activity of the saliency map’s top-layer neurons controls the activity of the gain control for the interneurons. Thus, a noisy image is gain controlled locally using the gradient of excitation in a local group controlled by a single interneuron for that group.

Contours are sharpened and extended using the dopaminergic-like priming described previously. The outputs from the three different scaled sub-corresponding groups are merged together using a weighted average. The end effect is a combined saliency map from across scales, which is the final output from CINNIC.

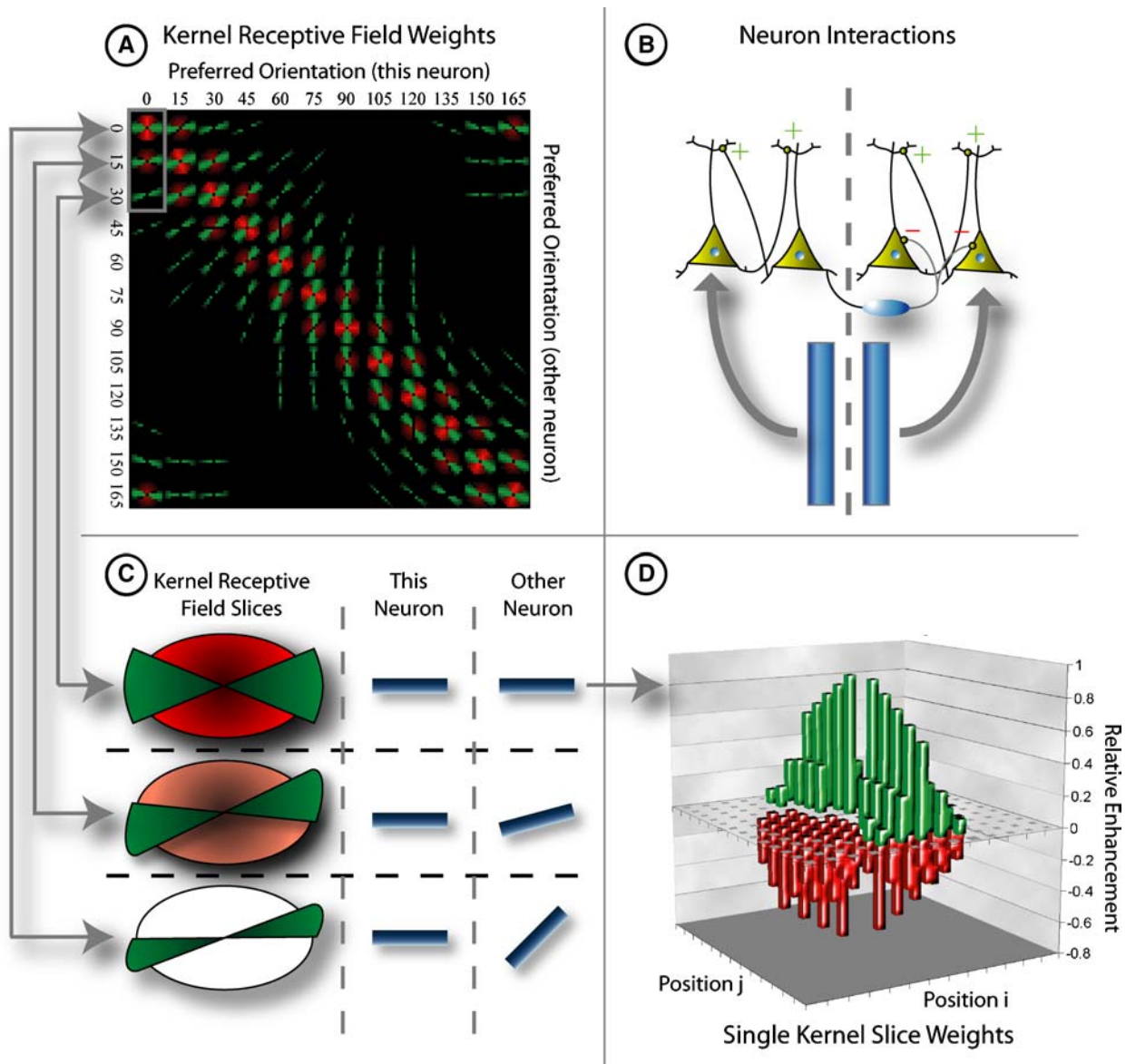


Fig. 5 **A** A Kernel is generated that dictates the base strength of the connections between neurons in the network. Each kernel slice shown represents the interaction between two neurons given their preferred orientations. Red represents inhibition while green represent excitation. **B** If two neurons are parallel in preference but not collinear, then they inhibit each other. **C** Parallel bars excite if they are close to collinear in preference. The three kernels shown (the same as highlighted in **a**) show the interaction if elements are related to each other as shown by the bars. For instance, if two elements are totally co-linear they would use the first kernel. The next kernel would be used if one element is offset by 15°. **D** This is a side view of the 0° offset kernel. The kernel has modest second- and third-order polynomial curvature, which can be observed on close inspection

2.2 Kernel

As mentioned the hyper-kernel is defined that contains both excitation and inhibition in it. However, excitation e is defined in the kernel in a slightly different way than inhibition s . As can be seen in Fig. 5, excitation is strongly sensitive to the preferred orientation between two neurons, while inhibition is mostly sensitive to the spatial location between two neurons. That is, excitation is sensitive to the preferred orientation of both neurons in an interaction, while inhibition is

only sensitive to the orientation of the operating neuron so most of its effect is from the distance between neurons. The excitation term can be seen in Eq. 1. Here a_α is a term for the collinear disjunction (how much this neurons preferred orientation points to the other neurons) between this neuron and the other neuron. a_β oppositely describes how much the other neuron points to this one. The planar Euclidian distance between these neurons is expressed as d^e , this can be thought of more in terms of the distance between the hyper-columns a neuron resides in and not the direct distance between two

neurons in space. The excitation output expression to the kernel is $K_{\alpha\beta}^e$, this is the excitation that will be expressed by the kernel from the preferred orientation α of the neuron that is operating (this neuron) and orientation β of the neuron to be operated on (the other neuron). In simplest terms, a_α^e and a_β^e describe how much two neurons point toward each other in a collinear fashion. That is, a_α^e is the angle from the other neuron to this one, and a_β^e is the angle from this neuron to the other as seen in Fig. 2. Thus as Eq. 1 shows, the excitation part of the kernel is the average over a collinearity term and distance.

$$K_{\alpha\beta}^e = (d^e + (a_\alpha^e \cdot a_\beta^e))/2 \quad (1)$$

The output angles are derived as:

$$a_\alpha^e = l_{f_e} \cdot A^e + P_2^e \cdot (A^e)^2 + P_3^e \cdot (A^e)^3 + 1 \quad (2)$$

$$a_\beta^e = l_{f_e} \cdot B^e + P_2^e \cdot (B^e)^2 + P_3^e \cdot (B^e)^3 + 1 \quad (3)$$

The terms P_2^e, \dots, P_3^e are constants used to curve the kernel's shape with a third-order polynomial. That is, as preferred orientation a^e differences increase and the distance d^e between neurons increases, excitation tapers off along a slightly flat, but in this case, an almost monotonically decreasing polynomial function. The polynomial is used since its ability to take on a variety of shapes is very strong. Additionally, since it is applied radially, it can take on shapes similar to a Gaussian, but we are able to avoid explicitly making such assumptions. B^e and A^e are expressions for how far off collinearity is in this interaction. Basically, this ranges from 1 to 0 with 1 being if two neurons are collinear and 0 if two neurons are non-collinear to a degree that surpasses a threshold. l_{f_e} simply normalizes B^e and A^e to be within the 0 to 1 threshold. Here normalization is used to constrain values used in the kernel manufacture so that initially values for inhibition fall within the same range as excitation. Inhibition is expressed in more simple terms as

$$K_{\alpha\beta}^s = W \cdot (d^s + (a_\alpha^s \cdot c))/2 \quad (4)$$

In this equation the major difference from excitation is c which is the difference between preferred angles in the two layers being interacted (remember, inhibition is only sensitive to the operating neurons orientation α and not the receiving neurons orientation β). That is, it is less important how much another neurons preference points at this neuron compared with how much this neurons points at it during inhibition. Spatial location is thus more important than strict collinearity for inhibition. The reason for this is because originally, better results were obtained early on by removing the a_β^s term between elements and replacing it with c . This also has the effect of making inhibition more purely center-surround in its effects.

Just as with excitation d^s is the distance between this neurons column and the other neurons column and a_α^s is based upon the orientation of the operating neuron. Again note Fig. 5, which shows the general shapes of the kernel. The most obvious result of the difference between excitation and inhibition is that inhibition is strongly symmetric over both principal axis. Thus, the shape of its field of influence

stays ellipsoidal. W is a constant that gives a gain to the inhibition, either making it stronger, or weaker than excitation depending on what value we decide is suitable. Again, a_α^s is expressed as

$$a_\alpha^s = -1 \cdot (l_{f_s} \cdot A^s + P_2^s \cdot (A^s)^2 + P_3^s \cdot (A^s)^3 + 1), \quad (5)$$

where again l_{f_s} is a normalizer and B^s and A^s range between 1 and 0 depending on the angle offset of this neuron and the other neuron. Similar but orthogonal to excitation, a_α^s is equal to 1 if the operating neuron and the neuron being operated on are parallel, but not collinear. It becomes 0 if the two neurons are orthogonal. Thus, an important note about this system is that preferentially orthogonal neurons do not have direct influence on each other for either excitation or inhibition, but do carry indirect influence as will be discussed later in our discussion of junction finding.

Values for a_α^s and a_α^e are derived such that they are mutually exclusive causing both excitation and inhibition to zero at the same angle. Thus, when $K_{\alpha\beta}^s$ and $K_{\alpha\beta}^e$ are combined into a single kernel it is a simple matter of mapping one over the other. This can be thought of as having computed the hill and the valley separately and then bringing the two together. Since the system is discrete, any minor disjoint is not noticed.

2.3 Psuedo-convolution

The main process of CINNIC lies in the mechanisms of the corresponding group. Interactions in a corresponding group, which defines how collinear sensitive neurons work, uses a pseudo-convolution. The major difference between CINNIC's hyper-kernel convolution and traditional convolution is that the results from the operation are stored at the other pixel, not the pixel being operated on. This was done earlier on when we were experimenting with other features that were later removed. Equation 6 shows the basic pseudo-convolution operation, which is also illustrated in Fig. 6. Here x is an orientation processed image pixel at image location i, j in one of the 12 different orientation layers α . Each processed image pixel, which becomes represented as a neuron, is multiplied by the sum of its interactions with other pixels (neurons) in its receptive field at the relative location k, l with respect to the neuron i, j, α , with a field size of m by n . That is, k, l is the location of the other neuron relative to this neuron. The main interaction of this pixel-neuron ($x_{ij\alpha}$) and the other pixel-neuron in its receptive field ($x_{kl\beta}$) is described by their weights from the kernel ($K_{\alpha\beta(k-i)(l-j)}$) described earlier (where $(k-i)(l-j)$ is the corresponding hyper-kernel slice pixel mapped onto the field n by m). An approximation for the dopamine-like fast plasticity term is described as $(f_{kl\beta})^t$ which is derived in Eq. 9. Thus, this neuron ($x_{ij\alpha}$) will dopamine prime the neuron at location k, l, β . Further, iff the interaction is inhibitory (the neural activity is computed as *less than zero*), $(g_{kl})^t$ represents an addition to suppression from the gain control group suppression term from $(x_{kl\beta})$'s group (Eq. 7) at time t which is the last complete iteration. Thus, this represents the GABA-based group

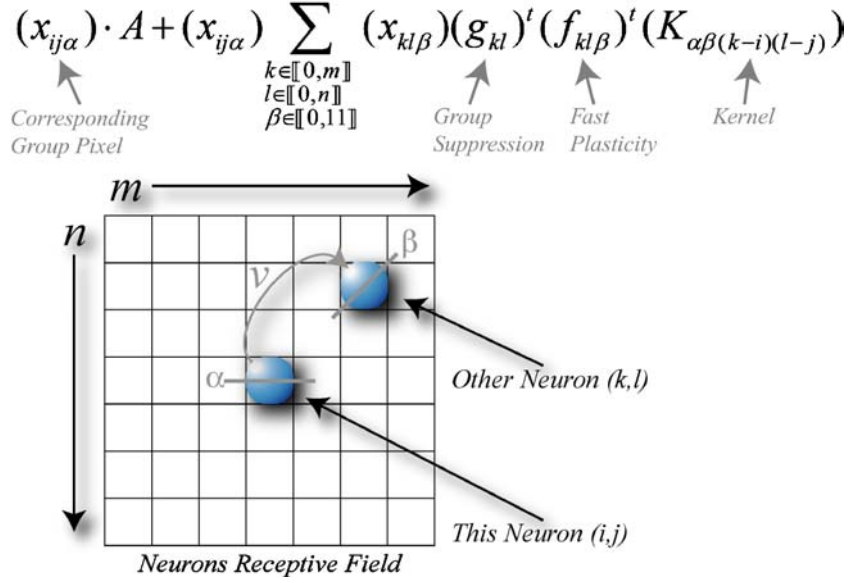


Fig. 6 This graph illustrates the way in which neurons interact with neurons in other hypercolumns. By mapping the hyper-kernel K over the neuron α, i, j we can find the base synaptic current generated that should be sent to another neuron at the relative position β, k, l .

suppression mentioned earlier. This interaction is combined with the base excitation to this neuron times a constant gain $(x_{ij\alpha})A$ with a pass through term. That is, the sum excitation of this neuron also includes the input pixel intensity from the orientation image as well as the activity from other neuron interactions in its corresponding group. The linear output from this neuron is stored in $(v_{ij\alpha})^t$, which is the total activity for this pixel-neuron after a single pseudo-convolution iteration at time t .

$$(v_{ij\alpha})^t = (x_{ij\alpha}) \cdot A + (x_{ij\alpha}) \times \sum_{\substack{k \in [0, m] \\ l \in [0, n] \\ \beta \in [0, 11]}} (x_{kl\beta})(g_{kl})^t (f_{kl\beta})^t (K_{\alpha\beta(k-i)(l-j)}) \quad (6)$$

$$(g_{kl}) = \begin{cases} (g_{kl}) & \text{iff } (K_{\alpha\beta(k-i)(l-j)}) \leq 0 \\ 1 & \text{otherwise} \end{cases} \quad (7)$$

The resulting potential is sent to an upper level of leaky integrator neurons (Eq. 8). This is the neuron that rests at the top of the hyper-column and along with the other neurons at the top of their respective hyper-columns forms a saliency map for this scale. A simple leak is approximated here with a constant leak term L with the sum being placed in $(V_{ij})^{t+1}$ as a quick, but sufficient leaky integrator approximation, with the down side of not being proportional to potential. In essence, this sums the potential of all 12 neurons in this column that receive input from the same pixel in the image.

$$(V_{ij})^t = \sum_{\alpha \in [0, 11]} (v_{ij\alpha})^t - L \quad (8)$$

Dopamine-like fast plasticity $(f_{ij\alpha})^t$ is approximated as Eq. 9. Here a neuron is primed to have a greater weight if it received

input during the last iteration $(v_{ij\alpha})^{t-1}$, which is proportional to that input. A constant F controls the gain on this effect. A ceiling is placed by Eq. 10 which limits this effect to be no less than 1 (no effect) or greater than 5 (strong effect). In this case, the selection of a ceiling of 5 is slightly arbitrary and dependant on observations that it worked well in our early test cases.

$$(f_{ij\alpha})^t = (v_{ij\alpha})^{t-1} \cdot F \quad (9)$$

$$1 \leq (f_{ij\alpha})^t \leq 5 \quad (10)$$

Group suppression (Eq. 11) is based upon the gradient of the increase in excitation for all neurons in this group and approximates the GABAergic gradient circuit previously described. That is, all the neurons that are in this group (V_{pq}) have their output summed, with the finite difference determining the gradient. A gain v is applied and the constant T is a resistance threshold term that assures that group suppression can only occur when excitation has reached a certain level. N_i and N_j express the boundary of this local group which is $1/8\text{th} \times 1/8\text{th}$ of the total image size. In other words, if the image is 64×64 pixels, a local suppression group is 8×8 pixels in size. This size makes the range of this inhibition roughly the same size as the kernel and assures even division.

$$(g_{ij})^t = v \left[\left[\sum_{(p,q) \in N_i \times N_j} (V_{pq})^t - (V_{pq})^{t-1} \right] - T \right] + (g_{ij})^{t-1} \quad (11)$$

$$N_i = \llbracket i - (m/8); i + (m/8) \rrbracket \quad (11a)$$

$$N_j = \llbracket j - (m/8); j + (m/8) \rrbracket \quad (11b)$$

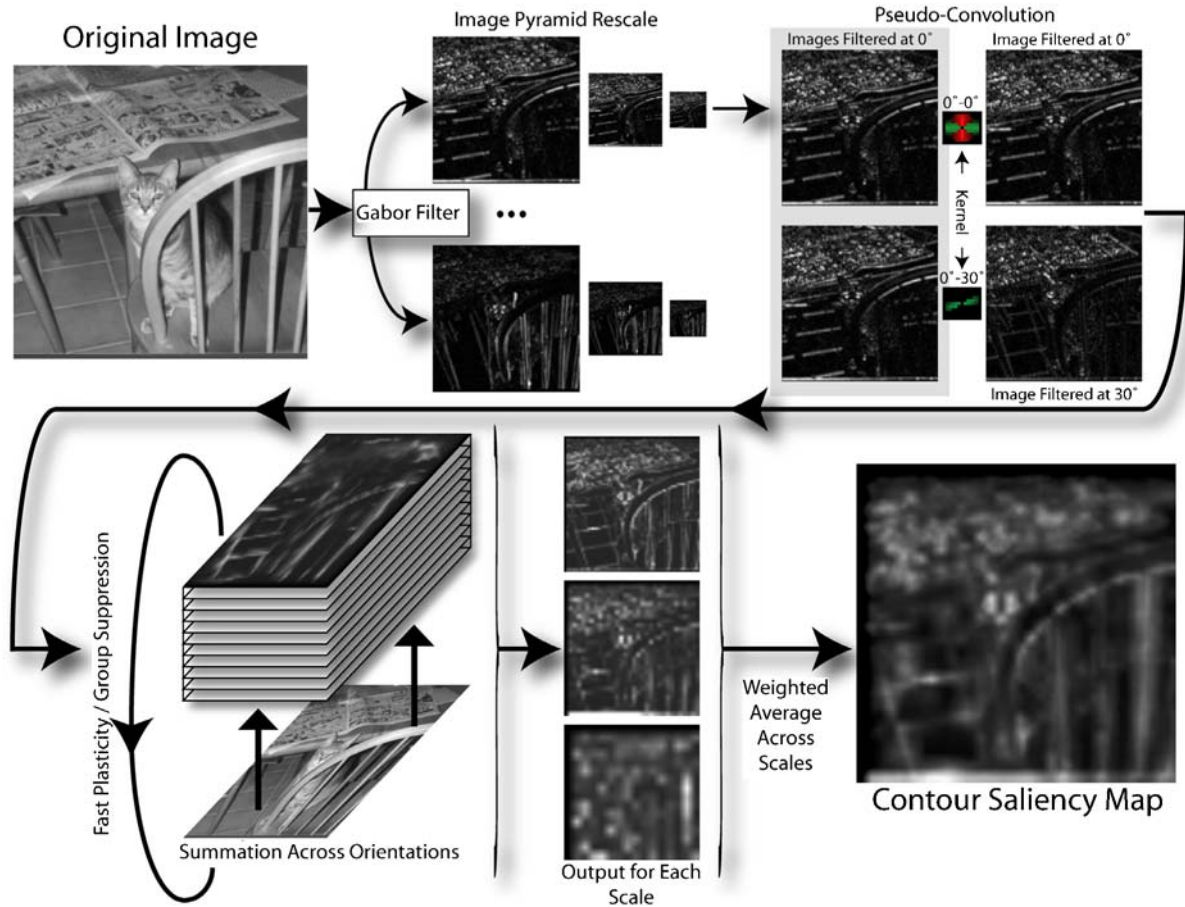


Fig. 7 CINNIC works in several phases. The first is to take in a real world image. Gabor filters are applied that creates 12 orientation selective images. The image is then rescaled using an image pyramid into three different scales. The 12 orientation selective images are then pseudo-convolved and the corresponding region is run with dopamine-like fast plasticity and group suppression over several iterations. The three different scales are then brought back together using a weighted average and combined into a contour saliency map

All the potential is run through a logistic sigmoid (Eq. 12), which simulates firing rates. Thus, the final top most saliency map for contours at this image scale is taken from Eq. 13.

$$S(x) = 1 / (1 + \exp(-2\beta v)) \quad (12)$$

$$I'_{ij} = S((V_{ij})^t) \quad (13)$$

The final saliency map for all scales is created by taking a weighted average of all the scales (sub-corresponding groups), as can be seen in Eq. 14 (Fig. 7). Here I_{iju} is the saliency map for this sub-corresponding group at its own scale u while w_u is the weight bias given to this scale (a number from 0 to 1). n_u is the number of scales analyzed (in this case 3) and M_{ij} is the final saliency map derived from across all differently scaled sub-corresponding groups.

$$M_{ij} = \frac{\left(\sum_u I_{iju} \cdot w_u \right)}{n_u} \quad (14)$$

Thus, M_{ij} represents a saliency map of what parts of the image are most salient based on contour information. If the

algorithm is effective then M_{ij} should have a large value corresponding to a contour segment at location i, j in the input image. It should correspondingly have a low value where no contour segment or a noise segment lies. The most salient point or points are the pixels from M_{ij} which have the highest or maximum values (Fig. 8). Additionally, it should be noted that while the saliency map, that is, output shows clearly the contours, since the goal of this work is to simulate visual saliency, the most important component of the output should be the salient points that draw attention to the contours.

3 Experiments

To investigate the validity of our model we followed a multi-tier approach. The idea was that our model should be viable at several levels. First we looked at how our model worked with simple element interactions. For instance, how would our model work on a Gabor patch with two flankers only. In this we should see saliency enhancement with greater collinear alignment as observed in humans (Polat

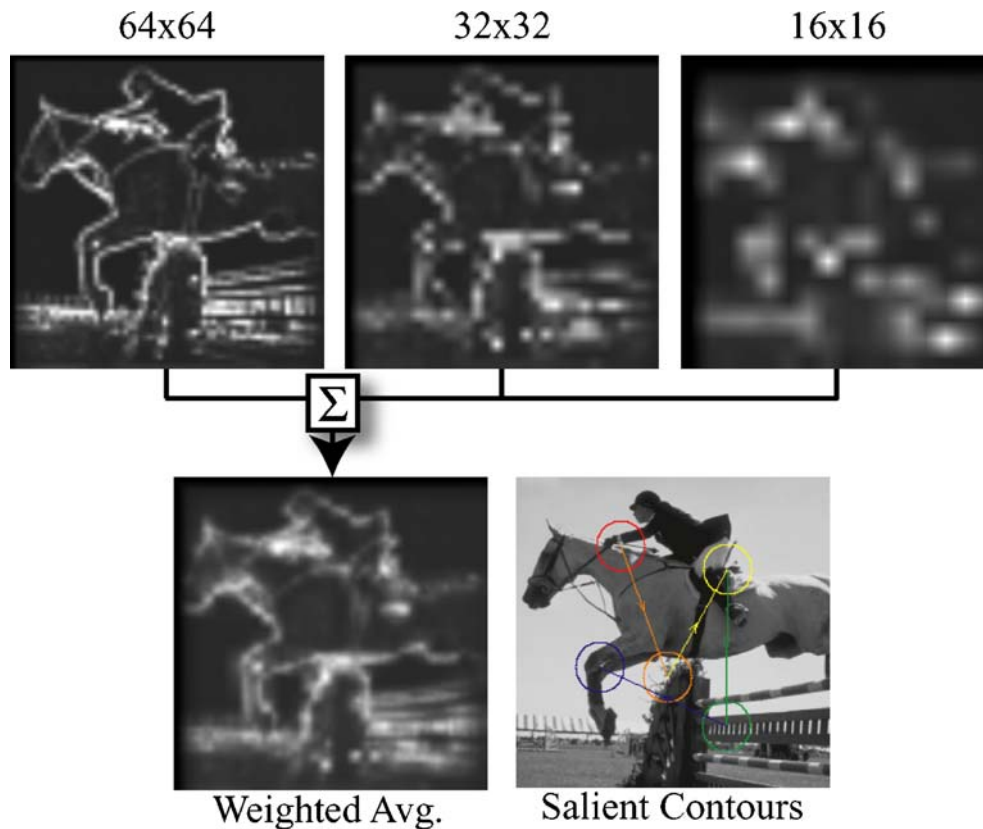


Fig. 8 The top three images show the results of pseudo-convolution at each of the three scales used. The *bottom left image* shows the weighted average of the three images. The *circles* represent what the program feels are the five most salient points. The *bottom right image* is the input image with the most salient points shown with the *red circle* on the most salient point and the *blue circle* on the least salient of the top five

and Sagi 1993a,b). Additionally, enhancement should extend beyond a small number of elements. That is, we needed to check if our model worked on chains of Gabor elements. This would validate our model against data that shows that enhancement is formed for collinear Gabor elements against background noise Gabor elements along paths extending beyond the receptive field of V1 neurons (Braun 1999). The third level of validation involved real images. This was the next logical increment. That is, we first test if our model works on a few simple Gabor elements (simple, local), then we test longer chains of Gabor elements with Gabor noise (simple, nonlocal), then next we test on natural images (complex, nonlocal). We should expect to find validity of our model at all three levels if we are to claim that it could be a reasonable approximation to contour integration in humans. Additionally, we also report on results that suggest that the CINNIC model is also sensitive to junctions and end-stops. This is to illustrate the generalization of the CINNIC model as well as demonstrate possible efficiencies in visual cortex for finding junctions with the same or a similar mechanism as used for contour integration. Additionally, a unified mechanism that finds contours and junctions may help explain some psychophysiological observations made by others, which we discuss later.

3.1 Local element enhancement

As has been discussed, contour integration behavior can be seen in cases where only a few Gabor or other directionally specific element, such as a line segment, flank one element (Polat and Sagi 1993a,b; Kapadia et al. 1995; Gilbert et al. 1996; Kapadia et al. 2000; Freeman et al. 2003). We attempted to replicate work by Polat and Sagi (1993a,b) showing that a Gabor element when flanked by one collinear Gabor on either side can be enhanced from this arrangement. That is, the ability to detect the Gabor element in the center is increased or in some cases decreased as two flanking Gabors are altered for distance from the central Gabor. Enhancement changes should also be observed with alterations in contrast/amplitude for the Gabors. The results they obtained show that when the flanking elements are moved away from the central Gabor in increments of λ (which is the frequency size for the Gabor wavelet and is used as the measure for the separation between Gabor elements), at very close distances, flanking Gabors seem to make it harder to detect the central Gabor. Maximal enhancement is obtained when the flanking Gabors are separated from the central Gabor by approximately 2λ . However, as the flankers are moved even further away, the enhancement effect seems to be completely

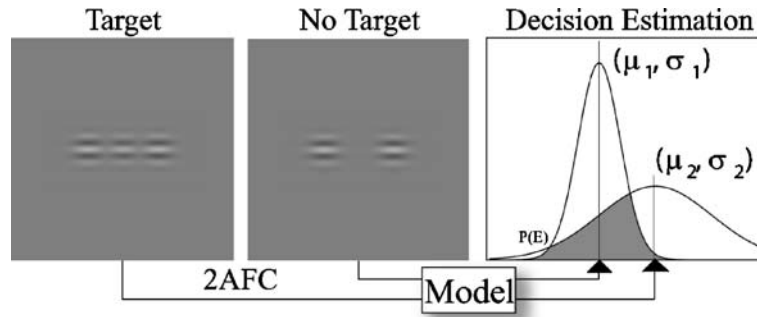


Fig. 9 The program makes a decision as to which of the two images has the target in it. The model estimates this decision by taking the probability of a decision as the Poisson of the output at the target. The error is the error function (EFC) of the two distributions for both target and nontarget. Target amplitude is changed until error rate is 25%. This marks the relative enhancement

diminished. This reaches a total diminishment of enhancement when separation reaches about 12λ .

Using this experiment as a guide, we optimized the kernel parameters of our model to create an outcome that resembled theirs as closely as possible. This was done by creating Gabor images with flankers at 0, 1, 2, 3, 4 and 12λ . We created our Gabor images as closely as possible to the ones used in their experiments (Polat and Sagi 1993a,b). Additionally, the images could have alterations for the amplitude of the target Gabor in the same way they altered their image targets. In their experiments, they found the amplitude of enhancement for a center Gabor element when flanked by two collinear Gabor elements of the same size using a two alternative forced choice paradigm. Thus, they did this by showing two images and forcing a participant to choose which one had the central Gabor in it and which one had an image with only the flankers and no central element. When the amplitude of the central element yielded a 75% correct rate, that was considered the threshold amplitude of detection for that particular separation of Gabor condition. They then mapped the relative enhancement of the target Gabor in the condition by comparing it with a single stand alone Gabor with no flankers which served as the baseline for detection threshold.

We achieved a similar result by estimating the error rate using the error function from the Poisson obtained from the output of the target/no-target conditions (Fig. 9). This method used previously by our group (Itti et al. 2000) and others estimates the error from physiological observations since noise and error in the brain follows a Poisson distribution. By modeling this, from Eq. 15 we could show that given the output stimulus in the target/no-target condition, what would be the probability that it would pick one image over the other. This method was used because it gives us dramatically increased performance over using a Monte Carlo simulation for determining error in a two alternative forced choice paradigm which was pivotal to train our model as will be described.

$$P(\text{error}) = \frac{1}{2} \operatorname{erfc} \frac{\mu_1 - \mu_2}{\sqrt{2(\sigma_1^2 + \sigma_2^2)}}. \quad (15)$$

What this means is that we showed our algorithm the target and no target images. An intensity value from the saliency

map at the location in M_{ij} (Eq. 14) where the target Gabor from the input location corresponded to was obtained. The value from M_{ij} was then considered to be a mean value with the expected standard deviation of outputs defined from the Poisson distribution. Using an iterative technique, amplitude was adjusted for the central Gabor using a hill climbing method with momentum, until the error rate was $75\% \pm 1$. The amplitude at threshold was then compared with the output from an image with a single unflanked element, to measure relative enhancement just as in the study by Polat and Sagi. Our results were then compared with their results. The error was tallied and used to drive a second custom gradient descent search algorithm whose goal it was to minimize the error between our results and theirs by adjusting kernel parameters. As can be seen in Fig. 10, error was reduced substantially and fit — Polat and Sagi’s experimental output for subject AM almost perfectly with a maximum error at less than 2 standard errors off of subject AM’s results (as estimated for this experimental paradigm in (Polat and Sagi (1993b) p. 76 and Polat and Sagi (1993a) p. 995). These results fare particularly well for our model because not only do they fit the experimental result of Polat and Sagi, but they have the same eccentric nature of reducing enhancement for Gabors that are particularly close.

To illustrate why we observed the result of decreased enhancement at very close distance between Gabors, kernel slices from CINNIC were extracted and interacted with targets of different sizes to measure the enhancement when two targets are moved closer or further away. What we discovered is that with larger targets of approximate size, 4λ , when compared with the 64×64 scale kernel, had the ability to contact neurons that were in inhibitory regions as well as the excitatory regions. This stimulus is about the same size as the Gabors used in our study that were about 3.5λ in size. This occurred as the elements moved closer to each other. Figure 11 shows that as target objects get larger, they begin to have far stronger inhibitory ability at close distances. Thus, for enhancement, given a wedge-shaped excitation range, there is an optimal distance for enhancement between two elements, with that distance being closer for smaller Gabors. Also note that enhancement begins to fall off between 2.4λ

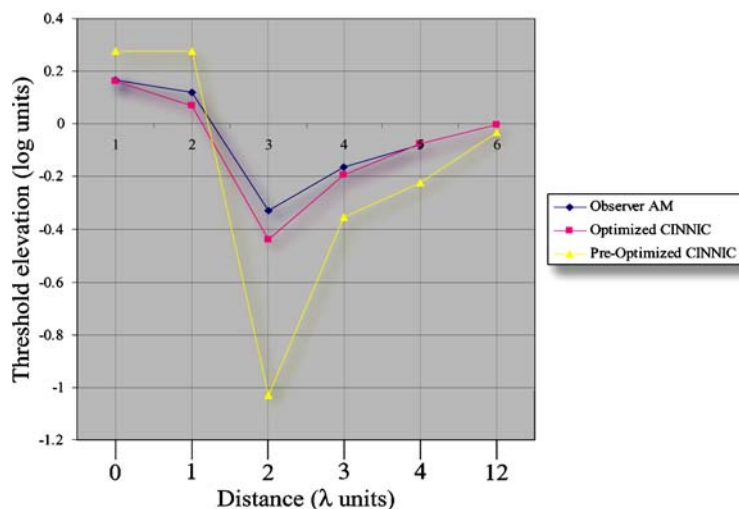


Fig. 10 The algorithm was optimized against observer AM. The pre-optimized output has a similar shape, but approaches the performance results from observer AM following optimization of CINNIC using hill climbing. The decision process from the program yields results that are within 2 standard errors (0.05) at its greatest difference found at a separation of 2λ .

and 1.6λ . This is where you would expect it to fall given the current reviewed psychophysical data and the outcome of CINNIC.

3.2 Nonlocal element enhancement

Further testing of CINNIC was done using a special program called Make Snake provided and created by Braun (1999). This was used to generate test images in which a salient closed contour is embedded among noise elements. Using these stimuli, we tested under which conditions our algorithm would detect the contour elements as being the most salient image elements.

Make Snake creates images like the one presented in Fig. 1. The output is several Gabor patches aligned with randomized phase into a circular contour. The circle itself is carefully morphed by the program using energy to flex the joints of an “N-gon” to create a variety of circular potato-like contour shapes. The circles made up of foreground elements are controlled for the number of elements as well as the spacing in λ sinusoidal wavelengths. The elements can also be specified in terms of size and wave period. Background noise Gabors are added randomly and are of the same size as foreground elements but may be at different separation distances. They are placed in such a way that they are moved like particles in liquid to a minimum spacing specified by the user. Gabors are added and floated until minimum spacing requirements are satisfied. The end result can also create accidental smaller contours among the noise background elements.

Test images were created 1024×1024 pixels in size and corresponded to a simulated total visual angle of 7.37×7.37 degrees. Test images were created using two different Gabor sizes, a small Gabor (70 pixels wide with a 20 pixel Gabor wave period) and a large Gabor (120 pixels wide with a

30 pixel wave period). The background elements were kept at a constant minimum spacing (48 pixels for the smaller Gabors and 72 pixels for the larger Gabors). Spacing for larger foreground Gabors (120 pixel size) was varied between 2λ and 3.5λ in steps of 0.1666λ . This was constrained since values above 3.5λ made the contour circle larger than the image frame itself. The smaller Gabors (70 pixel size) had more leeway and could be varied from 1.5λ to 6λ in steps of 0.5λ . For both Gabor sizes, the minimum spacing is set the way it is because below this, the foreground elements begin to overlap. It should be noted that the ratio of foreground separation to the minimum background separation was the same for both large and small Gabor patch conditions given the same λ . That is, the background elements had the same constant λ separation for *all* images. The smaller Gabors in these tests were the same size in pixels as the Gabors used in the experiment in 3.1. This size corresponded to a visual size of 0.5° .

For each condition, Gabor size and foreground spacing, 100 images were created. An output mask was also created representing where foreground elements were positioned. This was used for later statistical analysis. In all, 2,000 images were created and tested.

Statistical analysis was done by taking the output saliency map from CINNIC, which always ran with identical model parameter settings for all images, and comparing it to the mask; this was done by looking for the top most salient points in the combined saliency image map M . When a salient point was found, the local region was concealed by a disk to prevent the same element area from being counted twice. Salient points were marked as first, second, third and so on depending on its value in the saliency map. That is, the most salient point was ranked first, and the second most salient point was ranked second and so on. Analysis was done by finding the most salient point in an image, which was also found within

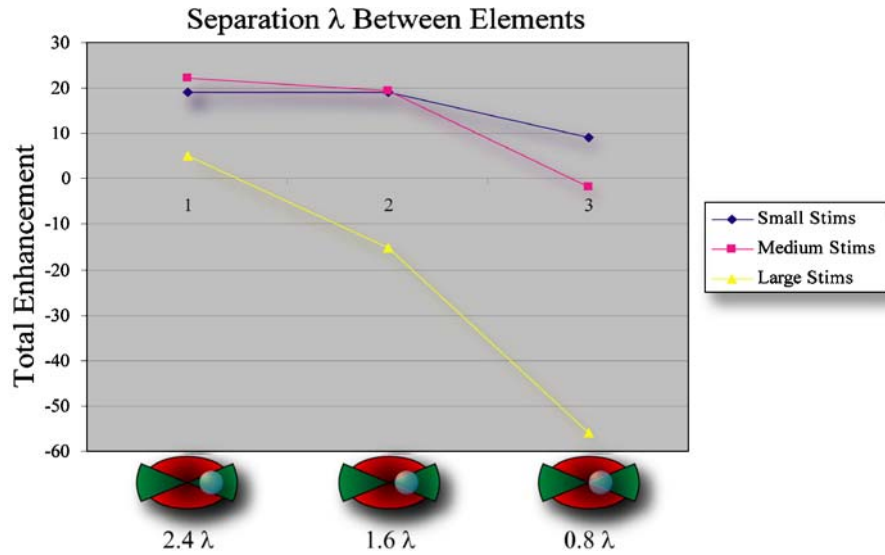


Fig. 11 As a collinear element draws closer, its receptive field begins to overlap another element’s region of surround inhibition (*red*). Here the stimulus element sizes may be compared with the kernel at the 64×64 pixel scale, which are 2.396λ (3 pixels), 4λ (5 pixels) and 5.597λ (7 pixels). The separations for elements shown are at 2.4λ , 1.6λ and 0.8λ . Here we interacted two single elements with a kernel. As elements get larger and closer, it can be seen that enhancement dips. Careful analysis shows that this is due to overlap of elements into inhibition zones, in the surround, as they move closer. Thus, no special kernel, or neural structure is necessary to create inverse enhancement at very close distances between two elements. This explains the dip in enhancement at close distances observed in CINNIC and by Polat and Sagi

the foreground element mask. The rank of the most salient point, also within the mask, was the rank given to the image. For instance, if the most salient point CINNIC found that also corresponded to a real contour element as indicated by the mask was the second most salient point, that image was given a rank of second. The number of images of each rank was summed to find out, for instance, how many images had their most salient point also lie within the mask (ranked as 1st). Figure 12 illustrates how images looked and the subsequent saliency map looked after processing as compared with an example of the masks used to rank the contour images.

As can be seen in Fig. 13, for the larger Gabors of size 120, the top five most salient points fall on a contour in a minimum of 95 of 100 images for all conditions. For half the conditions, all 100 images have a top five salient point falling on the contour. Further, we analyzed the probability of obtaining these results at random. This was done by counting the number of pixels in the mask and the number of pixels not in the mask. This determined the probability at random of a salient point falling on the mask. Given 100 images and five samples per image we could then use a Bernoulli binomial probability distribution and ascertain the probability of our results. This was done using Eq. 16 where from Hayes (1994, p 139), in sampling from a stationary Bernoulli process, with the probability of a success equal to p , the probability of achieving exactly r successes in N independent trials is:

$$p(r \text{ successes}; N, p) = \binom{N}{r} p^r q^{N-r} \quad (16)$$

From Table 1 we see that the p of obtaining these results at random for larger Gabors is at maximum 3.1×10^{-05} . The

results for smaller Gabors of size 70 is not as potent. The top five salient points fall on a contour element between 75% and 80% of the time. However, the probability of obtaining these results is still very small and is at a maximum of $p 1.3 \times 10^{-05}$ for conditions where foreground element separation ranges between 1.5λ to 5.5λ . Only in the condition at a separation of 6λ do the results come out as non-significant at a p of 0.078. This is understandable since at larger separations of foreground elements, detectability of contours seems to become less tangible as can be seen in Fig. 14.

A question raised by our results is that of why there seems to be an optimal separation distance in the data while an optimal distance is not explicitly defined in the neural connection weights. This is due to two factors. The first as explained in our first experiment is that as elements get too close, they tend to inhibit each other as the elements overlap with inhibitory regions. The second seems to be that group suppression begins to over activate and has a greater likelihood of treating real foreground contour Gabors as noise background Gabors. That is, at closer distances, the gain for a foreground Gabor may be high enough to trip its own suppression. This we believe creates the slight dip in the Gabor size 70 results. Additionally, suppression from over facilitation of local Gabor elements should be expected since it has been found to exist by neurophysical experiments (Polat et al. 1998). The final tapering off on the size 70 Gabor results seems to come as the Gabor separation becomes too large for the kernel in the 64×64 pixel scale sub-corresponding group to connect them. Thus, at 5.14λ , the first kernel can no longer bridge between two Gabor elements and its stimulus ends all together in the final saliency map (Fig. 15).

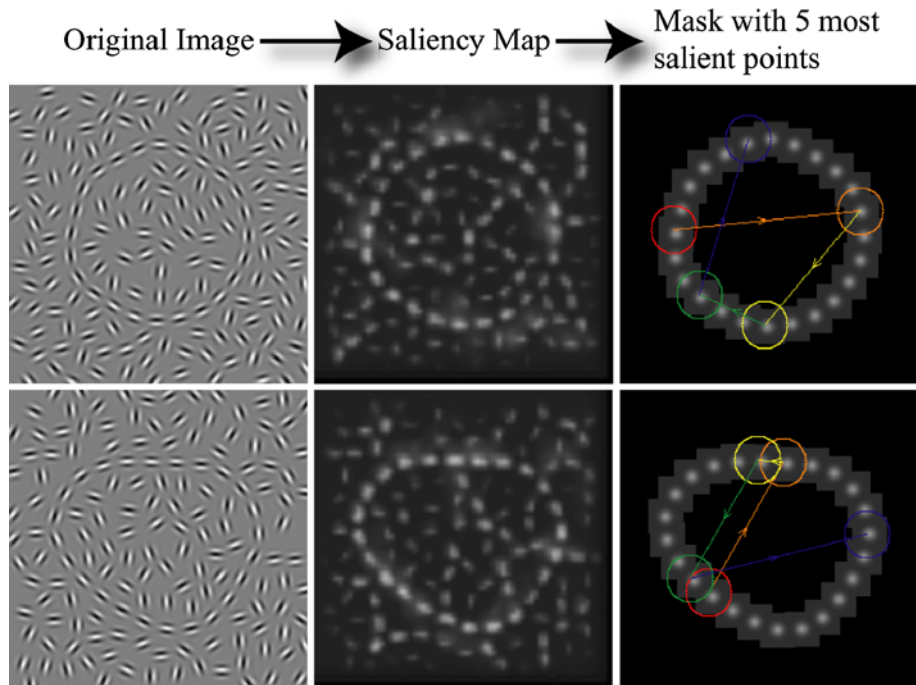


Fig. 12 Input images created by Make Snake are run through CINNIC. The output saliency map is processed to find the five most salient points. These five points are compared with a mask that represents the position of foreground contour elements. This allows the ground truth for such images to be determined with greater ability since foreground elements are controlled

It should also be noted that using this same display Braun (1999) noticed that one of the two subjects showed slightly improved threshold when the ratio of foreground element distance λ to background distance λ was increased from 1 to 1.25. The ratio 1.25 corresponds to 3λ to 4λ of foreground separation in our results, which is slightly less than the peak at 4.5λ in the data presented in Table 1. That is, our results peak near a ratio of 1.25. As such it is not a perfect fit, but it does display an increase of enhancement at about the same ratio and drops off near a ratio of 1.6, which is between 3.8λ and 5.1λ . This corresponds with the drop off in threshold of human subjects, which occurs at a ratio of about 1.6. As such, enhancement of contours by CINNIC is within a similar range for drop off in threshold observed in human subjects.

3.3 Sensitivity to non-contour elements

3.3.1 Sensitivity to junctions

In addition to selectivity for contour elements we have found that CINNIC is sensitive for junctions and conditionally for end-stops which has been described in the visual cortex (Gilbert 1994). This is important since junctions seem to hold important visual information, especially for reconstruction of geometric interpretation of objects (Rubin 2001; Biederman et al. 1999). For instance, following a Geon theoretical construct for object identification, simple lines without junctions may lack certain necessary information since it may be harder to determine where line segments connect to each

other. However, junctions hold more information than single lines since they contain the line projections as well as the determined junctions. Thus, a junction is a line plus its intersection and thus holds more information.

It is also interesting to note this sensitivity to junctions since it creates a possibility that the mechanisms described in this paper are generic enough to be applied to not only contour finding, but junction finding as well. That is, it is interesting to think that only mild augmentation of a corresponding group can change it from a contour detector to a junction detector or that one corresponding group may detect both junctions and contours at the same time. From a functionally simplistic standpoint this is an attractive idea. Especially since the most interesting junctions are probably found at the end of longer contours rather than shorter contours, such a synergy may also prove advantageous. For instance, when not wanting to walk into a desk, the corners and the center of the contour edges are very important to notice.

CINNIC was not designed explicitly to filter for junctions and end-stops. However, analysis of processed images seemed to reveal this ability as can be seen in Fig. 16. For *plus* and *T*-junctions it is easy to show that CINNIC should be sensitive to these type of image features. This is because CINNIC was designed without orthogonal suppression. Thus, two orthogonal lines will not cancel out. Additionally, since two orthogonal lines are processed in two separate layers in the corresponding group which are summed, the junction of two line segments are additive. This can be seen in Eq. 8. Thus, if each pixel element in two intersecting lines is equal to

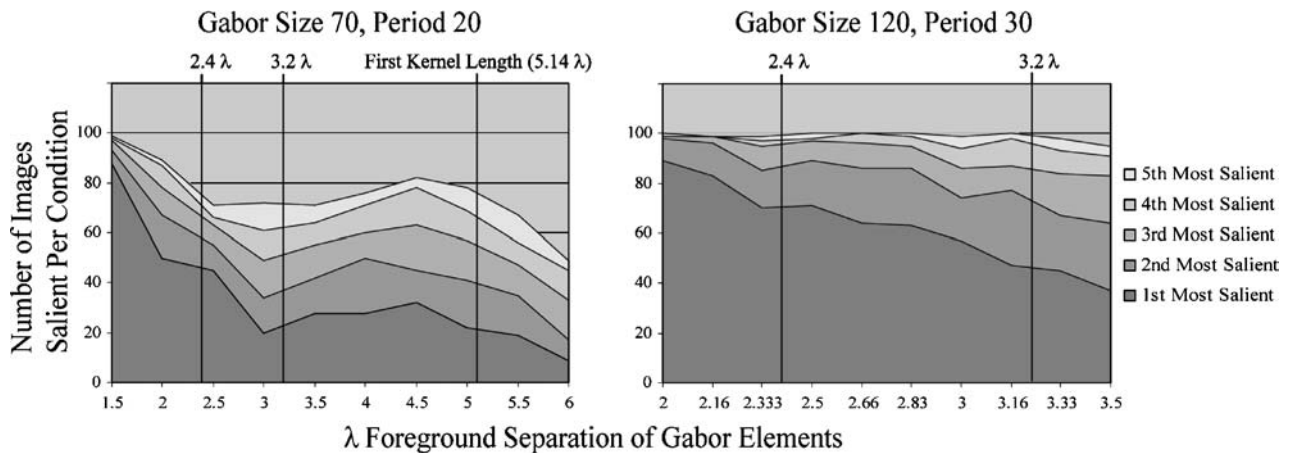


Fig. 13 The results of from processing 2,000 images from Make Snake by CINNIC are shown. The sum of all images where the most salient point was on a foreground contour is shown in dark gray for each of the λ separation conditions. In the experiment all images where the second most salient point was on a foreground element but the first was not are labeled second and are in a lighter shade of gray. In each condition, the general saliency result can be seen by summing the number of images where a foreground element is among the five most salient points found. At separations between 2.4λ and 3.2λ foreground and background element separation is about the same. At 5.14λ , elements fall beyond the reach of enhancement defined by the finest resolution kernel. Thus, we expect to begin to see a drop off here. There is a slight pick up in enhancement between 3.2λ and 5.14λ perhaps due to optimal separation where elements do not overlap each other's inhibition regions.

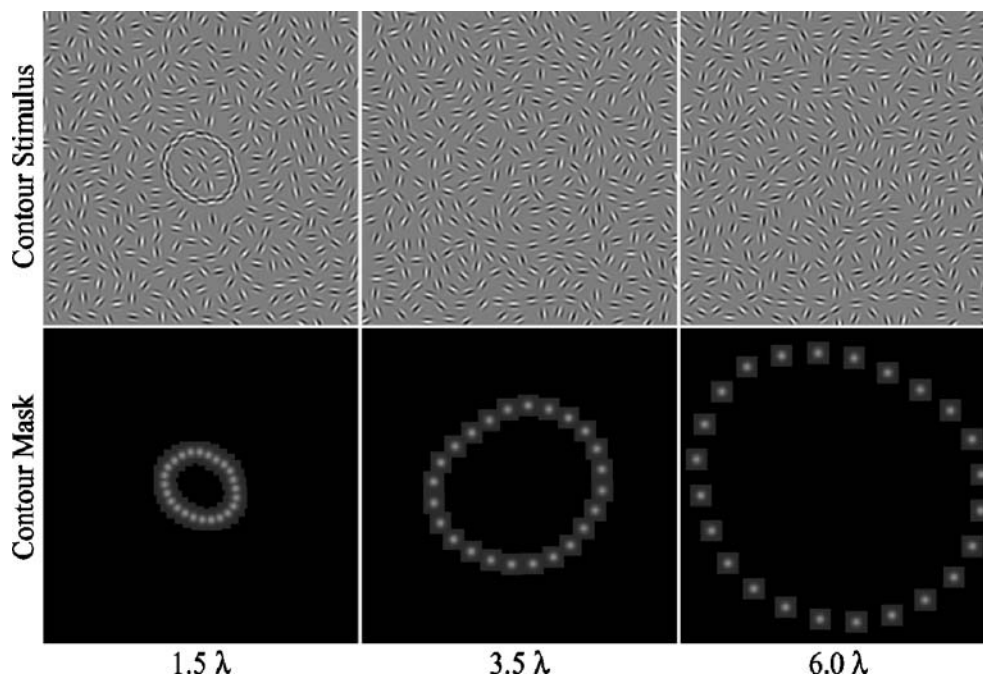


Fig. 14 The declining performance of CINNIC at increasing λ separation is easy to understand by inspecting the contour images at 1.5λ , 3.5λ and 6.0λ of foreground separation. Casual observation shows that saliency decreases with larger separation of contour elements. At 6.0λ contour elements are almost invisible

one, the saliency map at the point of intersection would be equal to two. This can also be seen for T-junctions. Again, the enhancement of the junction should be 1.5 times that of elements on either line segment. This is because a half line segment that joins a full line segment should enhance less than a full line segment. Thus a T-junction would intuitively have 1.5 times the excitation of a single line rather than 2 times for a plus junction.

Another interesting facet of these results is that they suggest a possible explanation for the reduced enhancement when the gestalt continuity of a line is violated. For instance studies have shown that when a line is presented with two flanking lines its enhancement is greater than if one of the flankers is in the shape of a T (Kapadia et al. 1995). Such a result might be predicted by our model since the flanking T would then be promoted to have a higher saliency value

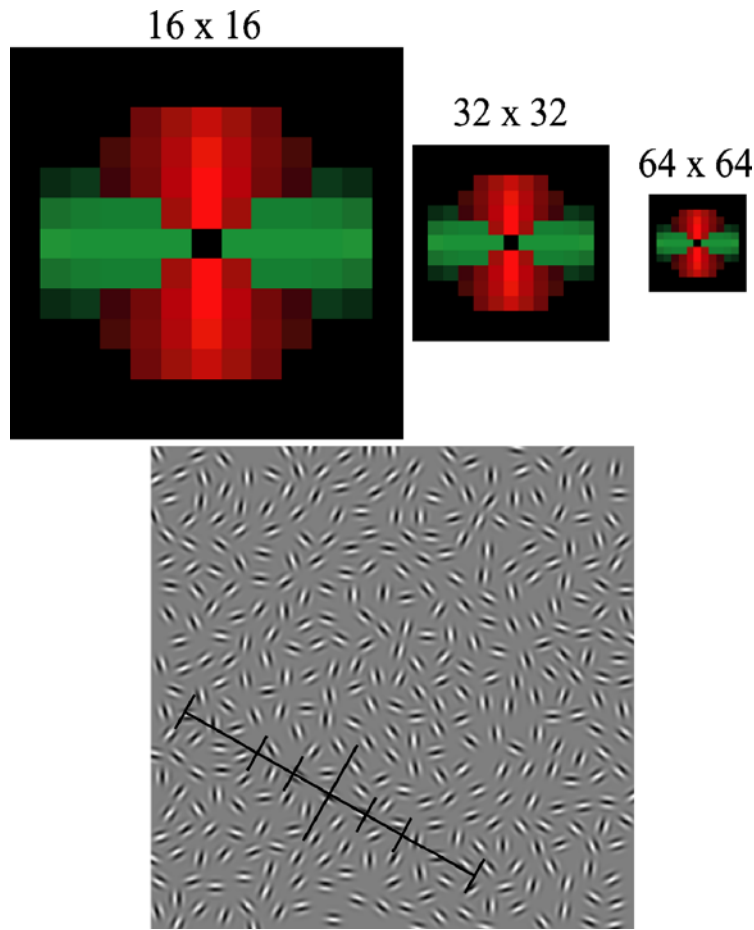


Fig. 15 The size of kernels at each of the three scales is shown compared with Make Snake image. The line on the Make Snake image shows the width of each kernel for close reference against an image with foreground separation of 4.5λ , which is the same separation as the peak observed in Fig. 13. As can be seen, when the image is reduced to 16×16 , the kernel stretches across much of the image, but with little specificity of effect on the image due to the scale reduction

Table 1 As λ separation increases between foreground elements, saliency decreases

Gabor size 70, period 20			Gabor size 120, period 30		
λ separation	Salient images	p	λ separation	Salient images	p
1.5	99	2.3×10^{-99}	2	100	2.5×10^{-32}
2	89	8.6×10^{-48}	2.16	99	4.1×10^{-26}
2.5	71	1.0×10^{-19}	2.333	99	4.2×10^{-22}
3	72	4.0×10^{-15}	2.5	100	4.2×10^{-21}
3.5	71	1.3×10^{-11}	2.66	100	8.3×10^{-19}
4	76	6.8×10^{-13}	2.83	100	1.0×10^{-16}
4.5	82	2.8×10^{-16}	3	99	6.3×10^{-13}
5	78	2.8×10^{-12}	3.16	100	6.6×10^{-13}
5.5	67	1.3×10^{-05}	3.33	98	6.6×10^{-09}
6	49	0.078	3.5	95	3.1×10^{-05}

For the smaller Gabor sized image, around 75% of all images with a foreground separation of 1.5 to 5λ have a foreground element as one of the top five most salient. The probability of obtaining such a result at random is less than .005 percent. For images with larger Gabor elements, almost all the images contain a foreground element that is highly salient. Again the probability is very low suggesting that the null hypothesis should be rejected

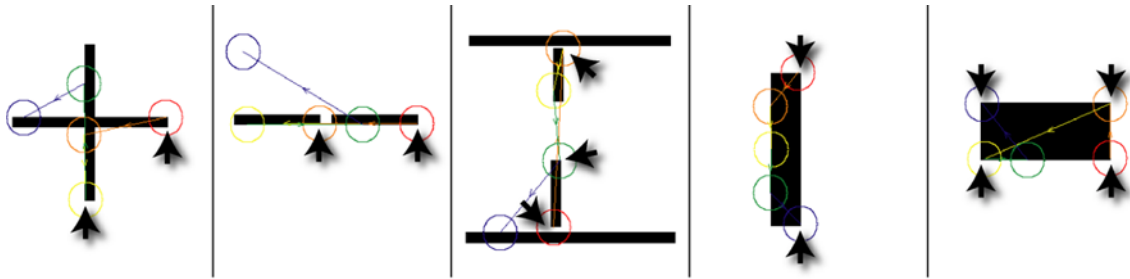


Fig. 16 The five images shown above demonstrate CINNIC's sensitivity to junctions in the elemental shapes seen. Here, the most salient point is always on a junction (*red circle*) and there is always another point of very high saliency (in the top 5) on a junction. When not falling on a junction, the most salient point is near the center between two junctions which is quite possibly the second most important part to find salient. Some of the anomalies observed such as a saliency point in blank space are due to the algorithm blanking out the saliency map as it selects points to prevent it from picking the same point more than once

than the central stimulus element. As such, continuity is not broken by suppression from the T so much as it is broken by having a lower saliency than the T. It should be noted additionally that we do not predict enhancement of a central line element with two orthogonally oriented elements if they are divided by a large enough gap. That is, it is important to note that enhancement of junctions here likely relies on a joint overlap of orthogonal lines.

While the evidence for plus and T-junctions is intuitive, it is not so much for L junctions. Thus, we tested L junctions against the CINNIC kernel. In this case we assumed perfect colinearity on the kernel. This allows us to test elements against only one kernel slice, which keeps analysis much simpler. Figure 17 shows the results of passing two types of L joints in front of the CINNIC kernel. Each L joint can be thought of as infinitely long. That is, the end-stops on the L junctions will never pass in front of the kernel. Two types of L junction line segments are used. One is a two pixel wide line while the other is one pixel wide. To determine the enhancement of a junction, we compare the enhancement of the pixel that lies on the junction compared with other pixels on the line. That is, we move the L over the kernel. Then each pixel will report some enhancement level. If CINNIC could have sensitivity to junctions, we would expect that the junction pixel would be more enhanced than other pixels on the line not on the junction.

For the one pixel width ($0.12\text{--}0.46^\circ$ of visual field depending on the image scale), it can be seen that the kernel will enhance the junction pixel more strongly than neighboring pixels along the line as far away as 5 pixels ($0.575^\circ\text{--}2.38^\circ$). When the kernel is moved to a point, 6 pixels ($0.69\text{--}2.78^\circ$) in distance from the junction pixel, enhancement is the same as for the junction pixel. This can be considered intuitively this way: a line segment that is half way through the kernel will enhance one half as much as a full line passing all the way through the kernel. However, at the junction, two halves sum to the same enhancement as a full line. Thus, by the 6th pixel in, enhancement is the same since the junction has moved outside of the kernels field and is now essentially a simple bar. So for any L junction, enhancement will be higher at the junction pixel than any other part in the line segment

for a radius of 5 pixels. Very similar results are found with L junctions of width 2 ($0.23^\circ\text{--}0.93^\circ$). However, the maximal enhancement is found at the inner elbow junction and not the outer junction. That is, an L junction of two pixels in width has two pixel junctions, one on the inside and the other on the outside of the joint. The inner junction seems to have more enhancement for a radius of 5 pixels.

Since the enhancement of the junction is isolated, this means that even if it has a similar enhancement of a line segment six pixels in, it may be enhanced more since it will not push the local region activity higher and increase the group suppression. Thus, enhanced lines are more likely to create levels of excitement that will trip group suppression than junctions, which are more isolated in their activity. From this it might be hypothesized that group suppression may aid in the discovery of L junctions in CINNIC.

3.3.2 Conditional sensitivity to end-stops

Using the same procedure for analysis of L junctions, we checked the sensitivity of CINNIC to end-stops. We found that there was some elevated sensitivity to end-stops, but only under certain conditions. Three conditions were tested. The first involved the outline of bars. Enhancement was tested for the junction area on the outline of a bar versus an edge in the middle of the bar. The results in Fig. 17 show that when the bar is wide enough in width, sensitivity is increased for the end-stop junction. Additionally, this affect is increased as group suppression effects are added. Thus, the junctions on the end of bars are enhanced over elements in the middle of the bar by even greater amounts as group suppression is added. Further, the bar of width 4 ($0.46^\circ\text{--}1.86^\circ$) becomes stronger than a middle segment when group suppression reaches 50% above normal.

The second and third test involved passing a bar of width 2, in front of the Kernel. As can be seen the second bar was sharply pointed at its tip Fig. 17. The kernel shows no enhancement for the end of the plain bar even if group suppression is increased to 250%. However, for the pointed bar, enhancement is seen over the other 4 segments tested once group suppression reaches 150% above normal.

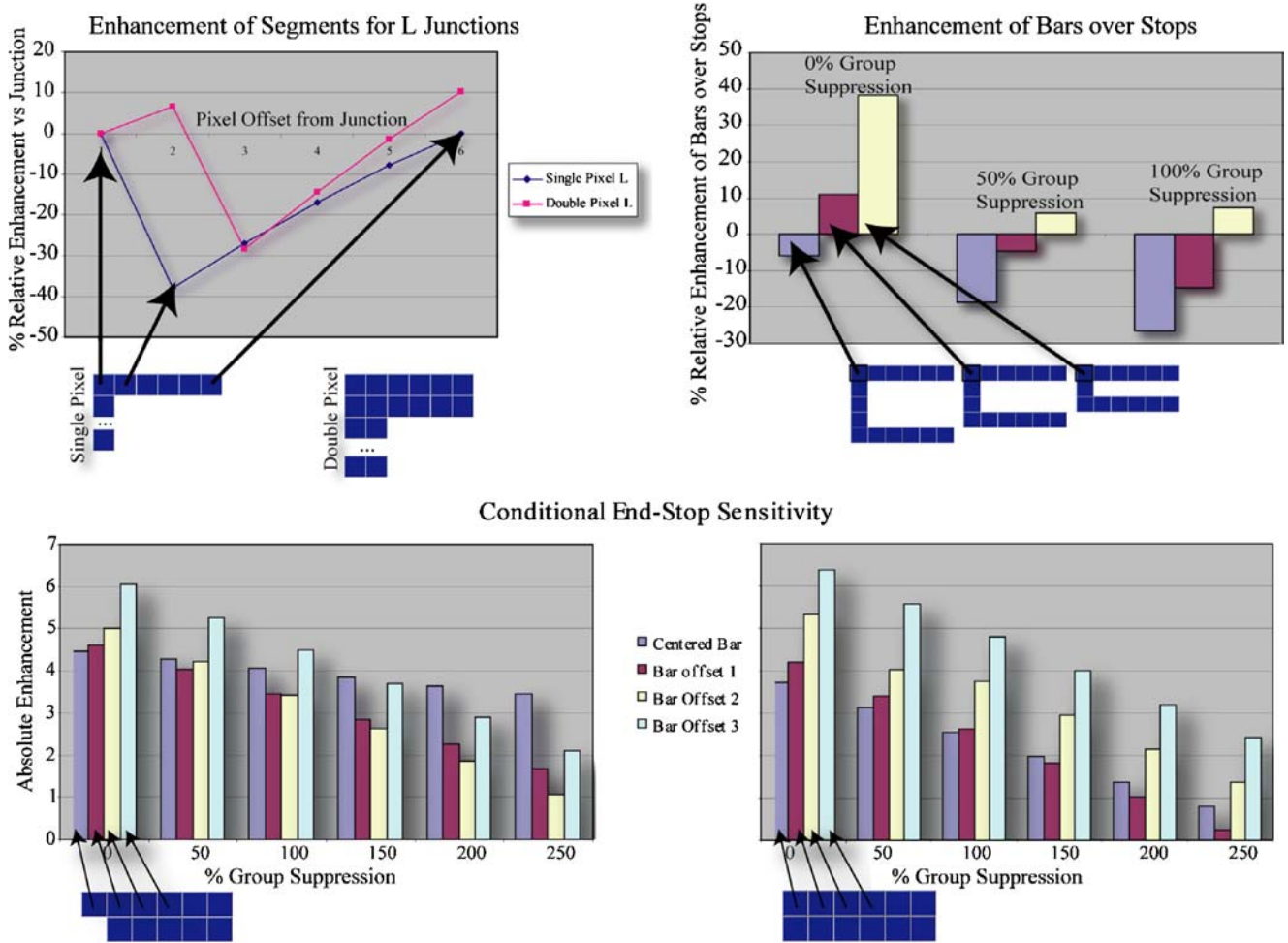


Fig. 17 These graphs show enhancement of pixels from an image when convolved with an orthogonal slice from the CINNIC kernel. As can be seen, in the *top left graph*, the corners on L junctions, both 1 and 2 pixels wide, are enhanced more than their neighbor pixels and other pixels along the L out to a distance of 4.8λ . Additionally, in the *top right*, we can see that the corners on *bars* are enhanced over pixels outside of their receptive field ($>4.8 \lambda$) along the same bar as the two parallel edges are separated and additionally as group suppression is added. The bottom row shows that end-stops with a point are not enhanced at base group suppression, but as suppression is added, the end point overtakes its three closest neighbors (0.8λ , 1.6λ , 2.4λ) when group suppression reaches 200%. This effect is not seen for the non-pointed bar. Thus, the current version of CINNIC is only conditionally sensitive to end-stops. Note, each pixel corresponds to a width of 0.8λ with the 64×64 scale kernel

Thus it can be seen that CINNIC has sensitivity for some types of end-stops. This agrees well with research on V1 neurons which shows that most neurons there have some sensitivity to end-stops (Jones et al. 2001; Sceniak et al. 2001; Pack et al. 2003). Additionally it follows a very similar pattern of behavior seen in end-stop neurons in the cat visual cortex. In this case, end-stop sensitive neurons were found to detect end-stops after an initial saturation period. Thus, the neurons for a brief interval (<30 ms) were sensitive to non-end-stopped elements, but built up to end-stop sensitivity (Pack et al. 2003). Our model agrees with these observations since build up of group suppression increases end-stop detection and would also create a delay for such detection as suppression builds. This is also similar to the model by Rao and Ballard (1999), which used a predictive feedback suppression mechanism to facilitate end-stop detection.

However, the primary difference is that suppression in CINNIC comes from activity in the corresponding group and not from a higher level process.

3.4 Real world image testing

Real world testing was conducted by inspecting the output of CINNIC on 132 real world images. We did this by inspecting each image and cataloging the results by hand. This was done due to the fact that classifying contours in an image *a priori* is extremely difficult due to the subjective nature of classifying image elements in a natural image. However, this has a new subjective drawback in that the efficacy is based on a post hoc analysis, which may carry a different expectation bias. In either case, the results are difficult to not bias. Either

Table 2 *Post Hoc* analysis of CINNIC for its sensitivity to certain kinds of features again suggests that it is not only sensitive to contours, but junctions as well. This can be seen as the most salient point in 42% of random real world images analyzed lies on a contour junction. Prior probability is not supplied since it is not known by us what the real incidence of contour junctions is in real world images. Thus, the true posterior significance is unknown

Type of feature	Number	Likelihood
Contours, no junctions	46	0.348
Contours, with junctions	56	0.424
Contours, end-stops	13	0.098
Contours, short	10	0.075
None	7	0.053
Total	132	1.0

the experimenter subjectively leaves out or includes contours before the analysis is done, or the experimenter sees results in the post analysis due to a personal bias. The latter approach provides room for analytical reanalysis and careful evaluation which we believe to be a strength in a situation where there seems to be a bias no matter which method is used. It is hoped that the reader will consider the previously presented material as showing some of the more controlled evidence of the model's efficacy and take this as evidence of real world applicability.

Each image was analyzed for salient content of contours, junctions, end-stops and short contours, which often include, for example, eyes or mouths. For each image it was noted what the nature of the most salient location was. So for instance, if the most salient contour location was on a junction, then that image was counted as having its most salient contour on a junction. Each image was thus counted in one of five exclusive groups (a) contours without a junction (b) junctions between two contours (c) end-stopped points from a contour (d) short contours that tended to be eyes and mouths or (e) none of the above, which tended to mean it was a poor result. Table 2 shows the results. As can be seen, these results agreed with the analysis provided from junctions. In essence, it was observed that most of the top salient locations, as determined by CINNIC, seem to lie on a junction. Additionally, the conditional end-stop sensitivity can be seen in about 10% of all real world images. Thus, CINNIC has a strong sensitivity to contours at junction points and additionally has some sensitivity to end-stops, which is to be expected since most neurons in V1 have some end-stop sensitivity.

Since there seems not to be any studies which suggest the real prior probability of junctions in natural images, we are forced to read these results from a worst case hypothetical framework. Thus, the significance of these results may be interpreted as follows, since each junction in an image requires at least one line segment edge pixel, there can never be more junction pixels than contour non-junction pixels. Thus, in a worst-case scenario, at most 50% of all detected contours would be on junctions if the likelihood of falling on a junction versus a non-junction was totally random. However from our image analysis, contour junctions are more likely to be detected as the most salient object in an image

than contours not on a junction. Thus, this analysis again suggests that CINNIC is indeed more sensitive to junctions than contour segments without junctions.

Additionally, it can be seen from Fig. 18¹ that in many images CINNIC finds facial features salient. In the 27 images where human or animal facial features are visible, CINNIC finds 14 to have salient facial features in the top five most salient points. Here we define facial features as noses, mouths, eyes or ears. That means that based on contour analysis alone, half of all faces have a highly salient feature. This suggests that CINNIC may be able to play a role in a face finding algorithm. It also suggests that contour integration mechanisms may be involved in a dual role that includes not only landscape contour finding but face finding as well. Here CINNIC seems sensitive to facial features such as short contours since they are isolated from other similar parallel lines on smooth faces. Thus, even though they are short, they are not suppressed by anything else.

The reason why we believe that face feature finding is interesting in that it suggests that CINNIC may approximate more generic mechanisms in visual cortex, and as such may be a closer fit to what processes actually occur in the brain. For instance, it is suggested that the interaction of simple horizontal and vertical lines derived from important facial objects such as eyes and noses play a part in facial categorization (Peters et al. 2003). If this is correct then a neural device that finds such features and can describe them in terms of lines may be necessary. Thus, it may be possible to augment the simple butterfly kernel connection with some of the other mechanisms described here to find a variety of different useful features.

4 Discussion

The CINNIC model performs contour integration and seems to satisfy the criteria of its design. First it uses simple biologically plausible mechanisms for its actions. Second, it performs its action with enough speed that a real time implementation is within our grasp. Third, it helps to illuminate what processes are at work in human contour integration and fourth, current examination of CINNIC show its performance to be within parameters of human contour integration as shown from psychophysical data.

The model is biologically plausible because all neural connections within the network are of types that are known to exist in the human brain. For instance, no neuron should connect to any neuron that is outside its reach. This means that no global mechanisms were introduced to control the gain of the network. Indeed, each neuron is independent from any other neuron for which it is not connected from its kernel interactions or through its group suppression. Our model then uses dopamine-like priming to connect neurons that do not directly connect. While this may not have been directly observed in V1, the actions of dopamine priming as well as

¹All results for real world images may be viewed at <http://www.cinnic.org>

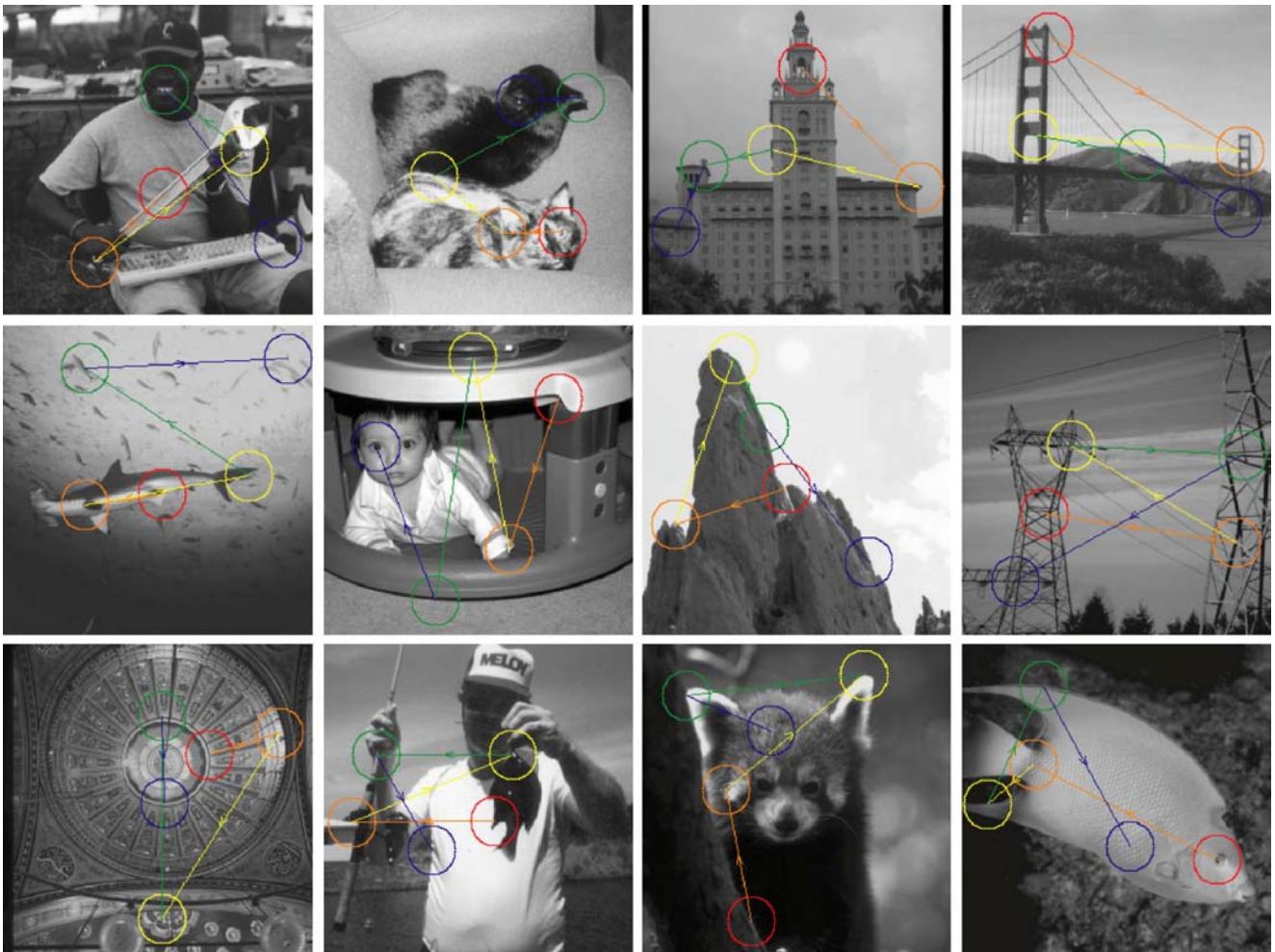


Fig. 18 The five most salient points are shown in 12 real world images processed by CINNIC (*red is most salient, next is orange etc.*). Notice the prevalence of representation by facial features, junctions and end-stops

other types of priming are well known to exist in the human brain (Schultz 2002). Other models have explained linking using neural synchronization. While this has been observed in human neural networks, its observation and importance in the neocortex has been open for debate.

Further, other computational models have shown dopamine modulation to be effective at linking sequences (Suri et al. 2001). Since visual contours are spatial sequences, this would show yet another way in which dopamine-like priming would be feasible in the long-range connection of contours. More evidence for the dopamine-priming hypothesis can be seen in the degradation of contour integration in patients with schizophrenia (Silverstein et al. 2000). This lends support to a dopamine hypothesis since dopamine, is one of the neurotransmitters suspected of playing a major role in schizophrenia (Kapur and Mamo 2003), with such an effect seen in striatal dopamine neurons as well (Laruelle et al. 2003).

The group suppression we have used is also plausible because GABAergic interneurons of many types are found

throughout the brain. Interneurons are also known to connect to many neurons at the same time, sending inhibitory synaptic currents to a possibly large population of pyramidal dopamine neurons (Durstewitz et al. 2000; Gao and Goldman-Rakic 2003). The firing of these neurons has also been shown to have dramatic effects on the neurons they connect since they can exhibit spikes at very high rates (100 Hz) (Bracci et al. 2003) and can have low firing thresholds as well as a need for few inputs (Krimer and Goldman-Rakic 2001). Also, the group suppression in our model uses an axonal reach that is about the same size as the reach for pyramidal neurons created by our kernel. Thus, it fits well within spatial constraints.

It should also be noted that another feature which makes our model unique is that it not only works in saliency for contours, but also for junctions. As mentioned this was an unexpected result. However, it is very interesting for several reasons. The first is that it suggests that V1 and V2 neurons can have dual or multiple roles and that the fea-

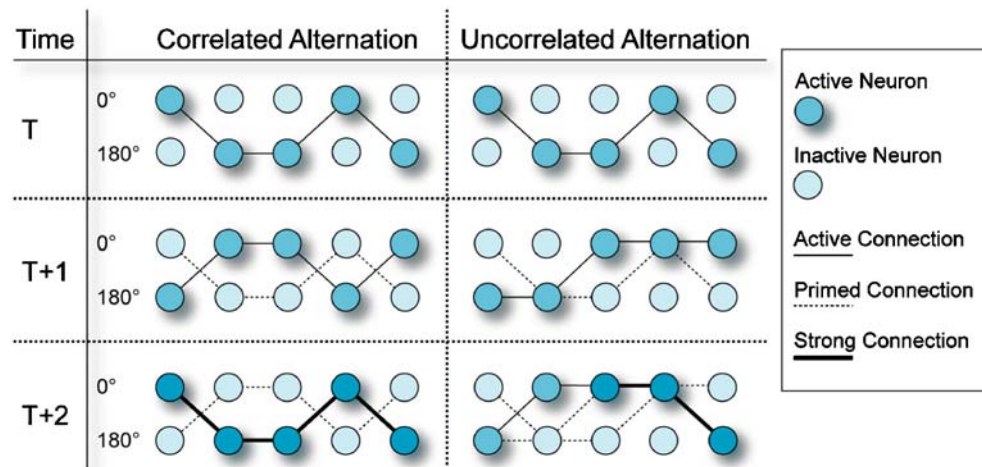


Fig. 19 Temporal grouping can be explained by fast-plasticity mechanisms. If alternation is strongly correlated then plastic connections are strong and less ambiguous, also by the second alternation, all connections are primed unlike uncorrelated alternation where only some connections are primed. As such, correlated temporal alternation would facilitate neurons more strongly than a less correlated temporal structure if it used fast-plasticity based priming

ture detection dimensionality within various processing units in visual cortex may be higher than is generally considered the case. Thus, following the logic behind the utilization of Gabors in vision, neural structures may exist, which have a broad utility. The structure for CINNIC may shed light on a structure that allows neurons to become sensitive to many different visual features but yet not be exotic from each other. That is, contours, end-stops and junctions may be detected by the same mechanisms, but the detectors are different due to subtle variations such that a base neuron is taken in infancy and morphed subtly to its new function through learning. However, a morphed neurons structure is still very similar to its original structure and is similar to other feature detectors that operate on seemingly unrelated features. Such a theory would be in agreement with observations that natural images can be described with a relatively small number of Gabor derived kernels very efficiently (Olshausen and Field 1996). As such one might expect the flora of feature detectors to be somewhat constrained at this level of cortex.

Additionally, the analysis here lends support to the notion of the importance of the temporal domain in perceptual dimensionality. That is, as has been suggested, (Prodöhl et al. 2003) perception may not just be a matter of the 3D structure of neurons, but may also hinge on the pattern of the working of neurons. As such, an end-stop detector is only an end-stop detector after a certain interval of suppression from interneurons. Prior to that its role may be different and it may be a simple contour detector. Since most neurons in V1 show end-stop sensitivity and end stopped neurons take extra time to register those end-stops, it seems feasible that a neuron may detect different features at different times.

4.1 Extending Dopamine to temporal contours via TD (dimensions)

In addition to static contours, dynamic contours may also be enhanced by mechanisms of fast plasticity. For example, covert object tracking (in the absence of eye movements) could be enhanced by similar mechanisms as have been proposed here. This can be hypothesized since any neuron that receives an input in our model will attempt to prime its neighbors. When an object moves to the next neuron, it maintains a saliency enhancement (imagine the phosphors on an old TV still glowing in a trail as a dot moves across the screen). Additionally, neurons along the trajectory of the object will receive the greatest enhancement, which will maintain the saliency along that path. Because of Dopamine's involvement in fast temporal difference correlation (Suri et al. 2001; Schultz 2002) it may be a natural candidate for such actions. Thus, the key to understanding temporal contours and smooth pursuit may merely lie with the basic contour integration mechanisms.

Additionally, it is easy to imagine that the dopamine-like priming mechanism we have hypothesized here not only enhances contours, but may play an integral part in training the system in a similar manner as suggested in Rao and Ballard (1999). For instance, it has been proposed that observed movement of objects trains neurons to recognize contours (Prodöhl et al. 2003). As such, following our hypothesis, a dopamine-like priming may not only enhance contours, but may train contour integrating neurons. Since dopamine is known to play a role in reinforcement learning (Suri et al. 2001; Schultz 2002) it is an excellent candidate for such a mechanism, and since it is already in place for the purpose of learning, an occam's razor reasoning would state that if it can also fulfill the role of nonlocal interaction for contour

integration, it is the most reasonable candidate to do so since that would be the simplest explanation.

4.2 Explaining visual neural synchronization with fast plasticity

It is important to note that temporal synchronization in vision does not necessitate correlated firing as a cause. For instance, Lee and Blake (2001) observed that alternating motion of Gabor patches allowed greater facilitation of contours if the Gabor motion alternates in a correlated manner. That is, they displayed Gabor contour patterns much like the Make Snake patterns. However, the Gabors were given visual motion by changing the wave phase in a direction that created an orthogonal motion to the Gabor patches. The direction of the motion was randomized, but switching the direction of Gabor elements could be correlated. As such, in the highly correlated condition direction was shifted simultaneously while in the low correlation condition switching was somewhat random. Facilitation was observed when switching was correlated.

We believe this can be explained by fast plasticity as follows. Due to collinear relation, neurons with different motion sensitivities will prime. For instance, two collinear Gabor patches, one moving in the direction of 0° and one moving in the direction of 180° will prime neurons in a hebbian fashion. When the Gabors switch, two completely different sets of motion sensitive neurons will prime. Through this alternation, it will create two sets of mutually exclusive linked sets. By removing correlation it will begin to create cross-linked pairs of neurons and increase the number of primed synapses which will increase noise in the network. As such the more synchronous the alternation of motion is, the more crisp the plastic connections will be (Fig. 19).

4.3 Contours + Junctions, opening a new dimension on visual cortex

The research thus far agrees with work to date that suggests that V1 neurons are extremely powerful for extracting data from a scene (Olshausen and Field 1996). Additionally, it also helps to validate hypotheses that suggest neurons in V1 have a high dimensionality for visual processing. That is, a neural group may not be responsible for just sensitivity to one feature, but may have sensitivity to multiple features. Additionally, interaction between partially sensitive neurons may create complete sensitivity. So for instance, if two or more groups have some sensitivity to end-stops then, the combination of their sensitivities may yield full sensitivity to end-stops.

It should also be noted that at least in terms of junctions, one would expect that the same mechanism would be responsible for finding L, T and + shaped junctions. This is due to recent research that suggests that searches for L versus T versus + junctions is inefficient (Wolfe and DiMase 2003). That is, because we are unable to find different types of junctions faster among noise of different junction types. From

a saliency stand point, one would expect that V1 or other saliency centers do not differentiate them and thus, would be explained by the brain using the same mechanism to find junctions irrespective of the type.

4.4 Model limitations

Like most computation models of biological systems CIN-NIC has its limitations. The first is that the model does not include effects on contour integration from color (Mullen et al. 2000). One reason for not accounting for color is that it would most likely add another dimension to the pseudo-convolution computation. That is, in addition to orientation and position as dimensions, color would become a third set making the hyper-kernel six dimensional with the addition of blue-yellow and red-green channels. The model also does not account for enhancement of parallel elements. This, as mentioned previously, is when Gabor elements are aligned like the rungs on a ladder. The primary question on parallel enhancement is where it occurs. For instance, is there a second set of contour integrators for parallel elements or do parallel elements enhance in the exact same corresponding group as collinear elements? Such questions still need to be answered. If they do enhance in the same corresponding group, then the shape of neural receptive fields in a contour integration model may need to be rethought since the classic butterfly shape used in most contour integrators cannot account for such enhancements.

An additional limitation is that inhibition and excitation are treated with temporally similar dynamics at the kernel level. This may be considered a weakness of the model. However, it should be remembered that inhibition does have a build up pattern via the group suppression mechanism. As such, temporal differences between excitation and inhibition mechanisms are partially addressed. Indeed, as mentioned, the key to detection of L junctions and end-stops by contour integrators may be the temporal difference between excitation and inhibition.

5 Conclusion

We believe we have created a reasonable model and simulation of contour integration in visual cortex for saliency. As the results have shown, we have fit the results of human observers to within two standard errors for a single Gabor element with two flankers. We have also achieved reasonable results for images with multiple Gabor elements, which are statistically significant. Taken with our results from real world images we suggest that this makes our model a reasonable approximation of human contour integration. Additionally, we believe that our model demonstrates how the neural mechanisms for contour integration may be extended into other types of feature processing.

Acknowledgements We would like to thank Robert Peters, Jochen Braun, Christof Koch, Vidhya Navalpakkam, Irving Biederman, and

Mike Olson for their invaluable help and suggestions. This research is supported by the National Imagery and Mapping Agency, the National Science Foundation, the National Eye Institute, the Zumberge Faculty Innovation Research Fund, the Charles Lee Powell Foundation and Aerospace Corporation.

Appendix

Parameter values

Max range for collinear separation for excitation	$0^\circ\text{--}31^\circ$
P_2^e (kernel polynomial parameter)	-0.75
P_3^e (kernel polynomial parameter)	0.095
W (kernel inhibition multiplier)	0.65
P_2^s (kernel polynomial parameter)	0.16
P_3^s (kernel polynomial parameter)	-0.1
A (pass through multiplier)	30.0
L (constant leak)	94.0
F (fast plasticity gain)	1.0001
Max group size	768 neurons ($8 \times 8 \times 12$)
T (max group suppression threshold)	50,000
v (group suppression gain)	0.0003
w_u , 64×64 scale weight	0.58
w_u , 32×32 scale weight	0.85
w_u , 16×16 scale weight	0.35

References

- Ben-Shahar O, Zucker S (2004) Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Comput* 16:445–476
- Biederman I, Subramaniam S, Bar M, Kalocsai P, Fiser J (1999) Subordinate-level object classification reexamined. *Psychol Res* 62:131–153
- Bracci E, Centonze D, Bernardi, Calabresi P (2003) Voltage-dependant membrane potential oscillations of rat striatal fast-spiking interneurons. *J Physiol* 549(1):121–130
- Braun J (1999) On detection of salient contours. *Spat Vis* 12(2):211–225
- Burt PJ, Adelson EH (1983) The Laplacian pyramid as a compact image code. *IEEE Trans Commun* 31:532–540
- Choe Y, Mikkilainen R (2004) Contour integration and segmentation with self-organized lateral connections. *Biol Cybern* 90:75–88
- Durstewitz D, Seamans JK, Sejnowski TJ (2000) Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J Neurophysiol* 83(3):1733–1750
- Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for local “association field”. *Vision Res* 33(2):173–193
- Field DJ, Hayes A, Hess RF (2000) The roles of polarity and symmetry in the perceptual grouping of contour fragments. *Spat Vis* 13(1):51–66
- Freeman E, Driver J, Sagi D, Zhaoping L (2003) Top-down modulation of lateral interactions in early vision: does attention affect integration of the whole or just perception of parts. *Curr Biol* 13:985–989
- Gao W, Goldman-Rakic PS (2003) Selective modulation of excitatory and inhibitory microcircuits. *PNAS* 100(5):2836–2841
- Gilbert CD (1994) Circuitry, architecture and functional dynamics of visual cortex. In: Bock GR, Goode JA (eds), *Higher-order processing in the visual system* (Ciba Foundation symposium 184), Wiley, Chichester, pp 35–62
- Gilbert CD, Das A, Ito M, Kapadia M, Westheimer G (1996) Spatial integration and cortical dynamics. *PNAS* 93:615–622
- Gilbert CD, Ito M, Kapadia M, Westheimer G (2000) Interactions between attention, context and learning in primary visual cortex. *Vision Res* 40:1217–1226
- Guy G, Medioni G (1993) Inferring global perceptual contours from local features. In: *Proceedings IEEE CVPR* 786–787
- Grigorescu C, Petkov N, Westenberg MA (2003) Contour detection based on non-classical receptive field inhibition. *IEEE Trans image process* 12(7):729–739
- Hayes WL (1994) *Statistics*, 5th edition. Harcourt Brace, Fort Worth
- Hempele CM, Hartman KH, Wang X-J, Turrigiano GG, Nelson SB (2000) Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex. *J Neurophysiol* 83:3031–3041
- Hess R, Field D (1999) Integration of contours: new insight. *Trends Cogn Sci* 3(12):480–486
- Hubel D, Weisel T (1977) Functional architecture of macaque monkey visual cortex. *Proc R Soc London Ser B* 198:1–59
- Itti L, Koch C, Braun J (2000) Revisiting spatial vision: towards a unifying model. *J Opt Soc Am JOSA-A* 17(11):1899–1917
- Itti L, Koch C (2001) Computational modeling of visual attention. *Nat Rev Neurosci* 2(3):194–203
- James W (1890) *Princ Psychol*. Harvard University Press, Cambridge
- Jones HE, Grieve KL, Wang W, Silito AM (2001) Surround suppression in primate V1. *J Neurophysiol* 86:2011–2028
- Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–856
- Kapadia MK, Westheimer G, Gilbert CD (2000) Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J Neurophysiol* 84:2048–2062
- Kapur S, Mamo D (2003) Half a century of antipsychotics and still a central role for dopamine D2 receptors. *Prog Neuropsychopharmacol Biol Psychiatry* 27(7):1081–1090
- Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiology* 4(4):219–227
- Koffka K (1935) *Princ Gestalt Psychol*, Lund Humphries, London
- Kovács I, Julesz B (1993) A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *PNAS* 90:7495–7497
- Krimer LS, Goldman-Rakic PS (2001) Prefrontal microcircuits: membrane properties and excitatory input of local, medium and wide arbor interneurons. *J Neurosci* 21(11):3788–3796
- Laruelle M, Kegeles LS, Abi-Dargham A (2003) Glutamate, dopamine, and schizophrenia: from pathophysiology to treatment. *Ann N Y Acad Sci* 1003:138–158
- Lee SH, Blake R (2001) Neural synergy in visual grouping: when good continuation meets common fate. *Vis Res* 41:2057–2064
- Li W, Gilbert CD (2002) Global contour saliency and local collinear interactions. *J Neurophysiol* 88:2846–2856
- Li Z (1998) A neural model of contour integration in the primary visual cortex. *Neural Comput* 10:903–940
- Miniussi C, Rao A, Nobre AC (2002) Watching where you look: modulation of visual processing of foveal stimuli by spatial attention. *Neuropsychologia* 40(13):2448–2460
- Mullen KT, Beaudot WH, McIlhagga WH (2000) Contour integration in color vision: a common process for the blue-yellow, red-green and luminance mechanisms? *Vision Res* 40:639–655
- Mundhenk TN, Itti L (2003) CINNIC, a new computational algorithm for modeling of early visual contour integration in humans. *Neurocomputing* 52–54:599–604
- Navalpakkam V, Itti L (2002) A goal oriented attention guidance model. *Lect Notes Comput Sci* 2525:453–461
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature* 381:607–609
- Pack CC, Livingstone MS, Duffy KR, Born RT (2003) End-stopping and the aperture problem: two-dimensional motion signals in Macaque V1. *Neuron* 39:671–680

- Pernberg J, Jirrmann KU, Eysel UT (1998) Structure and dynamics of receptive fields in the visual cortex of the cat (area 18) and the influence of GABAergic inhibition. *Eur J Neurosci* 10(12):3596–3606
- Peters RJ, Gabbiani F, Koch C (2003) Human visual object categorization can be described by models with low memory capacity. *Vis Res* 43:2265–2280
- Peters RJ, Mundhenk TN, Itti L, Koch C (2003) Contour-facilitation in a model of bottom-up attention. In: *Proc Soc Neurosci Ann Meet (SFN'03)*
- Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM (1998) Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* 391(5):580–584
- Polat U, Sagi D (1993a) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiment. *Vis Res* 33(7):993–999
- Polat U, Sagi D (1993b) The architecture of perceptual spatial interactions. *Vision Res* 34(1):73–78
- Polat U, Sagi D (1994) Spatial interaction in human vision: from near to far via experience-dependant cascades of connections. *PNAS* 91:1206–1209
- Prodöhl C, Würtz RP, von der Malsberg C (2003) Learning the gestalt rule of collinearity from object motion. *Neural Comput* 15:1865–1896
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87
- Rubin N (2001) The role of junctions in surface completion and contour matching. *Perception* 30:339–366
- Sceniak MP, Hawken MJ, Shapley R (2001) Visual spatial characterization of macaque V1 neurons. *J Neurophysiol* 85:1873–1887
- Schultz W (2002) Getting formal with dopamine and reward. *Neuron* 36:241–263
- Shashua A, Ullman S (1988) Structural saliency. In: *Proceedings of the International conference on computer vision*, pp 482–488
- Silverstein SM, Kovács I, Corry R, Valone C (2000) Perceptual organization, the disorganization syndrom, and context processing in chronic schizophrenia. *Schizophr Res* 43:11–20
- Shevelev IA, Jirrmann KU, Sharaev GA, Eysel UT (1998) Contribution of GABAergic inhibition to sensitivity to cross-like figures in striate cortex. *Neuroreport* 9(14):3153–3157
- Suri RE, Bargas J, Arbib MA (2001) Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience* 103(1):65–85
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12(1):97–136
- Usher M, Bonnef Y, Sagi D, Herrmann M (1999) Mechanisms for spatial integration in visual detection: a model based on lateral interaction. *Spat Vis* 12(2):187–209
- Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Neslon SB (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci* 17(20):7926–7940
- von der Malsberg C (1981) The correlation theory of brain function. Internal Report 81–2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany
- von der Malsburg C (1987) Synaptic plasticity as basis of brain organization. *The Neural and Molecular Basis of Learning*, S. Bernhard, Dahlem Konferenzen, pp 411–432
- Wang XJ, Tegner J, Constantinidis C, Goldman-Rakic PS (2004) Division of labor among distinct subtypes of inhibitory neurons in cortical microcircuits of working memory. *PNAS* 101(5):1368–1373
- Wertheimer M (1923/1950) Law of organization in perceptual form. In: Ellis WD (ed) *A source book of gestalt psychology* pp 71–88 The Humanities Press, New York
- Wolf JM (1994) Visual search in continuous, naturalistic stimuli, *vision Res.* 34(9) 1187–1195
- Wolfe JM, O'Neill P, Bennett SC (1998) Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Percept Psychophys* 60(1):140–156
- Wolfe JM, DiMase JS (2003) Do intersections serve as basic features in visual search? *Perception* 32:645–656
- Yen S, Finkel LH (1998) Extraction of perceptually salient contours by striate cortical networks. *Vision Res* 38(5):719–741
- Yu C, Levi DM (2000) Surround modulation in human vision unmasked by masking experiments. *Nat Neurosci* 3(7):724–748
- Zenger B, Sagi D (1996) Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection. *Vision Res* 36(16):2497–2513