ANALYSIS AND SYNTHESIS OF VISUAL IMAGES IN THE BRAIN: EVIDENCE FOR PATTERN THEORY*

TAI SING LEE†

Abstract. At each moment in time, we perceive a very small fragment of the world through our retinas, yet our subjective perception of the visual world in front of us is rather clear, coherent and complete. Often we see things that are not even there. This is because what we perceive is actually a 'virtual' visual world that is created in our minds – a product of the interaction between our experience, prior knowledge and the incoming sensory data. This world is dynamic and plastic. It depends on the behavioral demands imposed on us and the statistics of our experiences. In this lecture, I will present neurophysiological evidence that suggests that the early visual cortex participates in many levels of visual processing underlying the generation and the representation of this subjective visual world in our brain.

Key words. Vision, neurobiology, computational vision.

AMS(MOS) subject classifications. 68T45, 92C20.

1. Theory.

1.1. The nature of perception. The visual world we perceive is a mental construction inside our brain, rather than the raw spots and dots that photons create on our retinas. This mental construction is so real and compelling that we rarely question or think twice about it. We realize this fact through the painful examples of patients who have 'visual form agnosia'. These patients lost their ability to organize and construct objects in the virtual visual world in their minds because of lesions in their visual areas. For example, Benson and Greenberg (1969) reported a patient whose vision is normal in discriminating fine features, color and motion but couldn't put all the tiny details back together to experience and perceive coherent objects.

The raw image sampled by the retina at each of our glances provides a very impoverished image of the outside world. Figure 1 illustrates a sequence of images approximating what one of our retinas see at several fixations. These images are severely limited – they are high-resolution only in the fovea but are very blurry in the periphery. Yet, we do not 'feel' the fuzziness in the surround, we realize it only when we pay attention to it and ponder about it. However, by making saccadic eye movements three or four times a second to constantly scan the visual scene, we somehow are able to obtain enough samples of the external world to create an apparently stable and complete visual world inside our brain.

*The work was supported in part by NSF CAREER 9720350 and NIH EY08098.

†Center for the Neural Basis of Cognition and Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213.

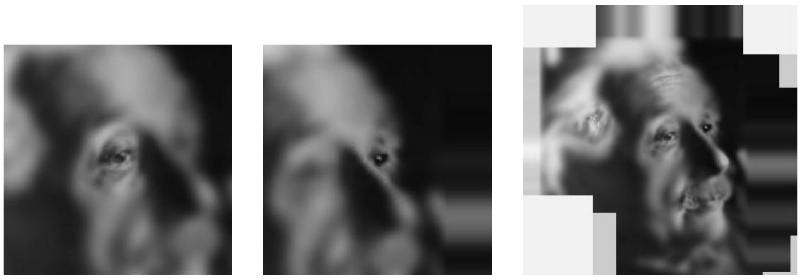


FIG. 1. *(a) and (b) Raw input sampled by the retina in two fixations. (c) A 'mental' image created in our perception by integrating the retinal images from several fixations (see also Lee and Yu 2000).*

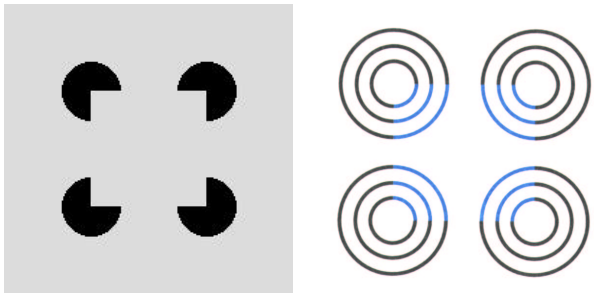


FIG. 2. *Illusory squares: we see subjective contour and surface at places where there are actually no direct visual evidence for it.*

Another piece of evidence supporting the idea of constructive processes in vision is beautifully illustrated by Kanizsa with his famous visual illusion. When viewing the display shown in Fig. 2a, we perceive a subjective square and we see vivid borders of the square even in regions of the image where there is no direct visual evidence for them. This is one example of the phenomenon of illusory figure. Figure 2b show an even more stunning example presented by Hoffman (1998). In this display, when parts of the rings change their color from black to blue, a subjective perception of a ghostly blue square surface is induced over the empty space.

1.2. Generative processes in unconscious inference. Helmholtz (1867) had argued that perception is a product of *unconscious inference*: what we perceive is our visual system's best guess as to what is in the world. This guess is based both on our prior experience and the retinal image. Can this unconscious inference be accomplished simply through association and memory? Or does it require the generation of an explicit representation in our brain?

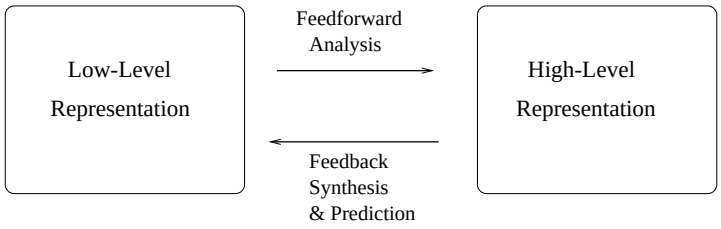


FIG. 3. *According to Grenander, inference is made through a combination of analysis and synthesis loops.*

Marr (1981) would have favored an explicit representation, but he would also say that it can be computed by a feedfoward chain of computational modules, each computing a more complex and abstract perceptual structure. The perceived abstract structure, such as the illusory surface and contour of the square, could be computed and represented by higher visual areas. There is no need to reconstruct and represent them explicitly in the early visual cortex.

On the other hand, Grenander (1976-81) would have argued that inference is accomplished through the interaction of analysis and synthesis. From this point of view, as articulated by Mumford (1992) and Lee et al. (1998), vision is a series of interactive hypothesis testing. Prediction and expectation continuously generated by the higher visual areas are tested and matched with the representations in the earlier visual areas (Figure 3). This feedback synthesis serves two purposes. First, it is useful for analyzing ambiguous images in which a dialog between knowledge and perception is required to disambiguate the scene, as in the example devised by R.C. James in Figure 4. Second, having top-down expectation and prediction is important for speeding up the inference process in real time. That is, if we know what we are going to see, it is much easier and more efficient to verify objects in a visual scene than to deduce them from sketch at each moment in time. Vision is then considered an active process of generating and testing hypotheses, very much like conducting a scientific experiment. This construction is *unconscious* and the hypotheses constructed are our perception of the visual world. Current important theories on brain functions such as Grossberg's adaptive resonance theory (1987), and McClelland and Rumelhart's (1981) interactive activation theory, Dayan et al.'s Helmholtz machine (1995), and Rao and Ballard's (2001) predictive coding model basically advocated the same fundamental view, particularly for the purpose of disambiguation.

1.3. High-resolution buffer hypothesis. Mumford and I (Lee and Mumford 1996, Lee et al. 1998) have suggested a new framework for con-

FIG. 4. *An image devised by R.C. James to illustrate how the interpretation of some images relies on top-down knowledge.*

ceptualizing the role of the primary visual cortex (V1) in visual processing from this perspective. This very large region in the occipital cortex has been traditionally considered the first stage of visual processing, extracting edges and other low-level cues. We think that it might play a far more important role than previously imagined. Because the receptive fields of neurons in V1 are much smaller and more spatially localized than those of neurons in the extrastriate cortex, V1 could furnish a unique *high resolution buffer* or a sketch pad for the whole visual cortex to make detailed geometric calculations and synthesize images through the generative processes. For example, suppose we want to explicitly construct the precise contour of the illusory square (Figure 2), V1 is an ideal place to do so because it furnishes a precise representation for integrating the bottom-up information from the raw images and the top-down hypotheses, generated from prior experience, to construct and represent the sharp subjective contour. As another example, suppose the brain needs to compute the axis of symmetry of an object for shape discrimination; V1 could provide an appropriate buffer for representing the axis explicitly. As in the case of illusory contour, the axis of symmetry is a computation that requires the integration of local information and global scene context.

More generally, we think that V1 is not limited to curvilinear geometric computation as illustrated by the examples of illusory contour and symmetry axes. Rather it serves as a high-resolution buffer for even more general computation. We know that the information output by V1 is channeled into the dorsal stream (or commonly known as the *where* pathway) for motion processing and spatial analysis, as well as the ventral stream (or commonly known as the *what* pathway) for form processing and object analysis. These two streams are further subdivided into multiple modules or areas, each responsible for processing different aspects of the visual scene: color, form,

motion, stereo, and spatial locations of objects in various coordinates. A major question is how the brain combines all the processed information back together to form an unified percept. There are at least three possible loci of interaction for such an unification to occur. First, they could be mediated by the intercortical connections between modules in the two streams (Baizer et al. 1991). Second, they could be mediated in the prefrontal cortex such as area 46 where both the dorsal and ventral streams converge to (Rao et al. 1997). Third, with the massive feedforward and feedback connections it has with many extrastriate areas, V1 can potentially serve as a sketch pad for integrating the higher level information, derived from the different modules, including color, shape, depth, object identity and spatial location. The high-resolution buffer hypothesis basically argues for the importance of this third possible locus of interaction and emphasizes that V1 participates in all levels of perceptual computations that require high resolution image details and spatial precision.

2. Experiments. What evidence supports the high-resolution buffer hypothesis and, more generally, the generative and constructive processes in the brain? I will describe three experimental observations we made that are in part both supportive and suggestive of these ideas.

2.1. Illusory contour. The first experiment was to examine whether V1 represents subjective contours of the Kanizsa square as shown in Figure 2. Seventeen years ago, von der Heydt and his colleagues (von der Heydt et al, 1984) found that neurons in macaque V2 are sensitive to an illusory bar moving across their receptive fields. This discovery was seminal because it showed neurons possess a direct physiological correlate of a perceptual phenomenon. Curiously, they didn't found V1 neurons responding to the illusory contour. Hence, they proposed a feedforward model that integrates end-stopping signals in V1 to produce the illusory contour responses in V2. In short, their evidence caused a problem for all the interactive models of visual processing, as well as the high resolution buffer hypothesis which predicts that we should be able to observe the emergence of illusory contour at the later part of V1's responses because of the generative feedback processes in the visual system.

We decided to put the high-resolution buffer hypothesis to test. We (Lee and Nguyen 2001) studied the responses of V1 and V2 neurons to five sets of of stimuli, as shown in Figure 5. The set of test stimuli included a Kanizsa figure with illusory contours (Figure 5a), an amodal figure in which the contours are partially occluded (Figure 5b), and a variety of control and comparison stimuli (Figures 5c-i). In each trial, while the monkey fixated, a sequence of 4 stimuli was presented. The presentation of each stimulus in the sequence lasted for 400 msec. Over successive trials, one of the contours (real, amodal or illusory) in the figure was placed at 10 different locations relative to the center of the receptive field, 0.25^o apart, spanning a range of 2.25^o, as shown in Figure 6. It is important to bear in mind

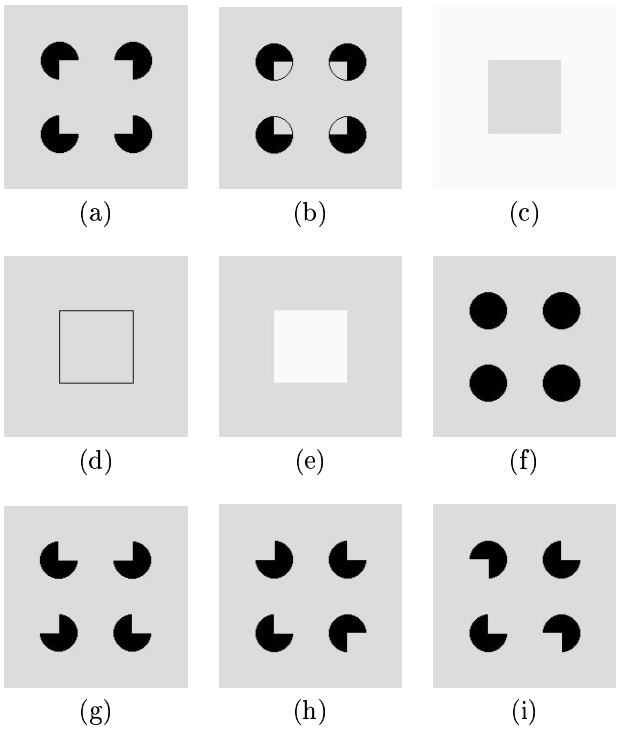


FIG. 5. *A subset of the stimuli used in the illusory contour experiment by Lee and Nguyen (2001).*

that the receptive fields of the neurons, as plotted by small oriented bars, was less than 1 degree at that eccentricity (about 2 -3 degree away from the fovea). The gap between the pac-men was 2 degree wide. The neurons are considered to be sensitive to illusory contour if their response to the illusory contour, at the precise location of that contour, was significantly larger than their response to the amodal contour or other controls. We found that a significant number of V1 neurons at the superficial layer of V1 exhibited sensitivity to the illusory contour under our experimental manipulation (Lee and Nguyen 2001).

Figure 7 presents the findings from a V1 neuron. This cell responded significantly more to the illusory contour than to the amodal condition, or the rotated corner disc configuration. The illusory contour elicited a response precisely at the same location at which a real contour elicited the maximum response. However, the response to the illusory contour appeared at 100 msec after the onset of the subjective square, as compared to 45 msec after the appearance of the square defined by lines or luminance contrast. The averaged temporal response of 50 neurons in the superficial layers of V1 to the illusory contour and to the controls (Figure 8) demon-

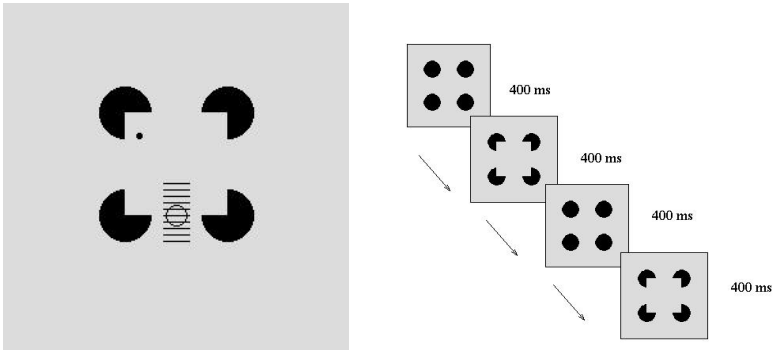


FIG. 6. *(a) The 10 different positions where the receptive field of a cell was placed relative to the subjective contour. (b) Sequence of presentation: Abrupt onset of the subjective square in front of the four discs helps to call attention to the square (Lee and Nguyen, 2001).*

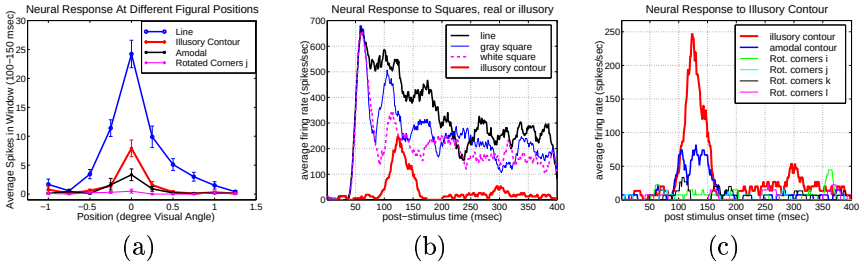
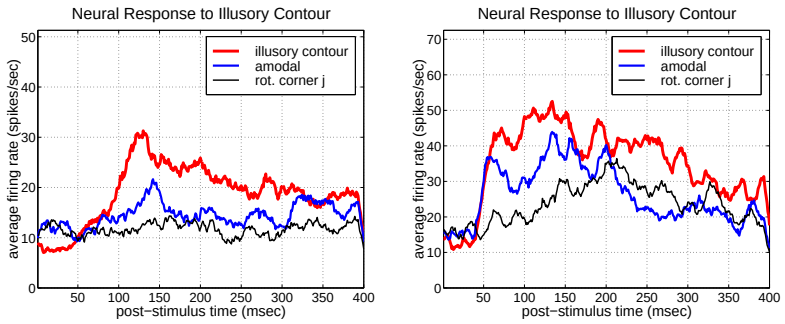


FIG. 7. *Response of a V1 neuron. (a) Response at different spatial location relative to the illusory contour. (b) The onset of response to illusory contour emerges at 100 msec, 60 msec later than the response to the real contours. (c) Response to the illusory contour (Figure 5a) is significantly greater than the response to the amodal contour (Figure 5b) and the other rotated pac-men controls (Figure 5g–i) (Lee and Nguyen, 2001).*

strated a significant response to the illusory contours at the population level. The average temporal response of 39 neurons in the superficial layers of V2 showed an earlier onset of illusory contour response. We will refer the readers to our paper (Lee and Nguyen 2001) for the details of the experiment. Suffice to say, the experimental finding suggests the illusory contour is being explicitly 'constructed' and represented in V1.

If V2 neurons are already detecting and encoding information about illusory contours, what is the advantage of feeding it back to V1? One reason is that V2 neurons' receptive fields are twice the diameter of V1 neurons at the same spatial location. Hence, V2 neurons can integrate information more globally, but it can no longer represent precise spatial



(a) V1S (n = 49) **(b) V2S (n = 39)**

FIG. 8. *Comparing the responses of the neurons at the superficial layers of V1 and V2 to the stimuli, we can see the illusory contour signal, which is indicated by the difference between the response to illusory contour over the response to amodal or rotated pac-men control, emerges 100 msec after stimulus onset for V1 neurons, and at 65 msec for V2 neurons (Lee and Nguyen, 2001, with permission from PNAS).*

information. That is, it can detect the existence of the contour but cannot know explicitly and precisely where the contour is. The feedback from V2 to V1 is spatially diffuse but orientation-specific. It only informs V1 of the existence and orientation of a long contour, but not its precise spatial location. The feedback of this global information, when combined with the bottom-up cues that are represented precisely in V1 (i.e. the edges of the pac-men), will enable the neural circuit within V1 to complete a spatially precise and complete contour (Figure 9).

2.2. Axis of symmetry. The second observation I now describe concerns evidence suggestive of a possible medial axis representation at V1. Medial axis transform, or skeletonization, is a powerful way of representing shape. Blum (1973) proposed to describe the complex biological forms using the skeleton and a small finite set of shape primitives. The skeleton links these elementary parts together hierarchically like the clauses in a sentence. This method is particularly useful for encoding the infinite variety of biological forms, in which relationships between body parts possessing flexible joints can change drastically along with changes in view point and motion. Blum suggested a region-based description using skeletons or axis of symmetry of the objects might be more robust and stable than a boundary based description against such variations (see Figure 14).

Our experiment (Lee et al. 1998) was designed to investigate the neural representation of texture contours and surface. We examined the responses of the neurons to the strip stimuli defined by texture contrast (Figure 10). The texture defined strip was presented in a randomized series of sampling positions relative to the cell's receptive field. The cell's response at one position was sampled at each trial. In each recording session, the cell's response was sampled along a horizontal cross-section of the image at a

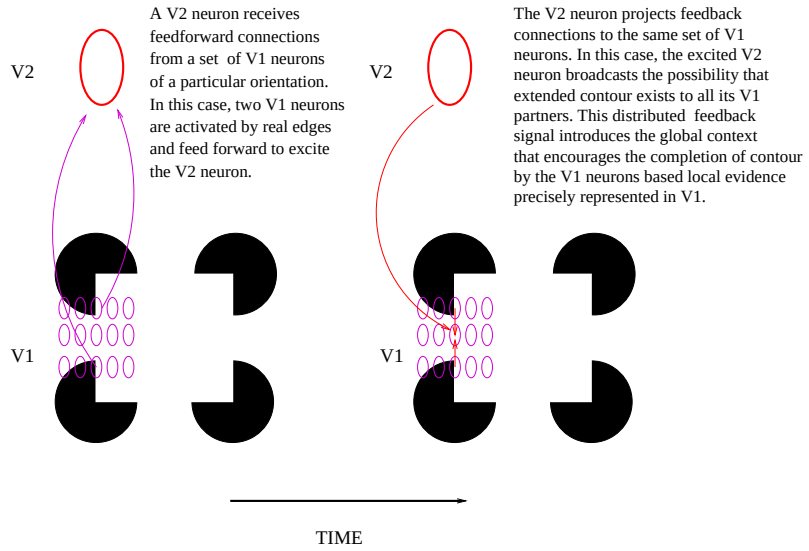


FIG. 9. *How V2 facilitates precise contour computation in V1.*

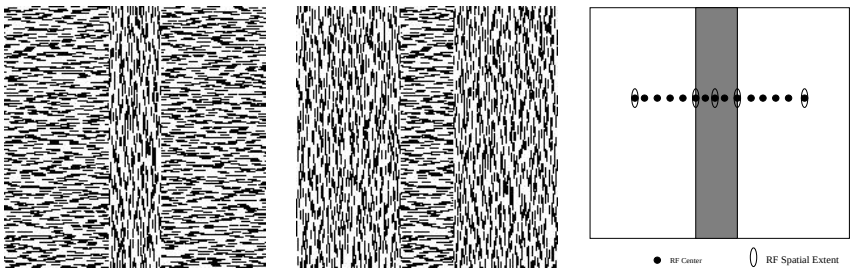


FIG. 10. *(a) A vertically textured strip in front of a horizontally textured background. (b) A 4 degree width horizontally textured strip in front of a vertically textured background. Both strip widths were 4 degree visual angle, which is 4–6 times larger than the diameter of the classical receptive fields of the cells.(c) The receptive fields of the neurons were placed at 16 locations in each of the stimulus images.*

0.5^o visual angle step interval (Figure 10c). The size of the receptive field of the cell ranged from 0.5^o to 0.8^o in visual angle.

We found that there were several stages in the responses of V1 neurons, each with distinct spatial response profiles at different temporal windows. Typically, V1 neurons started to respond about 40 msec after the stimulus was displayed on the screen. From 40 to 60 milliseconds after stimulus onset, the cells behaved essentially as local feature detectors or linear filters (Hubel and Wiesel 1978, Pollen et al 1989). The responses to the texture

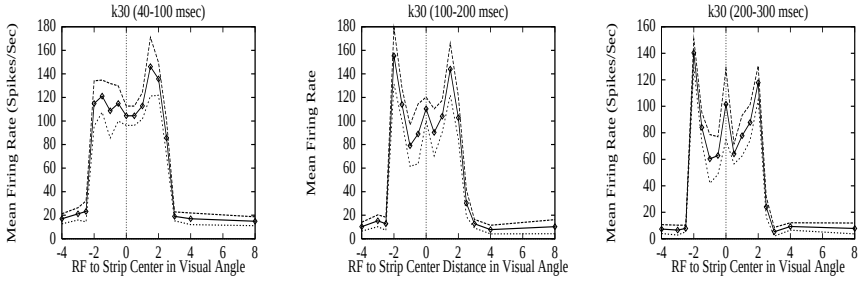


FIG. 11. *Spatial response profiles of a vertically oriented V1 neuron to different parts of the vertically textured strip (Figure 10a). The abscissa is the distance in visual angle from the RF center to the center of the figure. The solid lines in these graphs indicate the mean firing rate within the time window, and the dashed lines depict the envelope of standard error. The later response of the neuron exhibits response peaks at the boundaries and at the axis of the strip.*

stimuli were therefore initially uniform within a region of homogeneous texture, based upon the orientation tuning of the cells. In the example shown in Figure 11, the neuron showed preference for features of vertical orientation. In the initial period, it responded uniformly well within the interior of the vertically-textured strip, but responded very poorly to the horizontal texture outside the strip. At 60 milliseconds following stimulus onset, boundary signals started to develop at the texture boundaries. By 80 msec, the responses at texture boundaries have become sharpened, consistent with the psychophysical time course of texture segmentation (Julesz 1975). Interestingly, beginning at 80 msec, as the responses at the boundary became more localized, a response peak was sometimes observed at the center or the axis of the strip. The spatial response profiles in successive temporal windows in Figure 11 show the development of this central peak. In another dramatic example (Figure 12), cell m32, which was also vertically oriented, at first did not respond at all within the horizontally textured strip but became active at the axis of the strip after 80 msec.

Statistically significant central peaks were observed in 14 out of the 50 neurons tested with the vertically-textured strip and 10 cells with the horizontally-textured strip (T-test, $p < 0.05$). Figure 13 shows the average spatial temporal response of the 14 neurons that are sensitive to the positively textured strips, revealing a subtle response peak at the axis of symmetry of the strip.

Whether this signal that we accidentally observed has anything to do with the axis of symmetry representation is still open to question. There were two major problems related to this interpretation. First, the axis response seems to be dependent on the width of the strips, i.e. the axis response of a cell disappeared when the width of the strip was increased in all cases. This suggests it could be explained as a product of some global surround lateral inhibition and dis-inhibition mechanism. We have pointed

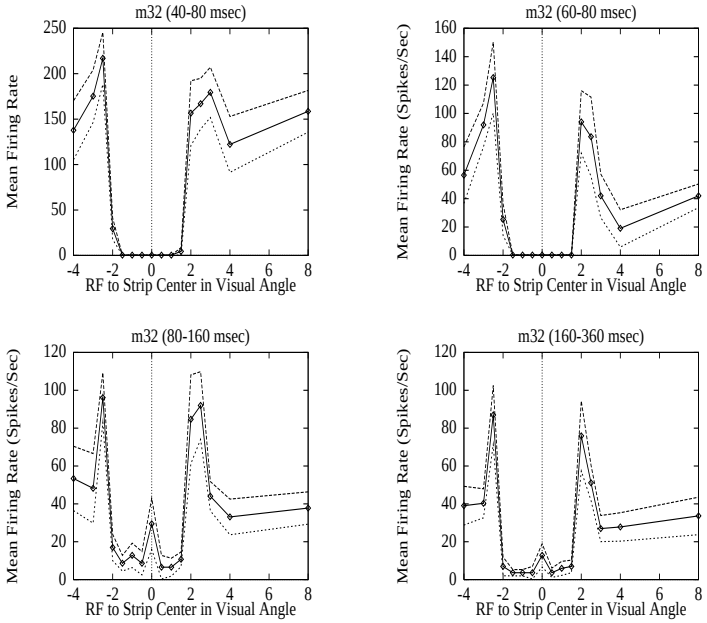


FIG. 12. *Another neuron's (cell m32) response to the horizontally textured strip (Figure 10b). Approximately 40–60 msec after stimulus onset, the cell responded uniformly to the background but did not respond to the texture strip at all because it was not tuned to the texture inside. From 60 to 80 msec, the boundary started to sharpen, but there was still no response within the strip. Interestingly, 80 msec onward, a pronounced response peak gradually developed at the axis of the texture strip.*

this out (Lee et al. 1998), and Zhaoping Li (1999) has also demonstrated this possibility with a model based on lateral inhibition. The fact that the neuron is sensitive to the width of the strip might not be as devastating as it might seem, for a medial axis neuron can be sensitive to the diameter of the inscribing disk as well (Figure 14). One can conceive that a group of these cells, each tuned to a particular width, could together provide an invariant representation of the medial axis. A bit more troubling is the observation that the axis response seems to be absent in black and white strips (see Lee et al. 1998 for details). We speculate that perhaps the top-down feedback is weak in this case and excitation by local features is required in order for the sub-threshold axis signal to manifest itself. However, if the signal is so weak, could it possibly serve any purpose? Could we be seeing a little too much in this response peak at the center? This experiment demonstrates that texture contour is computed and represented in V1. The evidence hints at the possibility of a neural correlate of medial axis computation, even though the data do not provide solid proof that medial axis is represented in V1 explicitly. More carefully designed experiments are required to clarify this issue.

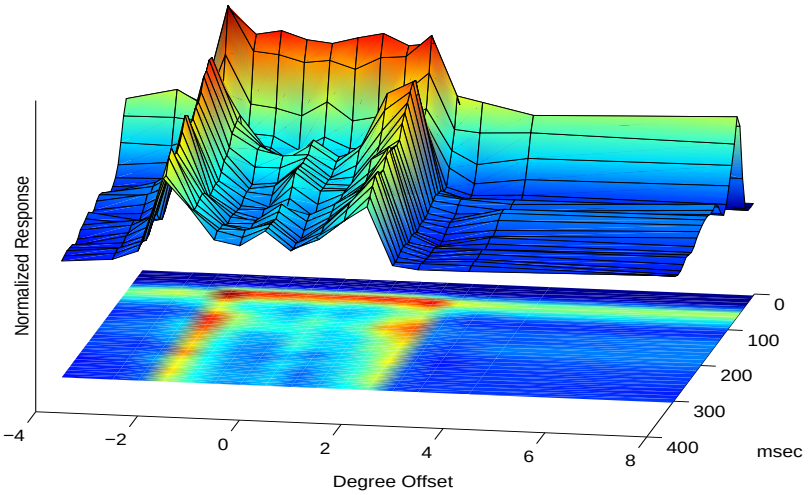


FIG. 13. *The spatiotemporal population average response profile of the 14 axis-positive cells to the different locations horizontally across the vertically textured strip. Locations −2 and 2 in the spatial offset indicate the locations of the texture boundaries. Location 0 in the spatial offset indicates the center of the texture strip. Easily observable is the strong and sharp boundary responses and an axis response of smaller magnitude in the later part of the response (see Lee et al. 1998 for details).*

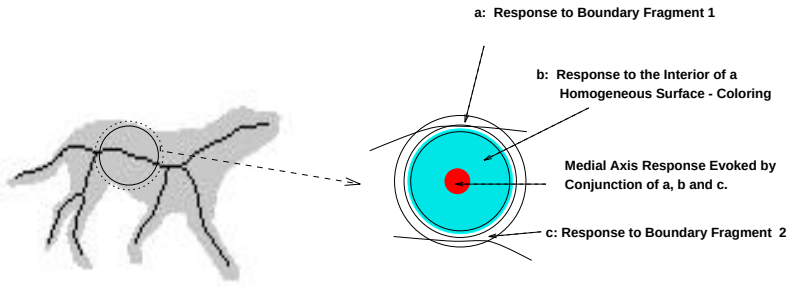
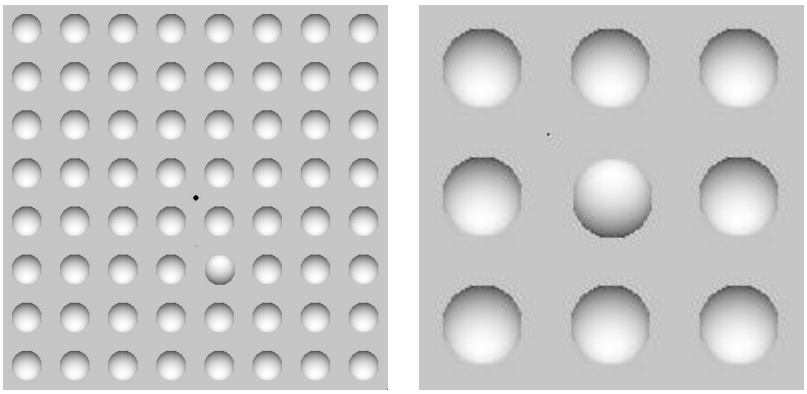


FIG. 14. *Medial axis is a descriptor that integrates local and global information. It encodes information about the location of the skeleton and the diameter of the inscribed disk. This figure illustrates how a cell may be constructed so that it fires when located on the medial axis of an object. The conjunction of three features has to be present: at least 2 distinct boundary points on a disk of a certain radius, and the homogeneity of surface qualities within an inscribing disk. Such a response is highly nonlinear, but can be robustly computed.*

2.3. Shape from shading. A third experiment, yet to be published, provides another piece of evidence in support of the generative processes. Here, we pushed the high-resolution buffer hypothesis one step further by testing whether V1 neurons are sensitive to 3D shapes. There has been



(a) (b)

FIG. 15. *(a) a convex odd-ball stands out in a field of concavities. (b) a concave odd-ball stands out in a field of convex balls. Psychological study shows that this pop-out can be detected 'preattentively' despite the fact that the stimulus element is defined by 3D shape from shading.*

earlier physiological studies showing V1 neurons are sensitive to perceptual pop-out in terms of oriented bar or oriented texture (Knierim and Van Essen 1993, Lamme 1995, Lee et al 1998). That is, a cell responded better when its receptive field was seeing a bar(s) of optimal orientation as it was surrounded by bars of an orthogonal orientation than when it was surrounded by bars of the same orientation. But perceptual pop-out is not limited to oriented bars. Figure 15 showed examples in which a 3D convex shape pops out from a set of 3D concave shapes and vice versa (Ramachrandron 1988, Sun and Perona 1996). This experiment was to designed to examine whether V1 neurons are sensitive to odd-ball pop-out, defined by 3D shape from shading, at a relatively high level construct.

We trained the monkeys to make eye movement towards the odd-ball which could appear at one of the four random positions. The pop-out stimuli include 3D shape stimuli as well as 2D luminance contrast stimuli, in which stimulus elements were arranged in an 8×8 array (3×3 arrays are shown Figure 16 as iconic examples). Then we tested whether a V1 neuron, when its receptive field was placed inside one of the stimulus elements, is sensitive to the difference in the surrounding context; i.e., the V1 neuron responded differently when it was surrounded by dissimilar elements (pop-out condition) as opposed to when it was surrounded by similar elements (homogeneous condition).

We found that V1 neurons exhibit neural pop-out response, defined as the relative increase in response to the pop-out condition over the homogeneous condition of the shape from shading stimuli (Figure 17). The pop-out response for 3D shapes (Figure 17a,b) was significantly greater than that for 2D luminance contrast stimuli (Figure 17c,d). The pop-out

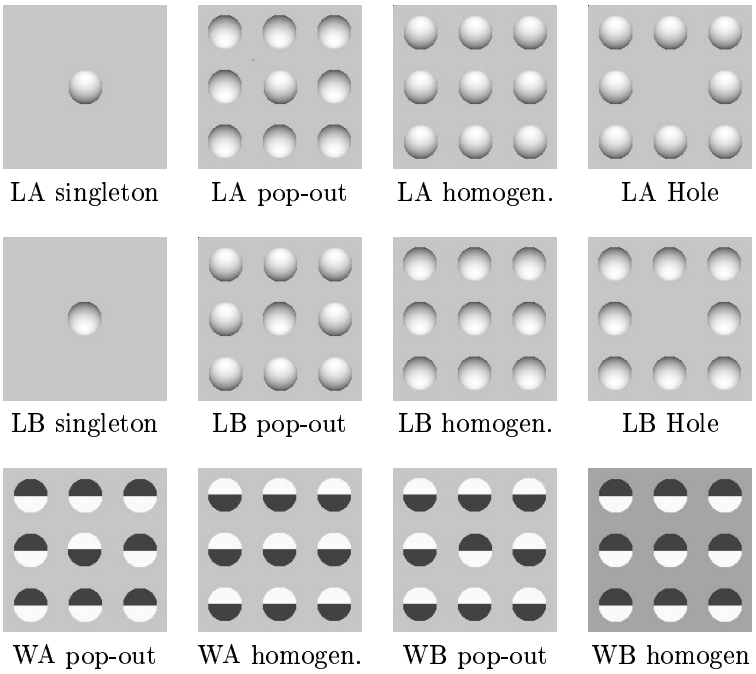


FIG. 16. *LA: lighting from above; LB: lighting from below; WA: white above; WB: white below. These are iconic representations of the stimuli. The actual stimulus display is an array of 12 × 12 stimulus elements as shown in Figure 15. In this iconic representation, the receptive field of the neuron is placed inside the stimulus at the center element (or the hole). The sizes of the classical receptive fields of the neurons (minimum responsive area) we studied are smaller than the center element.*

response is correlated with the monkey's performance in the pop-out detection test. When the pop-out signal was stronger, the monkey reacted faster and more accurately (see behavioral performance data below the figures). More interestingly, when we manipulated the statistics of the occurrence of the pop-out stimulus – specifically, when the convex pop-out was presented more frequently than any other odd-balls, we found that the behavioral preference of the monkeys shifted, and the relative neural pop-out response among the different stimuli also changed accordingly (Figure 18). This finding suggests that 1) V1's pop-out response is sensitive to 3D shape from shading information, 2) the behavioral relevance of the pop-out stimulus can determine the level of the pop-out response. From this observation and from similar observations from other monkeys, we think that when the monkey is looking for the concave object, as in this case, it is in part because the concavity was preattentively more salient than the other odd-balls. Hence the V1 neurons' activities are relatively more enhanced for the concave pop-out as well. This 'pop-out' response enhancement signal thus may be related to the so-called object attention.

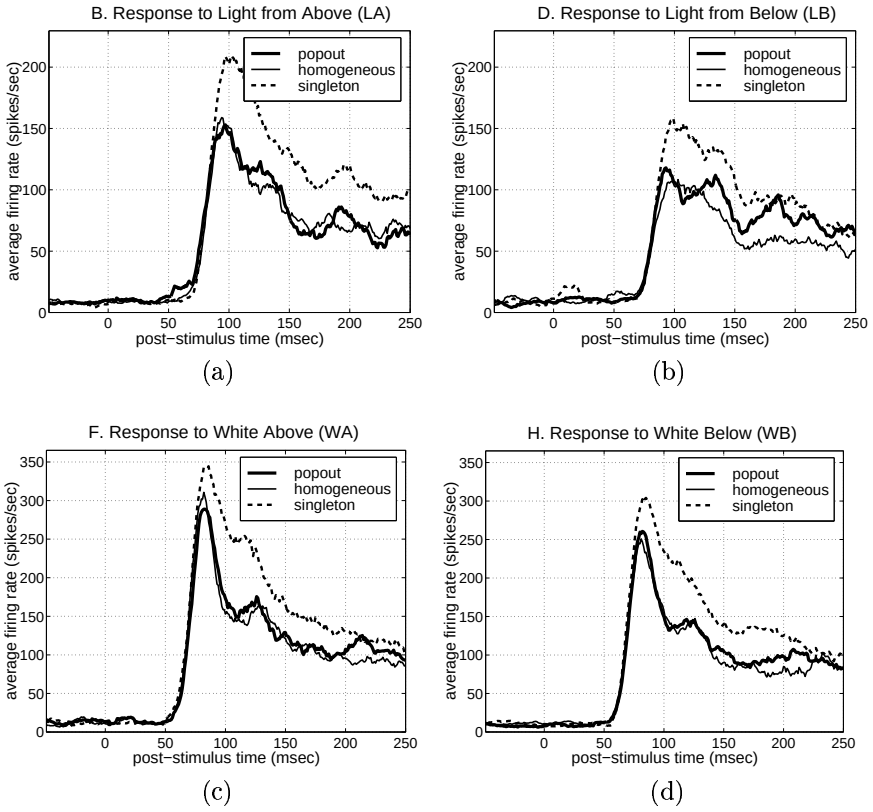


FIG. 17. *Population averaged response (PSTH) depicting the temporal evolution of the neuronal activity in V1 neurons to the four stimulus sets. Response to the pop-out is compared against the response to singleton and to a homogeneous field of convex objects. (a) convex (or lighting from above) pop-out; (b) concave (or lighting from below); (c) white above 2D luminance contrast; (d) white below 2D luminance contrast. We can observe that for this monkey, the response to the pop-out condition is stronger than the homogeneous condition in the LB set, and much less so for the LA, WA and WB sets. The behavioral performance measure associated with each pop-out condition shows that the monkey reacts faster and more accurately to the LB condition.*

Object attention is generated when an animal is searching for a particular object or feature over a large visual space in parallel (James 1890). Object attention is sometimes also called feature attention. This is in contrast with the spatial attention proposed by Helmholtz (1867) and Treisman (1982). Spatial attention can be visualized as a spotlight that 'illuminates' a certain location of visual space for focal visual analysis. When the monkeys are performing visual search for a particular object, they basically function in the object attention mode. Since object attention is object specific but spatially distributed, one can imagine that object templates are being sent down from IT (Inferotemporal cortex – an object encoding area)

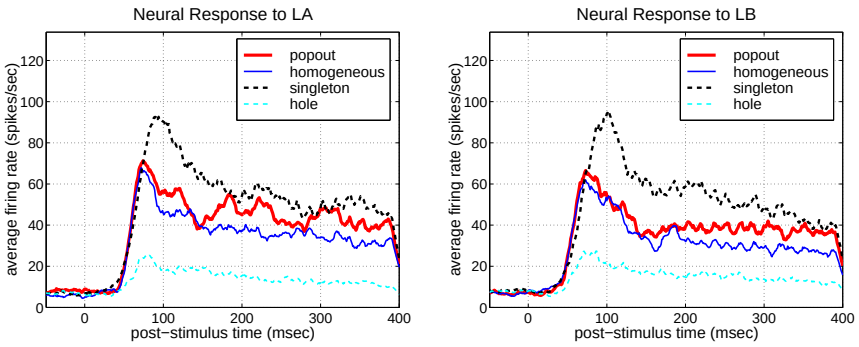


FIG. 18. *When the occurrence of the LA pop-out was made more frequent, the monkey started to exhibit better detection performance for the LA set, and a stronger neural pop-out signal for the LA pop-out stimulus.*

and broadcast to all the hypercolumns in the early visual cortex. When a match occurs, the activity in the area of V1 that contains the the image of the searched object will become elevated. The 'pop-out' response enhancement we observed is more likely this object attention effect rather than a bottom-up saliency effect. This is because we did not observe these pop-out responses before the monkeys were trained to do the detection task!

3. Model. Now, if the task is to look for a certain searched object, what is the purpose of elevating the response of particular neurons in V1? This is best understood in the context of a neural model that Gustavo Deco and I (Deco and Lee 2002) developed to explain how object attention and spatial attention would function in a unified system to accomplish translation-invariant object recognition and visual search. The premise of this model is that the early visual cortex serves as a high resolution buffer for the dorsal stream and the ventral stream to interact, combine and coordinate their information in a set of feedforward/feedback loops (Figure 19). The model is formulated in the framework of biased competition. Basically, within each cortical area, there is inhibitory competition among neurons. However, there is also excitatory facilitation between corresponding neurons across the different areas. This long range facilitation from one area can serve as a bias that will tilt the balance of competition within each module. The conceptual framework of biased competition has been proposed by Duncan and Humphrey (1989). This scheme has been implemented to explain attentional phenomenon observed in IT by Usher and Niebur (1996) and in V4 by Reynolds et al. (1999). Our contribution is to bring the dorsal stream and the early visual cortex into the picture in explaining how attentional modulation can be mediated in the ventral stream. Furthermore, our neural model also provides a functional rationale for attentional modulation. It shows attentive object recognition and vi-

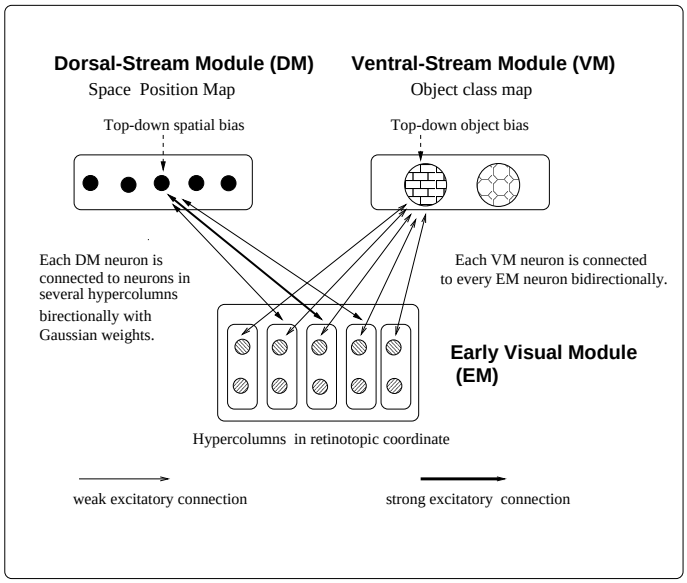


FIG. 19. *A schematic diagram of the model. The model contains three modules: the early visual module (EM), the ventral stream module (VM) and the dorsal stream module (DM). The early visual module contains orientation-selective complex cells and hypercolumns, as in the primary visual cortex. The ventral stream module contains neuronal pools encoding specific object classes, as in the inferotemporal cortex. The dorsal stream module contains a map encoding positions in the retinotopic coordinate. The early module and the ventral module are connected with symmetrical connections developed with Hebbian learning. The early module and the dorsal module are connected with symmetrically localized connections modeled with Gaussian weights. Competitive interaction within each module is mediated by inhibitory pools. Connection between modules are excitatory, providing biases for shaping the competitive dynamics within each module. Concentration of neural activities to an individual pool in the ventral module corresponds to object recognition. Concentration of neural activities to a small number of nearby pools in a dorsal module corresponds to object localization. The early module provides a buffer for the ventral and the dorsal modules to interact.*

sual search can be accomplished through the interaction of the two streams via the early visual cortex in an unified system.

The system basically has three modules: a dorsal module, which contains a spatial map for coordinating competition and coding of the location of spatial attention; a ventral module, which contains a set of neurons for coding different object classes, and an early visual module, modeled after V1. Spatial attention is initiated by introducing a top-down bias to a particular neuronal pool in the dorsal map, which will help to elevate the activities of a particular area in V1, effectively gating information from V1 to higher processing areas in the ventral stream. Alternatively, object attention – the search for a particular object is initiated by introducing a top-down bias to a particular neuronal pool in the ventral module (IT).

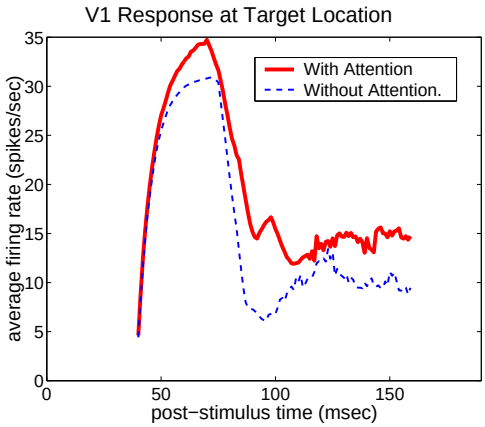


FIG. 20. *Population average firing rate of neuronal pools in the early module at the location of the searched target compared to the response of the same neuronal pools when the system was not looking for anything. Significant and sustained enhancement was observed in the later part of the neuronal pool's response due to object attention. However, given that all three modules are always engaged in interaction in either of the spatial or object attentional modes, the attention-induced response elevation in the early visual module or any other area cannot be considered as a purely spatial or a purely object attentional effect. From this point of view, all attentional effects observed necessarily possess both spatial and object attentional components.*

IT will then send down its expectation of V1 activities corresponding to the particular target object down to V1. A match will produce a bias in favor of that area of V1, enhancing its activities (Figure 20). The mutual coupling between V1 and the dorsal stream will allow the competition within each of these modules to help each other synergistically, thus eventually leading the contraction of activities in both V1 and the dorsal spatial map to a particular spatial location. The contraction of response to a localized area in the retinotopic space corresponds to the localization of the searched target. The reason V1 is needed in this visual search task is that in order for the monkey to make a saccadic eye movement towards the target, it has to be able to localize the object with spatial precision – provided explicitly at the level of V1. This explains why before the monkey utilized the stimuli in its behavior, we did not observe the contextual elevation in V1's activities. Only after the monkey had performed the detection task did it start to finely appreciate the stimuli both in space and in feature, leading to the sensitivity of its V1 neurons to the 3D pop-out context. This finding is again consistent with the generative processes suggested by Pattern theory.

4. Conclusion. The neurophysiological evidence presented here strongly suggests that during perception at the appropriate context explicit local and global representations might be actively constructed at the

level of V1 in conjunction with the extrastriate cortices. Other neurophysiological findings, particularly the curve tracing experiment of Roelfsema et al. (1998) and a recent experiment of Paradiso and his colleagues on the effect of expectation, also point in the same direction. Visual inference is as much a generative and synthesis process as an analysis and deduction process. Our evidence suggests that low level vision and high level vision are intimately intertwined. The primary visual cortex, rather than simply contributing to the first stage of early visual processing, might play a more important role by providing a high-resolution buffer to mediate the interaction among the different expert extrastriate modules and streams. Furthermore, the primary visual cortex might actively participate in many computations that are responsible for constructing the illusion of stable and complete visual world in our mind.

REFERENCES

[1] J.S. Baizer, L.G. Ungerleider, and R. Desimone, *Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J. Neuroscience*, **11**(1): 168–190, 1991.

[2] D.F. Benson, D.F., and J.P. Greenberg, *Visual form agnosia: A specific defect in visual discrimination.* Archives of Neurology, **20**: 82–89, 1969.

[3] H. Blum, *Biological shape and visual science.* J. Theoretical Biology, **38**: 205–287. 1973.

[4] P. Dayan, G.E. Hinton, R.M. Neal, and R.S. Zemel, *The Helmholtz machine.* Neural Computation, **7**(5): 889–904, 1995.

[5] G. Deco and T.S. Lee, *An unified model of spatial and object attention based on inter-cortical biased competition.* Neural Computing, in Press.

[6] J. Duncan, J. and G. Humphreys, *Visual search and stimulus similarity. Psychological Review*, **96**: 433–458, 1989.

[7] U. Grenander, *Lectures in Pattern Theory I,II and III: pattern analysis, pattern synthesis and regular structures*, Springer-Verlag, 1976–1981.

[8] S. Grossberg, *Competitive learning: from interactive activation to adaptive resonance.* Cognitive Science **11**: 23–63, 1987.

[9] H.V. Helmholtz, Handbuch der physiologischen Optik. Leipzig: Voss, 1867.

[10] D.D. Hoffman, Visual intelligence: how we create what we see. W.W. Norton and Company, 1998.

[11] D.H. Hubel and T.N. Wiesel, *Functional architecture of macaque monkey visual cortex. Proc. Royal Soc. B*, (London), **198**: 1–59, 1978.

[12] B. Julesz, *Experiments in the visual perception of texture. Sci Amer.*, **232**: 34–43, 1975.

[13] J.J. Knierim, J.J. and D.C. Van Essen *Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J. Neurophysiology*, **67**: 961–980, 1992.

[14] V.A.F. Lamme, *The neurophysiology of figure-ground segregation in primary visual cortex. J. Neuroscience*, **10**: 649–669, 1995.

[15] T.S. Lee, and D. Mumford, *The role of V1 in scene segmentation and shape representation.* Society of Neuroscience Abstract, Vol. 22, 117.7, p. 283, 1996.

[16] T.S. Lee, D. Mumford, R. Romero, and V.A.F. Lamme, *The role of the primary visual cortex in higher level vision. Vision Research*, **38**(15–16): 2429–54, 1998.

[17] T.S. Lee and M. Nguyen, Dynamics of subjective contour formation in the early visual cortex. PNAS, **98**(4): 1907–1911, 2001.

[18] T.S. Lee, and S. Yu, *An information-theoretic framework for understanding saccadic eye movements.* In Advance in Neural Information Processing Systems, 12. Ed. S.A. Solla, T.K. Leen, K-R. Muller, MIT Press, 834–840, 2000.

[19] Z. Li, *Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex.* Network **10**(2): 187–212, 1999.

[20] D. Marr, *Vision.* N.J: W.H. Freeman & Company, 1982.

[21] J.L. McClelland and D.E. Rumelhart, *An interactive activation model of context effects in letter perception. Part I: An account of basic findings.* Psychological review: **88**: 375–407. 1981.

[22] D. Mumford, *On the computational architecture of the neocortex II.* Biological Cybernetics, **66**: 241–251, 1992.

[23] D.A. Pollen, J.P. Gaska, and L.D. Jacobson, *Physiological constraints on models of visual cortical functions.* Models of brain function, ed. Rodney M., Cotterill, J. Cambridge University Press, 115–135, 1989.

[24] V.S. Ramachandran, *Perception of shape from shading. Nature*, **331**: 163–166 (1988).

[25] R. Rao and D.H. Ballard, *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.* Nature Neuroscience. **2**(1): 79–87, 1999.

[26] S.C. Rao, G. Rainer, and E.K. Miller, Integration of what and where in the primate prefrontal cortex. Science, **276**(5313): 821–824.

[27] J. Reynolds, L. Chelazzi, and R. Desimone, *Competitive mechanisms su bserve attention in macaque areas V2 and V4. Journal of Neuroscience*, **19**: 1736–1753, 1999.

[28] P.R. Roelfsema, V.A. Lamme, and H. Spekreijse *Object-based attention in the primary visual cortex of the macaque monkey. Nature*, **395**(6700): 376–81. 1998.

[29] J. Sun and P. Perona, *Where is the sun? Nature Neuroscience*, **1**(3):183–4, 1998.

[30] A. Treisman and G. Gelade, A feature-integration theory of attention. *Cognitive P sychology*, **12**: 97–136 (1980).

[31] M. Usher and E. Niebur, *Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. Journal of Cognitive Neuroscience*, **8**: 311–327, 1996.

[32] R. von de Heydt, E. Peterhans, and G. Baumgarthner. *Illusory contours and cortical neuron responses. Science* **224**(4654): 1260–1262, 1984.