



Computations in the early visual cortex

Tai Sing Lee *

Center for the Neural Basis of Cognition and Department of Computer Science, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Abstract

This paper reviews some of the recent neurophysiological studies that explore the variety of visual computations in the early visual cortex in relation to geometric inference, i.e. the inference of contours, surfaces and shapes. It attempts to draw connections between ideas from computational vision and findings from awake primate electrophysiology. In the classical feed-forward, modular view of visual processing, the early visual areas (LGN, V1 and V2) are modules that serve to extract local features, while higher extrastriate areas are responsible for shape inference and invariant object recognition. However, recent findings in primate early visual systems reveal that the computations in the early visual cortex are rather complex and dynamic, as well as interactive and plastic, subject to influence from global context, higher order perceptual inference, task requirement and behavioral experience. The evidence argues that the early visual cortex does not merely participate in the first stage of visual processing, but is involved in many levels of visual computation.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Primate electrophysiology; Vision; Saliency; Segmentation; Symmetry; Shape

1. Introduction

1.1. The interactive nature of perceptual computation

Visual processing may be conceptualized as what Helmholtz [29] called the unconscious inference, or in more recent times what is referred to as Bayesian inference. That is, we must rely on contextual information and our prior knowledge of the world to make inferences about the world based on retinal data. Consider the image patch depicted in Fig. 1a. Seen alone, it is merely a collection of spots and dots. However, when placed in a larger scene context (Fig. 1b), the same image patch assumes a more specific and richer meaning. The image context in this case is also ambiguous, as it will take unfamiliar viewers a few minutes to perceive the object in the scene. However, once they are given information as to what they are supposed to see, i.e. a Dalmatian dog sniffing the ground near a tree, the image will begin to

crystallize in their minds. The spots and dots are transformed into the surface markings of the dog's body. Furthermore, if the same image is presented again to the same viewers in the future, they will be able to see the dog instantly. This terrific example by R.C. James illustrates the important role of both global context and prior knowledge play in perceptual inference.

How is visual inference carried out in the brain? Marr [53] proposed that there is a series of computational modules, each performing a relatively encapsulated computational step in image analysis, and that there is a rough correspondence between these modules and the areas in the visual cortex (Fig. 2). Subsequent theoretical and experimental work refined his analysis, sometimes modifying it, sometimes making it more precise, but still following the basic ideas. For instance, MT is considered to be the area where the aperture problem is solved by integrating local motion signals into a global whole, while V2 is the area where many Gestalt grouping operations are performed. A central tenet of this model, however, is the decomposition of visual processing into low, intermediate and high levels, and the belief that the visual cortex could likewise be divided these functional into these stages. In Marr's framework, the early visual area is the site of the *primal sketch*, where local features

* Address: Rm 115, Mellon Institute, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA. Tel.: +1-412-268-1060; fax: +1-412-268-5060.

E-mail address: tai@cnbc.cmu.edu (T.S. Lee).

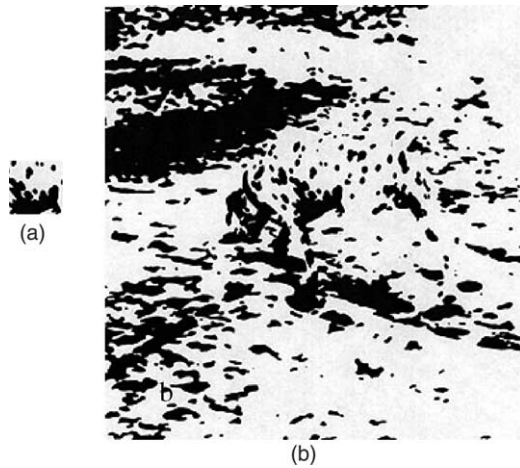


Fig. 1. (a) An image patch of spots and dots. (b) The image patch situated in a particular scene, as designed by R.C. James, depicting a Dalmatian dog sniffing the ground under a tree. This example elegantly illustrates how the interpretation of some images relies on top-down knowledge and contextual information.

are detected and grouped together into symbolic tokens and contours. Representations of surfaces and object models were thought to be computed and represented in the extrastriate cortices such as V4, MT and IT.

Object recognition in complex real world environments under multiple occlusions, perspectives and lighting conditions is a very difficult problem. Before recognizing the object, it is often hard to distinguish it from the background because on the one hand, its true contours are confused with local contrast edges caused by shadows, specularities and surface discontinuities; on the other hand, the true object edges can be irregular, faint and partially occluded. To find the boundary of an object, the first set of contrast edges must be discounted and the second set must be enhanced or interpolated. However, an object might need to be segregated from the background and its boundaries defined before one can compute its shape properties and recognize it. These

shape properties will have to be modified if the object is partly occluded or in shadow. An example is shown in Fig. 3: although the figure of the old man is extremely obvious to human perception, application of the popular edge detector makes mistakes in all of the above. We believe that this figure cannot be separated from the background without substantial reconstruction of the 3D structures and an understanding of the illumination of the scene. Curiously, the most recognizable object in the scene is the man's ear, which might, for example, entrain the process of matching next the face, and finally the body. In visual computation, it is often necessary to go back to the high resolution version of the stimuli because some details that were overlooked as noise in the first pass could turn out to be crucial in confirming the identity of an object. In other words, perceptual inference and object recognition are intertwined: they cannot progress in a simple bottom-up serial fashion, but have to take place concurrently and interactively in constant feedforward and feedback loops that involve

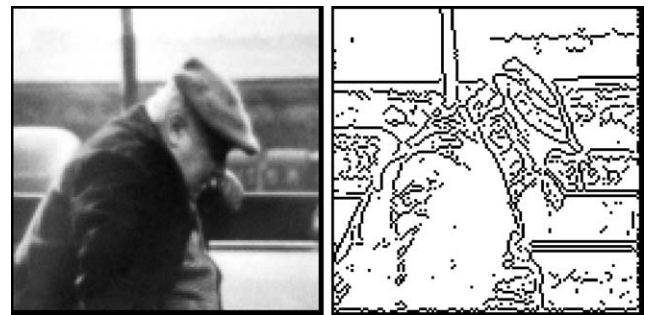


Fig. 3. An image of an old man and the edge signals produced by applying the popular Canny edge detector to the image. It illustrates that bottom-up edge signals are inherently difficult to interpret because of the ambiguities in local contrast edge information. Segmentation and recognition are necessarily intertwined, involving the entire hierarchical circuit of the visual system at the same time.

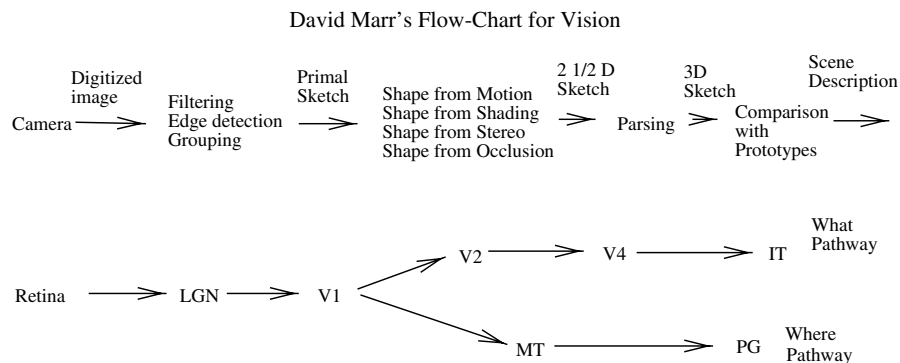


Fig. 2. Marr's framework of visual processing includes a series of computational modules chained together in a hierarchy, first computing local edges and tokens, then grouping them together to infer surface orientation of objects. With the 2.5D sketch, the different objects are then parsed and matched against stored 3D prototypes. It was thought that there is a rough correspondence between these modules and the different visual cortical areas.

the entire hierarchical circuit of the visual system [18,26,45,55,58,67].

1.2. High-resolution buffer hypothesis

In this context, the early visual area should not be considered simply as an “early module” that processes bottom-up information and then passes the results onto the other extrastriate areas, but rather as an integrated part of an inference engine that participates in many levels and varieties of visual computations, provided that these computations require specificity in spatial precision and image details that only the early visual machinery can furnish. This is the essence of the “high-resolution buffer hypothesis” of V1 that Mumford and I proposed [44,45,60]. This specific hypothesis predicts that *any level of visual reasoning that requires spatial precision and high resolution details necessarily engages the early visual areas.*

The motivation of this hypothesis is actually quite simple. In computer vision, algorithms for perceptual organization are often operated on arrays of numbers, and the results are represented in some arrays with resolution as high, or sometimes even higher, than that of the input arrays. For example, algorithms for computing contours use high-resolution edge maps to indicate the orientations and locations of the edge segments [1,6]. Algorithms for computing shape from shading often utilize a high-resolution needle map in which surface orientation is represented at each spatial location [30,42]. Fig. 4 provides some examples of high-resolution buffers to illustrate these ideas.

Representations in the early visual areas (LGN, V1 and V2) are precise in both space and feature domains because their receptive fields are small, arranged in retinotopic coordinates [31] and decomposed into oriented, wavelet-like local representations that can capture the full details of the image [16,43]. The size of the receptive fields of neurons increases dramatically as one traverses the successive visual areas along the two visual streams (the dorsal “where” and the ventral “what” streams). For example, the receptive fields in V4 or MT are about four times larger in diameter than those in V1 at the corresponding eccentricities [21]; and the receptive fields of neurons in the inferotemporal cortex (IT) tend to cover a large portion of the visual field [25]. This dramatic increase in receptive field size indicates a successive convergence of visual information that not only is necessary for extracting invariance and abstraction (e.g. translation, size), but also results in the loss of spatial resolution and fine details in the higher visual areas.

How can higher order inference that relies on information with spatial precision and fine details be accomplished? One possible strategy is that of coarse coding, in which the precise information might be

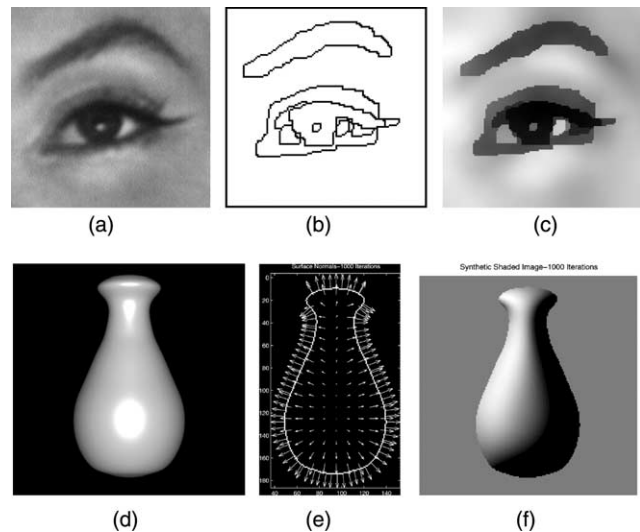


Fig. 4. High-resolution buffers are commonly utilized in computer vision. (a–c) Image segmentation using Blake and Zisserman’s [6] weak-membrane model. The model formulates segmentation as an energy minimization problem. It takes an image (a) and produces two outputs: a boundary process map (b) which indicates the location of boundary, and a surface attribute map (c) which indicates the smoothed (interpolated) luminance values on the surface of an object. Both the line process map and the surface attribute map are arrays of the same resolution as that of the input image and could be considered as high-resolution buffers. (d–f) Shape from shading computed using Horn’s [30] classic algorithm, which also formulates the problem in terms of energy minimization. It takes in an image (d) and computes the contour of the object. The surface normal at the contour is perpendicular to the viewer. The algorithm uses the shape from shading constraint to propagate the surface orientation estimates in from the border. The needle map (e) is a high-resolution buffer to represent surface orientation at each location. Each “needle” is the projection of a surface normal vector erected on the patch of the surface corresponding to a particular pixel, (f) illustrates the rendering of the vase model, represented by the needle map, with another light source.

implicitly encoded in the differential responses of higher cortical neurons in a distributed fashion. However, there are at least two difficulties associated with this solution. First, the higher order areas have to implicitly and redundantly encode the precise information and then force the burden of decoding onto even higher areas, as cortical modules do have a tendency to make the representation of certain information or attributes explicit. Second, without an explicit common forum in which all the expert modules can communicate to each other, a tremendous amount of wirings would be needed to fully connect all of the individual modules.

A simpler alternative solution is to exploit the known reciprocal connection between higher order cortical areas and the early visual areas (LGN, V1 and V2), so that the high-resolution computational and representational machinery in the early visual areas can be harnessed to compute and represent higher order information. In this scenario, LGN, V1 and V2 could serve as high-resolution buffers for the rest of the visual

cortex utilizing the reciprocal connections between them and the extrastriate cortical areas. This perspective dictates that the early visual areas do not merely perform simple filtering [17] or feature extraction operations [31], then turning their results over to higher areas for further processing. Rather, the early visual cortex likely continues to be involved in all levels of perceptual computations, insofar as such computations require the utilization of their intrinsic computational machineries. Computations concerning geometry such as contours, surfaces and shapes likely involve V1 and V2 as the precise orientation and disparity tuning of the neurons in these areas make them ideal “geometric computational engines”. This hypothesis thus predicts that, when presented with static images, the early part of V1 neurons’ response should reflect the results of feedforward filtering operations, while the later part of their response should reflect higher order constructs emerging from the interactive computation across the visual hierarchy.

This conjecture then suggests that the neural activities in the early visual cortex could provide a clue for us to understand the evolution of many visual computational algorithms. In this paper, we will examine several pieces of neurophysiological evidence in the early visual cortex in relation to geometric computations in vision: the inference of contours, surface, shapes, and forms. The evidence, some solid and some anecdotal, provides tentative support for the high-resolution buffer hypothesis, and argues for a much broader functional and a richer computational role for the early visual cortex.

2. Segmentation and contour completion

Since Hubel and Wiesel’s [31] discovery of the orientation selectivity of V1 neurons, it has been generally agreed upon that these neurons should play an important role in edge detection. Oriented edge filters are widely utilized in edge detection algorithms [12] as well as many neural models for contour detection and completion [27,38,51]. It was further assumed that the horizontal connections in V1 [24] provide the computational infrastructure for tracing and completing boundary and curves. Although the receptive field properties of V1 neurons have been studied extensively using drifting bars and sinewave gratings, few studies have examined the neurons from a more abstract visual processing perspective. Creutzfeldt and Nothdurft [13] were the first to explore the responses of LGN neurons, as well as V1 simple and complex cells to different parts of a complex and natural image. Their conclusion was that the early visual neurons were reducing complex stimuli, such as a photograph of a natural environment, to its contours. Almost 15 years passed before Gallant et al. [20] studied the neural activity in V1, V2, and V4 in response to natural scene during free viewing. In 1995,

Lamme [40] examined the response of V1 neurons to different regions of a texture image and found that V1 responded better when their receptive fields were located inside the figure than when they were in the background, providing a clue that V1 neurons might be computing or reflecting something that might be more abstract in nature. In this context, we [45,46] carried out a series of experiments to study the detailed dynamics of V1 neurons in contour processing. Here, I will review some of our basic findings on the dynamics of early cortical responses to a variety of contour stimuli.

All the experimental results discussed in this paper came from single-unit recordings in the primary (V1) and secondary visual cortex (V2) of awake behaving rhesus monkeys (*Macaca mulatta*), and have been previously documented [45–47]. Recordings were conducted transdurally with epoxy-coated tungsten electrodes or glass-coated platinum electrode through a surgically implanted well overlying the operculum of area V1. The eye movements of the monkeys were monitored using implanted scleral search coils [76] and sampled at a rate of 200 Hz.

The typical paradigm of these experiments was to present different parts of the stimulus image to the receptive field of a neuron in successive trials, while the neuron was being recorded. In each trial, while the monkey was fixating at a red dot on the screen within a $0.5^\circ \times 0.5^\circ$ window, a full screen sample stimulus was presented for a period of time. The duration of presentation varied depending on the specific experiments, but typically around 350–450 ms. Each stimulus was presented at a randomized series of sampling positions relative to the cells’ classical receptive fields (CRF) (Fig. 5). The receptive fields of the V1 and V2 neurons we studied were located at 2–5° away from the fovea. Their CRFs, as plotted by small oriented bars, were less than 1° (typically around 0.6° in diameter) for V1 neurons, and less than 1.6° for V2 neurons. Most earlier experiments in the early visual cortex [48,77] used moving stimuli that drifted across the receptive field of the neuron. The temporal response of neurons in those studies was a function of both the dynamics of the neurons in response to the stimuli, as well as the temporal change of each stimulus. This paradigm of step presentation of static stimuli allowed us to study the temporal evolution of the neuronal responses to specific parts of the stimulus.

2.1. Luminance contour

V1 neurons responded to luminance contrast borders at 40 ms after stimulus onset (Fig. 6)—that is approximately the time for the retinal signals to reach V1. When the orientation of the luminance border was aligned with the preferred orientation of the cells, the neuronal response was focused at the luminance border in the

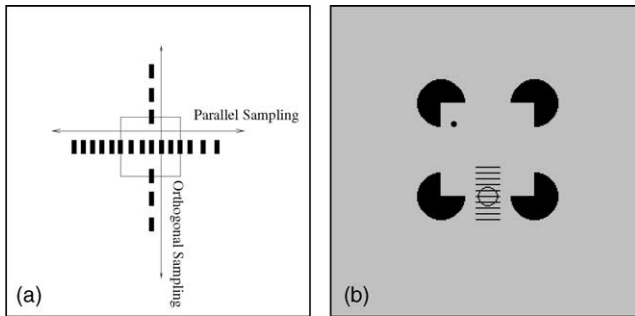


Fig. 5. Stimulus presentation paradigm. (a) Luminance and texture boundary experiment: the displays illustrate the placement of the luminance or texture boundary relative to the receptive field in two testing conditions. In the *parallel* sampling scheme, the preferred orientation of a cell was parallel to the boundary, and in the *orthogonal* sampling scheme, the preferred orientation of the cell was perpendicular to the boundary. Here, the square figure is shown to be displaced horizontally over successive trials so that the cell's receptive field was placed at the center, the boundary and outside of the figure in different trials. In the parallel sampling scheme, the preferred orientation of the cell was parallel to the figure boundary it encountered. In the orthogonal sampling scheme, the preferred orientation of the cell was orthogonal to the figure boundary it encountered. The *sampling line* is defined as the line on which the receptive field of the cell is translated over trials. The response to the image was sampled with 16 steps across the image. The step size was 0.5° inside the figure and close to the boundary. (b) Illusory contour experiment: The spatial relationship between the fixation spot (black dot), the cell's receptive field (circle), and the stimulus figure. The cell's receptive field was placed in the middle of the illusory contour at the bottom of the Kanizsa figure, for a cell selective for horizontal orientation. For cells of other orientations, the stimuli were rotated accordingly so that the contour was parallel with the preferred orientation of the cell. The stimulus was presented in 10 locations 0.25° apart in successive trials relative to the receptive field so that the illusory contour was collinear with the 10 line segments as shown in the figure.

stimuli from the very beginning. When the orientation of the luminance border was orthogonal to the preferred orientation of the cells, the cells did not give marked

responses. In this case, the local contrast feature and the global perceptual border is of the same orientation, we cannot tell whether the response was due to the local feature and to the global perceptual structure.

2.2. Texture contour

To dissociate the influence of local features and that of global perceptual constructs, we study the responses to the neurons to texture stimuli with borders defined by orientation contrast (Figs. 7 and 8). The temporal response of V1 neurons to these stimuli can be classified into at least two stages: the initial (40–80 ms after stimulus onset) response to the local oriented feature and the later response (80 ms onwards) to the global perceptual structure. In the examples shown, since the neurons' preferred orientation was vertical, the vertical line segments inside the central vertical strip (4° in width) (Fig. 7) stimulated the neurons well, producing an uniformly high response within the vertically texture strip and a uniformly low response to the horizontally textured background during the initial stage. This reflected the orientation tuning property of the neurons. Starting at 70 ms, however, the neural responses started to 'contract' spatially toward the boundary, resulting in a sharp and sustained response at the texture contrast border by 80 ms. The responses of vertical neurons to vertical texture border defined by slanted textures were more dramatic (Fig. 8): the cells' initial responses were relatively mild because both the textures inside and outside the central texture strip were not of the preferred orientation of the cells. Remarkably, 80 ms after stimulus onset, the cells started to respond to the texture boundary, and at a rate higher than the initial response. This evidence suggests that the later response of the V1 neurons is not merely a habituated version of the initial response, but is signalling something possibly more

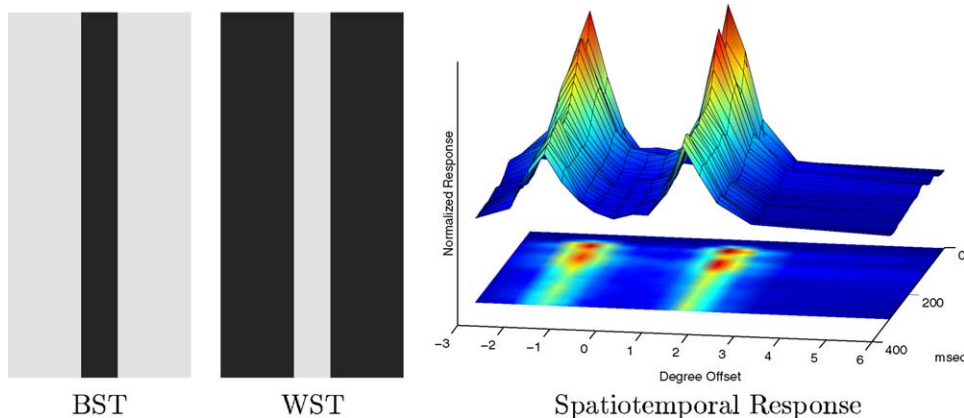


Fig. 6. Average spatial response profile of a population of 12 vertical complex cells responding to BST (black strip) and WST (white strip), showing almost instantaneous boundary formation (40 ms) and a slight spreading of the border excitation signals inside and outside the strip during the resurgent period (80–120 ms after stimulus onset). The response to the BST and the response to the WST was summed together at corresponding spatial locations (see [45] for details).

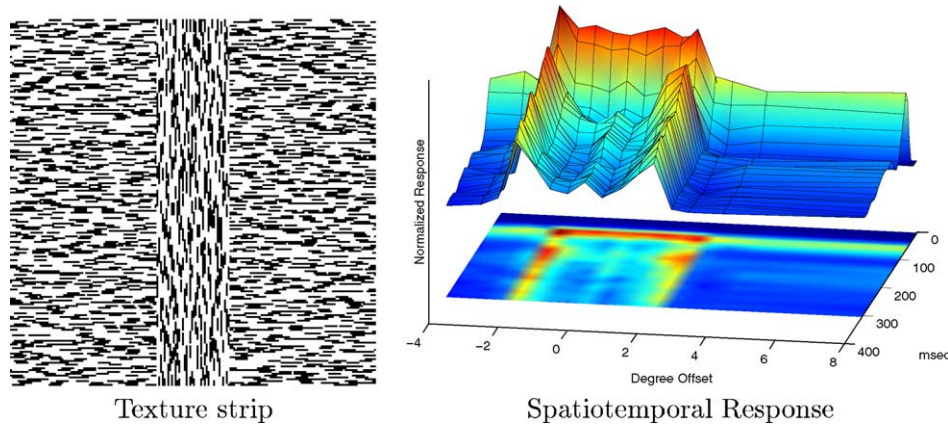


Fig. 7. Spatiotemporal response of a population of V1 neurons to the different parts of the texture strip stimulus, tested with the parallel sampling scheme, i.e. along a horizontal sampling line. The abscissa is the distance in visual angles from the RF center to the center of the strip. The texture boundary is located at -2.0° and 2.0° visual angles away from the center. The responses to the texture stimuli were therefore initially uniform within a region of homogeneous texture based on the orientation tuning of the cells. At 60 ms after stimulus onset, boundary signals started to develop at the texture boundaries. By 80 ms, the responses at the texture boundaries had become sharpened [45].

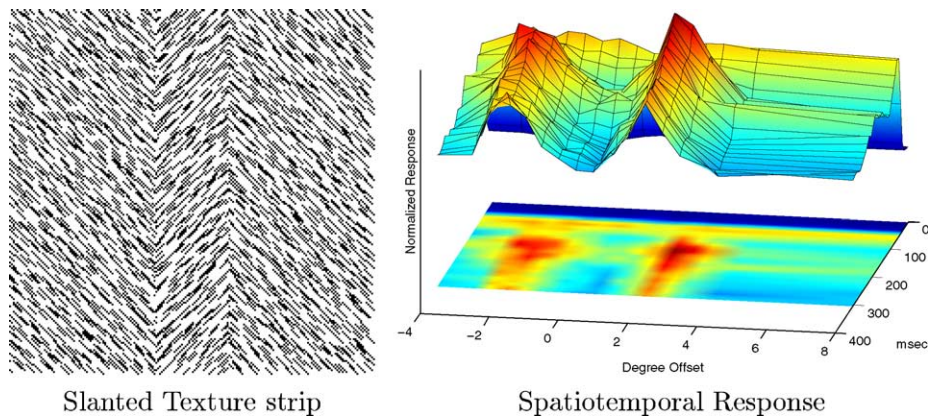


Fig. 8. Spatiotemporal response of a population of V1 neurons to the different parts of the slanted texture strip stimulus, tested with the parallel sampling scheme. The abscissa is the distance in visual angle from the RF center to the center of the strip. The texture boundary is located at -2.0° and 2.0° visual angles away from the center. The cells' initial responses were relatively mild because both the textures inside and outside the figure were not of the cells' preferred orientation. Remarkably, texture boundary signals that emerged during the resurgent period were stronger than their initial responses, suggesting that the later responses of the cells were sensitive to the orientation of the global boundary than to the local texture features [45].

abstract such as the existence of a global vertical perceptual contour.

Texture contrast enhancement, or boundary sharpening, theoretically could be mediated by lateral inhibition among neurons through the horizontal collaterals in the superficial layer of V1, as demonstrated in several modeling studies [51,72,73]. Lateral inhibition can be understood in several functional terms, either as a mechanism for non-maximum suppression [6,12], or a mechanism for sparse [49,63], or predictive coding [67,71]. In the context of contour processing, the idea boils down to selecting the most probable boundary and suppressing the spurious ones. However, lateral inhibition and non-maximum suppression is only a part of the

contour processing algorithm; the other part is contour continuation and completion [1,22,27,51,59,61,78,82]. Most of these approaches for contour completion assume implicitly or explicitly an association field that integrates edge information to interpolate a smooth and 'elastic' contour. The existence of such an association field or local facilitation connections have been derived either theoretically [78], or from psychophysical experiments [19] or edge co-occurrence statistics in natural images [23]. Gilbert and coworkers [35] recently demonstrated that the response of a neuron to a bar was enhanced when another bar was placed in a longitudinal direction in the surround—an enhancement pattern that is consistent with the facilitatory component of the association field.

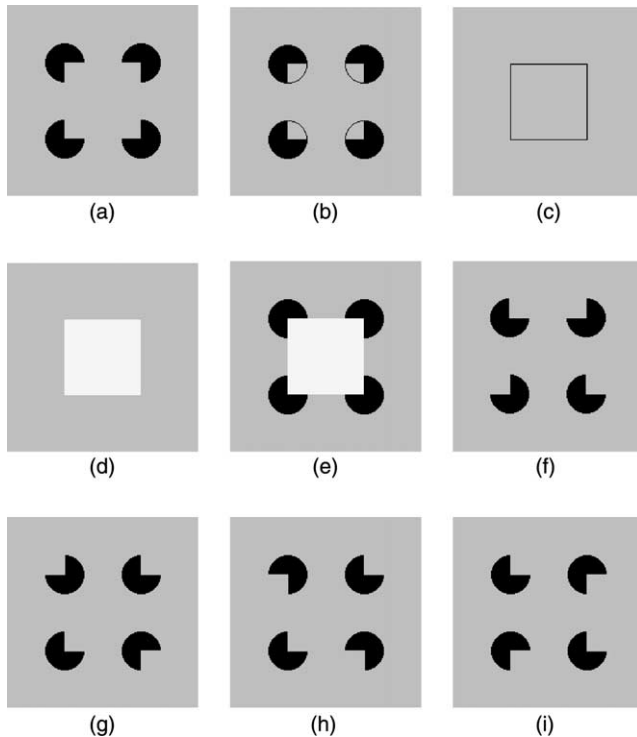


Fig. 9. Selected stimuli used in the illusory contour experiment. (a) Kanizsa square with illusory visual contour, made up of 4 corner discs (pac-men). (b) Amodal contour. The illusory contour was interrupted by intersecting lines. It can be perceived as a gray square in a black background seen through four holes; the contour is amodal as it exists only cognitively but not visually. (c) Square with line contours. (d) Square with contrast borders. (e) Square with contrast borders plus occlusion. (f-i) Various configurations of the rotated corners.

2.3. Illusory contour

To establish that V1 participates in contour continuation and completion, we [46] tested whether V1 and V2 neurons are sensitive to the illusory contour of the Kanizsa square (Fig. 9). The classic experiment by von der Hedyt et al. [77] demonstrated that the neural correlate of subjective contour can be found in V2. However, they did not see significant signals at V1. This, in some way, was in direct contradiction with the high-resolution buffer hypothesis and has been taken by many as evidence that visual computation is primarily a feedforward process. However, the stimulus in the original experiment was an illusory bar sweeping across the receptive field. This might have provided insufficient reason for the feedback signals to go back to V1. We reasoned that the perception of a sharp subjective contour might necessitate the participation of V1. In a procedure similar to the texture experiment, we tested V1 and V2 neurons' responses to illusory contour, and various types of real contours and controls, when they were placed at different locations relative to the receptive fields of the cells (Fig. 5b). The static presentation of the

stimulus allowed us to monitor the temporal dynamics of the response.

With additional necessary manipulations designed to capture the monkey's attention to the illusory figures (see [46] for details), we found that the neural correlate of an illusory contour emerged in V1 at precisely the location it was supposed to be, as determined by the cell's responses to a line or a luminance contrast boundary (Fig. 10a). The response to the illusory contour was delayed relative to the response to the real contours by 55 ms (Fig. 10b), emerging at about 100 ms after stimulus onset. The response to the illusory contour was significantly greater than the response to the controls, including the stimuli in which the illusory contour was interrupted by lines, or in which the pac-men were rotated (Fig. 10c). At the population level, we found that sensitivity to illusory contour emerged at 65 ms in V2, and 100 ms at the superficial layer of V1 (Fig. 10d and e). One possible interpretation is that V2 could be detecting the existence of an illusory contour by integrating information from a more global spatial context, and then feeding the information back to V1 to facilitate contour completion. Since V2 receptive fields are larger, and their feedback to V1 is more spatially diffused [56], it seems rather improbable that V2 could actually synthesize a precise contour and then feed it back to V1. A more likely scenario, as shown in Fig. 11, is that an oriented V2 neuron, once activated by the illusory contour, feeds back to excite the V1 neurons of the appropriate orientation under its spatial coverage, i.e. in a spatially non-specific, but feature-specific, manner. This effectively provides a top-down prior that allow the V1 circuitry to integrate with the bottom-up cues to construct the precise and spatially sharp contour.

3. Saliency and surface coloring

A second class of computational phenomena that has been observed in the early visual cortex is related to the coloring of salient objects in a visual scene. This is one of the visual routines proposed by Ullman [82]. Coloring of an object serves several functions: it facilitates the computation of shape; and it attracts spatial attention to the salient object. The 'intensity' of the color is proportional to the perceptual saliency of the object.

The computation of object saliency in a visual scene is important for directing the spotlight of attention and for guiding eye movements during scene analysis. Earlier single-unit studies in awake and anesthetized monkeys have implicated the primary visual cortex in the mediation of bottom-up pop-out saliency computation of oriented bars and contrast gratings. Consider the stimuli in Fig. 12: the vertical bar appears to be perceptually more salient or pop-out when it is surrounded by hori-

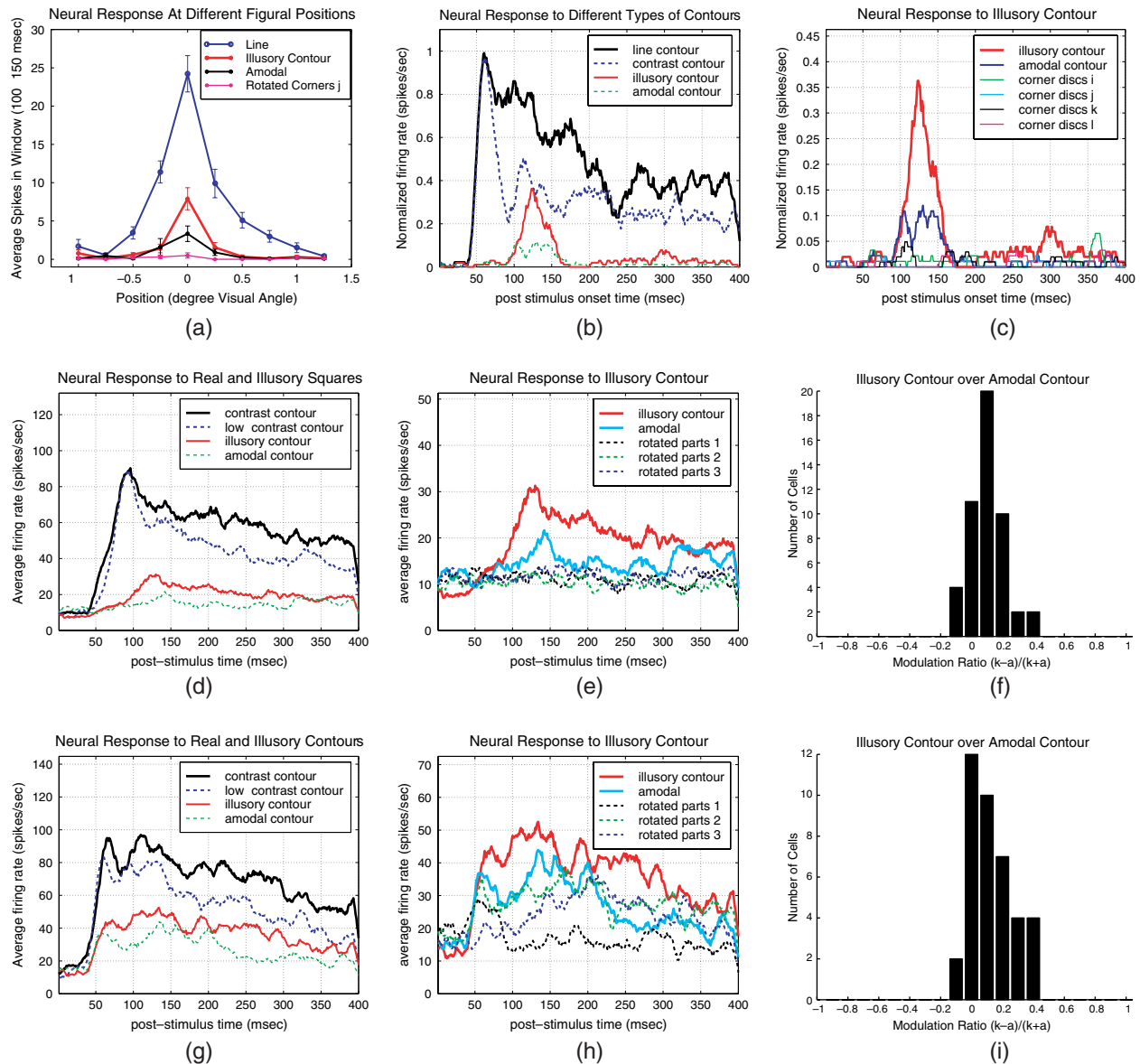


Fig. 10. (a) The spatial profile of a V1 neuron's response to the contours of both real and illusory squares, in a temporal window 100–150 ms after stimulus onset. The real or illusory square was placed at different spatial locations relative to the receptive field of the cell. This cell responded to the illusory contour when it was at precisely the same location where a real contour evoked the maximal response from the neuron. This cell also responded significantly better to the illusory contour than to the amodal contour (T -test, $p < 0.003$) and did not respond much when the pac-men were rotated. (b) Temporal evolution of this cell's response to the illusory contour, the amodal contour and the various rotated corner disc controls at the location where the real contour elicited the maximum response. The response to the illusory contour emerged at about 100 ms after the illusory square appeared. The cell responded slightly to the amodal contour and did not respond to any of the rotated corner discs. (c) The cell's response to the illusory contour compared to its response to the real contours of a line square, or a white square. The onset of the response to the real contours was at 45 ms, about 55 ms before the illusory contour response. (d) Population averaged temporal response of 50 V1 neurons in the superficial layer to the real and illusory contours. (e) Response to the illusory contour of the 50 V1 neurons compared to the response to the controls. (f) Histogram of the illusory contour modulation ratio, which is computed by $(R_i - R_a)/(R_i + R_a)$, where R_i is the response to the illusory contour, and R_a is the response to the 'amodal contour' control. (g–i) Same as (c–e) but for 39 V2 neurons in the superficial layer.

zontal bars than when it is surrounded by vertical bars. What is the neural correlate of this enhanced perceptual saliency? Knierim and Van Essen [37] found that a certain percentage of the V1 neurons responded better to the pop-out condition than to the uniform condition even though the CRF of the neuron was seeing exactly the same stimulus, indicating a sensitivity of the indi-

vidual neurons' response to the global context of the surround. This sensitivity to surround context was considered a neural correlate of perceptual pop-out. Analogous perceptual pop-out effect has also been observed due to orientation contrast in sine wave gratings (Fig. 12b) in various cortical areas [7,24,48,50,52,54]. These phenomena have been observed in both anesthe-

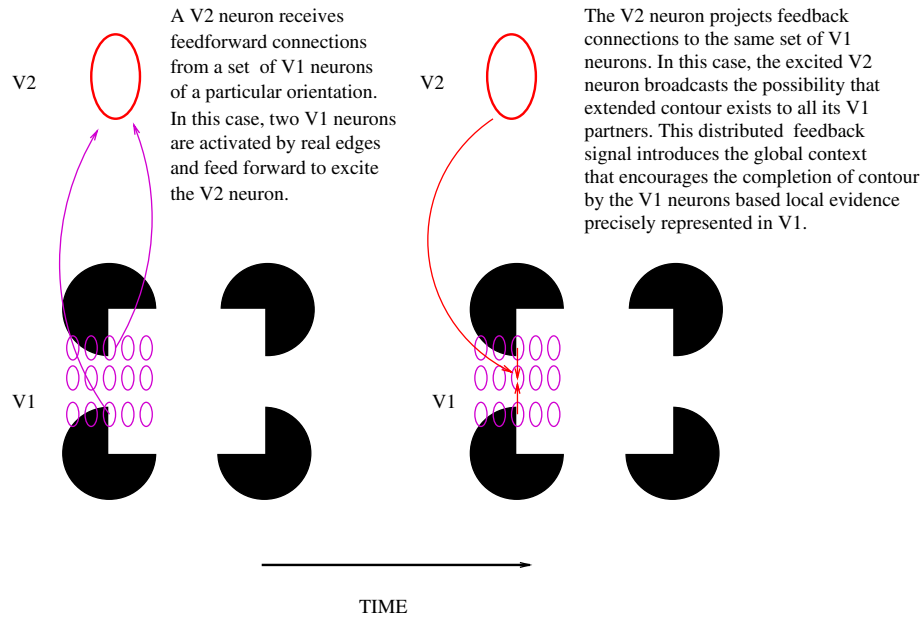


Fig. 11. A scenario of the interactive computation between V1 and V2.

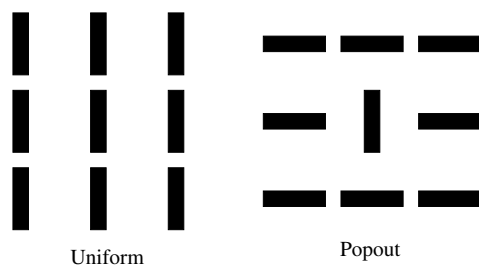


Fig. 12. The phenomenon of pop-out. An oriented bar is more salient when it is surrounded by bars of the orthogonal orientation than when it is surrounded by bars of the same orientation. This increase in saliency allows the bar to pop out—the time required to detect it is independent of the number of distractors.

tized and awake animals [62], though there was some evidence that awareness or consciousness, and intactness of the extrastriate visual cortex might be necessary for the texture contrast figure-ground effect in V1 [41,75]. Theoretically speaking, since all these stimuli are constructed from oriented line segments, the computation of pop-out could be mediated purely by lateral inhibition using the horizontal collaterals in the primary visual cortex without necessarily involving the recurrent feedback [32,51,72,73]. In these texture and bar studies, perceptual saliency and orientation contrast of the stimuli were correlated, making it difficult to assess the contributions from feedforward/local mechanisms and that from intercortical feedback. In order to separate the contribution of bottom-up and top-down processes to neural phenomena in the early visual areas, we [47] examined the shape from shading pop-out phenomena.

3.1. Shape from shading

Shape from shading stimuli have been used to demonstrate that parallel pop-out can occur with “high level” perceptual constructs [8,66,74]. It has long been a mystery as to why shape from shading stimuli, which presumably require processing by higher order perceptual areas, segregates readily into groups (Fig. 13a) and pop out ‘preattentively’ as if they were textons (Fig. 14a). Ramachandran [66] demonstrated that shape from shading stimuli pop out more readily than two-dimensional contrast patterns (Figs. 13b and 14b). The degree of perceptual pop-out saliency depends on both the 3D shape as well as the lighting direction. LA (lighting from above) and LB (lighting from below) stimuli in Fig. 13b pop out readily because we have this built-in bias of assuming lighting come from above. The perception of the convexity and concavity of elements are immediate.

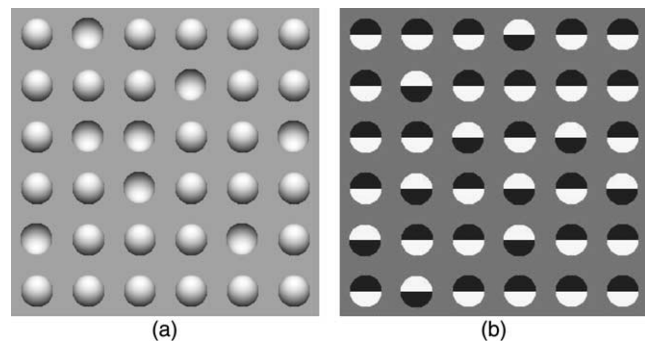


Fig. 13. (a) Convex and concave shapes defined by shading segregate instantly into two groups. (b) Patterns of different contrast polarities did not segregate as readily.

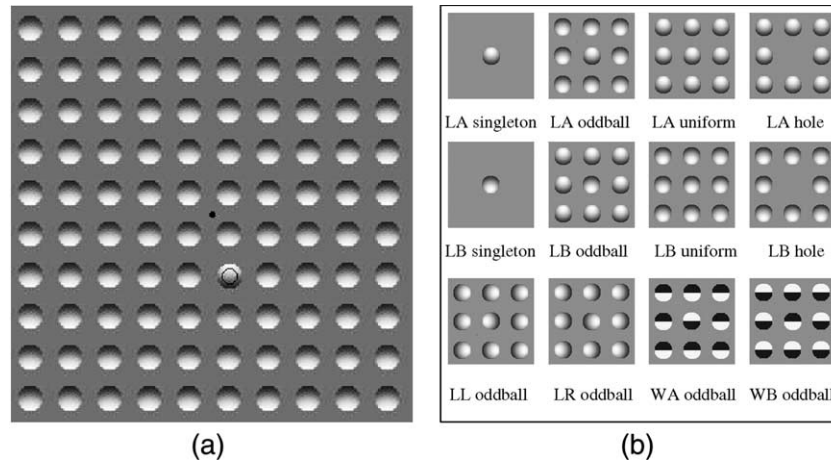


Fig. 14. (a) Stimulus set for the shape from shading pop-out experiment. A typical stimulus display was composed of 10×10 stimulus elements. Each element was 1° visual angle in diameter. The diameter of the CRF of a typical cell at the eccentricities tested ranged from 0.4° to 0.8° visual angle. Displayed is the LA (lighting from above) oddball condition, with the LA oddball placed on top of the cell's receptive field, indicated by the open circle. The solid dot indicates the fixation spot. (b) There are six stimulus sets, each defined by the receptive field element, i.e. LA (lighting from above), LB (lighting from below), LL (lighting from left), LR (lighting from right), WA (white above), and WB (white below). Each stimulus set had four conditions: singleton, oddball, uniform and hole. Displayed are the iconic diagrams of all the conditions for the LA set and the LB set, as well as the oddball conditions for the other four sets. The center element in the iconic diagram covered the receptive field of the neuron in the experiment. The surround stimulus elements were placed outside the RF of the neuron. The comparison was between the oddball condition and the uniform condition, while the singleton and the hole conditions were controls. The singletons measured the neuronal response to direct stimulation of the RF alone, and the holes measured the response to direct stimulation of the extra-RF surround only.

LL and LR (lighting from left, right) stimuli pop out to a less degree because the interpretation of lighting direction, and hence the 3D shape of the elements, is ambiguous. WA and WB (white above, below) stimuli, even though they are stronger in luminance contrast, are much weaker in perceptual saliency. The computation of pop-out for the shape from shading stimuli likely involves both the parallel machinery in the early visual area and extrastriate areas for inferring 3D shapes. The stimuli WA and WB thus are good control stimuli for dissociating bottom-up contrast saliency and top-down perceptual saliency.

We examined whether V1 and V2 neurons responded differently to the oddball condition and the uniform condition when the receptive field of the tested neuron was covered by the same stimulus element. Six types of stimuli (LA, LB, LL, LR, WA and WB) were tested, each with the four types of conditions (singleton, oddball, uniform and hole). A singleton stimulus and a hole stimulus were used as controls for each stimulus types. The singleton stimuli measured the neuronal response to direct stimulation of the RF alone, without any surround stimulus, while the hole stimuli measured the response to direct stimulation of only the extra-RF surround. In each trial, one of the conditions was displayed on the screen for 350 ms while the monkey fixated at a red fixation dot (shown as a black dot in Fig. 14a). The monkeys were performing a fixation task during all the neuronal recording sessions.

Before the monkeys had to utilize the stimuli in their behavior, V2 neurons, but not V1 neurons, were found

to be sensitive to shape from shading pop-out [47], as shown in Fig. 15a–f. After we trained the monkeys to perform an odd-ball detection task, in which the monkey had to make a saccade to the oddball stimulus appearing at one of the four random locations, V1 began to exhibit pop-out sensitivity to the shape from shading pop-out stimulus but not to the WA and WB contrast stimuli in a fixation task (Fig. 15g–i). In both stages (before and after behavioral training), V1 and V2 neurons were not sensitive to the difference in the surround in the oddballs and the uniform conditions for the WA or the WB stimuli, i.e. these stimuli did not elicit a pop-out response in the early visual cortex.

When we changed the statistics of the stimuli in the environment, in this case the presentation frequency of different types of oddball targets, the behavioral performance of the monkeys in detecting the oddball stimuli adapted accordingly: a target that appeared more frequently was detected with shorter reaction time and high accuracy, reflecting a change in the perceptual saliency of specific target. A corresponding change in neural pop-out responses was also observed in V1 and V2 neurons, suggesting that perceptual saliency is not static, but dynamic and malleable, contingent on the animals' experience and behavioral relevance of the stimuli. Given that the observed effect was not restricted to the retinotopic location of the target during training, plasticity component of the observed effect was likely distributed over multiple memory and perceptual areas or in the feed-forward/feedback connections between cortical areas, although the changes in V1 circuitry were also possible.

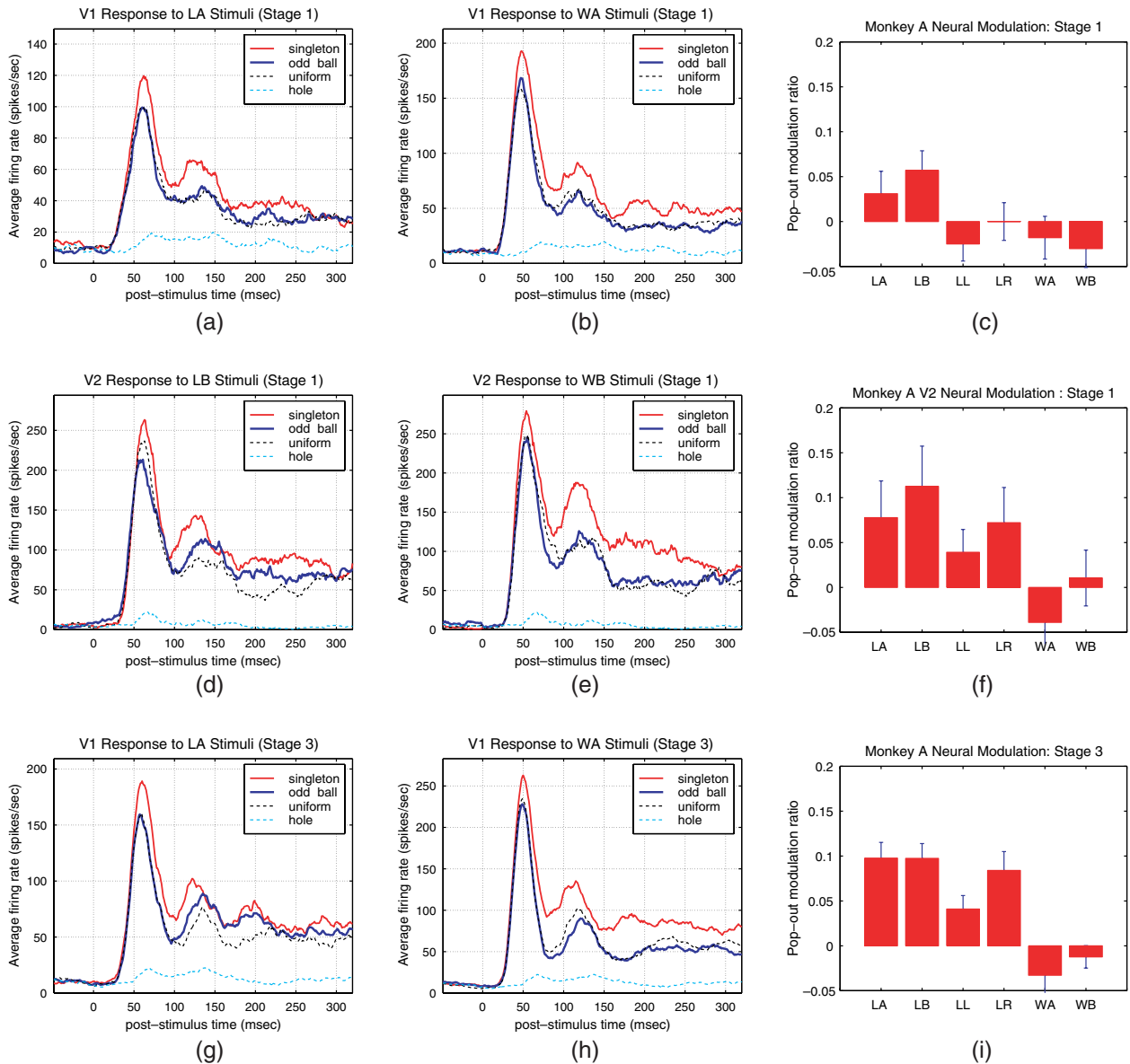


Fig. 15. (a,b) Temporal evolution of the normalized population average response of 30 V1 units from monkey A to the LA set (a) and the WA set (b) in a stage prior the monkey had utilized the stimuli in its behavior. Each unit's response was first smoothed by a running average within a 15 ms window, then averaged across the population. A very small difference (pop-out response) was observed between the population average response to the oddball condition and that to the uniform condition in the LA set. No pop-out response was observed in the WA set. (c) Mean pop-out modulation ratios of 30 units for all six stimulus sets in this 'preattentive' stage. Pop-out enhancements were significant for LA and LB ($P = 0.011, 0.007$), but not for the other stimuli. The error bars were the standard errors of the means. (d-f) Data are from 15 V2 units in the preattentive stage showing more significant shape from shading pop-out in V2. (g-i) Data from V1 45 units after the monkeys have utilized the stimuli in their experience. Pop-out enhancement increased and became highly significant for stimuli LA, LB, LL, and LR ($p = 10^{-6}, 10^{-6}, 0.0045, 10^{-4}$), but not for WA and WB. A second monkey's data were similar.

When the behavioral measurements during the various training stages were regressed against the neural pop-out responses obtained in the adjacent recording stages, the two were found to be highly correlated (Fig. 16): a stronger pop-out response (the enhancement of response to the oddball condition relative to the uniform condition) was matched to a shorter reaction time and higher accuracy in detection performance. This indicates that the neural pop-out signals were correlates of the

subjective perceptual pop-out saliency. The observation that pop-out due to 3D shapes emerged at the level of V1 and V2 suggests that the computation of perceptual saliency involves both bottom-up processing in V1 and the top-down spatial attentional modulation from the dorsal stream, as well as top-down 3D shape inference feedback from the ventral stream areas [34,65].

What is the role of V1 in this computation? Could it simply be reflecting passively the attentional effect

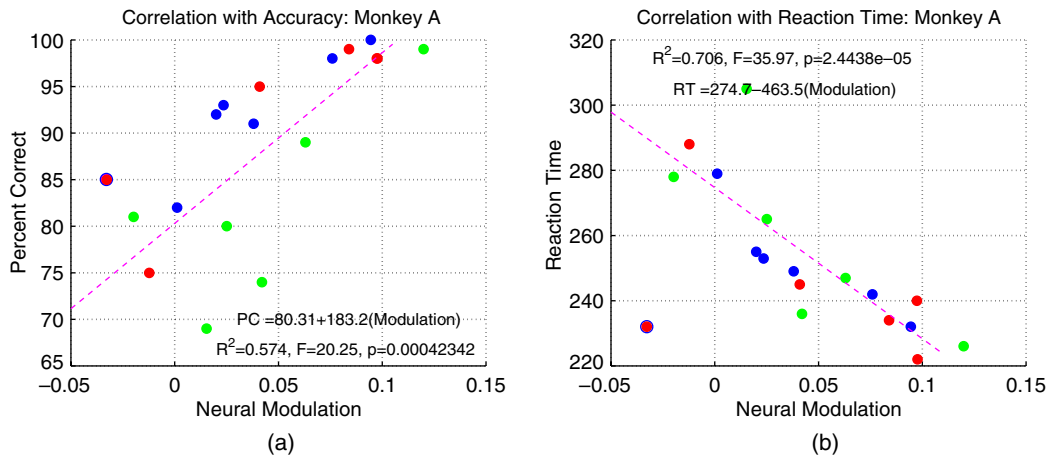


Fig. 16. Correlation between performance and neural response. In the shape from shading experiment, we trained the monkeys to do oddball detection task with stimuli with different presentation frequencies. The monkeys' behaviors and neural responses adapted after each stage of training. Here, behavior performance measurements (percent correct and reaction time) in three different training stages: 2 (red), 6 (green), and 8 (blue) were paired with neural pop-out modulation data in three recording stages: 3 (red), 5 (green), and 7 (blue) (see [47] for details). Each pair of stages produced six data points (corresponding to the six stimulus types). Eighteen points are shown in each graph relating a behavioral measure with neural pop-out modulation. Reaction time and percent correct was regressed on the pop-out modulation independently. A linear regression line, their equations and statistical significance are shown in each plot. An out-lier, which was over 2.5 standard deviations away from the regression line, was discarded in each graph (the red dot with blue outline). The out-lier could have arisen from interference due to other top-down influences. Both accuracy and reaction time were correlated with the neural pop-out modulation. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

higher up in the extrastriate cortex? Granted the signal we observed was related to perceptual saliency, it is computationally advantageous to distribute this signal over the surface of the salient object, i.e. making the whole object pop out from the background. We [45] had suggested earlier that this may be related to Ullman's coloring operation, which might require the spatially precise topological map in V1.

3.2. Texture figure saliency

Evidence of surface coloring can be traced back Lamme's [40] figure-ground effect. By examining the neurons' responses to different parts of the texture stimuli, he found that V1 neurons' responses were enhanced when their receptive fields were placed inside a texture-defined figure relative to when they were placed in a textured background, and that this enhancement was uniform within the spatial extent of the figure. When we [45] repeated his experiment (Fig. 17), we found that the magnitude and spatial extent of figure enhancement signal was smaller than what Lamme had reported earlier (see also [70,80]), and that the interior enhancement effect was superimposed by a very strong texture boundary signal particularly when the preferred orientation of the neurons was aligned with the orientation of the texture boundary (Fig. 17a and c). Nevertheless, there was indeed a general enhancement of the neural response within the figure against the background (Fig. 17d). Although surround inhibition was known for 20 years, the discontinuity in the enhance-

ment effect between figure and ground is new and important because it suggested a form of object-based attention—an effect that has been associated with an improvement in people's ability in discriminating features on the same surface of an object (even in partial occlusion) than across the surfaces of different objects [3,64]. Apparently, it is easier to compare features within the same layer or within a similarly colored surface.

While the illusory contour response of V2 neurons preceded that of V1 neurons by 35 ms, the shape from shading pop-out responses seemed to emerge in both areas at roughly the same time, at about 100 ms for the LA and LB stimuli. This time framework of 100–130 ms is close to that of target selection signals observed in the frontal eye field during a visual search task [5], suggesting that the computation and representation of perceptual saliency of objects in a visual scene might be distributed across multiple cortical areas and its computation might be inherently interactive and concurrent in nature. While saliency maps might be distributed across multiple cortical areas, the early visual area (LGN, V1 and V2) seems to provide the highest resolution saliency map for spatial localization and coloring of object surface.

4. Shapes, forms and symmetry

Once an object's boundary is delineated and its surface colored, it is possible then to compute the shape

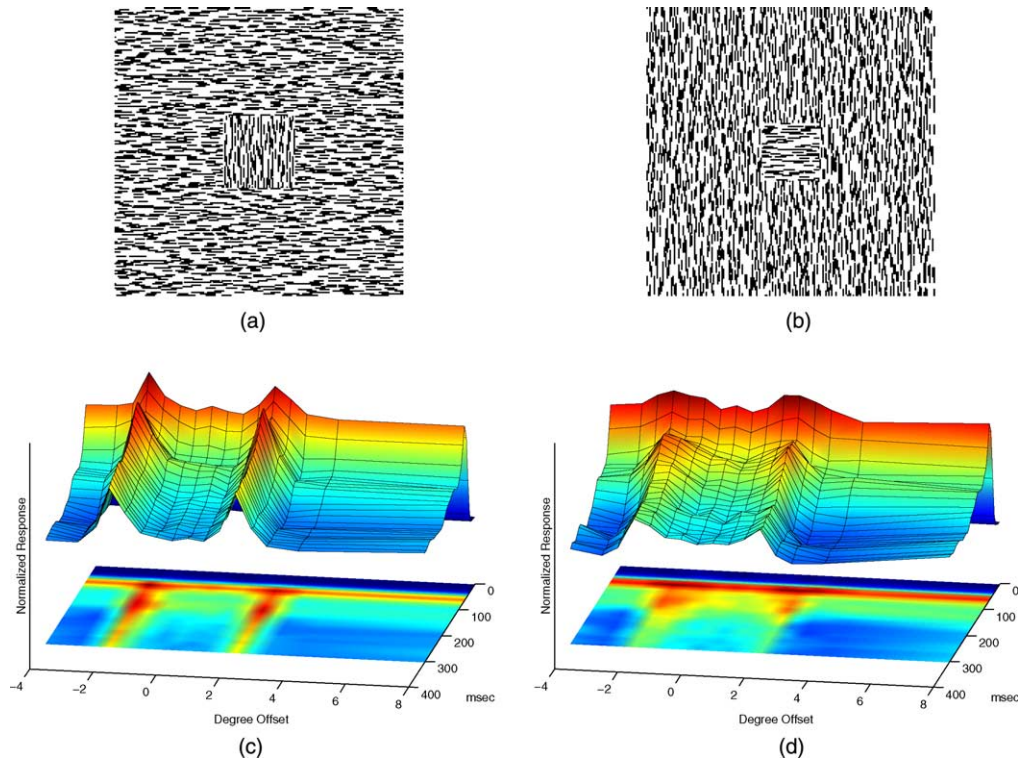


Fig. 17. Saliency coloring of the surface of an object: V1 neurons' responses were found to be enhanced inside the figure relative to outside the figure. (a) and (b) show figures defined by texture contrast. (c) shows the population averaged response of neurons in the parallel sampling scheme with the response to (a) and the response to (b) summed at each corresponding spatial location. (d) shows the population averaged response of neurons in the orthogonal sampling scheme. In the parallel sampling scheme, we found a large persistent texture edge response, which was absent in the orthogonal sampling scheme. This suggests that the cells were sensitive to the orientation of the texture boundaries. The response inside the figure showed an enhancement (about 15%) for both sampling schemes. The response inside the figure in the orthogonal sampling scheme appeared to be relatively flat, similar to Lamme's [40] original finding (see [45] for details).

and form of a segmented surface. Could the early visual cortex also play a role in the representation of local 3D surface shapes as well as the computation of the global shape (form) of an object's outline?

4.1. Surface shape primitives

An interesting observation in the shape from shading experiment [47] was that while V1 neurons responded much better to the WA and WB stimuli over the shape from shading stimuli in the initial (feedforward) phase (40–100 ms post-stimulus onset) of their response, V2 neurons responded equally well to both types of stimuli in their initial (feedforward) responses even though the shape from shading stimuli are much weaker in contrast. This evidence, together with the fact that the shape from shading pop-out was observed in V2 preattentively, but not in V1, led us to think that V2 might be the first cortical area that has the infrastructure for explicitly representing 3D surface shapes, allowing them to compute shape from shading pop-out. These ideas are consistent with the recent evidence provided by the studies of Von der Heydt and his colleagues that V2 neurons are sensitive to convex and concave stereo shapes [65] and

that V2 neurons are sensitive to border ownership (hence figure-ground) [79]. Taken together, it is reasonable to believe that V2 neurons might indeed furnish the infrastructure for the explicit representation of local 3D surface shape, with each V2 neuron's receptive field serving to encode a local patch of differential geometric surface, and that a combination of these elementary surface shapes can then represent any arbitrary surface shapes locally.

4.2. Axis of symmetry

A colored surface could also support another computation: the inference of the boundary shape or the form of an object. Given V1's circuitry is ideal for curve tracing, it is possible that it might also participate in the tracing of the axis of symmetry. Tracing the axis of symmetry of shape is a popular approach in computer vision. Biological bodies with flexible joints can change drastically with motion and changes in view point. Blum [4] observed that under such changes, a region-based description from the skeletons of the objects is much more stable than a boundary-based description. He proposed that complex biological forms could be de-

scribed efficiently using the skeletons and a small finite set of shape primitives. The skeleton transform, also called the medial axis transform, is formally defined as the locus of centers of the largest circles inside the region, hence touching its boundary in two distinct points (Fig. 18). It provides a grammatical or structural approach to object representation: an object is decomposed into primitive parts [2,4,11,15,53,81] and are linked together using a hierarchical framework as in the parse tree of a sentence.

Given that the computation of medial axis involves curve tracing which requires precise spatial precision and orientation resolution in a 2D topological map, one would expect that the early visual cortex should also play a role in this computation. There is some indirect evidence pointing toward this direction. In the texture strip experiment described in Fig. 7, we [44,45] had observed that some cells exhibited an enhancement peak at the center of the texture strip. For a subpopulation of neurons, it manifested as a weak, but sustained, central peak that persisted in the later period of the response (Fig. 7). These central peak could be stronger and more significant for individual neurons. Fig. 19a–c show the responses of a neuron to different parts of the texture strip at different time windows. It can be observed that as the boundary response became more localized at the later phase of the response, a response

peak sometimes also appeared at the center or the axis of the strip. The spatial response profiles of a neuron in successive temporal windows in Fig. 19a–c show the development of this central peak. In another dramatic example (Fig. 19d–f), a vertically oriented neuron at first did not respond at all within the horizontally textured strip, but became active at the axis of the strip after 80 ms.

However, we observed that the central response peak at the texture strip was sensitive to the width of the strip, i.e. it tended to disappear as the width of the strip increased. This suggests that the observed phenomenon itself could be a mere epiphenomenon emerged from the disinhibition and inhibition of the local circuitry in V1 [50,51]. However, it also hints at the possible existence of a medial axis computation, as revealed by Kovacs and Julesz's psychophysical experiment [39]. It is worthwhile to consider the several mechanisms for computing and representing medial axis transform and look for evidence for it in the early visual cortex. One mechanism is the grass-fire algorithm originally proposed by Blum [4], which mimics a grass fire that propagates in from the boundary of a figure. The points of collision of the fire-fronts form the skeleton of the figure, and the fire itself colors the figure (Fig. 18). This idea is modeled in an elegant formulation by Kimia et al. [36] which could be implemented by the local connectivity structure within

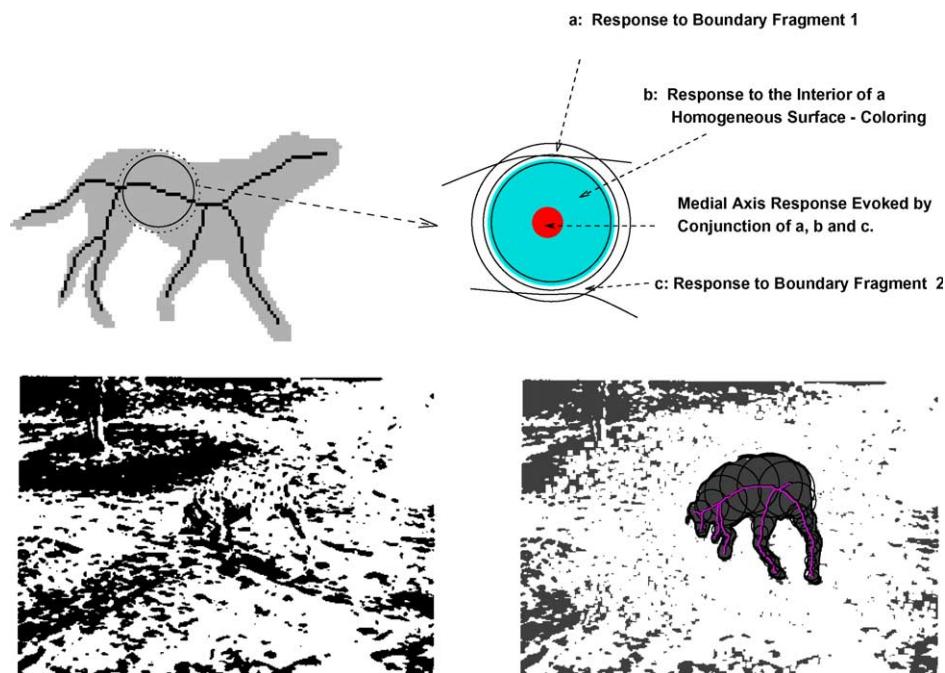


Fig. 18. Medial axis is a descriptor that integrates local and global information. It encodes information about the location of the skeleton and the diameter of the inscribed disk. This figure illustrates how a cell may be constructed so that it fires when located on the medial axis of an object. The conjunction of three features has to be present: at least two 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. A hypothetical view on the representation of the Dalmatian dog image in V1 is shown. Saliency processes enhance the figure at the expense of the background. Skeleton of the object can be computed from the pop-out figure in conjunction with the coloring process.

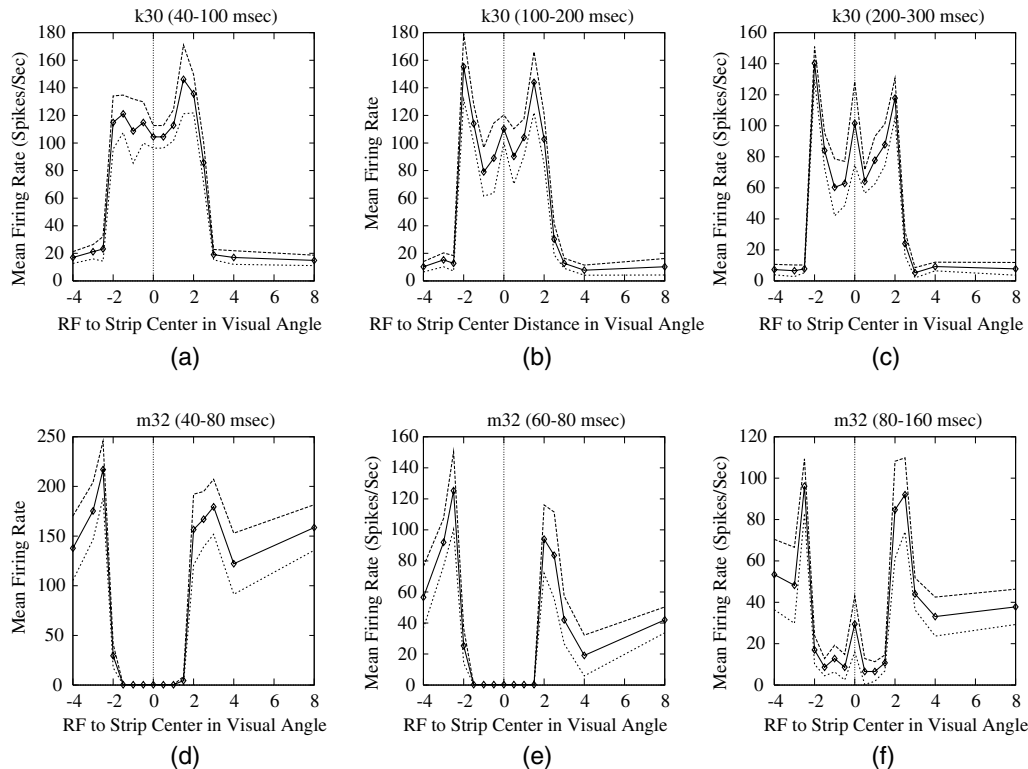


Fig. 19. (a–c) Spatial response profiles of a vertically oriented V1 neuron to different parts of the vertically textured strip. The abscissa is the distance in visual angle from the RF center to the center of the figure. The solid lines in these graphs indicate the mean firing rate within the time window, 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. (d–f) Another neuron's response to the horizontally textured strip—the complement of the stimulus shown in Fig. 7. Approximately 40–60 ms 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 ms, the boundary started to sharpen, but there was still no response within the strip. Interestingly, 80 ms onward, a pronounced response peak gradually developed at the axis of the texture strip.

V1. An alternative approach is to use a set of large center-surround filters [15] to compute the axis in a feedforward manner. An ensemble of such neurons, each 'tuning' to an inscribing circle of a particular diameter can potentially represent an invariant medial axis. Even though the medial axis is a very attractive computational idea, at this stage, the evidence is anecdotal and inconclusive. More carefully designed experiments need to be carried out to carefully evaluate this hypothesis. It is possible that the grassfire might be mediated by subthreshold activities, and might be more visible to optical imaging techniques.

5. Conclusions

In this paper, we present recent neurophysiological evidence that supports the idea that the early visual cortex should not be considered as merely a filter bank in the first stage of processing, but as a high-resolution buffer that supports a rich variety of visual computations. We have shown that the early visual cortex potentially takes part in the computation and repre-

sentation of (1) perceptual contours, (2) surface shapes, (3) object saliency and (4) possibly medial axis of forms. We have shown while the early part of V1 neuronal response is correlated to their orientation selectivity to local features, the later part of V1 neuronal response is correlated to higher order contextual processing. This provides us a window to monitor the evolution of a variety of visual computational algorithms.

Traditionally, the different stages of visual processing were mapped to the different cortical areas, with V1 responsible for early vision, and IT responsible for late vision. The evidence reported here suggests that the early visual areas are involved in many levels of visual inference. The different stages of processing might occur in multiple areas concurrently, which might be tightly coupled together through the recurrent feedforward/feedback connections (Fig. 20). Bullier et al. [9] have also proposed, based on the timing of the response onset of neurons in the different cortical areas, that information might not be serially processed through successive cortical areas, but simultaneously in multiple areas. They have also recently proposed an integrated model of visual processing in which V1 and V2 serve as an 'active

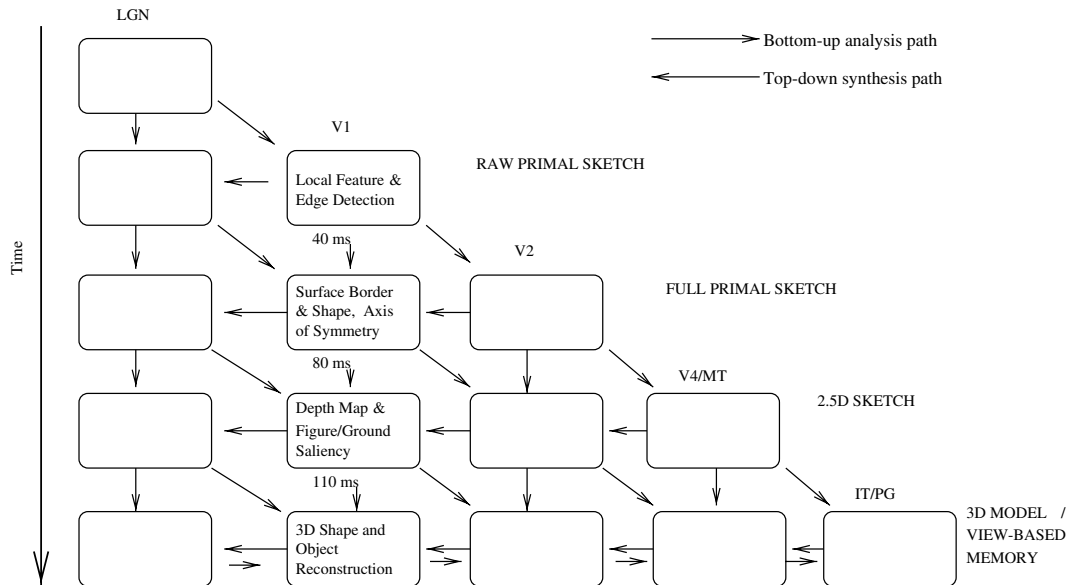


Fig. 20. Our ideas on how the different cortical areas become engaged in different levels of visual processing at different time post-stimulus onset. There is still a division of labor among the cortical areas with V1 responsible for detecting edges and completing contours; V2 responsible for the computation and representation of surface and surface shapes; V4 responsible for coding higher order conjunctions or invariant structures and IT for storing in a distributed fashion object prototypes. Each column depicts the involvement of a particular area in the various computations over time. Each row depicts the completion of computation across the multiple areas. As time progresses, the activities in V1 should reflect the computations at the higher cortical areas through the recurrent feedback connections. The rightward arrows indicate the bottom-up analysis path, and the leftward arrows indicate the top-down synthesis path. The top-down synthesis path might or might not necessarily carry a synthesized expectation of the image at the lower level. Rather it likely carries priors for influencing the Bayesian inference in the earlier areas.

black-board' for the rest of the visual cortical areas [10], in much the same spirit as the high-resolution buffer hypothesis. Numerous other labs [28,33,68,69,57,80] have also demonstrated spatial attention and object attention effects in the early visual cortex at the level of V1, particularly when the monkeys were engaged in tasks involving curve tracing [69] and line alignment [14,33]. These data, together with the findings reviewed in this paper [45–47], support this new perspective on the functional role of the early visual cortex in visual processing.

While we found attention does play a very important role in mediating both the illusory contour response [46] and the shape from shading pop-out response in V1 [47], we believe the feedback plays a more profound role than merely selecting the relevant local features from the early visual areas for downstream processing. Our evidence from the shape from shading experiment demonstrates that attention not only selects information to feed forward, but might also actively gate feature-specific information back to the early visual areas. The recurrent feedback between cortical areas might provide the top-down priors for geometric inference in the early visual areas in a hierarchical Bayesian inference framework. Elucidating the nature of computations in each cortical area, and the nature of the top-down priors is central to understanding the fundamental principles of visual inference.

Acknowledgements

This work is supported by NSF CAREER 9984706 and NIH Vision Research core grant EY08098 and Siemens, AG. The author thanks many of his colleagues, in particular David Mumford, Peter Schiller, Carl Olson, Carol Colby, James McClelland, Gustavo Deco, Song-chun Zhu, Victor Lamme, My Nguyen, Cindy Yang, Rick Romero, Stella Yu, Julie Rollenhagen, Karen Medler and Kae Nakamura for advice and assistance.

References

- [1] J. August, S.W. Zucker, The curve indicator random field: curve organization via edge correlation, in: K. Boyer, S. Sarka (Eds.), *Perceptual Organization for Artificial Vision Systems*, Kluwer Academic, Boston, 2000, pp. 265–288.
- [2] I. Biederman, Recognition-by-components: a theory of human image understanding, *Psychol. Rev.* 94 (2) (1987) 115–147.
- [3] M. Behrmann, R.S. Zemel, M.C. Mozer, Object-based attention and occlusion: evidence from normal participants and a computational model, *J. Exp. Psychol. Hum. Percept. Perform.* 24 (4) (1998) 1011–1036.
- [4] H. Blum, Biological shape and visual science, *J. Theoret. Biol.* 38 (1973) 205–287.
- [5] N.P. Bichot, J.D. Schall, Effects of similarity and history on neural mechanisms of visual selection, *Nature Neurosci.* 2 (6) (1999) 549–554.

- [6] A. Blake, A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
- [7] R.T. Born, R.B.H. Tootell, Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex, *Proc. Natl. Acad. Sci.* 88 (1991) 7071–7075.
- [8] J. Braun, Shape from shading is independent of visual attention and may be a 'texton', *Spatial Vision* 7 (4) (1993) 311–322.
- [9] J. Bullier, L.G. Nowak, Parallel versus serial processing: new vistas on the distributed organization of the visual system, *Curr. Opin. Neurobiol.* 5 (4) (1995) 497–503.
- [10] J. Bullier, Integrated model of visual processing, *Brain Res. Rev.* 36 (2–3) (2001) 96–107.
- [11] C.A. Burbeck, S.M. Pizer, Object representation by cores: identifying and representing primitive spatial regions, *Vision Res.* 35 (13) (1995) 1917–1930.
- [12] J.F. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* (1986) 679–698.
- [13] O.D. Creutzfeldt, H.C. Nothdurft, Representation of complex visual stimuli in the brain, *Naturwissenschaften* 65 (1978) 307–318.
- [14] R.E. Crist, W. Li, C.D. Gilbert, Learning to see: experience and attention in primary visual cortex, *Nature Neurosci.* 4 (5) (2001) 519–525.
- [15] J. Crowley, A.C. Parker, A representation for shape based on peaks and ridges in the difference of low-pass transform, *IEEE Trans. Pattern Recognit. Mach. Intell.* 6 (2) (1984) 156–170.
- [16] J. Daugman, Complete discrete 2D-Gabor transforms by neural networks for image analysis and compression, *IEEE Trans. Acoustic, Speech, Signal Process.* 36 (1988) 1169–1179.
- [17] R. De Valois, K. De Valois, *Spatial Vision*, Oxford University Press, New York, 1988.
- [18] P. Dayan, G.E. Hinton, R.M. Neal, R.S. Zemel, The Helmholtz machine, *Neural Comput.* 7 (5) (1995) 889–904.
- [19] D. Field, A. Hayes, R. Hess, Contour integration by the human visual system: evidence for a local "association field", *Vision Research* 33 (2) (1993) 173–193.
- [20] J.L. Gallant, C.E. Connor, D.C. Van Essen, Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing, *Neuroreport* 22 (9) (1998) 2153–2158.
- [21] R. Gattass, A.P. Sousa, C.G. Gross, Visuotopic organization and extent of V3 and V4 of the macaque, *J. Neurosci.* 8 (6) (1988) 1831–1845.
- [22] D. Geiger, H. Pao, N. Rubin, Organization of multiple illusory surfaces, in: *Proceedings of Computer Vision and Pattern Recognition*, Santa Barbara, 1998, pp. 118–124.
- [23] W.S. Geisler, J.S. Perry, J. Super, D.P. Gallogly, Edge co-occurrence in natural images predicts contour grouping performance, *Vision Res.* 41 (2001) 711–724.
- [24] C.D. Gilbert, T.N. Wiesel, The influence of contextual stimuli on the orientation selectivity of cells in the primary visual cortex of the cat, *Vision Res.* 30 (1990) 1689–1701.
- [25] C.G. Gross, Visual function of inferotemporal cortex, in: R. Jung (Ed.), *Handbook of Sensory Physiology*, 7.3B:L, Springer-Verlag, Berlin, 1973, pp. 451–482.
- [26] S. Grossberg, Competitive learning: from interactive activation to adaptive resonance, *Cognitive Sci.* 11 (1987) 23–63.
- [27] S. Grossberg, E. Mingolla, W. Ross, A neural theory of attentive visual search: interactions at boundary, surface, spatial and object attention, *Psychol. Rev.* 101 (3) (1994) 470–489.
- [28] P.E. Haenny, P.H. Schiller, State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks, *Exp. Brain Res.* 69 (2) (1988) 225–244.
- [29] H.V. Helmholtz, *Handbuch der physiologischen Optik*, Leipzig, Voss, 1867.
- [30] B. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1986.
- [31] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular integration and functional architecture in the cat's visual cortex, *J. Physiol.* 160 (1962) 106–154.
- [32] J.M. Hupe, A.C. James, P. Girard, J. Bullier, Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2, *J. Neurophysiol.* 85 (1) (2001) 146–163.
- [33] M. Ito, C.D. Gilbert, Attention modulates contextual influences in the primary visual cortex of alert monkeys, *Neuron* 22 (1999) 593–604.
- [34] P. Janssen, R. Vogels, G.A. Orban, Three-dimensional shape coding in inferotemporal cortex, *Neuron* 27 (2) (2000) 385–397.
- [35] M.K. Kapadia, G. Westheimer, C.D. Gilbert, Spatial distribution of contextual interactions in primary visual cortex and in visual perception, *J. Neurophysiol.* 84 (4) (2000) 2048–2062.
- [36] B.B. Kimia, A.R. Tannenbaum, S.W. Zucker, Shapes, shocks and deformation. I: the components of shape and the reaction-diffusion space, *Int. J. Comput. Vision* 15 (1995) 189–224.
- [37] J.J. Knierim, D.C. Van Essen, Neuronal responses to static texture patterns in area V1 of the alert macaque monkey, *J. Neurophysiol.* 67 (1992) 961–980.
- [38] C. Koch, J. Marroquin, A.L. Yuille, Analog "neural" networks in early vision, *Proc. Natl. Acad. Sci. USA* 83 (1986) 4263–4267.
- [39] I. Kovacs, B. Julesz, Perceptual sensitivity maps within globally defined visual shapes, *Nature* 370 (1994) 644.
- [40] V.A.F. Lamme, The neurophysiology of figure-ground segregation in primary visual cortex, *J. Neurosci.* 15 (2) (1995) 1605–1615.
- [41] V.A.F. Lamme, K. Zipser, H. Spekreijse, Figure-ground activity in primary visual cortex is suppressed by anesthesia, *Proc. Natl. Acad. Sci., USA* 95 (6) (1998) 3263–3268.
- [42] M.S. Langer, S.W. Zucker, Shape-from-shading on a cloudy day, *J. Opt. Soc. Am., A* 11 (2) (1994) 467–478.
- [43] T.S. Lee, Image representation using 2D Gabor wavelets, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (10) (1996) 959–971.
- [44] T.S. Lee, D. Mumford, S.C. Zhu, V.A.F. Lamme, The role of V1 in shape representation, in: J. Bower (Ed.), *Computational Neuroscience*, Plenum Press, New York, 1997, pp. 697–703.
- [45] T.S. Lee, D. Mumford, R. Romero, V.A.F. Lamme, The role of the primary visual cortex in higher level vision, *Vision Res.* 38 (15–16) (1998) 2429–2454.
- [46] T.S. Lee, M. Nguyen, Dynamics of subjective contour formation in the early visual cortex, *Proc. Natl. Acad. Sci.* 98 (4) (2001) 1907–1911.
- [47] T.S. Lee, C. Yang, R.D. Romero, D. Mumford, Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency, *Nature Neurosci.* 5 (6) (2002) 589–597.
- [48] J.B. Levitt, J.S. Lund, Contrast dependence of contextual effects in primate visual cortex, *Nature* 387 (1997) 73–76.
- [49] M.S. Lewicki, B.A. Olshausen, Probabilistic framework for the adaptation and comparison of image codes, *J. Opt. Soc. Am. A* 16 (7) (1999) 1587–1601.
- [50] C.Y. Li, W. Li, Extensive integration field beyond the classical receptive field of cat's striate cortical neuron—classification and tuning properties, *Vision Res.* 34 (19) (1994) 2577–2598.
- [51] Z. Li, Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex, *Neural Comput.* 13/8 (2001) 1749–1780.
- [52] L. Maffei, A. Fiorentini, The unresponsive regions of visual cortical receptive fields, *Vision Res.* 16 (1976) 1131–1139.
- [53] D. Marr, *Vision*, W.H. Freeman & Company, NJ, 1982.
- [54] J.H. Maunsell, W.T. Newsome, Visual processing in monkey extrastriate cortex, *Ann. Rev. Neurosci.* 10 (1987) 363–401.
- [55] J.L. McClelland, D.E. Rumelhart, An interactive activation model of context effects in letter perception. Part I: an account of basic findings, *Psychol. Rev.* 88 (1981) 375–407.

- [56] A. Morel, J. Bullier, Anatomical segregation of two cortical visual pathways in the macaque monkey, *Visual Neurosci.* 4 (1990) 555–578.
- [57] B. Motter, Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli, *J. Neurophysiol.* 70 (1993) 909–919.
- [58] D. Mumford, On the computational architecture of the neocortex II, *Biol. Cybernet.* 66 (1992) 241–251.
- [59] D. Mumford, *Elastica and Computer Vision*, in *Algebraic Geometry and its Application*, Springer-Verlag, 1994, pp. 491–506.
- [60] D. Mumford, Patter theory: a unifying perspective, in: D.C. Knill, W. Richards (Eds.), *Perception as Bayesian Inference*, Cambridge University Press, 1996, pp. 25–62.
- [61] M. Nitzberg, D. Mumford, T. Shiota, *Filtering, Segmentation and Depth*, Springer-Verlag, Berlin, 1993.
- [62] H.C. Nothdurft, J.L. Gallant, D.C. Van Essen, Response modulation by texture surround in primate area V1: correlates of “pop-out” under anesthesia, *Visual Neurosci.* 16 (1) (1999) 15–34.
- [63] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature* 381 (1996) 607–609.
- [64] C.R. Olson, Object-based vision and attention in primates, *Curr. Opin. Neurobiol.* 11 (2) (2001) 171–179.
- [65] F.T. Qiu, K. Endo, R. von der Heydt, Neural representation of 3-dimensional shape primitives in monkey visual cortex, *J. Vision* 1 (2001) 273a.
- [66] V.S. Ramachandran, Perception of shape from shading, *Nature* 331 (1988) 163–166.
- [67] R. Rao, D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nature Neurosci.* 2 (1) (1999) 79–87.
- [68] J. Reynolds, L. Chelazzi, R. Desimone, Competitive mechanisms subserve attention in macaque areas V2 and V4, *J. Neurosci.* 19 (1999) 1736–1753.
- [69] P.R. Roelfsema, V.A. Lamme, H. Spekreijse, Object-based attention in the primary visual cortex of the macaque monkey, *Nature* 395 (6700) (1998) 376–381.
- [70] A.F. Rossi, R. Desimone, L.G. Ungerleider, Contextual modulation in primary visual cortex of macaques, *J. Neurosci.* 21 (5) (2001) 1698–1709.
- [71] E.P. Simoncelli, O. Schwartz, Image statistics and cortical normalization models, in: M.S. Kearns, S.A. Solla, D.A. Cohn (Eds.), *Adv. Neural Inform. Process. Syst.*, MIT Press, Cambridge, MA, 11 (1999) 153–159.
- [72] D.C. Somers, E.V. Todorov, A.G. Siapas, L.J. Toth, D.S. Kim, M. Sur, A local circuit integration approach to understanding visual cortical receptive fields, *Cerebral Cortex* 8 (3) (1997) 204–217.
- [73] M. Stemmler, M. Usher, E. Niebur, Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics, *Science* 269 (1995) 1877–1880.
- [74] J. Sun, P. Perona, Early computation of shape and reflectance in the visual system, *Nature* 379 (1996) 165–168.
- [75] H. Super, H. Spekreijse, V.A.F. Lamme, Two distinct modes of sensory processing observed in monkey primary visual cortex (V1), *Nature Neurosci.* 4 (3) (2001) 304–310.
- [76] D.A. Robinson, A method of measuring eye movement using a scleral search coil in a magnetic field, *IEEE Trans. Biomed. Electron.* 101 (1963) 131.
- [77] R. von de Heydt, E. Peterhans, G. Baumgarthner, Illusory contours and cortical neuron responses, *Science* 224 (4654) (1984) 1260–1262.
- [78] L. Williams, D. Jacobs, Stochastic completion fields: a neural model of illusory contour shape and saliency, *Neural Comput.* 9 (4) (1997) 837–858.
- [79] H. Zhou, H.S. Friedman, R. von der Heydt, Coding of border ownership in monkey visual cortex, *J. Neurosci.* 20 (17) (2000) 6594–6611.
- [80] K. Zipser, V.A.F. Lamme, P.H. Schiller, Contextual modulation in primary visual cortex, *J. Neurosci.* 16 (1996) 7376–7389.
- [81] S.C. Zhu, A.L. Yuille, FORMS: a flexible object recognition and modelling system, *Int. J. Comp. Vis.* (1995) 187–212.
- [82] S. Ullman, Visual routines, *Cognition* 18 (1984) 97–159.

Further Reading

- [1] M. Ahissar, R. Laiwand, S. Hochstein, Attentional demands following perceptual skill training, *Psychol. Sci.* 12 (1) (2001) 56–62.
- [2] E. Bienenstock, S. Geman, D. Potter, Compositionality, MDL priors, and object recognition, in: M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems* 9, vol. 9, MIT Press, 1997, pp. 838–844.
- [3] C. Connor, D. Preddie, J. Gallant, D. Van Essen, Spatial attention effects in macaque area V4, *J. Neurosci.* 17 (1997) 3201–3214.
- [4] G. Deco, T.S. Lee, A unified model of spatial and object attention based on inter-cortical biased competition, *Neurocomputing* 44–46 (2002) 769–774.
- [5] R. Desimone, J. Duncan, Neural mechanisms of selective visual attention, *Ann. Rev. Neurosci.* 18 (1995) 193–222.
- [6] D.J. Felleman, D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, *Cerebral Cortex* 1 (1991) 1–47.
- [7] U. Grenander, *Lectures in Pattern Theory I, II and III: pattern analysis, pattern synthesis and regular structures*, Springer-Verlag, Berlin, 1976–1981.
- [8] D.H. Groszof, R.M. Shapley, M.J. Hawken, Macaque V1 neurons can signal ‘illusory’ contours, *Nature* 365 (6446) (1993) 550–552.
- [9] J.M. Hupe, A.C. James, B.R. Payne, S.G. Lomber, P. Girard, J. Bullier, Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons, *Nature* 394 (6695) (1998) 784–787.
- [10] L. Itti, C. Koch, Computational modeling of visual attention, *Nature Rev. Neurosci.* 2 (3) (2001) 194–203.
- [11] W. James, *The Principles of Psychology*, Henry Holt, New York, 1890.
- [12] J.S. Joseph, M.M. Chun, K. Nakayama, Attentional requirements in a ‘preattentive’ feature search task, *Nature* 387 (1997) 805–807.
- [13] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* 42 (3) (1993) 300–311.
- [14] T.S. Lee, D. Mumford, The role of V1 in scene segmentation and shape representation, *Soc. Neuro. Abstract* (1996).
- [15] N.K. Logothetis, Object vision and visual awareness, *Curr. Opin. Neurobiol.* 8 (4) (1998) 536–544.
- [16] J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex, *Science* 229 (1985) 782–784.
- [17] H.C. Nothdurft, J.L. Gallant, D.C. Van Essen, Response profiles to texture border patterns in area V1, *Visual Neurosci.* 17 (3) (2000) 421–436.
- [18] D.A. Pollen, J.P. Gaska, L.D. Jacobson, Physiological constraints on models of visual cortical function, in: M. Rodney, J. Cotterill (Eds.), *Models of Brain Function*, Cambridge University Press, England, 1989, pp. 115–135.
- [19] B.R. Sheth, J. Sharma, S.C. Rao, M. Sur, Orientation maps of subjective contours in visual cortex, *Science* 274 (5295) (1996) 2110–2115.
- [20] A.S. Tolias, T. Moore, S.M. Smirnakis, E.J. Tehovnik, A.G. Siapas, P.H. Schiller, Eye movement modulate visual receptive fields of V4 neurons, *Neuron* 29 (3) (2001) 757–767.

- [21] A. Treisman, G. Gelade, A feature-integration theory of attention, *Cognitive Psychol.* 12 (1980) 97–136.
- [22] L.G. Ungerleider, M. Mishkin, Two cortical visual systems, in: D.J. Ingle (Ed.), *Analysis of Visual Behavior*, MIT Press, Cambridge, MA, 1982, pp. 549–586.
- [23] S. Ullman, Sequence seeking and counterstreams: a model for bidirectional information flow in the cortex, in: C. Koch, J. Davis (Eds.), *Large-Scale Theories of the Cortex*, MIT Press, Cambridge, MA, 1994, pp. 257–270.