ELSEVIER

Physiology & Behavior 77 (2002) 645–650

PHYSIOLOGY & BEHAVIOR

Top-down influence in early visual processing: a Bayesian perspective

Tai Sing Lee*

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

Received 15 July 2002; accepted 16 August 2002

Abstract

Traditional views of visual processing suggest that early visual neurons are static spatiotemporal filters that extract local features by feedforward computation. The extracted information is then fed forward through a chain of modules to successively higher visual areas for further analysis. Recording from early visual neurons in awake behaving monkeys, we revealed there are many levels of complexity in the information processing of the early visual cortex. We found that the early visual neurons not only are sensitive to features within their receptive fields (RFs) but also to the global context of a visual scene, the behavioral relevance of the stimuli and the experience of the animals. These findings suggest that the early visual cortex (V1 and V2) is tightly coupled to and highly interactive with the rest of the visual system. The top-down interaction, mediated by recurrent feedback connections, introduces contextual information to influence the perceptual inference in the early visual cortex.

Keywords: Primate electrophysiology; Visual processing; Feedback; Saliency; Contours

1. Early visual cortex

Neurons in the primary visual cortex are known to be tuned to specific elementary local features in the visual scenes. These features include location, line orientation, stereo disparity, movement direction, color and spatial frequency [1,2]. It is also known that V1 neurons are also influenced by the surrounding context of the stimuli [3–6]. The interpretations of the contextual modulations in these studies have been mostly limited to low-level mechanistic description in terms of facilitation and inhibition, or to subjective perceptual interpretations such as the neural correlate of contour processing, pop-out or figure-ground saliency [4,5,7,8]. Some of the observed contextual modulations likely arise from the feedback mediated by the massive amount of recurrent connections from the extrastriate areas to V1. A plausible role of feedback is that of attentional selection based on the mechanisms of biased competition [9]. The idea of biased competition is that when multiple stimuli are presented in a visual field, the different neuronal populations activated by these stimuli will engage in competitive interaction. Attending to a stimulus at a particular spatial location or to a particular object feature, however, could bias the competition in favor of the neurons representing the attended features or locations, enhancing their responses and suppressing the responses of the other neurons. However, the intracortical interaction in biased competition models (e.g., Ref. [10]) is limited to lateral inhibition—a rather impoverished view on the computations being performed by the sophisticated machinery in the different visual areas.

2. Hierarchical Bayesian inference

Based on recent findings on top-down effect in V1 [5,7], Mumford and I have suggested that V1 can serve as a high-resolution buffer [7] that participates in many levels of visual computations through the recurrent feedback (see also Bullier's blackboard hypothesis [11]). In this context, we think that Bayesian inference can provide a more appropriate theoretical framework for reasoning about top-down visual processing in the brain [12–15]. This idea is not new, and can be traced back to the *unconscious inference* theory of perception by Helmholtz [16]. From the Bayesian perspect-

* Mellon Institute, Carnegie Mellon University, Room 115, 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).

0031-9384/02/$ – see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S0031-9384(02)00903-4

ive, the visual system arrives at the most probable interpretation of the visual scene by finding the *a posteriori* estimate S_i of the scene that maximizes $P(S_i|E,H)$, the conditional probability of a scene hypothesis S_i given a particular sensory evidence (E), and the information already known (H), which, by Bayes' theorem, is given by,

$$P(S_i \mid E,H) = \frac{P(E \mid S_i,H)P(S_i \mid H)}{P(E \mid H)}$$

where $P(E|S_i,H)$ is the conditional probability of the evidence given a particular scene hypothesis S_i and the prior or contextual knowledge about the scene H, based on either the viewer's prior assumption, information from the surrounding space or immediate past. $P(S_i|H)$ is the conditional probability of the scene hypothesis given H. $P(E|S_i,H)$ is the conditional probability of the evidence E given H and S_i, which can be factored into $P(E|S_i)P(H)$ if the influence of H is exerted directly only on S_i but not on E.

This basic formulation might allow us to conceptualize the interaction between two cortical areas, for example, V1 and V2, mathematically. Let E be the evidence furnished to V1 by the retina and the lateral geniculate nucleus (LGN); scene estimates S_i are the output of V1 neurons. H is a distribution of hypotheses generated by V2 based on its input from V1 as well as feedback from other higher order areas. The feedback from V2 to V1 is given by the distribution of H weighted by its prior $P(H)$, with $P(S_i|H)$ specifying the feedback connections. These are what we called *contextual priors*—the prior knowledge the system already has about the scene based on what has been observed in the past and in the surrounding. V1 tries to find the S_i that maximises $P(E|S_i)P(S_i|H)P(H)$, i.e., explaining E as well as being predicted by H optimally. This scheme can be applied again to higher areas recursively to form the whole hierarchy of inference. In this framework, each cortical area is an expert for inferring certain aspects of the visual scene, but its inference is made in consultation with the other brain areas, constrained by both incoming data and the top-down contextual priors. Unless the input image is simple and clear, each area normally cannot be sure of its inference and has to entertain a number of hypotheses simultaneously. The feed-forward input drives the generation of the hypotheses, the feedback from higher inference areas provides the contextual information (or priors) to shape the inference at the earlier levels. Hierarchical Bayesian inference is concurrent across multiple areas. Information does not flow back and forth between V1 and inferotemporal cortex (IT) in a big loop. Such a large loop requires a lot of time per iteration and is infeasible for real time inference. Rather, successive cortical areas in the visual hierarchy can constrain each other's inference in small loops instantaneously and continuously. Such a system, as a whole, might converge rapidly to an interpretation of the visual scene.

We carried out a series of neurophysiological experiments on awake behaving monkeys to test these possibilities. The first experiment examined the neural representation of the famous Kanizsa illusion (as shown in Fig. 1b) that illustrates how a strong prior assumption about occlusion relationship between surfaces at different depths could make us see things that actually do not exist. The second experiment examined a saliency effect that emerged from the interaction of the

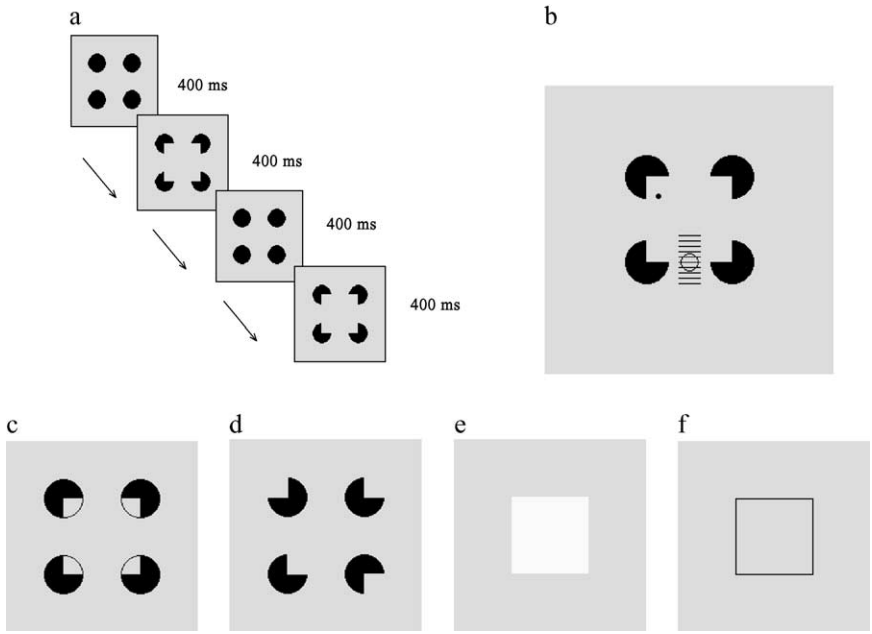

Fig. 1. Selected stimuli in the subjective contour experiment. (a) An example sequence of stimulus presentation in a single trial. (b) Receptive field of the tested neuron was 'placed' at 10 different positions across the illusory contour, one per trial. (c) Amodal contour—the subjective contour was interrupted by intersecting lines. (d) One of the several rotated corner disc stimuli. The surround stimulus was roughly the same, but there was no illusory contour. (e) One of the several types of real squares defined by luminance contrast. (f) Square defined by lines.

brain's inference of three-dimensional (3D) surface shapes based on shading information (Fig. 3a,b). Both are wonderful examples that illustrate the influence of higher order contextual information on early visual inference.

3. Evidence I: construction of subjective contours

When viewing the display of stimulus sequence shown in Fig. 1a, we perceive an illusory square that abruptly appears in front of four circular discs with vivid subjective borders even in regions of the image where there is no direct visual evidence for the borders. Could we see evidence of this illusory contour in the early visual cortex? Is there any evidence that this illusion is generated by feedback from higher areas?

We [17] recorded the responses of over 200 V1 and V2 neurons of awake behaving monkeys to an illusory contour in contrast to their responses to real contours or to conditions where there are no contours. Among the stimuli tested in each recording session (Fig. 1), the most important stimulus is the illusory square (Fig. 1b). Other stimuli were also tested as control. These include the amodal figure (Fig. 1c), the stimuli with rotated corner discs (Fig. 1d) and a variety of real squares defined by contrast and lines (Fig. 1e and f). Each stimulus was presented for 120 trials (10 conditions and 12 trials per condition). The monkey's task was to fixate a spot on the screen during stimulus presentation. In each trial, a sequence of four stimuli, 400 ms each, was presented. Fig. 1a illustrates the presentation of the subjective square stimuli. First, four circular discs were presented. Then, they were turned into corner discs, creating an illusion that a white square had abruptly appeared in front of the discs, partially occluding them. Over successive trials, the receptive field (RF) of the cell being recorded was placed at 10 different locations relative to the center of each stimulus, 0.25° apart, spanning a range of 2.25° (across the illusory contour in the case of illusory square) as shown in Fig. 1b. The RFs of the neurons, elucidated using a small oriented bar, were typically less than 0.8° at that eccentricity (about 2–3° away from the fovea). The gap between the corner discs was 2° wide. The neurons were considered to be sensitive to illusory contour if their responses to the illusory contour, at the precise location of that contour, were significantly larger than their responses to the amodal contour (Fig. 1c) and the conditions in which the corner discs were rotated (e.g., Fig. 1d).

We found that 26% of the V1 neurons in the superficial layer of V1 exhibited sensitivity to the illusory contour under our experimental paradigm. The neural correlate of the illusory contour signal, defined as the extra response above the response to the amodal contour, emerged in a V1 neuron at precisely the same location where a line or luminance contrast elicited the maximum response from the cell (Fig. 2a). The response to the illusory contour was

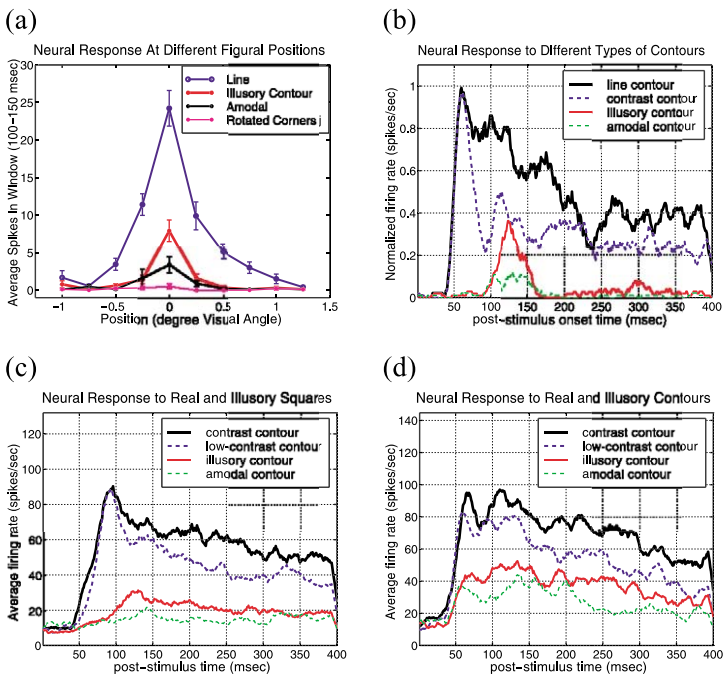

Fig. 2. (a) The spatial profile of a V1 neuron's response to the contours of both real and illusory squares, in a temporal window 100–150 ms after stimulus onset. The real or illusory square was placed at different spatial locations relative to the receptive field of the cell. This cell responded to 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, $P < .003$) and did not respond much when the corner discs were rotated. (b) Temporal evolution of the cell's response to the illusory contour compared to its response to the real contours of a line square, a white square and to the amodal contour. The onset of the response to the real contours was at 45 ms, about 55 ms ahead the illusory contour response. (c) Population averaged temporal response of 49 V1 neurons in the superficial layer to the real and illusory contours. (d) Population averaged temporal response of 39 V2 neurons in the superficial layer to the real and illusory contours.

delayed relative to the response to the real contours by 55 ms (Fig. 2b), emerging about 100 ms after stimulus onset. The response to the illusory contour was significantly greater than the response to the controls, including the amodal contour or when the corner discs were rotated. At the population level, we found that sensitivity to illusory contours emerged at 65 ms in V2 and 100 ms in the superficial layer of V1 (Fig. 2c and d). Our interpretation is that V2 detects the existence of an illusory contour by integrating information from a more global spatial context, and then generates top-down feedback signal to constrain the more precise contour inference in V1. Since the feedback connection is known to be rather diffuse spatially, feedback likely provides only a general guidance that is specific in feature domain, but nonspecific in spatial domain, helping the V1 circuitry construct and complete a precise representation of the subjective contour.

4. Evidence II: modulation by shape from shading

In the second experiment [8], we asked two questions. First, does 3D shape from shading information, presumably computed in the higher visual areas, influence the processing in V1? Second, when we bias the monkey to look for a certain object, does this top-down bias have an impact on V1 inference?

We first trained monkeys to perform an odd-ball detection task. In this task, the monkey was presented with a stimulus in which one element (odd-ball) is different from the others, as in Fig. 3a, and the monkey was required to make a saccadic eye movement to the oddball to get juice reward. Then we recorded from 550 V1 and V2 neurons, while the monkeys were performing the fixation task. In the fixation task, the monkey was required to fixate at the black dot on the screen during stimulus presentation. No saccadic eye movement to the odd-ball was required. Shape from shading oddballs (Fig. 3a) is known to pop out, or readily segregate into different groups, while the 2D contrast patterns (WA, Fig. 3e; WB, Fig. 3h) do not. The main difference between the two types of patterns is that the shading stimuli afford a 3D shape interpretation. Our task is to examine whether V1 neurons are modulated by shape from shading information.

We studied the response of V1 and V2 neurons to a variety of stimuli. For each stimulus type, we compared the response of the neuron to four different conditions. The most important comparison is the response to the odd-ball condition and the uniform condition. In these two conditions, the RF of the tested neuron was covered by an identical stimulus element. An increase in neural responses to the odd-ball condition relative (e.g., Fig. 3a) to the uniform condition (e.g., Fig. 3b) can be considered a neural correlate of pop-out or perceptual saliency.

For each stimulus type, four conditions (singleton, odd-ball, uniform and hole) were tested. In the singleton stimulus,

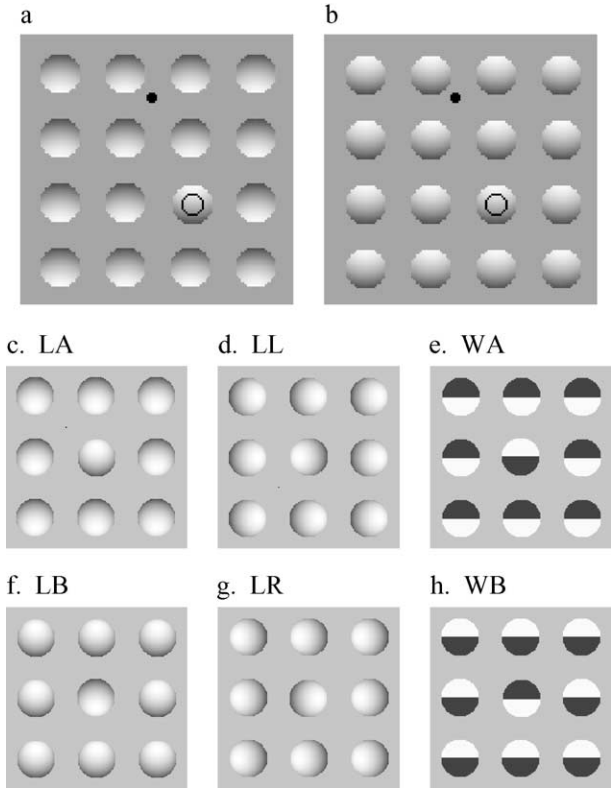

Fig. 3. Higher order perceptual pop-out. We compared two conditions from each stimulus set: an odd-ball condition where the receptive field element is an odd-ball and a uniform condition where the RF element is one of the elements of the background. Lighting from above (LA) oddball (a) and LA uniform (b) conditions are shown for illustration. The black dot was the fixation dot for the monkey to stare at during stimulus presentation. Six sets of stimuli were tested (c–h), i.e., lighting from above (LA), below (LB), left (LL) right (LR) and white above (WA) and white below (WB). In the actual experiment, a singleton stimulus (only the RF element) and a hole stimulus (background only, without the RF element) were also tested for each stimulus set for comparisons (see Ref. [8]).

there was only one stimulus element, covering the RF. It was used to measure the neuronal response to direct stimulation of the RF alone, without any surround stimulus. The hole stimulus was the same as the uniform condition except the stimulus element on the RF was absent. It was used to measure 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 dot (shown as black dot in Fig. 3a and b).

We found that, after the monkeys were trained to perform the odd-ball detection task, V1 and V2 neurons responded better to the odd-ball condition than the uniform condition for the lighting from above (LA) or the lighting from below (LB) stimuli (Fig. 4a), but the difference in their responses to the two conditions of the white-above (WA) type or the white-below (WB) type was weaker or absent (Fig. 4b). These response differences, or the pop-out signals, were found to be inversely correlated with the reaction time of the monkeys in detecting the odd-ball of the various types of stimuli (Fig. 4c,d), and hence could be considered a neural correlate of

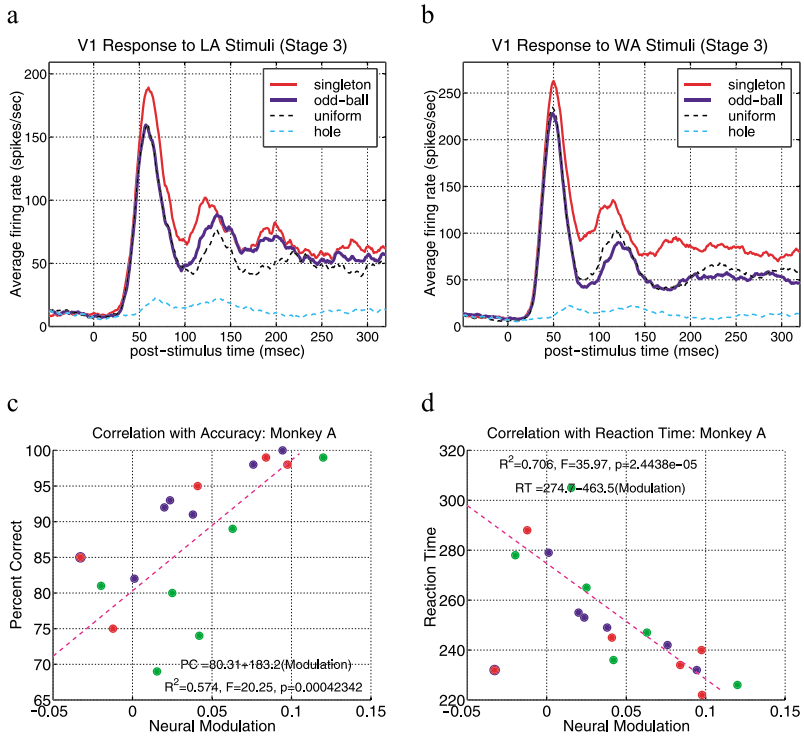

Fig. 4. Temporal evolution of the normalized population average response of 30 V1 units from a monkey to the LA set (a) and the WA set (b) in a stage after the monkey had utilized the stimuli in its behavior. Each unit's response was first smoothed by a running average within a 15-ms window and then averaged across the population. A significant difference (pop-out response) was observed between the population average response to the odd-ball condition and that to the uniform condition in the LA set. No pop-out response was observed in the WA set. (c, d) 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, each to the six types of stimuli, were regressed against the pop-out response. The significant correlation between the neural pop-out responses and the behavioral performance, suggesting the neural response is a correlate of subjective perceptual saliency of an object (see Ref. [8] for details).

perceptual saliency of the odd-ball stimulus. Interestingly, before the odd-ball detection training, V2 but not V1 neurons exhibit sensitivity to shape from shading pop-out. This suggests that V2 may be the first cortical area where 3D shape inference is made about object surface. The shape information was fed back to V1 after the monkeys used the stimuli in their behavior. Another interesting observation is that when we changed the relative presentation frequency of the stimuli to bias the monkeys' preference to a specific stimulus, for example, the LB odd-ball, the neural pop-out response became much stronger for LB at the expense of LA. When we reversed the relative frequency, the pop-out response reversed correspondingly. Our interpretation is that when the monkey developed a preference of looking for a certain stimulus as a result of the training, the extrastriate ventral stream might provide a top-down feature (object) expectation (or in Bayesian term, feature prior) to facilitate the processing or detection of that particular stimulus in V1.

5. A new perspective

The findings of these two experiments are consistent with the concept that visual processing in the brain can be conceptualized in terms of hierarchical Bayesian inference. Feedback from a higher order area to an earlier area can be conceptualized as providing top-down priors to bias the early inference. The impact of feedback is often subtle and becomes evident only when there is ambiguity in the visual stimuli, which is true in both of our experiments.

From this perspective, attention should not be conceptualized in terms of biased competition, but maybe more appropriately in terms of *biased inference*, or providing top-down priors in a hierarchical Bayesian inference framework. This conceptualization casts attention in a more mathematically tractable light. Feedback from the posterior parietal cortex could provide a spatial prior, i.e., prior expectation of how informative or interesting a particular visual location is. The influence of this spatial prior is called spatial attention. On the other hand, feedback from the ventral stream areas would provide a top-down object or feature prior, telling the early visual area what object the system is looking for, or what features we are expected to see when we are inferring the existence of a particular object. This manifests as object attention or feature attention. The forms of attention depend on the task at hand, as different classes of priors would be required for different occasions. Theoretically speaking, priors can be derived from the statistics of stimuli and the processing constraints imposed by the computational tasks.

Priors thus provide a potential bridge between behaviors, environments and perceptual inference. Understanding them should therefore be a central question in the study of biological vision.

Acknowledgements

This work was supported by NSF CAREER 9984706 and NIH EY08098 and NIH 2P41RR06009-11. The author thanks his colleagues, in particular David Mumford, Peter Schiller, Carl Olson, Carol Colby, Rob Kass, Rick Romero, Stella Yu, Cindy Yang, My Nguyen, David Moorman and Kae Nakamura for advice and assistance; and David Moorman for proofreading this manuscript.

References

[1] Hubel DH, Wiesel TN. Functional architecture of macaque monkey visual cortex. Proc R Soc B (Lond) 1978;198:1–59.
[2] De Valois RL, De Valois KK. Spatial vision. New York: Oxford Univ Press; 1988.
[3] Maffei L, Fiorentini A. The unresponsive regions of visual cortical receptive fields. Vis Res 1976;16:1131–9.
[4] Knierim JJ, Van Essen DC. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 1992; 67:961–80.
[5] Lamme VAF. The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci 1995;15(2):1605–15.
[6] Kapadia MK, Westheimer G, Gilbert CD. Spatial distribution of contextual interactions in primary visual cortex and in visual perception. J Neurophysiol 2000;84(4):2048–62.
[7] Lee TS, Mumford D, Romero R, Lamme VAF. The role of the primary visual cortex in higher level vision. Vis Res 1998;38(15–16): 2429–54.
[8] Lee TS, Yang C, Romero R, Mumford D. Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency. Nat Neurosci 2002;5(6):589–97.
[9] Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annu Rev Neurosci 1995;18:193–222.
[10] Deco G, Lee TS. A unified model of spatial and object attention based on inter-cortical biased competition. Neurocomputing 2002;44–46: 769–74.
[11] Bullier J. Integrated model of visual processing. Brain Res Rev 2001;36(2–3):96–107.
[12] Mumford D. On the computational architecture of the neocortex II. Biol Cybern 1992;66:241–51.
[13] Mumford D. Pattery theory: a unifying perspective. In: Knill DC, Richards W, editors. Perception as Bayesian inference. Cambridge: Cambridge Univ Press; 1996. p. 25–62.
[14] Lee TS. A Bayesian framework for understanding texture segmentation in the primary visual cortex. Vis Res 1995;35(18):2643–57.
[15] McClelland JL. Connectionist models and Bayesian inference. In: Oaksford M, Chater N, editors. Rational models of cognition. Oxford: Oxford Univ Press; 1998. p. 21–53.
[16] Helmholtz HV. Handbuch der physiologischen Optik. Leipzig: Voss; 1867.
[17] Lee TS, Nguyen M. Dynamics of subjective contour formation in the early visual cortex. Proc Natl Acad Sci 2001;98(4):1907–11.