Front. Electr. Electron. Eng. 2012, 7(1): 85–93
DOI 10.1007/s11460-012-0184-0

REVIEW ARTICLE

Fang FANG, Yizhou WANG

Image understanding, attention and human early visual cortex

Abstract This paper reviews our recent fMRI and psychophysical finding on: 1) perceived size representation in V1; 2) border ownership representation in V2; and 3) neural processing of partially occluded face. These findings demonstrate that the human early visual cortex not only performs local feature analyses, but also contributes significantly to high-level visual computation with assistance of attention-enabled cortical feedback. Moreover, by taking advantage of recent findings on early visual cortex from neuroscience and cognitive science, we build a biologically plausible attention model that can well predict human scanpaths on natural images.

Keywords vision, attention, image understanding, early visual cortex

1 Introduction

One of the greatest mysteries in vision science is the human ability to understand visual scenes with high accuracy and lightening speed. Its importance is heightened by the fact that efforts to duplicate this ability in machines have not met with extraordinary success. It is well recognized that a natural way to make progress in machine vision is to study visual information representation, processing and transformation in the human visual system. The core part of the human visual system is the visual cortex, located in the posterior part of the brain [1] (Fig. 1). The visual cortex consists of tens of visual areas, each of which is relatively specialized for processing certain visual patterns. For example, the primary visual cortex (V1) is selective for orientation; the middle temporal area (hMT) is sensitive to motion and binocular disparity; the lateral occipital area (LO) responds to objects.

How these visual areas are organized to implement various visual functions is of central interest to vision scientists. A dominant theory is David Marr's model. Its central tenet is the decomposition of visual processing into successive feed-forward steps, into low, intermediate and high-level stages. Correspondingly, the visual cortex could be divided into areas implementing low, intermediate and high-level computations. Early visual cortex serves to analyze local features such as edge, spatial frequency and contrast. Object recognition and image understanding are thought to be functions of higher visual areas. In this paper, we review our recent findings and challenge this dominant view. In Sect. 2, we show that V1 could combine the retinal size and the depth of a stimulus to compute its real size. In Sect. 3, we demonstrate that V2 could represent border ownership and is critical for image understanding. In Sect. 4, we reveal that processing partially occluded face consists of two synergetic phases: early suppression in early visual areas and late enhancement in higher visual areas. The former process might contribute to figure-ground segmentation. In these three studies, functional magnetic resonance imaging (fMRI) was used to measure human brain activity and psychophysics was used to measure human behaviors. These studies provide converging evidence supporting that even early visual cortex could be involved in higher-order visual computations related to image understanding. Furthermore, we show that attention, which enables cortical feedback from higher areas to early visual cortex, is necessary for these computations. In Sect. 5, based on recent findings on early visual cortex from neuroscience and cognitive science, we build a dynamic attention model that can predict human scanpaths on natural images.

Received October 4, 2011; accepted November 22, 2011

Fang FANG (✉)
Department of Psychology, Peking University, Beijing 100871, China
Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China
Center for Life Sciences, Peking University, Beijing 100871, China
E-mail: ffang@pku.edu.cn

Yizhou WANG (✉)
National Engineering Lab for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China
E-mail: Yizhou.Wang@pku.edu.cn

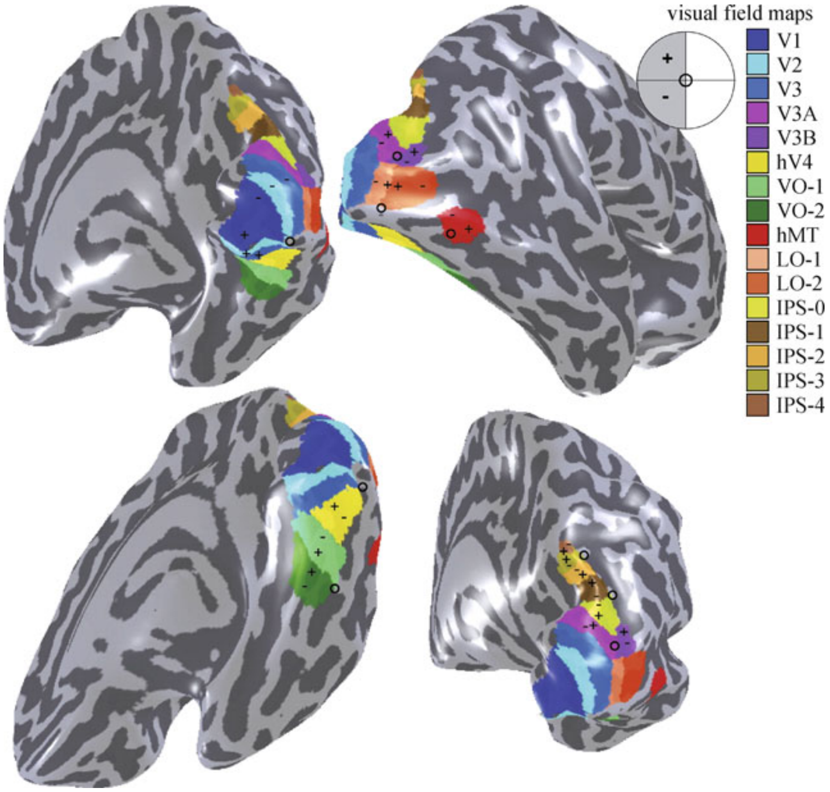

Fig. 1 Visual areas in human visual cortex (adapted from Ref. [1]).

2 Perceived size representation in V1

One of the most fundamental properties of V1 is its retinotopic organization, which makes it an ideal candidate for encoding spatial properties of objects such as size. Three-dimensional (3D) contextual information can often lead to size illusions. Figure 2 shows a size illusion. A rendered 3D scene of a hallway and walls was used to present two physically identical 3D rings: one was at a close ("front") apparent depth and the other was at a far ("back") apparent depth. The primary cues to depth include linear perspective, occlusion and relative texture size, which make the back ring perceived to be larger than the front ring. To assess the magnitude of the size illusion, we performed a psychophysical matching task. The result showed that, on average, the back ring appeared to be approximately 15% larger than the front ring.

With this size illusion, we performed a high-resolution fMRI experiment to address a critical question in vision sciences, that is, how complex 3D contextual information can influence spatial activity patterns in V1 [2]. We used a simple block design to present the stimuli in the fMRI experiment. Specifically, subjects fixated a small green point at the center of the front 3D ring (i.e., the intersection of the spokes) for 20 s, and then the front 3D ring was counterphase-flickered for 10 s. The green fixation point moved to the center of the back 3D ring

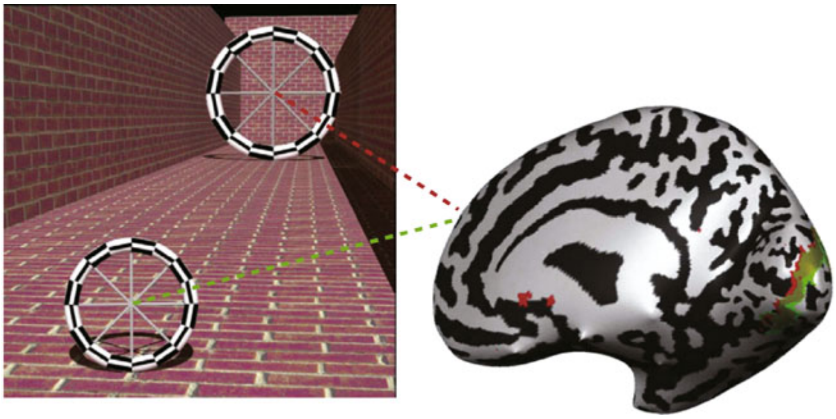

Fig. 2 Perceived size of rings affects retinotopic representations (adapted from Ref. [2]).

and the subjects were instructed to follow it and maintain fixation for 20 s, and then the back 3D ring was flickered for 10 s, including the spokes allowed the fixation point to always remain directed on the object (3D ring plus spokes). Otherwise, one fixation would have directed attention to the floor and the other to the back wall.

We hypothesized that the changes in the spatial distribution of activity in V1 because of a size illusion (if there is any) are the result of feedback of 3D contextual information from higher visual areas. One way to test the feedback model is to "silence" activity in areas that process 3D scene cues with an attention manipulation. Cortical areas in higher visual cortex are strongly modulated by attention. For example, withdrawing attention from face stimuli results in a 75% reduction in activity in face-selective brain areas in comparison to attending to faces [3]. Similarly, up to a 50% reduction in the fMRI signal occurs in shape-selective lateral occipital regions when attention is removed from the shape-related stimuli and focused on a demanding central fixation task [4].

To measure the effect of attention on the spatial distribution of activity in V1, each subject performed two tasks. In the attend-to-ring condition, attention was directed to the ring by asking subjects to detect a flickering pause of the 3D ring. In the attend-to-fixation condition, the focus of spatial attention was narrowed to minimize the perception of the 3D scene. Subjects were asked to perform a very demanding fixation task in which they needed to press one of two buttons to indicate the luminance change (increase or decrease) of the fixation point as soon as possible.

We found that the spatial distribution of V1 activity induced by the far ring was shifted towards a more eccentric representation of the visual field, while that induced by the close ring was shifted towards the foveal representation, consistent with their perceptual appearances (Fig. 2). Significantly, the spatial shift in peaks in the fMRI response on the cortical surface for each subject appears to predict the magnitude of each subject's size illusion. However, this effect was significantly reduced when the focus of spatial attention was narrowed with a demanding central fixation task. These results demonstrated that, with the feedback from higher cortical areas, the retinotopic property of V1 can be significantly modified by 3D context and is consistent with the perceptual appearance of the stimulus. They suggest that V1 could combine the retinal size and the depth of a stimulus to compute its real size.

3 Border ownership representation in V2

Border ownership is a term for the phenomenon that a visual border between two image regions is normally perceived to belong to only one of the regions. Border ownership assignment determines the figure-ground organization in a visual image and it is a critical aspect of object recognition [5,6]. Figure 3 shows an image of an old man and the edge signals generated by applying the Canny edge detector to the image [7]. It illustrates that edge signals are inherently difficult to interpret because of the ambiguity of the edge (border) ownership. For example, the border highlighted in red could be owned by either the foreground (i.e., the old man) or the background. Only after determining that the border belongs to the foreground, our visual system can comprehend the image in a right way. Primate electrophysiological studies [8,9] have shown that neurons in the early visual cortex encode the side to which a border belongs. Human functional imaging studies [10,11] have demonstrated that higher-level visual areas lateral occipital complex (LOC) and fusiform face area (FFA) are sensitive to a change of border ownership, but to date have provided no evidence regarding border ownership selectivity in human early visual cortex.

To study border ownership representation in human early visual cortex, we designed our stimuli (Fig. 4(a)) by modifying a bright/dark square-wave radial grating annulus. In the stimuli, either the bright or dark stripes

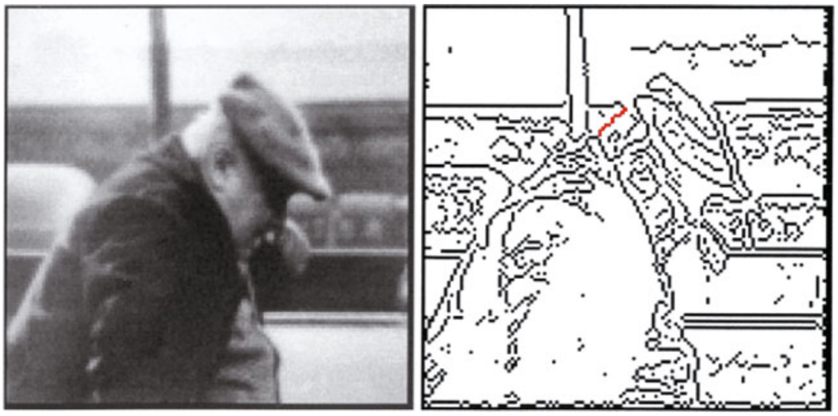

Fig. 3 An image of an old man and the edge signals generated by applying the Canny edge detector to the image (adapted from Ref. [7]). It illustrates that edge signals are inherently difficult to interpret because of the ambiguity of the edge (border) ownership.

(sectors of a disk) are slightly longer in the radial direction, both inward and outward. This provides contextual information that causes the borders between the bright and the dark stripes to appear to belong to either the bright stripes or the dark stripes, respectively. Although the image difference (the contextual information) between the two stimuli is very small, it dramatically changes the border ownership of locally identical edges along the edges of the stripes. With these two stimuli, we attempted to address three specific questions: 1) Are neurons in human early visual cortical areas selective for border ownership because of contextual modulation? 2) If so, is there any selectivity difference between the striate cortex (V1) and extrastriate cortical areas (e.g., V2)? 3) What is the role of attention in the processing of border ownership?

Because the border ownership selective neurons, if any, are very likely to mix with each other below fMRI spatial resolution, we used fMRI adaptation to overcome this difficulty. In a typical fMRI adaptation experiment, an adapting stimulus is presented that is presumed to adapt the population neurons sensitive to that stimulus. After removal of the adapting stimulus, a test stimulus is presented that is either identical to the adaptor or transformed in some dimension (e.g., orientation). If the fMRI signal is larger for the transformed stimulus as compared with the identical stimulus, it is inferred that neurons in the region have selective sensitivity to that manipulated dimension because the transformed stimulus is thought to be accessing a separate, unadapted neural population. fMRI adaptation is especially useful for exploring the neuronal selectivity of a subpopulation of neurons within an imaging voxel. For example, neurons in human early visual cortex (e.g., V1) tuned to different orientations are mixed within individual voxels and they cannot be resolved spatially by standard fMRI techniques, which makes fMRI adaptation necessary for studying orientation representation. Fang et al. [12] found that, after adaptation to an oriented pattern, the fMRI response in V1 and other early visual areas to a test stimulus was proportional to the angular difference between the adapting and test stimuli. In a separate psychophysical adaptation experiment, they measured contrast detection thresholds after adaptation and found that the fMRI signal in V1 closely matched the psychophysically derived contrast detection thresholds.

In this study, subjects were instructed to adapt to one of the two stimuli in Fig. 4(a). After adaptation, one of the two stimuli was briefly presented as a test stimulus. Regions of interest (ROI) in V1 and V2 were defined as areas that responded more strongly to the flickering ring (Fig. 4(b)) than the blank screen. Blood oxygenation level-dependent (BOLD) signals from the ROIs were extracted and analyzed. According to the logic of fMRI adaptation, if a cortical area could represent border ownership, the area should have a higher signal response to the test stimulus when the adapting and the test stimuli are different than when they are the same. The larger the differential response, the stronger the adaptation effect. To address the third question, we used two distinct attentional tasks to examine how manipulating attention modulates the border ownership selectivity of early cortical areas. Subjects were asked to attend to either the stimulus or a fixation point.

We found that border ownership selectivity in V2 was robust and reliable across subjects, and it was largely dependent on attention (Fig. 4(c)) [13]. Our study provides the first human evidence that V2 is a critical area for the processing of border ownership and that this processing depends on the modulation from higher-level cortical areas. Consistent with previous studies [14], our results suggest that neurons in early visual cortex integrate the image context far beyond the classical receptive field, and add weight to the claim that high-level visual computations and representations involve neural activity in early visual cortex [7].

4 Neural processing of partially occluded face

Great strides have been made in understanding the

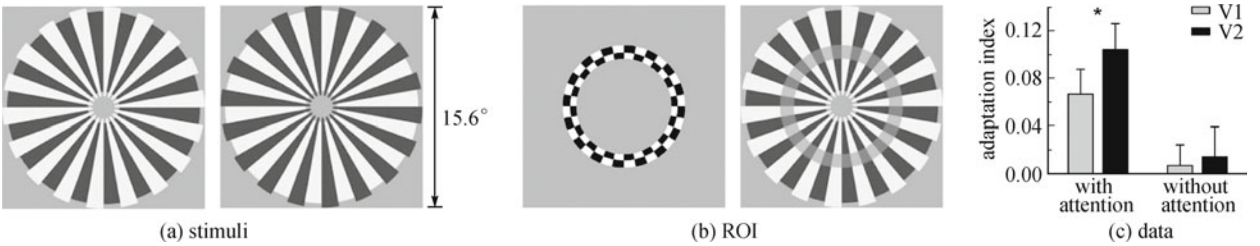

Fig. 4 Border ownership selectivity in human early visual cortex (adapted from Ref. [13]). (a) Stimuli used in the experiment. The interior part of the stimuli was locally identical across the two stimuli, but as a consequence of the difference in the contextual information, the borders between the bright and the dark stripes were perceived to belong to either the bright or the dark stripes. (b) Region of interest (ROI) definition. The checkered ring in the left panel was used to define ROIs in V1 and V2. The transparent gray ring in the right panel shows the size of the checkered ring relative to the stimulus. (c) Adaptation indices of V1 and V2 in the with-attention condition and the without-attention condition. Asterisks indicate a statistically significant difference between the adaptation indices of V1 and V2.

neural mechanisms of visual object recognition [15,16]. So far, object recognition has been studied mainly using individual objects presented alone. However, visual objects rarely occur in isolation in natural scenes. It is common for one object to occlude another object in natural images. A striking ability of human vision is the recognition of objects even when the sensory information specifying objects is optically incomplete due to occlusion. We have little difficulty completing occluded objects and seeing whole, uninterrupted objects.

How object completion is implemented in the visual cortex remains elusive. Evidence from human brain imaging studies suggest that (only) high-level visual areas selective for objects are likely candidates to mediate the operation of object completion effects [17–19]. However, psychophysical and electrophysiological studies suggest that early visual cortical areas also play an important functional role in object completion [20–22]. The discrepancy could be due to sluggish temporal response of the fMRI method. In the fMRI studies by Lerner et al. [23] and Hegde et al. [18], occluded objects were presented for hundreds of milliseconds. This completion-related activation in object selective areas as detected by fMRI might have reflected the perceptual consequence of object completion, rather than the process of object completion.

Psychophysical and event-related potential (ERP) studies have shown that object completion is not instantaneous; instead, it manifests its effect within a temporal window shortly after stimulus onset [24,25]. In this study, we attempted to investigate the temporal evolvement of face completion at different levels of the visual hierarchy. To circumvent the low temporal resolution deficit of the fMRI method, we used a backward masking paradigm to present occluded faces with various durations, which rendered it possible to interrupt the visual processing of occluded faces and follow in some detail the temporal evolvement of face completion [26,27].

Visual stimuli were constructed by presenting identical face fragments stereoscopically either behind or in front of a textured occluder (Fig. 5). In the first condition, the stimuli were perceptually completed and organized into a coherent face. In the second condition, they were perceived as disjoint face fragments hovering above the textured occluder [5,29]. In the psychophysical experiment and the fMRI experiment, by contrasting subjects' behavioral and BOLD responses in these two conditions, we measured the psychophysical time course of the face completion and its underlying cortical dynamics.

A face stimulus was presented with duration of 50, 150, 250, 350, or 450 ms, followed by a 300 ms mask. Subjects pressed one of the two response keys to indicate the view direction of the face stimulus, either left or right. The performance in the first condition, compared to the second condition, can be taken as a measure of face completion. When the performance in the first condition is better than that in the second condition, we attribute this to face completion. When the performance in the first condition is no better than that in the second condition, we take this to mean that face completion has not occurred. The extent of face completion as a function of stimulus duration was measured and defined as the time course of face completion [24]. We found that the face completion started to manifest its effect after 50 ms. To investigate when face completion terminated, we compared the performance at different duration conditions. Our data show that subjects' performance in the first condition saturated at 250 ms and suggest that face completion terminated before 250 ms. Overall psychophysical data suggested that face completion took effect between 50 ms and 250 ms after stimulus onset.

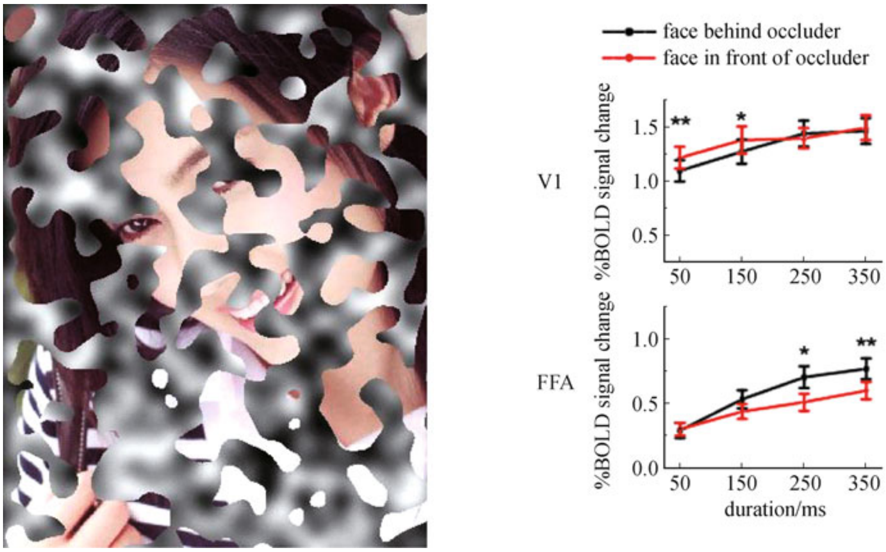

Fig. 5 Neural processing of partially occluded face consists of two synergetic phases: early suppression in early visual areas (e.g., V1) and late enhancement in later visual areas (e.g., FFA) (adapted from Ref. [28]).

fMRI results showed that V1 and V2 responded stronger in the second condition than in the first condition at short durations (50 and 150 ms), while occipital face area (OFA) and FFA responded stronger in the first condition than in the second condition at long durations (250 and 350 ms) (Fig. 5) [28]. These data provide evidence that both early visual cortical areas (V1 and V2) and high-level face selective areas (OFA and FFA) were involved in face completion, but they responded at different temporal phases with opposite activation patterns. Early suppression in early visual areas might be involved in figure-ground segmentation and late enhancement in higher visual areas might be related to perceptual grouping and face recognition. We also show that, when attention was directed to do a very demanding fixation task, all these effects were abolished, which demonstrated that attention was necessary to these neural events.

5 From image understanding to simulating human saccadic scanpaths on natural images

Human saccade is an important kind of eye movement to obtain visual information. In a highly dynamic and cluttered world, to acquire visual information efficiently and rapidly, it is important for our brain to decide not only where we should look at, but also the sequence of fixations. Indeed, both of them are essential for us to understand human saccadic behavior. We propose a computational model to simulate human saccadic scanpaths on natural images. Investigating this topic is not only helpful in understanding the computational aspects of visual perception, but also beneficial to many important applications such as image and video compression, object detection, and web-page design.

Our saccade model integrates three factors that drive human attention reflected by eye movements: reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. Figure 6 shows the framework of the proposed model [30]. The three modules highlighted with a grey-blue background are the key factors.

1) The reference sensory responses of an image are multi-band filter responses of sparse coding functions extracted from the stimuli. They are considered as a representation of the raw input signal and serve as the reference information in the proposed system. Neurophysiological evidence shows that the receptive fields of neurons in V1 are similar to sparse codes learned from natural images [31].

2) The fovea-periphery resolution discrepancy provides detailed information around a fixation location, but coarse information in periphery. This information discrepancy directly leads to sequential fixation transitions. In Fig. 6, a foveal image at fixation Q_t is generated and multi-band filter response maps of this foveal image are extracted.

3) Visual working memory retains perceptual information for a certain period of time and inhibits immediate return of attention. The content in working memory is also represented with multi-band filter response maps. To simulate the decay of working memory, we multiply the filter response maps in the memory with a forgetting factor ε $(0 \leqslant \varepsilon \leqslant 1)$. We update visual working memory by taking the maximal filter response values of the foveal image filter responses and the current decayed filter responses in the memory at each location. The resulted filter responses encode the so far obtained perceptual information of the scene in the brain.

Then, by subtracting the updated filter response maps in the working memory from the reference sensory responses, we obtain multi-band residual filter response maps. This simulates the dynamic interaction among the three factors along with eye movements. Consequently,

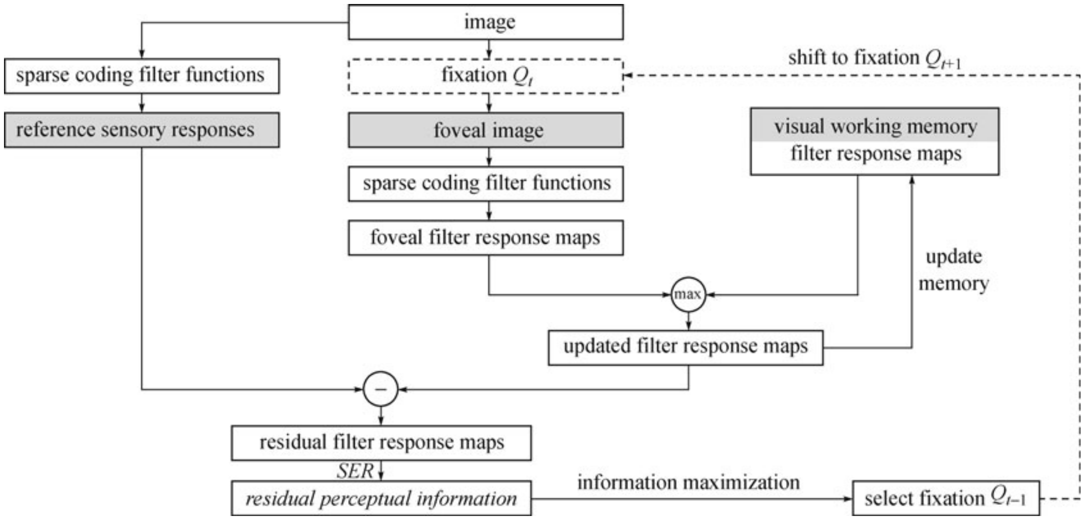
image
sparse coding filter functions
fixation Q_t
shift to fixation Q_{t+1}
reference sensory responses
foveal image
visual working memory
filter response maps
sparse coding filter functions
foveal filter response maps
update
memory
max
updated filter response maps
−
residual filter response maps
SER
residual perceptual information
information maximization
select fixation Q_{t-1}

Fig. 6 The proposed framework for our dynamic attention model (adapted from Ref. [30]).

we compute the site entropy rate (SER) [32] on residual filter response maps to obtain a residual perceptual information (RPI) map. The resulted RPI map represents the spatial distribution of the discrepancy between the amount of information stored in the brain and that contained in input stimulus. To pursue the maximal amount of "new" information of the scene and reduce this perceptual discrepancy, according to the information maximization principle, we choose the spot of the maximal SER value as the next fixation Q_{t+1}, and then start the next iteration.

To test the performance of the proposed model, we collected human eye movement data from a natural image dataset, and compare scanpaths generated by our model and two other approaches against the eye movement data in two aspects. One is the dynamic aspect of saccades such as fixation orders. The other is the static property of scanpaths such as fixation spatial densities. We randomly collected a dataset of 20 color images from the Internet including natural scenes, street and buildings, indoor images, etc. Eye movement data were collected from 24 subjects with this dataset using a high-speed SMI eye-tracker with a 500 Hz sampling rate. Subjects were positioned 0.53 m away from a 21-inch CRT monitor. The images were presented in a random order. Each was displayed for 3 s and followed by a blank screen for 1 s. A cross was placed at the center of the blank screen to engage the first fixation at the center of the images. The subjects were given no particular instruction except asking them to observe the images.

Itti and colleagues propose a scanpath generation method from static saliency maps based on winner-takes-all (WTA) and inhibition-of-return (IoR) regulations [33–35], which is the most referred method in literature as far as we know. Hence, we compare our model with Itti et al.'s approach. Moreover, to demonstrate that the advantage of the proposed model does come from incorporating the dynamic process of information pursuit, we substitute the static saliency computation module in Itti et al.'s work [33] with Wang et al.'s static saliency model [32], but still using the WTA and IoR regulations, then we compare generated scanpaths by our method and by the updated model of Wang et al.'s [32].

When simulating eye movements, we first decide the length of the scanpath for an image. The length is sampled from the statistics of eye movement data on that image (usually 8 to 10 fixations). Then, we generate three fixation sequences of that length using our model, Itti et al.'s model [33] and Wang et al.'s model [32], respectively. To quantitatively compare the stochastic and dynamic scanpaths of varied lengths, we propose to employ time-delay embedding, which has been used widely in the study of dynamical systems [36]. Specifically, we divide scanpaths into pieces of length k ranging from 2 to 5, and use two distance measures: 1) Hausdorff distance (H-Distance) and 2) the mean minimal distance (MM-Distance), to evaluate the scanpaths generated by different models and human data. The smaller such a distance, the closer/more similar to human scanpath, thus a better prediction. Figures 7(a) and 7(b) show comparison results between our model and the other two methods in terms of the Hausdorff distances and the mean minimal distances, respectively. We compare scanpaths at different lengths $k = 2$ to 5. From the comparison results, we can see that our model performs better in simulating the dynamics of saccadic scanpaths.

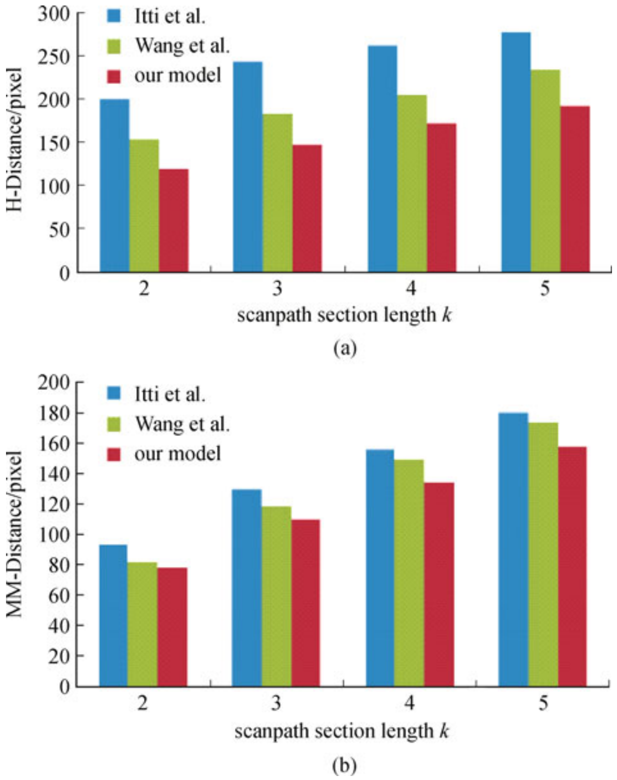

Fig. 7 Comparison results between different models using Hausdorff distance (a) and the mean minimal distance (b) at different scanpath length k (adapted from Ref. [30]).

We also compare our model with Itti et al.'s [33] and Wang et al.'s [32] static saliency model using ROC areas. ROC areas are generated by classifying the locations in a saliency map into fixations and non-fixations with varying quantization thresholds [37]. The larger the ROC area is, the better prediction the model makes. The results show that the ROC area in our model is 0.7183, which is higher than Itti et al.'s (ROC area = 0.6706) and Wang et al.'s (ROC area = 0.7081) models, indicating that our model predicts human fixations more accurately than these two models.

In summary, we propose a computational model to simulate human saccadic scanpaths on natural images without a specific task based on the principle of information maximization. The proposed model identifies and integrates three important factors as the driving forces to guide eye movements based on a coherent multi-band sparse code filter response representation. The computational model embodies several well-known psychological phenomena, such as center-surround saliency, inhibition of return, the forgetting effect of working memory, etc. Extensive experiments show the advantage of the proposed model in predicting human saccadic scanpaths on natural images.

6 Conclusion

This paper reviews our recent finding on the functions of human early visual cortex in image understanding. These findings demonstrate that human early visual cortex not only perform local feature analyses, but also contribute significantly to high-level visual computation with assistance of attention-enabled cortical feedback. Furthermore, we show that a biologically plausible attention model can well predict human scanpaths on natural images.

Acknowledgements This work was supported by the Ministry of Science and Technology of China (Nos. 2011CBA00405 and 2010CB833903), the National Natural Science Foundation of China (Grant Nos. 30925014, 90920012, and 30870762) and the Fundamental Research Funds for the Central Universities.

References

1. Wandell B A, Dumoulin S O, Brewer A A. Visual field maps in human cortex. Neuron, 2007, 56(2): 366–383
2. Fang F, Boyaci H, Kersten D, Murray S O. Attention-dependent representation of a size illusion in human V1. Current Biology, 2008, 18(21): 1707–1712
3. Wojciulik E, Kanwisher N, Driver J. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. Journal of Neurophysiology, 1998, 79(3): 1574–1578
4. Murray S O, He S. Contrast invariance in the human lateral occipital complex depends on attention. Current Biology, 2006, 16(6): 606–611
5. Nakayama K, Shimojo S, Silverman G H. Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. Perception, 1989, 18(1): 55–68
6. Driver J, Baylis G C. Edge-assignment and figure-ground segmentation in short-term visual matching. Cognitive Psychology, 1996, 31(3): 248–306
7. Lee T S, Mumford D, Romero R, Lamme V A F. The role of the primary visual cortex in higher level vision. Vision Research, 1998, 38(15–16): 2429–2454
8. Zhou H, Friedman H S, von der Heydt R. Coding of border ownership in monkey visual cortex. Journal of Neuroscience, 2000, 20(17): 6594–6611
9. Qiu F T, von der Heydt R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. Neuron, 2005, 47(1): 155–166
10. Kourtzi Z, Kanwisher N. Representation of perceived object shape by the human lateral occipital complex. Science, 2001, 293(5534): 1506–1509
11. Andrews T J, Schluppeck D, Homfray D, Matthews P, Blakemore C. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. NeuroImage, 2002, 17(2): 890–901
12. Fang F, Murray S O, Kersten D J, He S. Orientation-tuned fMRI adaptation in human visual cortex. Journal of Neurophysiology, 2005, 94(6): 4188–4195
13. Fang F, Boyaci H, Kersten D. Border ownership selectivity in human early visual cortex and its modulation by attention. Journal of Neuroscience, 2009, 29(2): 460–465
14. Albright T D, Stoner G R. Contextual influences on visual processing. Annual Review of Neuroscience, 2002, 25(1): 339–379
15. Grill-Spector K, Malach R. The human visual cortex. Annual Review of Neuroscience, 2004, 27(1): 649–677
16. Kanwisher N. Functional specificity in the human brain: A window into the functional architecture of the mind. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(25): 11163–11170
17. Stanley D A, Rubin N. fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. Neuron, 2003, 37(2): 323–331
18. Hegde J, Fang F, Murray S O, Kersten D. Preferential responses to occluded objects in the human visual cortex. Journal of Vision, 2008, 8(4): 16-1–16-16
19. Grutzner C, Uhlhaas P J, Genc E, Kohler A, Singer W, Wibral M. Neuroelectromagnetic correlates of perceptual closure processes. Journal of Neuroscience, 2010, 30(24): 8342–8352
20. Sugita Y. Grouping of image fragments in primary visual cortex. Nature, 1999, 401(6750): 269–272
21. Pillow J, Rubin N. Perceptual completion across the vertical meridian and the role of early visual cortex. Neuron, 2002, 33(5): 805–813
22. Bakin J S, Nakayama K, Gilbert C D. Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. Journal of Neuroscience, 2000, 20(21): 8188–8198
23. Lerner Y, Hendler T, Malach R. Object-completion effects in the human lateral occipital complex. Cerebral Cortex, 2002, 12(2): 163–177
24. Murray R F, Sekuler A B, Bennett P J. Time course of amodal completion revealed by a shape discrimination task. Psychonomic Bulletin & Review, 2001, 8(4): 713–720
25. Chen J, Liu B, Chen B, Fang F. Time course of amodal completion in face perception. Vision Research, 2009, 49(7): 752–758
26. Bar M, Tootell R B, Schacter D, Greve D, Fischl B, Mendola J, Rosen B, Dale A. Cortical mechanisms specific to explicit visual object recognition. Neuron, 2001, 29(2): 529–535
27. Lamme V A F, Zipser K, Spekreijse H. Masking interrupts figure-ground signals in V1. Journal of Cognitive Neuroscience, 2002, 14(7): 1044–1053
28. Chen J, Zhou T, Yang H, Fang F. Cortical dynamics underlying face completion in human visual system. Journal of Neuroscience, 2010, 30(49): 16692–16698
29. Fang F, He S. Viewer-centered object representation in the human visual system revealed by viewpoint aftereffect. Neuron, 2005, 45(5): 793–800
30. Wang W, Chen C, Wang Y, Jiang T, Fang F, Yao Y. Simulating human saccadic scanpath on natural images. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2011, 441–448
31. Olshausen B, Field D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 1996, 381(6583): 607–609

32. Wang W, Wang Y, Huang Q, Gao W. Measuring visual saliency by site entropy rate. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010, 2368–2375

33. Itti L, Koch C, Niebur E. A model of saliency based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254–1259

34. Yao J G, Gao X, Yan H M, Li C Y. Field of attention for instantaneous object recognition. PLoS ONE, 2011, 6(1): e16343

35. Hou X, Zhang L. Dynamic visual attention: Searching for coding length increments. Advances in Neural Information Processing Systems, 2008, 21: 681–688

36. Sauer T, Yorke J, Casdagli M. Embedology. Journal of Statistical Physics, 1991, 65(3–4): 579–616

37. Bruce N, Tsotsos J. Saliency based on information maximization. Advances in Neural Information Processing Systems, 2006, 18: 155–162

Fang FANG is a professor in the Department of Psychology at Peking University, Beijing, China. He earned his B.S. degree in psychology and M.E. degree in signal and information processing at Peking University. From 2001 to 2007, he did his Ph.D. and postdoctoral training in cognitive and biological psychology at the University of Minnesota at Twin Cities. He then moved back to Peking University and started his own lab as principle investigator. His research has been dedicated to human visual perception, attention and awareness, using functional brain imaging, psychophysics and computational modeling. He has authored or co-authored about 40 research papers published in prestigious journals, including Nature Neuroscience, Neuron, Current Biology, Proceedings of the National Academy of Sciences of the United States of America (PNAS), Journal of Neuroscience. He received the National Distinguished Young Scientist Award from the National Natural Science Foundation of China in 2009 and the National Award for Youth in Science and Technology from the China Association for Science and Technology in 2011. He serves as editorial board member of Experimental Brain Research and Frontiers in Perception Science.

Yizhou WANG is a professor of Department of Computer Science at Peking University, Beijing, China. He is a vice director of Institute of Digital Media at Peking University, and the director of New Media Lab of National Engineering Lab for Video Technology. He received his Bachelor's degree in electrical engineering from Tsinghua University, Beijing, China, in 1996, and his Ph.D. degree in computer science from University of California at Los Angeles (UCLA) in 2005. He worked at Xerox Palo Alto Research Center (Xerox PARC) as a research scientist from 2005 to 2007. Dr. Wang's research interests include computer vision, statistical modeling and learning, and digital visual arts.