Front. Electr. Electron. Eng. China 2011, 6(1): 120–136
DOI 10.1007/s11460-011-0133-3

RESEARCH ARTICLE

Dewen HU, Zongtan ZHOU, Zhengzhi WANG

Processing real-world imagery with FACADE-based approaches

Abstract This paper considers the processing of real-world imagery in the so-called Form-And-Color-And-DEpth (FACADE) framework, which features some superior mechanisms of the human vision system (HVS). FACADE framework was originally proposed by Grossberg et al. as an integrative model of the HVS to illustrate the possible procedures for visual perception of shape (the boundary contour), surface (luminance and color), and binocular depth. As a simplified, reasonable and mathematically full-fledged approach to the HVS, we saw FACADE as a promising infrastructure through which to construct a powerful image processing engine. However, in our attempts to use the approach in its original modality, to deal with real-world imagery, we found it to be inefficient and non-robust.

After re-introducing the model hierarchy and illustrating the involved cell dynamics of the FACADE framework, this paper reveals the crucial issues that lead to the deficiency and accordingly present our substitutive solutions by incorporating the mechanisms of anisotropic spatial- and diffusive orientational-competition to make the HVS-featured model efficient and robust. A computer system based on the improved FACADE engine has been implemented and tested not only with illustrative images to highlight the model characteristics, but also with some real-world imagery in both monocular and binocular situations, thereby demonstrating the ability of the FACADE-based image processing approach featuring the HVS.

Keywords visual perception, human vision system (HVS), Form-And-Color-And-DEpth (FACADE) framework, real-world imagery

Received October 28, 2010; accepted December 16, 2010

Dewen HU (✉), Zongtan ZHOU, Zhengzhi WANG
Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China
E-mail: dwhu@nudt.edu.cn

1 Introduction

Visual perception is the fundamental process in the human vision system (HVS) that allows us to analyze the surrounding situations efficiently and rapidly. Understanding the robustness and efficiency of human visual perception has been a critical issue of investigation in the cognitive sciences over the last few decades. Because humans outperform the best machine vision systems with respect to almost any measure, constructing computer systems featuring an HVS has always been an attractive goal. Currently, there are probably several hundred models of the HVS (or parts of the HVS), of which a large majority deal with specific visual phenomena (such as specific visual illusions) or with specific cortical areas (e.g., Gabor filters [1,2]). Some of these models have provided useful contributions to neuroscience, and a few have even had an impact on physiologists. In terms of applicability to constructing an HVS featured visual information processing system, several models have been more successful and are better known than others, such as the HMAX model by Poggio et al. [3–6], the VisNet model by Rolls et al. [7–10], and the series of visual perception models by Grossberg et al. [11–14], including the so-called FACADE framework, which is the main topic of this paper.

One of the most important features of the HVS is its ability to generate invariant perception from the rather variant visual input. The HMAX and VisNet models are probably two of the most important approaches to explain and emulate this sense of invariability. In the HMAX model [15], Poggio et al. introduced a fairly general framework to recognize complex visual scenes. They proposed a hierarchical feed-forward system that closely follows the organization of the visual cortex and built an increasingly complex and invariant feature representation by alternating between a template matching and a maximum pooling operation. The VisNet [16] is a

four-layered feed-forward network with competition between the neurons within a layer, and a trace learning rule to help it learn transform invariance.

Other amazing features of the HVS are its automatic selectivity and ability to realize self-organization from the input image, enabling the visual system to find the organized structures within the image automatically while at the same time, filtering out random noise or the disorganized details that could confuse the visual perception. From this perspective, there are also several HVS featured approaches, although the most important ones are those originally proposed by Grossberg et al. This series of HVS models developed continuously from the original form of the standalone boundary contour system (BCS) and the feature contour system (FCS) proposed in the early 1980s [17–19], to some rather integrated and specialized frameworks, such as the latest ARTScan [20]. The evolution of this series of HVS models began with the attempt to explain paradoxical data about brightness and form perception, when Grossberg et al. [17] tried to introduce the basic concepts of an integrated pre-attentional processing model of the HVS, i.e., the BCS and FCS. The BCS generates boundary segmentation by imposing neural mechanisms of short-range competition and long-range cooperation, while the FCS obtains surface perception through filling and diffusing features, such as brightness and color, or even higher level features, such as texture or motion.

To explore further the in-depth perception mechanism of binocular vision within the HVS, Grossberg and his colleagues introduced the idea of the FACADE (literally a combination of the words Form-And-Color-And-DEpth) model, initially combining BCS and FCS in 1994 [11], and then proposed an integrated FACADE model and established the FACADE theory, as the integrated neural network theory of 3-D biological vision [21,22]. The FACADE framework (Fig. 1) describes how 2-D pictures can be processed to generate the perception of a 3-D scene in which figure-ground separation of visual surfaces occurs. More specifically, the mathematically full-fledged formulas of FACADE made it possible to demonstrate through computer simulation how the visual cortex generates three-dimensional boundary segmentations and how these can be used to control the filling-in of 3-D surface properties in response to visual scenes.

Having proposed the FACADE theory, Grossberg and his colleagues presented the LAMINART model [12,23,24] to interpret how the laminar circuits of the visual cortex implement context-sensitive binding processes, such as perceptual grouping and attention. The LAMINART model, as claimed, refined the FACADE model and introduced some new features, and thus, could be employed to explain a much larger set of neurophysiological, anatomical, and psychophysical data

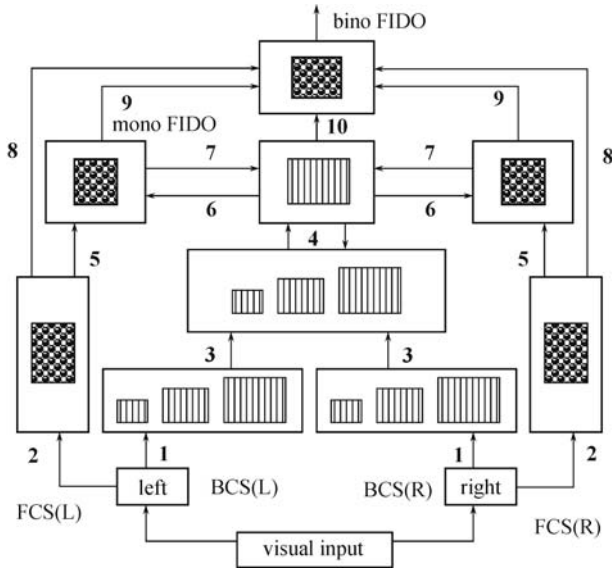

Fig. 1 Circuit diagram of the FACADE framework (adapted from Ref. [21])

about stereopsis, which had been previously impossible. In 2003, LAMINART was extended to a 3D LAMINART [13,25–28], which describes how initial stages of monocular and binocular oriented filtering interact with later stages of 3D boundary formation and surface filling-in in the lateral geniculate nucleus (LGN) and cortical areas of V1, V2, and V4.

Although the FACADE and 3D LAMINART models deal well with stereopsis and a variety of occlusion and transparency situations, they need to be further developed to account for other depth cues, including SFT (shape-from-texture). To achieve this aim, Grossberg et al. presented a neural model of SFT called theLIGHTness-and-SHApe-From-Texture (LIGHTSHAFT) model [29]. This model utilizes monocular visual texture information to produce a 3D percept of surface lightness. It is built upon and extends the FACADE and 3D LAMINART models, which explain and simulate a large amount of psychophysical and neurobiological data about 3D vision. More recently, Grossberg et al. proposed the ARTSCAN model [20] to explain how spatial and object attention work together to search a scene and learn what is in it. The ARTSCAN model provides a mechanistic explanation of how the concept of an attentional spotlight can be reconciled with concepts about object-fitting attentional shrouds.

The FACADE framework, as one of the simplified, reasonable and mathematically full-fledged models of the HVS ever proposed is, in our opinion, an appropriate and very promising infrastructure to be applied in constructing a corresponding computer vision system. As mentioned above, there has already been much progress since the original FACADE model was presented, thus making the ever outstanding FACADE theory seem somewhat outdated. However, with respect to researchers' practical interest in implementing a computer image process-

ing engine that features HVS mechanisms, these derived or improved models after the original FACADE are actually not necessarily superior to the more developed FACADE in dealing with real-world imagery. In fact, most of the above-mentioned models were proposed to explain specific neurophysiological, anatomical, and psychophysical data (e.g., contrast variations of dichoptic masking and the correspondence problem, the effect of interocular contrast differences on stereo-acuity, Panum's limiting case, the Venetian blind illusion, da Vinci stereopsis, and so on). These approaches are more delicate and dedicated, but may lack mathematical rigor and completeness, and do not necessarily consider computational rationality and applicability. To the best of our knowledge, relevant computer implementations based on the newer approaches that are descendants of FACADE are rare.

The FACADE framework, along with its building blocks the BCS and FCS models, has been implemented and utilized (although there are not too many cases of this) in image processing applications, for example, processing remote sensed imagery [30], and enhancing the imagery of range data gathered by a synthetic aperture radar (SAR) sensor [31]. Our research group applied FACADE in enhancing medical images and generating automatic segmentations of roadside scenery images obtained from our autonomous land vehicle (ALV) system. However, in these attempts, the framework in its raw form was found inefficient and non-robust, making it necessary to develop some refinements and enhancements accordingly. This paper discusses the issues around adapting and applying the FACADE framework for the purpose of processing real-world imagery.

Essentially, FACADE is a model for stereoscopic vision, which focuses more on the binocular fusing of the visual signals from each eye. Thus, it is reasonable for readers to expect a more in-depth discussion of the mechanisms of binocular fusing and 3-D surface segmentations of FACADE; nevertheless, this is beyond the scope of this paper, since the procedures are so complicated, and the length of text is limited. Only the monocular BCS and FCS subsystems in the FACADE framework are discussed in detail to facilitate the proposal of our substitutive paradigm for real-world imagery processing. So far, as that is concerned, the depth perception mechanisms for the binocular situation in FACADE were also implemented and improved by our research group [32–34] with even more adaptations; because once again, in its original form, the binocular processing is not robust enough and consumes too many computational resources, making it barely applicable.

The rest of this paper is organized in five major parts. Section 2 addresses the model hierarchy and illustrates the cell dynamics of the original FACADE framework involved in establishing the context of investigation of this paper (the related formulas of the original FACADE model as developed by Grossberg et al. are presented in Appendix A). Section 3 analyzes the robustness of the FACADE model and addresses those crucial issues leading to the deficiency of the original framework. Section 4 presents our corresponding substitutive solutions, incorporating the anisotropic spatial- and diffusive orientational-competition for the hyper-complex cell and a different cooperation mechanism for the bipolar cell, to improve the computational efficiency and make the framework more applicable, stable, and robust. Consequently, it is possible to implement and test a computer system based on the improved FACADE engine. Various images for illustration are introduced in Sects. 3 and 4 to compare the different paradigms and highlight the improvements brought about by our approach. In the last part of the paper (Sects. 5 and 6), some processing results of real-world imagery in both monocular and binocular situations are presented and discussed to demonstrate the characteristics and capability of the FACADE-based image processing engine, and to suggest a direction for possible future development.

2 Mechanism and framework of FACADE model

2.1 FACADE processing microcircuit

The global framework and the major visual information processing pathways of FACADE can be illustrated by the microcircuit diagram in Fig. 1. Here, we only present some brief and qualitative explanations to give the reader an overall view of the building blocks, and information flows inside this rather complicated HVS model. One may notice that there are different sized sub-squares inside the many blocks of the diagram; they denote the parallel multi-scale processing. The textured patterns of vertical bars inside a processing block denote the form-related features or what we call in this context, the orientational boundary contour signals, while the circle-fashioned patterns represent the surface features inside the visual input, such as luminance, color, and texture. Thus, it is clearly seen that the middle channels in this microcircuit deal with boundary contour signals and the left/right-hand side pathways manage luminance and color signals.

In FACADE, visual inputs (the images) from the left or right eyes are first fed into the left and right BCS, respectively, and then fused in the higher level processing blocks via different pathways to generate the depth perception and therefore, the 3D segmentation. To achieve the tuning and binding of different features, boundary contour signals and surface-feature signals from different pathways are pooled together at different processing lev-

els of the so-called filling-in domains (FIDOs), making it possible for them to interact with each other through the process of diffusive filling-in while generating visible outputs.

Detailed structures of one of the most important modules within the FACADE framework, i.e., the cells hierarchy of the BCS (left or right), are illustrated in Fig. 2. As a relatively standalone and representative subsystem of FACADE, the BCS consists of several layers of cells organized in a general feed-forward architecture with feedback pathways. The layers of cells from the bottom of the hierarchy to relatively higher levels include LGN cells, simple cells, complex cells, and hyper-complex cells, where each type has different neural-dynamics, and thus, different mechanisms for manipulating the visual information flows throughout the hierarchical feed-forward and feedback pathways.

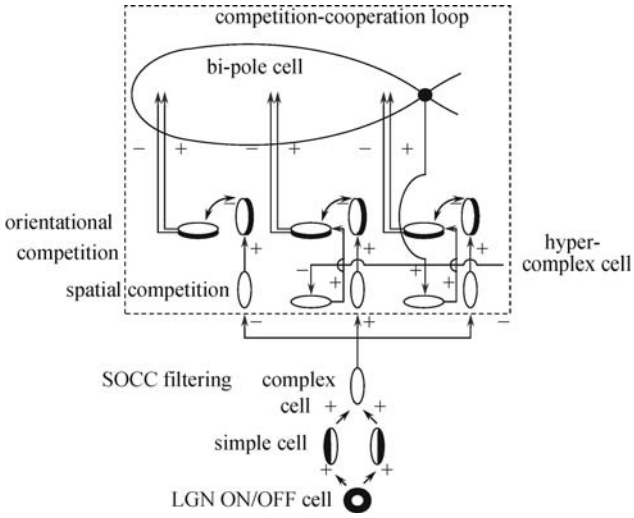

Fig. 2 Cell hierarchy of the monocular BCS inside FACADE [11]

The neural dynamics addressed below and the corresponding mathematical formulas (given in the Appendix) are only a kernel subset that we regarded as the most important and would have liked to have further manipulation. FACADE is a complicated framework that includes massive functional procedures for visual information manipulation. Thus, it is generally useful to point out and emphasize at the first opportunity, that the mechanisms for short-range competition and long-range cooperation, that is, the functions of the hyper-complex cells and bipolar cells, respectively, inside the BCS, along with the feedback induced push-pull dynamics, are the essential and unique mechanisms that empower this promising image processing framework and contribute most to making FACADE distinguishable from other HVS models.

2.2 SOCC filtering: LGN, simple, and complex cells

In the hierarchy of BCS processing, as illustrated in Fig. 2, the LGN, simple, and complex cells are located at the bottom of the feed-forward pathway, but outside the feedback loop called the competition-cooperation (CC) loop or static oriented competition-cooperation (SOCC) loop. The functions of these cells are to perform the necessary pre-processes on the raw visual inputs, i.e., ON/OFF spatial filtering of the LGN cells, orientation contrast filtering of the simple cells, and synthesizing the complex cells. These processes could easily be regarded as various kinds of image filters, and are thus referred to, as a whole, as SOCC filtering. If appropriately configured, SOCC filtering does not affect too much the robustness and efficiency of the whole BCS dynamics, and thus the overall FACADE processing, and consequently the corresponding formulas of the dynamics of these cells are omitted here. Readers who require more details of these stages of processing should refer to the original works of Grossberg et al. [21]. However, to establish the context for the fundamental formulas for further discussion, the output of a complex cell (which becomes the input of the next layer of cells, i.e., the hyper-complex cells, see Eqs. (13)–(15) in Appendix A-A) is henceforth given directly as C_{ijkd}, where (i, j) denotes spatial location in the receptive field of the complex cell, which in our image processing procedure is usually correspondent to the pixel-wise position of the input image; k represents the different orientations, where each k corresponds to the output of a specific orientated simple cell; and d is the index of the depth (disparity) layer in which the complex cell resides. In most cases of non-binocular image processing (where there is only one source for image acquisition, e.g., the processing of remote-sensory images and most medical images), we set $d \equiv 0$.

2.3 Hyper-complex cells: spatial competition

The hyper-complex cells in the BCS realize the mechanism of short-range competitions among the different locational- and orientational-contrasting input signals from the complex cells. The essential function of a competitive hyper-complex cell is to generate accurately localized (in terms of both spatiality and orientation) boundary contour signals. There are actually two stages of competitions in the hyper-complex cells; the first is spatial competition and the second is orientational competition, as described and illustrated here (Fig. 3) and in the next subsection (Fig. 4), respectively.

For spatial competition, the rectified output from each complex cell Y_{ijkd} (in Eq. (16), Appendix A-A) excites a like-oriented hyper-complex cell corresponding to its location and inhibits like-oriented hyper-complex cells corresponding to the nearby locations. This spatial inhibition occurs in a center-surround manner between cells coding the same scale, disparity, and orientation.

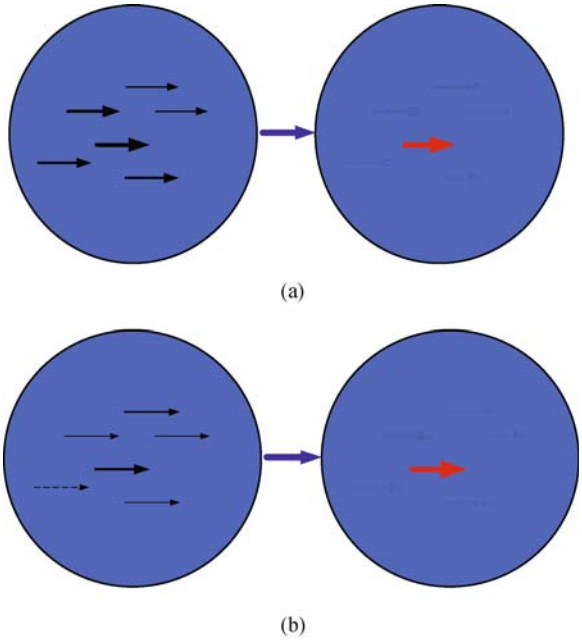

Fig. 3 Spatial competition occurs at the first layer of hyper-complex cells, where the outputs of those complex cells with the same orientation (C_{ijkd}) are competing with each other. As shown, there are many inputs in the reception field of a hyper-complex cell, but only the dominant one wins and survives. The effect of this competition in the sense of image processing, is to prune the ramified or messy contrast signals to obtain a precise localization of the boundary contour (a), while at the same time, adaptively amplifying and keeping those weakly-contrasted signals which would normally be filtered out in other image processing approaches (b)

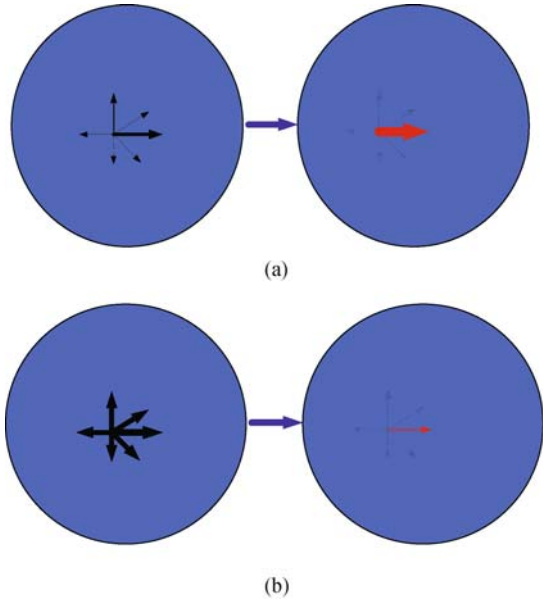

Fig. 4 Orientational competition as the second stage of interaction of hyper-complex cells. For the purpose of image processing, the effect of this second stage of competition is to obtain a deterministic and precise boundary contour orientation at each position of an input image (a), and suppress the noise-induced orientation signals, which could be thoroughly eliminated by the further push-pull dynamics after feedback (b)

The hyper-complex cells also receive feedback signals $g(O_{ijkd})$ (in Eq. (16), Appendix A-A) from the final layer of the cooperative grouping network.

2.4 Hyper-complex cells: orientational competition

In this stage of the competition, the hyper-complex cells compete across orientations at each position. Different oriented signals inhibit each other, and the greatest inhibition occurs between cells that code perpendicular orientations. Spatial and orientational competitions together realize a basic mechanism of push-pull activity (Appendix A-B). If one orientation is excited (or inhibited), then the cell coding the perpendicular orientation at the same spatial location is inhibited (or excited by dis-inhibition). Spatial competition causes the tuned cells just beyond the end of a thin line to be inhibited, while the hyper-complex cells that code perpendicular orientations at the same spatial position are then dis-inhibited (excited) by orientational competition. This mechanism for completing line ends, called end-cutting, helps to explain how illusory contours can be formed perpendicular to line ends.

The push-pull dynamics may also be modulated by the feedback signals within the CC loop, whereby the result of orientational competition, together with the signals of the long-range cooperation obtained by the bipolar cells, could be fed back to affect the next round of spatial competition further, and so on and so forth. Thus, the end-cutting effect may be grouped together by subsequent stages of the CC loop to form an illusory contour (as shown in Fig. 5). On the other hand, these global push-pull activities, if appropriately configured and controlled, could effectively filter out noise induced fake signals of boundary contours without the common side-effect of most traditional image de-noising approaches.

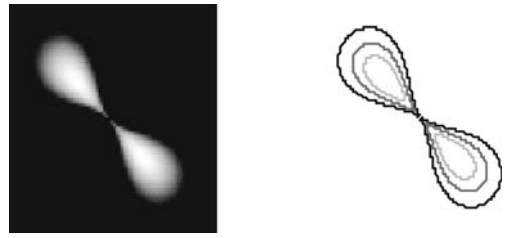

Fig. 5 Spatial distribution of the receptive field of a bipolar cell

As shown in Fig. 4, the orientational competition in the sense of image processing helps obtain a deterministic and precise boundary contour orientation at each position of an input image, and to suppress the noise-induced fake signals in the overall push-pull mechanism.

2.5 Cooperative bipolar cells

Outputs from the hyper-complex cells N_{ijkd} (Appendix A-B) are subsequently fed into a long-range cooperative

boundary completion process (Appendix A-C), which is responsible for the long-range grouping capabilities of the BCS. The long-range cooperation cells that receive this input are called bipolar cells because their receptive field is composed of two oriented lobes (Fig. 5). Each lobe receives input from a range of orientations and positions along the oriented axis of the lobe.

A bipolar cell is activated if both lobes receive large enough inputs, and then, the output is fed back to the hyper-complex cells within the CC loop, to initiate boundary completion and perceptive grouping (Fig. 6).

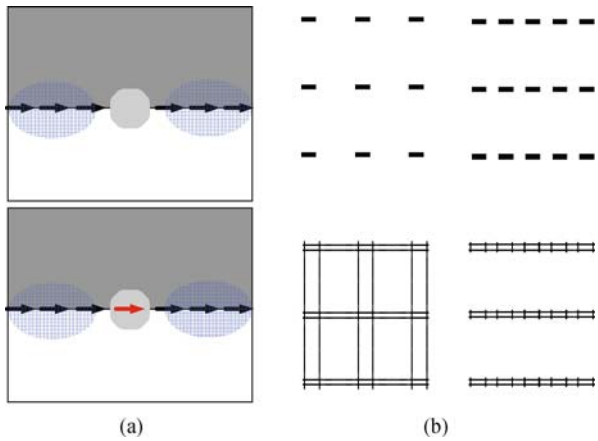

Fig. 6 Illustration of how the bipolar cells fill the gap in a broken boundary contour in processing the input images to generate a perceptive complete boundary outline (a), and how the push-pull mechanism affects the final result of boundary contour grouping (b). The top two images on the right hand side of these illustrative figures are the input images with some organized pattern, while the bottom images are the output of the layers of bipolar cells (with a rather larger receptive field). It can be seen that the self-organized grouping result could be affected by the dynamic push-pull mechanism, where rather weaker vertical groupings are suppressed by the stronger horizontal boundary contour cooperation

The term N_{ijkd} in Eq. (22) of Appendix A-C ensures that the ends of lines also strongly activate bipolar cells. This is reformulated in our approach below to obtain a more stable and robust boundary contour completion for real-world imagery processing.

As previously mentioned, the bipolar cells support feature grouping and boundary contour completion within a band of orientations. However, some of the cells receive weaker support from image contrasts than others, and thus, it is easy to reflect that further competitive interactions could help to remove this uncertainty. In the FACADE framework, an extra round of orientational competition and spatial competition are applied to the outputs of the bipolar cells before they are fed back to the hyper-complex cells to form the common sense CC loop (Appendix A-D). This feedback pathway is actually a mirror of the competitive stages of the feed-forward pathway.

After several iterations through these feed-forward/feedback pathways, the activities of the CC loop cells should converge (not considering the time varying transient item) to generate the final stable-state output (O_{ijkd}, Eq. (32) in Appendix A-D) of the BCS system, i.e., the boundary contour web to segregate the input image.

Considering the delicate push-pull mechanism provided by the competition and cooperation that can suppress unstructured noise while, at the same time, fixing a broken object boundary, the output of the BCS is supposed to be the precisely localized and semantically integrated (closed) boundary contours of those objects within the input image, without being influenced (too greatly) by the image noise or those disorganized structures that inevitably affect traditional image processing approaches for edge-extraction. Thus, the BCS could to some degree be regarded as an optimized image segmentation approach in the sense of image processing, which realizes the figure-ground and figure-figure segmentations (when interacting with the FCS, as discussed in the next subsection).

2.6 BCS/FCS interactive: filling-in

The boundary segmentation that was formed in the BCS, features no color or luminance qualities, since the complex cells only pool the orientational contrast signals from simple cells. The filling-in process within the FCS is thus responsible for binding the boundary contours and the surface features of luminance or color, to create the visible surface percepts. The BCS organizes these FCS filling-in events by topographically interacting with the FCS filling-in domains (FIDOs), whereby the boundary outputs from the BCS N_{ijkd} (Appendix A-E) act as filling-in generators and at the same time filling-in restrictive barriers to this diffusive process, as shown in Fig. 7. The filling-in process in the FCS can smoothen unstructured details (such as noise) within the segmentation, without the common side-effect of blurring the object boundary as is the case in most of the image denoising approaches.

3 Evaluation of efficiency and robustness of FACADE model

It is easy to conclude from the introduction in the previous section, that the FACADE framework, together with its full-fledged computational formulas, is a general-purpose and ready-made paradigm with which to implement a powerful computer image processing engine immediately. The competition and cooperation mechanisms combine perfectly within the uniform framework to enable self-organized 2D and 3D image segmentations featuring the perceptive characteristics of the HVS. As

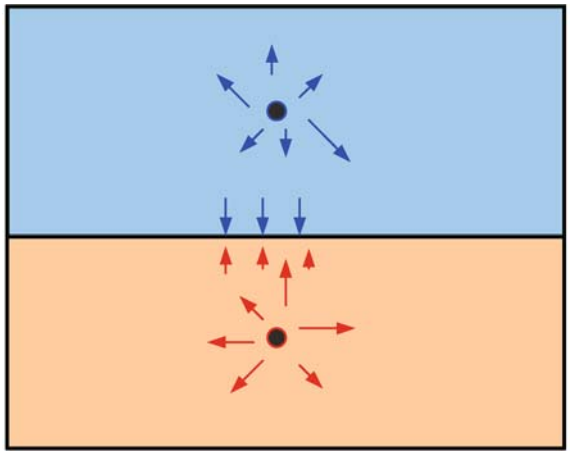

Fig. 7 The diffusive filling-in of features, such as luminance or color in the FCS, to generate visible image segmentations. The diffusions of these features are regulated by the boundary contour output of the BCS, which acts as a filling-in generator and barrier. Thus, for image processing, the interactions of the BCS/FCS filter out the noise along with those disorganized details inside the segmented elements, while at the same time retaining the precisely localized boundaries of the essential segmentations within the images

far as we know, such an image processing engine is far superior to the traditional approaches in terms of its ability to extract the distinctly localized boundary web of the structured objects within the input image, while simultaneously depressing the noise and other non-relevant image details to generate stable and perceptively reasonable image segmentations.

All these characteristics are exactly what we need for our practical goal of dealing with a variety of real-world imagery, but unfortunately things are not always that ideal. In our initial computer implementation of the FACADE framework, it was found that the actual cell activities, especially the boundary contour signals as the output of the BCS processing, are not exactly as flawless as expected. In actual fact, the system works well only at a fairly low noise (or those disorganized image details that we would prefer to neglect) level, at which the elaborately chosen model parameters may yield cells within the CC loop that collaborate to generate reasonable and stable image segmentations. Otherwise, the cell dynamics of the BCS may suffer from severe robustness issues, i.e., the system is considerably sensitive to the parameter configurations. Accordingly, the outputs of the BCS could be quite diversified with regard to the different choices of certain parameters, such as the shunting and decaying coefficient α of the hyper-complex cell, the feedback gain G, and the cooperative threshold Γ of the bipolar cell.

This initial evaluation reminds us that the FACADE framework is as yet (in its current modality) not appropriate for our real-world imagery processing tasks due to its instability and parameter sensitivity. Something must be done to overcome these issues of stability and robustness before we can apply the framework in the context of real-world imagery processing.

As an initial step, a rather simple but representative test image (as shown in Fig. 8) was deliberately constructed as input to our implemented FACADE processing engine, to illustrate the characteristics of FACADE processing and to help identify the possible causes of the stability and robustness issues. Using this test image it is usually easy to conclude at first impression (if one compares the filling-in output in Fig. 8 with the image filtering results shown and explained in Fig. 9) that the FACADE processing, if appropriately controlled, is far superior to the filtering based approaches in terms of its simultaneous abilities of object boundary retention and noise reduction. On the other hand, even with this simple test image, the FACADE processing could also yield poor results if the model parameters are not deliberately configured (Fig. 10). Here, as an example, we illustrate the effects of different choices of only two of the most crucial parameters in the BCS, i.e., the feedback gain G and the cooperative threshold Γ of the bipolar cell. Other parameters, such as the size of the receptive fields for the bipolar cell, the decaying coefficient in the dynamic equations of the hyper-complex cells that control the competitions, and the diffusive parameters of the

Fig. 8 (a) The test image containing a square object in a noisy background, where the integrity of the boundary contour of the square is disrupted by the strong noise; (b) the filling-in output of the FACADE-based image processing application in this paper with typical model parameters. The FACADE approach smoothes the strong noise while at the same time preserving the clarity of the object boundary and fixes the broken contour of the square object automatically, thus making it easier for further object recognition approaches to generate a reasonable image segmentation and separate the square object from the rather noisy background

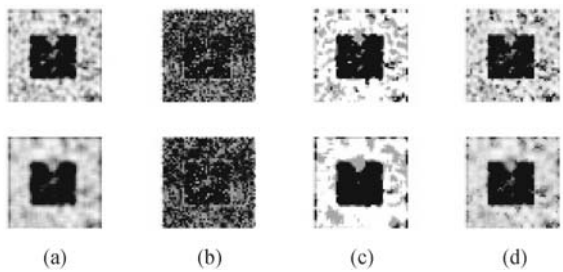

Fig. 9 Processing results of the same test image in Fig. 8 using various traditional filtering algorithms. (a) Low-pass filter; (b) high-pass filter; (c) mid-value filter; (d) Wiener filter. Results in the top line were obtained using a 3×3 template, while those in the bottom line made use of a 5×5 template. None of these filtering based approaches were able to recover the broken boundary contour of the square in the center, or filter out the noise effectively while preserving the visual clarity of the boundary of the square object

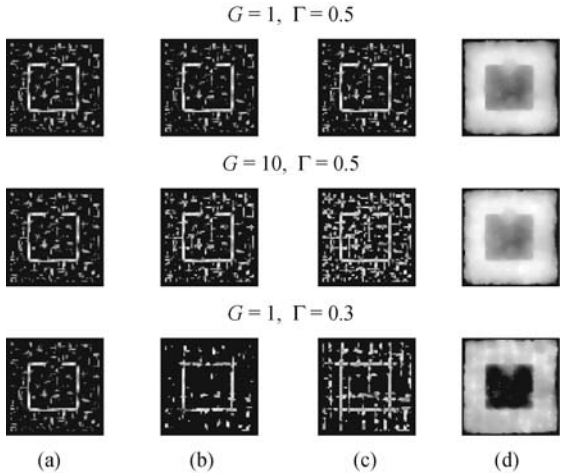

Fig. 10 Processing results with original formulas for the cell dynamics but different parameter configurations of the FACADE model; the input test image is the same as that shown in Fig. 8. (a) Output of the BCS without feedback from the CC loop; (b) after the first iteration of feedback; (c) after three feedback iterations; (d) the final filling-in output of the FCS. From top to bottom: different configurations of the two parameters: feedback gain G and threshold Γ for the bipolar cells. This figure indicates the bipolar dilemma: either a weak effect of the boundary contour completion, where the broken boundary is not perfectly padded and fixed when G is smaller and Γ is larger; or on the contrary, the many fake boundaries resulting from the bi-polarized cooperation when G is larger and Γ is smaller

FCS that controls filling-in, will likewise affect the final outputs of the FACADE processing as is to be expected.

The issue of robustness of the FACADE model, based on what is implied in Fig. 10, is caused by the overwhelming effect of the competition, or on the other hand, as a consequence of the excessive activities of cooperation within the CC loop, and thus the output of the closed-loop dynamic system of BCS could be unstable or even divergent. For example, if the competition activities are configured more weakly, while the cooperative activities are strong, the push-pull mechanisms could behave in an ineffective de-noising capability or worse still, the noisy signals could cooperate to generate many fake boundaries throughout the feedback pathways (as illustrated at the bottom of Fig. 10). On the other hand, if the competitive activities are too strong, the boundary contour completion activity could be inhibited or even blocked such that we obtain non-closed object boundaries causing the FCS filling-in process to overflow and fail. Figure 10 illustrates one of the cases of the parameter choice dilemma, referred to as the bipolar dilemma. Even if we could balance the activities of competition and cooperation for a specific situation, it is not actually easy, and may even be impossible to determine a rather general-purpose parameter configuration to cope with the large variety of input images, thereby somewhat restricting the potentially mighty FACADE framework, the design of which did not consider robustness, in dealing with real-world imagery.

After much effort in experimenting and speculating on the push-pull cell dynamics within the FACADE framework, we identified the fundamental causes, which could inevitably lead to the dilemma of balancing the activities of competition and cooperation. These center on three main issues: 1) the over-balanced push-pull dynamics causing a struggle between the end cutting and cooperative effects; 2) the inefficient noise (or disorganized image detail) reductions at the second competition stage of the hyper-complex cells, and 3) the improper cooperation manner of the bipolar cells.

For the first and the most critical issue, as illustrated in Fig. 11, it is easy to reason how the push-pull dynamics causes the inefficiency and difficulty for the cooperative feedback signals to fill the gap of the broken boundary and realize reasonable cooperative boundary completion. The end cutting effect, as mentioned before as a direct consequence of the competitive push-pull dynamics of the hyper-complex cells, is the very one to inhibit or block the boundary cooperation and thus the gap padding. Thus, around a possible gap in an object boundary, as show in Fig. 11, two contradicting yet balanced forces, namely, the end cutting dynamics of a hyper-complex cell to inhibit and break up a weaker boundary contour, and the cooperative effect of the bipolar cells managing to fill-in the broken gap, struggle to form the global push-pull equilibrium that makes the BCS dynamics unstable and uncertain. Moreover, it is not easy to decide which force to control to break the tie, because both the end cutting and gap filling effects, if not carefully controlled, tend to propagate to generate fake-boundary signals (as shown in Fig. 10), thus resulting in robustness issues and the bipolar dilemma.

As the second issue to overcome, Fig. 12 illustrates the

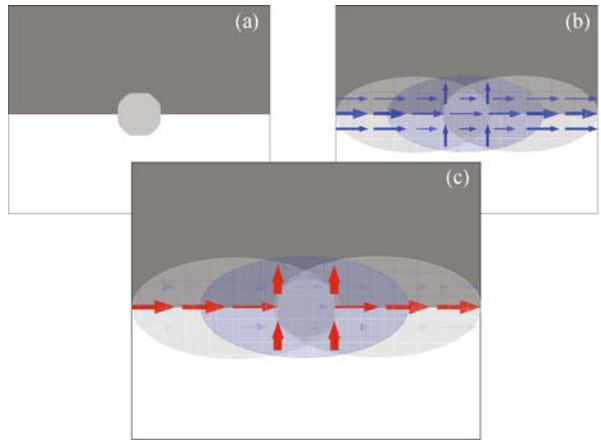

Fig. 11 The over-balanced push-pull dynamics in the process of boundary completion. In this illustrated situation the end cutting effect inhibits the cooperative boundary completion. (a) A broken boundary to be completed; (b) the output of the hyper-complex cell, incorporating some cooperative feedback signals; (c) the still broken boundary contour output of the BCS, if the end cutting effect dominates the push-pull dynamics where the cooperative feedback is not strong enough

inefficiency of noise reduction by the orientational competition. As was predicated, signals induced by random noise or the disorganized image details (which we would prefer to neglect), are generally non-orientationally-dominated and could be filtered out by the BCS in the second competition stage of the hyper-complex cell. Nevertheless, this is not exactly what happened in the FACADE processing, since the noise induced output of the simple cells, as illustrated on the left side of Fig. 12, are not really totally disorganized. These spatially rather patterned noise signals could not be completely eliminated by the orientational competition modeled by Eq. (20), because within one single hyper-complex cell the noise induced signals are still orientational-dominated. Moreover, the residual noisy signals have a further opportunity of being amplified by the push-pull dynamics when interacting with other structured signals, e.g., the end cutting signals and cooperative signals, which make the not so controlled BCS dynamics even more unstable and unpredictable.

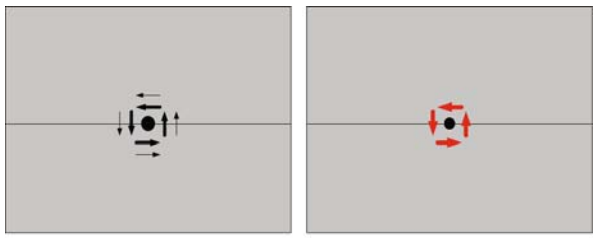

Fig. 12 Inefficient noise reduction of the hyper-complex cell. (a) Spatial distribution of the noise induced boundary contour as the output of the simple cells; (b) residual noise signals after the spatial and orientational competitions

The third issue involves the manner of cooperation of the bipolar cells, which is rather easy to explain. One may recall that the term N_{ijkd} in Eq. (22) was defined to ensure that the ends of lines strongly activate bipolar cells. Moreover, the manner (as indicated by Eq. (22)) in which a bipolar cell collects the activities from its two lobes is in a summing up fashion rather than in a logical AND fashion. These properties imply that a bipolar cell could easily be activated by an unilateral input, which may not be reasonable and may even be harmful to the stability of the overall BCS dynamics. Grossberg perhaps adopted the mechanism of line end activation because he was more concerned with the cooperation effect at the corner of a boundary contour; however, a harmful side effect called the line end extension illusion could be triggered and cause considerable damage to the stability of processing.

The dilemma of the bipolar cell in such an operational manner is that if a lower activating threshold Γ of the bipolar cells is chosen, the cell could be activated by a unilateral input, and cause the illusive line end extension effect (which could probably be further amplified); whereas if the threshold is too high, we do not obtain adequate cooperation output from the cell to fix the boundary gap (the weak feedback will be inhibited by the end cutting dynamics).

The easiest way to overcome the bipolar dilemma is to remodel the behavior of the bipolar cell to feature full-bilateral-activation, so that we can make use of cooperative feedback at the highest level without worrying about the line extension effect. In addition, this will overcome the issue of the over-balanced push-pull dynamics addressed before, because we can re-balance the activities of competition and cooperation freely now that the cooperative activities are well controlled. Commonly, we would like to configure a stronger cooperative feedback to facilitate the boundary contour completion without affecting the overall cells dynamics, for most image enhancing tasks. So far, the main issues of robustness and inefficiency have been ascertained, and the basic ideas of how to overcome them have been prepared. It is now time to remodel the corresponding cell dynamics within FACADE to incorporate these efforts, thereby focusing more on system stability and parameter robustness. However, there are enormous possibilities for remodeling the cell dynamics, and what is presented and described in the next section, is a rather compromised and minor revision of the original FACADE modality, to ensure moderate computational efficiency while keeping an eye on overcoming the robustness issues. This modality, despite not being at all complicated, has emerged after a lengthy iterative process of re-designing, testing, and comparing.

4 Remodelled dynamics of competition and cooperation

Our proposed approach only remodels the competition domain of the hyper-complex cell, the operational manner of the bipolar cell, and the entry point of the feedback signals, to stabilize the BCS dynamics and rebalance the local and global push-pull interactions. The modifications have been done in the same manner in both the feed-forward and feedback pathways of the BCS. All the modifications presented below were implemented and tested with our typical test image (Fig. 8) and a large variety of real-world images for stability and robustness issues.

4.1 Anisotropic spatial competition

Our first revision concerns the competition domain of the hyper-complex cell. In the original formulas, as indicated in Eqs. (14) and (15), one can deduce that the spatial competition of the hyper-complex cell occurs among

the neighboring hyper-complex cells (hence, the competition being called a "short-range" competition). A hyper-complex cell interacts with its neighboring hyper-complex cells in an isotropic circular range, which we call the competition domain of a hyper-complex. Here, we change the competition domain from isotropic (or circular) to anisotropic, as shown in Fig. 13, to re-balance the competition maneuvering along the boundary contour. The benefits from this modification are given as follows: 1) strengthening the boundary contour at the line end or the boundary corner directly, to reduce the reliance on cooperative feedback (such feedback signals require the unilateral activation of the bipolar cell, which we declared in the last section to be harmful) to close the boundary contour around the corner; 2) weakening the end cutting effect that causes the deficiency of the boundary completion. The remodeled dynamics of the spatial competition of the hyper-complex cell are

$$\frac{\mathrm{d}y_{ijkd}}{\mathrm{d}t} = -\alpha_4 y_{ijkd} + (U_4 - y_{ijkd})C_4 - (y_{ijkd} + L_4)E_4 + T, \tag{1}$$

$$C_4 = \sum_{p,q} \Phi_{pq} \left(C_{i+p,j+q,k,d} + g(O_{i+p,j+p,k,d})\right), \tag{2}$$

$$\Phi_{pq} = \frac{A}{2\pi\sigma_c^2} \exp\left[-\frac{1}{2}\left(\frac{(e_{s1}\bar{p})^2 + \bar{q}^2}{\sigma_c^2}\right)\right], \tag{3}$$

$$E_4 = \sum_{p,q} E_{pq} \left(C_{i+p,j+q,k,d} + g(O_{i+p,j+p,k,d})\right), \tag{4}$$

$$E_{pq} = \frac{B}{2\pi\sigma_s^2} \exp\left[-\frac{1}{2}\left(\frac{\bar{p}^2 + (e_{s2}\bar{q})^2}{\sigma_s^2}\right)\right], \tag{5}$$

where the $\bar{p}, \bar{q}$ are the indices of those neighbouring hyper-complex cells within the competitive field, which could be obtain by means of rotation as

$$\begin{bmatrix} \bar{p} \\ \bar{q} \end{bmatrix} = \begin{bmatrix} \cos(k\pi/K) & -\sin(k\pi/K) \\ \sin(k\pi/K) & \cos(k\pi/K) \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}. \tag{6}$$

And the output is

$$Y_{ijkd} = \frac{\left[\sum_{p,q} (U_4\Phi_{pq} - L_4 E_{pq})C_{ijpq} + T\right]^+}{\alpha_4 + \sum_{p,q} (\Phi_{pq} + E_{pq})C_{ijpq}}, \tag{7}$$

where

$$C_{ijpq} = \left(C_{i+p,j+q,k,d} + g(O_{i+p,j+p,k,d})\right). \tag{8}$$

Equations (1)–(5) restrict the domain of the spatial competition inside an orientational ellipse, whose shape and orientation are controlled by parameters e_{s1} and e_{s2}. Please note that in the new formulas, we changed the entry point of the feedback signals so that they are mixed to the output of the complex cell as the overall input of the hyper-complex cell. This is a minor yet critical revision to stabilize the dynamics of the BCS.

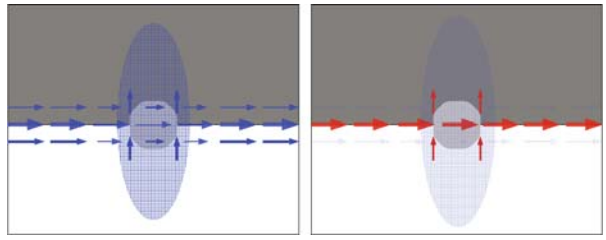

Fig. 13 The remodeled spatial competition in the anisotropic fashion. The vertical elongated ellipse illustrates the competition domain of a hyper-complex cell corresponding to the horizontal orientation. This anisotropic competition mechanism facilitates boundary completion

4.2 Diffusive orientational competition

The second revision, concerning the orientational competition, was made to overcome the inefficiency of noise reduction. This revision again changes the competition domain of the hyper-complex cell in a spatially diffusive fashion, but this time in the second stage of competition, making the orientational competition occur among spatially neighboring cells, and not only those hyper-complex cells corresponding to the same spatial position. The revised formulas for the orientational competition are written as

$$N_{ijkd} = \frac{\left[\sum_r (U_5 C_{kr} - L_5 E_{kr})\bar{Y}_{ijrd} - \sum_{e \geqslant d} F(\nu_{ijke})\right]^+}{\alpha_5 + \sum_r (C_{kr} + E_{kr})\bar{Y}_{ijrd}}, \tag{9}$$

$$\bar{Y}_{ijkd} = \sum_{p,q} C'_{pq} Y_{i+p,j+q,k,d}, \tag{10}$$

$$C'_{pq} = \frac{1}{2\pi\sigma_d^2} \exp\left[-\frac{1}{2}\left(\frac{\bar{p}^2 + (e_d\bar{q})^2}{\sigma_d^2}\right)\right]. \tag{11}$$

The definitions of $\bar{p}$ and $\bar{q}$ in the above formulas are the same as in Eq. (6).

As illustrated in Fig. 14, it is easy to understand why this new orientational competition mechanism performs better with respect to noise reduction. As discussed before, the noise induced signals are orientation-dominated in a single hyper-complex cell so that they cannot be completely eliminated. However, for a group of complex cells within the diffusive competition domain the noisy signal patterns as a whole are no longer orientation-

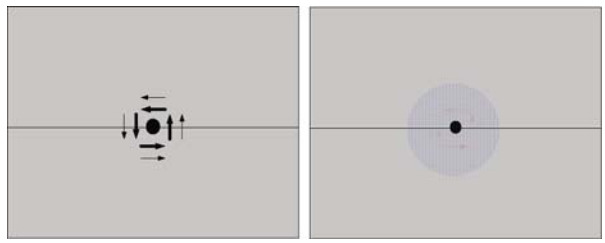

Fig. 14 Remodelled orientational competition, which facilitate the noise reduction

dominated, and thus they can be filtered out completely in the new mechanism.

4.3 Bilateral-activated bipolar cell

This remodeling has already been discussed at the end of the last chapter. The operational manner of a bipolar cell in collecting the activities from its two lobes is changed here from one that is summing up to a more selective logical AND. This revision ensures that the bipolar cell is not activated by the unilateral activities, thereby avoiding the harmful illusive line end extension effect. The drawback of this revision is that we no longer obtain competitive feedback at the end of the line or the boundary corner to help the process of boundary completion around the corner. However, introducing the new anisotropic spatial competition mechanism, as just argued, reduces the need for the feedback induced corner completion process, so that the unilateral activation is really not needed any more. The bipolar cell is formulated as

$$Z_{ijkd} = \sqrt{h\left(g(\sum_s A_s^b)\right) h\left(g(\sum_s A_s^b)\right)}, \qquad (12)$$

in which $g(x) = \frac{[x]^+}{D+[x]^+}$ and $h(x) = [x - \Gamma]^+$.

The formulas given above are the main remodeling approaches that we incorporated in the FACADE framework. These approaches were applied to stabilize the BCS dynamics, overcome the robustness issue, and make it more efficient in terms of noise reduction. Incorporating these modifications mainly in the BCS module, FACADE as an overall dynamical system is now stable and robust, since there is no other feedback pathway except the CC loop within the BCS. It is also desirable to apply some other modality reorganizations to the scale processing and the binocular fusion mechanisms, to improve the overall efficiency and suitability for computer implementation, but these approaches are beyond the scope of this paper.

Normally, in this improved modality of FACADE processing, we would like to select a low threshold Γ and a higher feedback gain G to facilitate boundary contour completion, and accelerate the dynamic process of the system, so that we obtain stable output in fewer computational iterations of the CC loop, which was impossible in the original FACADE modality. The same test image was processed with our improved FACADE engine to illustrate the robustness and efficiency (in terms of both noise reduction and boundary contour completion) of this new modality (Fig. 15). Comparing Figs. 10 and 15, it is easy to conclude that the reformulated FACADE processing is far superior to the original one in almost all aspects, especially robustness.

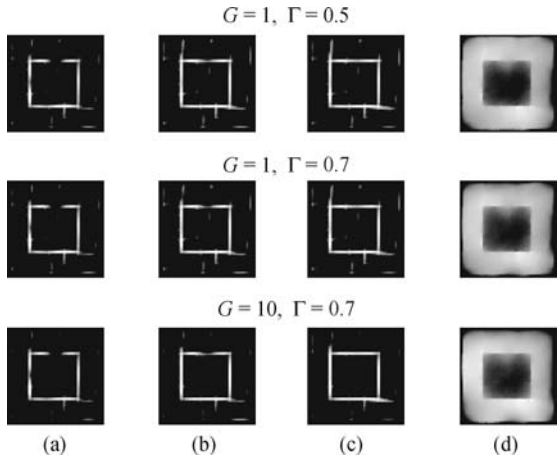

Fig. 15 Processing results of the same test image as in Fig. 8, using our reformulated FACADE engine. The image arrangement and processing conditions are the same as those illustrated in Fig. 10. In all parameter configurations, the boundary completion of the square object was achieved after three iterations of loop back computation without inducing fake boundary signals, while at the same time the strong noise was suppressed efficiently. What is most important though, and gives us the greatest satisfaction, is that the outputs are stable and coherent, indicating that the reformulated approach is seemingly robust enough to be applied in more situations

Speaking of computational efficiency of the remodeled FACADE, only the new mechanism of orientational competition may consume more computational resources than the original one, since we have increased the size of the computation domain. However, the overall efficiency is improved due to the rapid convergence dynamics achieved by increasing the feedback gain, as just mentioned. In the original modality we had to be very careful to control the cooperative feedback, to prevent the system dynamics from becoming unstable or divergent, as previously illustrated in Fig. 10.

So far, the FACADE framework has been remodeled to be more efficient and robust in the overall sense. We are now ready to apply and test this reconfigured image processing engine on our real-world imagery processing tasks. A computer system based on the new approach was constructed and various types of real-world imagery (satellite remote sensing images, aerial photography images, roadside scenery images acquired by our autonomous vehicle, and photographs of daily life) were tested to evaluate further the reformulated modality. In the next section, some representative results of our real-world imagery processing (enhancing) are addressed to highlight the features and capabilities of this unique and mighty HVS featured FACADE framework, with our robust enhanced new approaches.

5 Real-world imagery processing

The first task for the FACADE engine was to enhance and generate automatic segmentations from satellite remote sensing images. Satellite remote sensing images are

non-binocular, and thus only the monocular BCS and FCS within the FACADE framework were enabled. The image was first resized to 256×256 pixels. Three different scale channels and 12 simple cell orientations were configured to carry out the processing using our reformulated FACADE framework. Considering the typical requirement of retaining the spectral characteristics for further manipulations, we discriminated the LGN preprocessing for boundary contour generation in the BCS and the FCS filling-in of signals with different configurations.

A typical segmentation result of the FACADE processing of satellite remote sensing images is illustrated in Fig. 16. Hereafter for convenience, in any image processing case study, we only present the four most relevant images, which are arranged in a entire image-mosaic, to map the input image and those crucial outputs (or activities) of different layers of the cells within the FACADE framework.

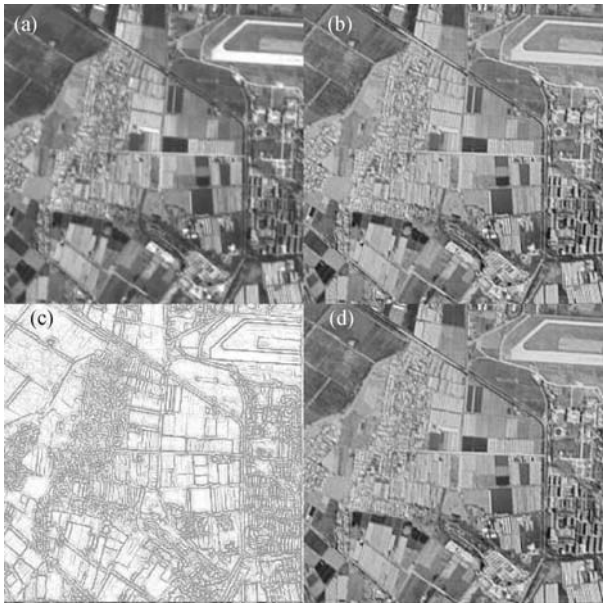

Fig. 16 Small-scaled FACADE processing a satellite remote sensing image. (a) Original image; (b) output of LGN layer, where the consequence of non-uniform illumination is somehow normalized; (c) boundary contour output of the BCS; this mapping image is generated by summing the different orientational stabilized activities of the super-complex cells (the mapping is inverted for the sake of clarity); (d) image segmentation as a result of the filling-in in the FCS

In Fig. 16 it is easy to assert that the structured boundary contours, either significant or not so, were all clearly extracted as the output of the BCS. This distinct boundary web facilitates the FCS filling-in to generate regional segmentation. The ability to extract rather weak boundaries, while avoiding the side effect of generating many noise (or disorganized image details) induced fake ones (similar to the HVS behavior), is one of the greatest advantages of FACADE in manipulating real-world imagery.

For quality degraded aerial photography imagery acquired in bad atmospheric conditions, the multi-scale scheme of FACADE processing exhibits its selective enhancing capability as shown in Fig. 17.

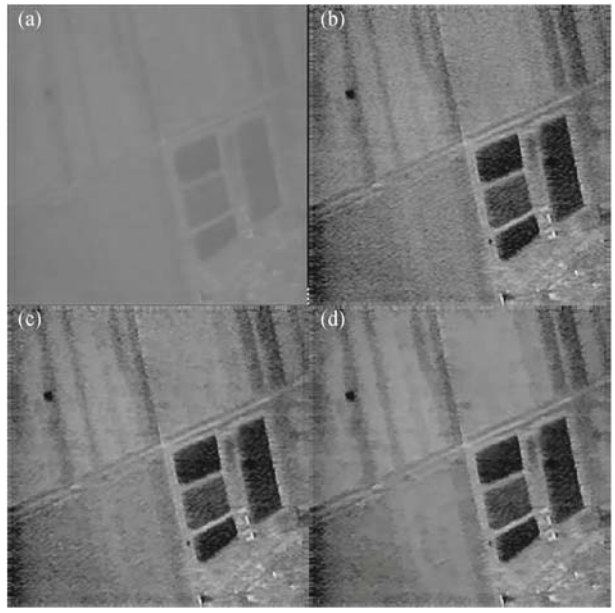

Fig. 17 (a) Blurry aerial photography image acquired under bad atmospheric condition; (b) the FACADE processing result in the small scale channel; (c) the output within the middle scale channel; (d) the large scale channel output

It can be seen in Fig. 17 that the small scale channel of FACADE detected and enhanced the blurred texture patterns, whereas the larger scale channels smoothed the noise and textured details to facilitate the generation of morphological segmentations. In this case, the integrated and multi-scaled FACADE processing captured both the structural information (the region boundaries) and the image details within the blurred image.

Beside the more structured remote sensing imagery as presented above, roadside scenery images acquired from experiments with our autonomous land vehicle (ALV), were also processed and tested with our FACADE-based computer system. The FACADE approach obtained far better results than our original image processing engine for the ALV system, in terms of extracting the important structure of road edges from the sophisticated background under various circumstances and conditions. Unfortunately, so far, it has still been impossible for us to apply the superior FACADE engine to a real-time system for the ALV due to the enormous computations of the cell dynamics in the FACADE framework. Figure 18 illustrates typical processing results for the ALV acquired roadside scenery images, where the disorganized image details (e.g., shadows of trees) were suppressed, while the disrupted structures (i.e., the road edges) were patched, enhanced, and extracted, and the figure of a man under the trees emerged more clearly.

To illustrate the ability of the suggested FACADE approach in generating binocular 3D segmentations, here

Fig. 18 FACADE processing of roadside scenery images. (a) Input image; (b) output of the LGN; (c) summed BCS output; (d) output after the FCS filling-in

we present a sample result without further explanation. In Figs. 19 and 20, it can be seen that FACADE can generate correct depth segmentation [33,34].

Fig. 19 A real-world stereo images pair, as the input of FACADE processing

6 Discussion

As analyzed, illustrated, and tested, the approach presented in this paper enhanced the robustness and improved the efficiency of FACADE processing, and exploited more of the potential of the mighty FACADE framework, so that we are now not so limited in making use thereof to explain how perceptions occur in the HVS merely with some illusive artificial images, as was the case when the FACADE theory was first put forward. In fact, our computer implementation of the remodeled FACADE approach, as a general-purpose image processing engine that features the great advantages of the HVS, is now robust and flexible enough to handle real-world imagery and able to obtain robust perceptions from a variety of natural situations.

The superiority of the FACADE-based approach as an image processing paradigm, in our opinion, is the

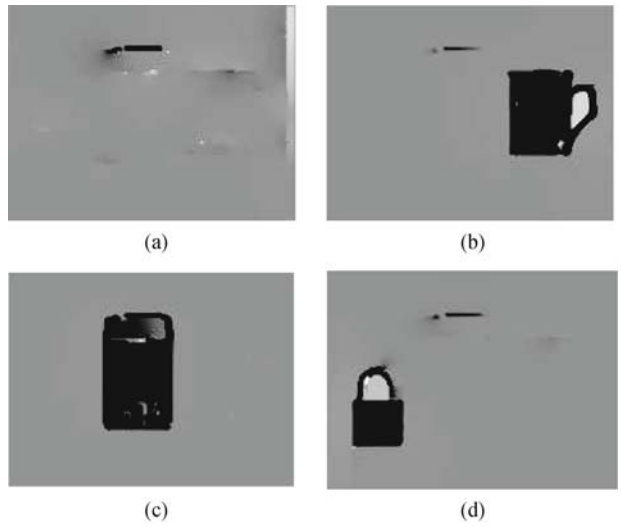

Fig. 20 Stabilized cells outputs within different depth layer, which represent a reasonable 3D depth segmentation

simultaneous ability to extract the distinctly localized boundary web of the structured objects within the input image, and to suppress noise and the non-relevant image details to generate stable and perceptively reasonable image segmentations. This capability is achieved by the dynamic interactions of the different cells within FACADE. The interactions of competition and cooperation were combined perfectly as the primary push-pull mechanisms within the uniformed framework to enable self-organized 2D and 3D image segmentations featuring the perceptive characteristics of the HVS. As far as we know, such an image processing engine is far superior to the traditional approaches, and that is the reason why we would like to make a great effort to reorganize and optimize this HVS framework, and to adopt it for real-world imagery manipulation. The main drawback of the FACADE-based approach is in that it can involve an enormous amount of computation to simulate the complicated cells dynamics. The computational inefficiency makes it difficult for a non-parallel computer to carry out the processing in real-time, thereby restricting the range of applications for the FACADE framework.

Although the main issues of robustness and inefficiency of the original FACADE modality were investigated and resolved in this paper, there is still something to desire. As mentioned in the introduction, another important and superior feature of the HVS, namely, invariability in object recognition, has still not obviously been embodied in the FACADE approach. Integrating both features of invariability and self-organization (as the main characteristics of the approaches by Grossberg et al. [11,19–22], including FACADE) of the HVS, is an appealing and ambitious idea for our ever pursuant goal of achieving a full HVS-featured computer vision system for the task of real-world image automatic segmentation

or object recognition.

As we already know, there were many other HVS approaches, such as HMAX [3–6] and VisNet [7–10], either featuring a transform invariant feature representation or specializing in invariant object recognition. Although the HMAX framework by Poggio et al. is a well-established HVS system to obtain such invariable object recognition, it is actually a rather simple feed-forward only infrastructure. This simplicity makes the HMAX approach easy to implement and light-weight in terms of computation; at the same time, it does not take too much advantage of the substantial neurodynamics in the HVS. In this sense, FACADE actually includes more neurodynamical mechanisms, including the feed-forward hierarchy, the within layer cell interactions (the competition and cooperation), and the feedback pathways. Thus, theoretically the FACADE framework should be more versatile and flexible in implementing an invariant feature representation than the rather simpler HMAX. It would therefore, be very interesting to incorporate some HMAX or VisNet featured mechanisms into the FACADE framework for further reinforcement. Our research group is now attempting to integrate these two approaches (HMAX and FACADE) to implement a more general-purpose and full HVS-feature computer vision engine for unrestricted real-world imagery recognition.

Appendix A Related formulas of FACADE model

The following formulas describe the most important cell dynamics within the FACADE model. These formulas were developed by Grossberg et al. and only those related to this paper are addressed; some have been adapted or combined in a compact form for simplification. Readers who are interested in the unabridged model descriptions or need more detailed explanation of the notations in the formulas below should refer to [17–22].

A-A Hyper-complex cells: spatial competition

The dynamics of the spatial competition of the hyper-complex cells were defined by

$$\frac{\mathrm{d}y_{ijkd}}{\mathrm{d}t} = -\alpha_4 y_{ijkd} + (U_4 - y_{ijkd})C_4 - (y_{ijkd} + L_4)E_4 + T + g(O_{ijkd}), \quad (13)$$

where

$$C_4 = \sum_{h,\nu} \Phi_{h\nu} C_{i+h,j+\nu,k,d}, \qquad \Phi_{h\nu} = \frac{A}{2\pi\sigma_c^2} \exp\left[-\frac{h^2+\nu^2}{2\sigma_c^2}\right], \quad (14)$$

$$E_4 = \sum_{h,\nu} E_{h\nu} C_{i+h,j+\nu,k,d}, \qquad E_{h\nu} = \frac{B}{2\pi\sigma_s^2} \exp\left[-\frac{h^2+\nu^2}{2\sigma_s^2}\right]. \quad (15)$$

In Eq. (13), α_4 is the decay parameter, U_4 and L_4 are the upper and lower bounds of hyper-complex cell activity, T is the baseline or tonic activity of the cell that contributes to generate so-called "end cutting" effect in the next stage of orientational competition. $g(O_{ijkd})$ is the feedback activity from the SOCC loop. Usually, we take $g(x) = Gx$, where the gain factor of the feedback G is a scale constant.

Solving Eq. (13) at equilibrium and then rectifying the result yields the steady-state spatial competition output Y_{ijkd} of the hyper-complex cell, as the input of the next stage of orientational competition.

$$Y_{ijkd} = \frac{\left[\sum_{k,\nu} (U_4\Phi_{k\nu} - L_4 E_{k\nu})C_{h\nu} + T + g(O_{ijkd})\right]^+}{\alpha_4 + \sum_{k,\nu} (\Phi_{k\nu} + E_{k\nu})C_{h\nu}}, \qquad C_{h\nu} = C_{i+h,j+\nu,k,d}. \quad (16)$$

A-B Hyper-complex cells: orientational competition

The activities of the second stage of hyper-complex cells were modelled as follows:

$$\frac{\mathrm{d}n_{ijkd}}{\mathrm{d}t} = -\alpha_5 n_{ijkd} + (U_5 - n_{ijkd})C_5 - (n_{ijkd} + L_5)E_5, \quad (17)$$

where

$$C_5 = \sum_r C_{kr} Y_{ijkd}, \qquad C_{kr} = \frac{c}{2\pi\sigma_c^2} \exp\left[-\frac{1}{2}\left(\frac{\sin(\frac{r-k}{2K}\pi)}{\sigma_c^2}\right)^2\right], \quad (18)$$

$$E_5 = \sum_r E_{kr} Y_{ijkd}, \qquad E_{kr} = \frac{c}{2\pi\sigma_s^2} \exp\left[-\frac{1}{2}\left(\frac{\sin(\frac{r-k}{2K}\pi)}{\sigma_s^2}\right)^2\right], \quad (19)$$

where c, s, σ_c, and σ_s are scale constants; r and k denote orientation indices ($0 \leqslant r, k < K$). Solving Eq. (17) at equilibrium and then rectifying the result yields the overall steady-state hyper-complex cell output:

$$N_{ijkd} = \frac{\left[\sum_r (U_5 C_{kr} - L_5 E_{kr})Y_{ijrd} - \sum_{e \geqslant d} F(\nu_{ijke})\right]^+}{\alpha_5 + \sum_r (C_{kr} + E_{kr})Y_{ijrd}}. \quad (20)$$

A-C Cooperative bipolar cells

The neural dynamics of a bipolar cell was defined as

$$\frac{\mathrm{d}z_{ijkd}}{\mathrm{d}t} = -z_{ijkd} + h\left(g\left(\sum_s A_s^b\right) + g\left(\sum_s B_s^b\right) + N_{ijkd}\right), \quad (21)$$

and the steady-state outputs are

$$Z_{ijkd} = h\left(g\left(\sum_s A_s^b\right) + g\left(\sum_s B_s^b\right) + N_{ijkd}\right), \quad (22)$$

where $h(x) = [x-\Gamma]^+$ with threshold Γ; $g(x)$ is the normalizing function to bound each lobe's activity, $g(x) = \frac{[x]^+}{D+[x]^+}$; s denotes the scale index; A_s^b and B_s^b correspond to the summed inputs from each of the lobes of the bipolar cell, which is defined generically as

$$A^b = \sum_{p,q,r}\left(\left(N_{i+p,j+q,r,d} - N_{i+p,j+q,R,d}\right)[J_{pqkr}]^+\right), \quad (23)$$

$$B^b = \sum_{p,q,r}\left(\left(N_{i+p,j+q,r,d} - N_{i+p,j+q,R,d}\right)[-J_{pqkr}]^+\right), \quad (24)$$

where N_{ijkd} is the output of the hyper-complex cells; r and R are the orthogonal orientations, $R \perp r$; J_{pqkr} and $-J_{pqkr}$ define the lobes of the bipolar cell's receptive field.

A-D Competition in feedback pathway

The orientation competition in the feedback pathway is achieved by

$$\frac{\mathrm{d}u_{ijkd}}{\mathrm{d}t} = -\alpha_6 u_{ijkd} + (U_6 - u_{ijkd})C_6 - (u_{ijkd} + L_6)E_6, \quad (25)$$

where

$$C_6 = \sum_r C_{kr} Z_{ijkd}, \qquad C_{kr} = \frac{c}{2\pi\sigma_c^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_c^2}\right)^2\right], \quad (26)$$

$$E_6 = \sum_r E_{kr} Z_{ijkd}, \qquad E_{kr} = \frac{s}{2\pi\sigma_s^2}\exp\left[-\frac{1}{2}\left(\frac{r-k}{\sigma_s^2}\right)^2\right]. \quad (27)$$

The stable-state output is

$$U_{ijkd} = \frac{\left[\sum_r (U_6 C_{kr} - L_6 E_{kr}) Z_{ijrd}\right]^+}{\alpha_6 + \sum_r (C_{kr} + E_{kr}) Z_{ijkd}}. \quad (28)$$

For the feedback pathway spatial competition, we have

$$\frac{\mathrm{d}O_{ijkd}}{\mathrm{d}t} = -\alpha_7 O_{ijkd} + (U_7 - O_{ijkd})C_7 - (O_{ijkd} + L_7)E_7, \quad (29)$$

where

$$C_7 = \sum_r C_{kr} U_{i+h,j+\nu,k,d}, \qquad C_{kr} = A\exp\left[-\frac{1}{2}\left(\frac{h^2+\nu^2}{\sigma_c^2}\right)\right], \quad (30)$$

$$E_7 = \sum_r E_{kr} U_{i+h,j+\nu,k,d}, \qquad E_{h\nu} = B\exp\left[-\frac{1}{2}\left(\frac{h^2+\nu^2}{\sigma_s^2}\right)\right]. \quad (31)$$

Again, the stable-state output is

$$O_{ijkd} = \frac{\left[\sum_{(h,v)} (U_7 C_{hv} - L_7 E_{hv}) U_{i+h,j+v,k,d}\right]^+}{\alpha_7 + \sum_{(h,v)} (C_{hv} + E_{hv}) U_{i+h,j+v,k,d}}. \quad (32)$$

In Eq. (32), O_{ijkd} is now the final feedback signal to be intervened in the activity of the corresponding hyper-complex cells (Eq. (13)) at the first competitive stage of the feed-forward pathways, thus to form the global recycling close loop, namely the CC of SOCC loop, of the overall BCS processing.

A-E BCS/FCS interactive: filling-in

The FCS filling-in obeys the following equations:

$$\frac{\mathrm{d}F_{ijd}^+}{\mathrm{d}t} = -M^m F_{ijd}^+ + \sum_{(p,q)\in N}\left(F_{pqd}^+ - F_{ijd}^+\right)\Psi_{pqijd}^m + X_{ijd}^+, \quad (33)$$

$$\frac{\mathrm{d}F_{ijd}^-}{\mathrm{d}t} = -M^m F_{ijd}^- + \sum_{(p,q)\in N}\left(F_{pqd}^- - F_{ijd}^-\right)\Psi_{pqijd}^m + X_{ijd}^-, \quad (34)$$

where the features are diffusing to the their nearest neighbouring cells, namely $N = \{(i, j-1), (i, j+1), (i, j-1), (i, j+1)\}$, whereas the boundaries attenuate the diffusion permeability coefficients as

$$\Psi_{pqijd}^m = \frac{\delta}{\kappa + \varepsilon(\bar{N}_{pqd} + \bar{N}_{ijd})}, \quad \bar{N}_{ijd} = \sum_k N_{ijkd}. \quad (35)$$

In Eqs. (33) and (34), F_{ijd}^+ and F_{ijd}^- are respectively the filled-in activity and output signal at position ij and disparity d of the ON FIDO and OFF FIDO; M^m is the decay parameter; X_{ijd}^+ and X_{ijd}^- are the ON channel and the OFF channel output from the preprocessing stage; ε, κ and δ are the scale parameters of blocking and spreading of the the boundary squashing function Ψ_{pqijd}^m.

At equilibrium, each F_{ijd} was computed as the solution of a set of simultaneous equations:

$$F^+_{ijd} = \frac{X^+_{ijd} + \sum\limits_{(p,q)\in N} F^+_{pqd}\Psi^m_{pqijd}}{M^m + \sum\limits_{(p,q)\in N} \Psi^m_{pqijd}},$$
$$F^-_{ijd} = \frac{X^-_{ijd} + \sum\limits_{(p,q)\in N} F^-_{pqd}\Psi^m_{pqijd}}{M^m + \sum\limits_{(p,q)\in N} \Psi^m_{pqijd}}. \quad (36)$$

Acknowledgements This work was supported by the National Basic Research Program of China (No. 2007CB311001), the National Natural Science Foundation of China (Grant Nos. 60835005, 60736018 and 90820304), Ministry of Education of China (No. NCET-08-0147), Hunan Provincial Innovation Team Project.

References

1. Gabor D. Theory of communication. Journal of the Institution of Electrical Engineers, 1946, 93: 429–459
2. Jones J P, Palmer L A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 1987, 58(6): 1233–1258
3. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience, 1999, 2(11): 1019–1025
4. Serre T, Riesenhuber M. Realistic modelling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. Technical Report CBCLPaper 239/AIMemo 2004-017. Cambridge: Massachusetts Institute of Technology, 2004
5. Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio T. A theory of object recognition: computations and circuits in the feed-forward path of the ventral stream in primate visual cortex. Technical Report AI Memo 2005-036/CBCL Memo 259. Cambridge: Massachusetts Institute of Technology, 2005
6. Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2005, 2: 994–1000
7. Wallis G, Rolls E T. A model of invariant object recognition in the visual system. Progress in Neurobiology, 1997, 51(2): 167–194
8. Deco G, Rolls E T. A neurodynamical cortical model of visual attention and invariant object recognition. Vision Research, 2004, 44(6): 621–642
9. Rolls E T, Stringer S M. Invariant visual object recognition: a model, with lighting invariance. Journal of Physiology, 2006, 100(1-3): 43–62
10. Stringer S M, Perry G, Rolls E T, Proske J H. Learning invariant object recognition in the visual system with continuous transformations. Biological Cybernetics, 2006, 94(2): 128–142
11. Grossberg S. 3-D vision and figure-ground separation by visual cortex. Perception and Psychophysics, 1994, 55(1): 48–120
12. Grossberg S. Linking the laminar circuits of visual cortex to visual perception: development, grouping, and attention. Neuroscience and Biobehavioral Reviews, 2001, 25(6): 513–526
13. Cao Y, Grossberg S. A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. Spatial Vision, 2005, 18(5): 515–578
14. Huang T R, Grossberg S. Cortical dynamics of contextually cued attentive visual learning and search: spatial and object evidence accumulation. Psychological Review, 2010, 117(4): 1080–1112
15. Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 411–426
16. Rolls E T, Milward T. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. Neural Computation, 2000, 12(11): 2547–2572
17. Grossberg S. The quantized geometry of visual space: the coherent computation of depth, form, and lightness. The Behavioral and Brain Sciences, 1983, 6: 625–692
18. Grossberg S. Outline of a theory of brightness, color, and form perception. In: Degreef E, van Buggenhaut J, eds. Trends in Mathematical Psychology. Amsterdam: North-Holland,1984: 59–85
19. Cohen M A, Grossberg S. Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. Perception and Psychophysics, 1984, 36(5): 428–456
20. Fazl A, Grossberg S, Mingolla E. View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. Cognitive Psychology, 2009, 58(1): 1–48
21. Grossberg S, Mcloughlin N. Cortical dynamics of 3-D surface perception: binocular and half-occluded scenic images. Neural Networks, 1997, 10(9): 1583–1605
22. Grossberg S. Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. Psychological Review, 1997, 104(3): 618–658
23. Raizada R, Grossberg S. Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. Cerebral Cortex, 2003, 13(1): 100–113
24. Grossberg S. How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. Behavioral and Cognitive Neuroscience Reviews, 2003, 2(1): 47–76
25. Grossberg S, Howe P D I. A laminar cortical model of stereopsis and three-dimensional surface perception. Vision Research, 2003, 43(7): 801–829
26. Grossberg S, Swaminathan G. A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention and bistability. Vision Research, 2004, 44(11): 1147–1187
27. Grossberg S, Yazdanbakhsh A. Laminar cortical dynamics of 3D surface perception: stratification, transparency, and neon color spreading. Vision Research, 2005, 45(13): 1725–1743
28. Grossberg S, Yazdanbakhsh A, Cao Y, Swaminathan G.

How does binocular rivalry emerge from cortical mechanisms of 3-D vision? Vision Research, 2008, 48(21): 2232–2250

29. Grossberg S, Kuhlmann L, Mingolla E. A neural model of 3D shape-from-texture: multiple-scale filtering, boundary grouping, and surface filling-in. Vision Research, 2007, 47(5): 634–672
30. Grossberg S, Mingolla E, Williamson J. Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. Neural Networks, 1995, 8(7-8): 1005–1028
31. Mingolla E, Ross W, Grossberg S. A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. Neural Networks, 1999, 12(3): 499–511
32. Peng X H, Zhou Z T, Wang Z Z. Acquiring disparity distribution from stereo images with monocular cues based on cell cooperation and competition. Acta Electronica Sinica, 2000, 28(8): 1–4 (in Chinese)
33. Song B Q, Zhou Z T, Hu D W, Wang Z Z. A new computational model of biological vision for stereopsis. Lecture Notes in Computer Science, 2004, 3174: 525–530
34. Chen Y. Research on attention in visual neural system. Dissertation for Master's Degree. Changsha: National University of Defense Technology, 2009 (in Chinese)

Dewen HU was born in Hunan, China, in 1963. He received the BSc and MSc degrees from Xi'an Jiaotong University, China, in 1983 and 1986, respectively. From 1986, he was with the National University of Defense Technology. From October 1995 to October 1996, he was a Visiting Scholar with the University of Sheffield, UK. He got his Ph.D degree from the National University of Defense Technology in 1999. He was promoted Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science. He is an action editor of Neural Networks.

Zongtan ZHOU was born in Henan, China, in 1969. He received the BSc, MSc and Ph.D degrees from National University of Defense Technology, China, in 1990, 1994 and 1998, respectively. From February 2010 to February 2011, He was a Visiting Scholar with the Eberhard Karls Universitt Tübingen. He was promoted Professor in 2007. His research interests include image/signal processing, computer/biological vision, neural networks, cognitive neuroscience and brain-computer interface.

Zhengzhi WANG gained his Ph.D in Electrical Engineering in 1984 from Rice University, USA. Since then he has been Professor at National University of Defence Technology, China, working on Neurocomputing and Robotics, Computer Vision, Bioinformatics and Remote Sensing.