Künstl Intell (2015) 29:19–29
DOI 10.1007/s13218-014-0341-0

DISCUSSION

Beyond Simple and Complex Neurons: Towards Intermediate-level Representations of Shapes and Objects

Antonio Rodríguez-Sánchez · Heiko Neumann · Justus Piater

Published online: 17 December 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Knowledge of the brain has much advanced since the concept of the neuron doctrine developed by Ramón y Cajal (R Trim Histol Norm Patol 1:33–49, 1888). Over the last six decades a wide range of functionalities of neurons in the visual cortex have been identified. These neurons can be hierarchically organized into areas since neurons cluster according to structural properties and related function. The neurons in such areas can be characterized to a first order approximation by their (static) receptive field function, viz their filter characteristic implemented by their connection weights to neighboring cells. This paper aims to provide insights on the steps that computer models in our opinion must pursue in order to develop robust recognition mechanisms that mimic biological processing capabilities beyond the level of cells with classical simple and complex receptive field response properties. We stress the importance of intermediate-level representations to achieve higher-level object abstraction in the context of feature representations, and summarize two current approaches that we consider are advances toward achieving that goal.

Keywords Computer modeling · Intermediate visual processing · Boundary grouping · Shape representation · Feedback connections · junctions

A. Rodríguez-Sánchez (✉) · J. Piater
Institute of Computer Science, University of Innsbruck,
Technikerstr. 21a, Innsbruck, Austria
e-mail: antonio.rodriguez-sanchez@uibk.ac.at

H. Neumann
Institute of Neural Information Processing, Ulm University,
James-Franck-Ring, Ulm, Germany
e-mail: heiko.neumann@uni-ulm.de

1 Introduction

Object recognition is a very hard problem, the best proof of this is to consider that the first works started in the 1960s (Robert's and Guzmán's theses) and up to now there are thousands of papers published each year with solutions to achieve that goal. Due to the complexity of this problem, many scientists and engineers have resorted towards neurophysiology for solutions on how the human visual system solves such a difficult task with astonishing efficiency and accuracy. The earliest models inspired by the primate visual system [24, 25] appeared little after the influential works of Hubel and Wiesel [31, 32] that revealed key mechanisms of the functional architecture of the visual cortex. Most importantly, the response characteristics of individual cells to stimulus probes in different areas were identified, characterized as, e.g., simple, complex, and hypercomplex (or, endstopped). Inspired by these findings, Fukushima [21] introduced the *Neocognitron* which has influenced many models after it, such as, e.g., convolutional nets [38], the four-layer hierarchical *Visnet* model [75], and until the most recent models of a layered architecture of generic sum/MAX [44, 58, 67] operations. Some of these models also serve as reference architectures in computer vision applications.

The mammalian visual cortex is hierarchically organized into different areas, of which V1 and V2 are the largest. Cortical processing along a hierarchy of stages is segregated in two different pathways with connections between them: The occipitotemporal pathway (V1, V2, V4, PIT and AIT) is concerned with object recognition features [39, 69], while the occipitoparietal pathway (V1, V2, V3, MT and MST) is associated with spatiotemporal characteristics of the scene [6]. The functional architecture of the cortical visual system can be characterized by some key characteristics. One is that neurons become increasingly more selective to complex stimuli and less sensitive to stimulus variation, covering larger areas of the visual field. At the bottom of the hierarchy, neurons in V1 are selective for edges (among other features), and at the top, AIT neurons respond to complex objects with significant variation in their orientation, size, illumination and foreshortening.

The distributed nature of sensory processing in cortical architecture is well suited to allow integration of local responses to specific features with more global scene-related context information [22, 40, 79]. For example, the receptive field properties to a first approximation can be described as filters. The output responses of a stage of signal filtering are nonlinear, reflecting a transformation of cell membrane potentials into output firing rates [12]. Output responses are normalized through mutual competition of neural responses between cells in local pools in the spatial surround of target cells. As a consequence, responses show spatial frequency interactions, cross-orientation inhibition, and contrast saturation properties [13, 14]. Grouping mechanisms enforce the integration of local responses in feature space to form more abstract prototypical items of boundary fragments. Such operations utilize cortical mechanisms such as lateral interactions between like-oriented feature preferences, as in V1 layers II/III for orientation [7]. However, the extent and the timing properties of such lateral interactions cannot explain the rapidness of integration as observed in experiments. It has thus been suggested that feedback signals from higher stages in the neural processing hierarchy mainly contribute to such feature grouping [53, 62]. Although feedback is a prevalent feature of cortical processing, its functional role is still a major topic of current research investigation [41, 66]. The two dominant theories consider feedback as a gain enhancement mechanism that biases competition [68, 73] or as one implementing predictive coding where higher-order cortical areas carry predictions to lower-order areas [56] (see discussion in [72]).

The recent approaches to object recognition are mainly driven by the ”edge doctrine” of visual processing pioneered by Hubel and Wiesel’s work. Edges, as detected by responses of simple and complex cells, provide important information about the presence of shapes in visual scenes. Their detection, however, is only a first step at generating an interpretation of images that is invariant against certain variances in scene properties and layout as well as their geometric transformations. Consider, for example, Fig. 1. In (a) localized image features, like oriented contrasts, need to be integrated to form prototypical boundary fragments for shape segmentation. Localized higher-order features need to be inferred that determine shape identities [43] or to determine the motion characteristics of such shapes [11].

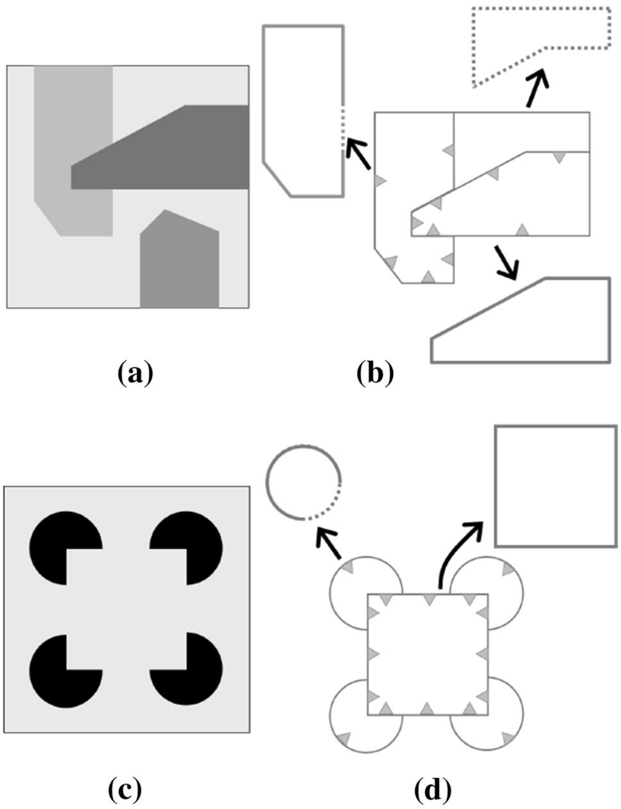

Fig. 1 Segmenting an image into figural parts and background (pattern in **a** is redrawn after [64]). Depending on the ownership assignment an image region is assigned to be a (foreground) figural shape (pattern **b**, *gray triangles* pointing towards the surface region that “owns” the boundary) or merely a part of the background (The region in the *upper right* is circumscribed completely by boundaries that are owned by different surface regions, displayed by a *dashed contour*). Therefore, the shape is occasional and not represented as independent shape. The wedge shaped region below is bounded by a contour that is completely owned by this surface patch (displayed by the continuous contour). The region on the upper left is partially occluded by the wedge-shape patch. A resulting boundary mainly consists of components owned by that region, while the occluding part (delimited by the T-junctions) is not owned. In this case, the “open ends” are suggested to be grouped and completed amodally to generate a most likely coherent shape. **c** The Kanizsa square shows similar arrangement with shape components composed of illusory contours and the segregation into occluding regions (pattern **d** with central square and four peripheral circular regions). Unlike the case in **a**, **b** the outline of the central occluder needs to be grouped by filling the illusory boundaries such that the ownership direction can be globally assigned in an unambiguous fashion (**d**)

However, edge integration is informative for the visual system only when the shape it bounds is assigned to be a figural component of the scene and not a mere occasion as part of the background. The assignment of border-ownership, or surface-belongingness [42] is therefore another key process operating in the above hierarchy (Fig. 1b). Taken together, the evidence about the functional organization of the mammalian visual system speaks in favor of the role of intermediate-level processes and representation. These establish interfaces linking spatially high resolution with object and more task related representations. Such intermediate processing stages that operate upon initial simple and complex cell outputs lead to the formation of neural representations of scene content that allow robust shape inference and object recognition.

The aim of this paper is to present some recent developments in the neural computational modeling of intermediate-level shape processing. We focus on the steps that computer models in our opinion must pursue in order to develop robust recognition mechanisms that mimic biological processing capabilities beyond the level of cells with classical simple and complex receptive field response properties. The goal is to provide the reader with some promising new directions of further investigation aiming to arrive at intermediate-level representations rich enough to provide the input to different behavioral tasks and analyses. In Sect. 2 we summarize two examples that build upon the original simple and complex cell models incorporating mathematical approximations to cells well known to exist in intermediate areas and which have been mostly neglected by many computer models in the past. In Sect. 3 we discuss and conclude on the aspects a computer model of the visual cortex should incorporate given recent evidence and findings.

2 Modeling Intermediate-level Processing Inspired by the Visual Cortex

Our focus here will be on works that include cells and mechanisms other than the classical simple and complex neurons utilized in mechanisms beyond initial stages of processing. In this section, we will briefly first present modeling efforts for V2 processing for feature extraction and boundary grouping and the functional consequences derived from this. We will then continue to present results of computational investigation for generating shape representations based on curvature-sensitive cell mechanisms in V4.

2.1 Boundary Grouping and Detection of Complex Localized Features: Modeling V1–V2 Interactions

Processing in V2 contributes to generating representations of boundaries as well as to localized contour features, or junctions. Why does the visual system bother to group oriented contrast to form extended boundaries? Why is it useful to complete fragmented contour segments and fill-in illusory contours? [52]. Consider the Kanizsa square pattern depicted in Fig. 1c: The square region in the center is defined by the long-range groupings of the like-oriented flanking contrast configurations generated by the pacmen patches. These are generated by the occlusion of peripheral circular patches through a central square with same luminance as the background. The appearance of such illusory contours depends upon the presence of symmetric bilateral input contrast to generate a neural boundary response filling the gap between them [29, 54].

Several neural computational models have been suggested to explain long-range grouping of contrasts and illusory contours [5, 26, 47, 77]. In [46, 48], it has been suggested a taxonomy of long-range grouping models which identifies the elements of how information is integrated from a space-feature domain representation to enhance the gain of target activations. In a nutshell, the spatial neighborhood is represented by spatial long-range filters (or weighting functions) and with the sub-fields combined non-linearly. The support from the feature domain (orientation in our case) is defined by a spatial relatability, or compatibility, measure that is evaluated based on geometric configurations of oriented contrast configured tangential to a smooth contour fragment [36, 49]. This relatability is represented as an independent weighting function that is combined with the spatial weighting to generate a 3D weight kernel in the $(\mathbf{x}, \theta)$-space with $\mathbf{x} = (\mathbf{x}, \mathbf{y})$. The computational mechanisms mainly utilize some weighting function (Λ) in which the feedforward input is matched against the expected configuration and generates a support measure:

$$support_{\mathbf{x},\theta} = input^{left}_{\mathbf{x},\theta} \circ input^{right}_{\mathbf{x},\theta} \tag{1}$$

with

$$input^{left/right}_{\mathbf{x},\theta} = \sum_{\mathbf{x}',\phi} r_{\mathbf{x}',\phi} \cdot relate_{\mathbf{x}\mathbf{x}',\theta\phi} \cdot \Lambda^{left/right}_{\mathbf{x}\mathbf{x}',\theta} \tag{2}$$

and $\circ$ to denote a proper combination of the subfield integration. In the model V2 grouping cell integration mechanism, subfields are combined multiplicatively. The relatabilty measure *relate* is defined by a geometric configuration of, e.g., co-circular arrangements of oriented contrasts (see examples and a discussion by [46]). The support measure yields an activation $z_{\mathbf{x},\theta} \propto support_{\mathbf{x},\theta}$ which, in turn, is used to combine it with the signal generated by bottom-up input. One such model is realized by a recurrent feedback mechanism that enhances those input features compatible with the activation calculated by the integration mechanism,

$$g_{\mathbf{x},\theta} \propto s^{FF}_{\mathbf{x},\theta} \cdot \left(1 + \lambda z^{FB}_{\mathbf{x},\theta}\right) \tag{3}$$

with g denoting the resulting gain enhanced response after the modulation of the driving input by top-down feedbacks driving feedforward input, and λ is a constant scaling factor (compare [47] and [71]). The feedback signal is meant to be delivered by V2 boundary integration cells, referred to as $z_{\mathbf{x},\theta}$ above. If feedback is missing the available feed-forward activity is left unchanged. In combination with a stage of activity normalization such feedback signals yield enhancement of activities that receive support from larger context information while those activities that have not received any feedback will be suppressed. The dynamic integration to yield an output activity r can be formally defined by

$$\begin{aligned}\dot{r}_{\mathbf{x},\theta} &= -\alpha r_{\mathbf{x},\theta} + g_{\mathbf{x},\theta} - \gamma r_{\mathbf{x},\theta} \cdot p_{\mathbf{x},\theta} \\ \dot{p}_{\mathbf{x},\theta} &= -p_{\mathbf{x},\theta} + \sum_{\mathbf{x}',\phi} r_{\mathbf{x}',\phi} \cdot \Lambda^{pool}_{\mathbf{x}\mathbf{x}',\theta\phi}\end{aligned} \tag{4}$$

where the activity normalization is accomplished by a pool of integrated activity from a neighborhood in the space-feature domain (represented by a separate state variable p); symbols α and γ are constants and $\Lambda^{pool}_{\bullet}$ denotes a weighting function for the pool integration. For an analysis of the computational properties of an even more elaborated version of such a circuit, refer to [9]. Activity normalization is realized by a mechanism of shunting inhibition in the rate equation for the output state r. The resulting effect of divisive inhibition leads to down-modulating the activities depending on context information. Such an effect has been dubbed biased competiton in attention selection [18, 57]. Related mechanisms of activity normalization have been successfully incorporated into vision algorithms for contrast detection [1, 27], keypoint detection [2, 76], and object recognition [34].

All such models mentioned above consider processing in the space-feature domain that arrives at a representation in which boundaries are enhanced and completed, while background clutter is much reduced. However, more computational intelligence is necessary to allow an intermediate-level interpretation of the layout of the visual scene components. While the illusory contour formation such as for stimuli in Fig. 1c takes time [59], the localized shape features (junctions of different configurations) undergo a reinterpretation over time.[1]

Several experimental investigations by Rubin [65] show compelling evidence that such junction features play an important role in figure-ground segregation and the interpretation of surface depth ordering. Figure 1d illustrates the tagging of surface boundaries according to their border ownership direction (indicating which side of the attached region is in front) such that surface regions can be reliably segregated. Also, occluded boundaries can be completed amodally (compare also Fig. 1b in the case of complex figural surface arrangements without illusory boundaries). Evidence suggests that such a representation of border-ownership is created at different stages along the visual hierarchy at areas V2, V4 and TE (the latter as part of the inferotemporal cortex, IT; [4, 64, 78]). In case of the Kanizsa figure presentation onset, the localized junctions of the pacmen segments are initially defined entirely by the localized corners of the figural contrast elements. However, when the illusory boundaries have been established via long-range grouping mechanisms, the L-junctions (corners) that define the lips of the pacmen mouths are reinterpreted as T-junctions in which the stem belongs to the circular boundary of the occluded region while the roof is generated by contrast as well as illusory segments of the occluding square region in front. Border-ownership assignment evaluates salient convex boundary features and utilizes T-junctions as additional evidence indicating surface occlusion directions.

Weidenbacher and Neumann [76] modeled the mechanisms of contour grouping and localized feature extraction incorporated in one coherent model. The reassignment of activation, or likelihood, for L- and T-junction configurations is demonstrated as a temporal process (Fig. 2): The dynamic interaction between different stages of contour and boundary representations in model areas V1 and V2 allow the creation of boundary representations, which, in turn, define context information for the interpretation of localized junction features. It also demonstrates how different filter mechanisms contribute to the extraction of key features and how the core processing principles, such as feedback, establish such representations robustly over time. Figure 3 shows how different responses by the different filtering mechanisms generate distributed representations of various boundary and junction features. While the boundary representations are made explicit, we make no explicit claim which of these localized features are explicitly represented in a specific feature map, such as corners of different opening angles [33]. We argue that the junction configurations can be reliably read out from the distributed responses of cells as in areas V2 and V3. The read out mechanism is also capable to assign transparent surface layout of occluded surface regions (compare [45]). While we acknowledge that the assignment of border ownership is reflected in a separate quality that tags boundary signal responses (see Discussion), we have generated qualitative depth segregation for overlapping surfaces that demonstrate the coherence of the assignment of the labeling of extended boundaries (see [71]).

In all, the results demonstrate that processing beyond initial simple/complex cell filtering is necessary to create boundary groupings and localized features together with

[1] Such reassignment of interpretation to meaningful figural surface layout coheres with different processing phases demonstrated in experiments by Roelfsema et al. [62]. While initial responses are mainly driven by image features, later response facilitations are generated as a consequence of selective feature enhancements generated during grouping and figure-ground segregation processes.

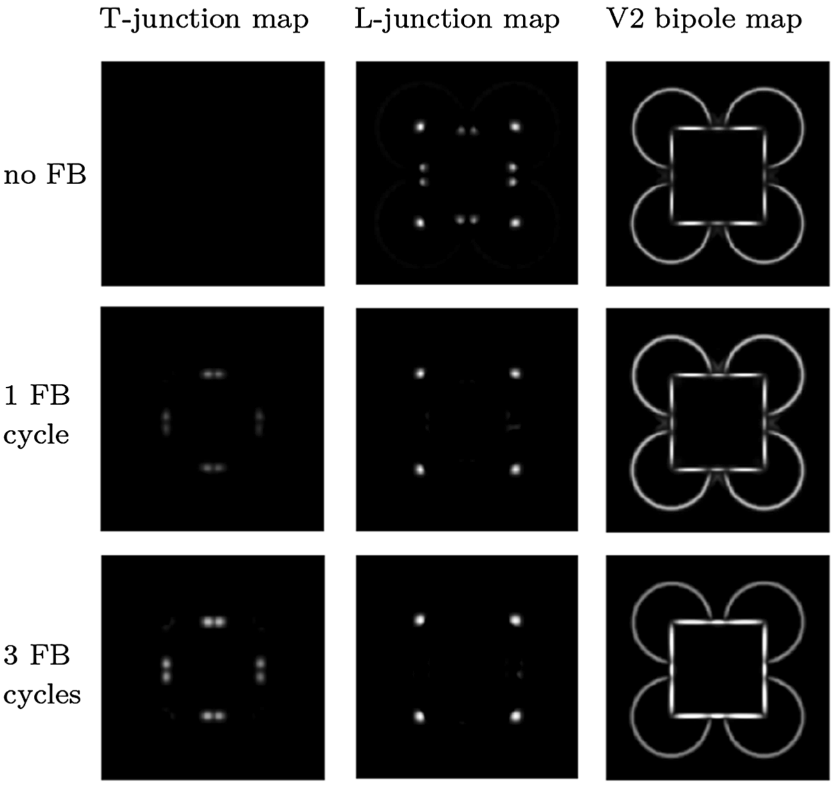

Fig. 2 Illusory boundary grouping and localized feature intepretation for the Kanizsa square pattern (Fig. 1). Bipole cell responses and T-/L-junction likelihoods from a readout of model cell activations in area V1 and V2 grouping. Initial signal responses are shown in the top row (no recurrent processing). Results of recurrent processing are shown after one cycle (*middle row*) and after three recurrent cycles (*bottom row*). Illusory boundaries are signalled by V2 bipole cell activities and are completed over time. A result of this grouping and completion process is that localized features are reinterpreted over time: L-junctions at the open mouths of the Pacmen are re-interpreted as T-junctions (corresponding to surface occlusions) once the illusory boundaries have been established in the neural representation. Results from [76]

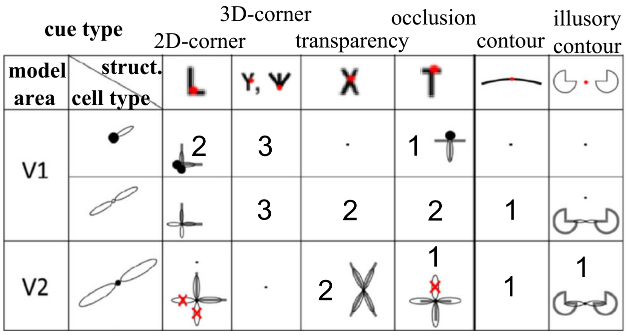

Fig. 3 Map of the distributed representation of boundary and feature configurations that uniquely define different structural configurations as a basis for various junction types. Small scale grouping and end-stopped cells in model area V1 and long-range grouping cells in model area V2 define the basic feature repository for relevant corner and junction types as well as boundary grouping computations. Map taken from [76]

proper scenic interpretations. Grouping mechanisms play an important role in filling illusory contour segments. The neural processing machinery can make use of such groupings as well as the representations generated to mark localized shape features such as corners or junctions. Below we show further modeling at such intermediate level range of processes that generates robust shape representations based on salient contour curvature features.

2.2 2DSIL: A Feedforward Model of Shape Representation Through Endstopped Neurons

Hubel and Wiesel [31] identified hypercomplex cells, which are currently referred to as endstopped. These were originally classified due to their response characteristics that depended on specific stimulus configurations that caused a target cell to fire. A large population of cells with such type of receptive field property has been identified in area V2. One such population of V2 neurons also respond to contours, both real and illusory [29], the latter of which requires bilateral input of flanking contrasts. Ito and Komatsu [33] and Boynton and Hegde [8] have found that other V2 neurons are selective for angles and corners, and that these showed submaximal responses for bars.

Dobbins et al. [19] provided physiological evidence that this type of cells may also be selective for curvatures and hypothesized a model which could fit those responses. Taken together, such evidence suggests that localized shape features are accessible for detailed processing of the visual scene configuration. It is not clear, however, whether different configurations of junctions are made explicit or whether their selectivity is rather represented in a distributed fashion.

Cortical area V4 follows after area V2 in the object recognition pathway. In the hierarchical object recognition pathway, V4 is the area just before the inferotemporal cortex (IT), where object recognition is achieved [69, 70]. Experiments in monkeys where area V4 was ablated showed that V4 is important for the perception of form and pattern/shape discrimination. Neurons in V4 selective to higher-order feature combinations characterizing shapes and their responses can be fit by a curvature-position function [50, 51]. In such a representation, the object's curvature is attached to a certain angular position relative to the object center of mass. Most V4 neurons represent individual parts or contour fragments [15].

Models of object recognition typically neglect the contributions of cells in intermediate areas of the visual cortex as are known from neurophysiology. These include endstopped cells which have been incorporated into a recent feedfoward hierarchical model of shape representation, named 2DSIL [60, 61]. Curvature is considered an important component in order to achieve object recognition in the human brain. The model is composed of simple, complex, endstopped, curvature and shape cells (Fig. 4).

Simple neurons of visual area V1 are sensitive to bar and edge orientations. In 2DSIL these are implemented using a Difference of Gaussians formulalism. These have been shown to provide a good fit to neuronal responses [28]. V1 simple model cells are organized into hypercolumns consisting of 12 orientations and 4 different scales (Fig. 4).

Complex cells respond to orientations but are invariant with respect to their location inside the receptive field. The best known hypothesis is that may be the result of the addition of simple cells along the axis perpendicular to their orientation. In 2DSIL a complex cell is the sum of 5 laterally displaced model simple cells within a column (Fig. 5a left).

Endstopped cells—having a great presence in area V2—have different properties from orientation-selective cells. Kato et al. [35] as well as later works have shown that this type of cells includes end-zone inhibitory areas. Model endstopped cells provide us with a coarse curvature estimation so that we can later divide contours into curvature classes. Dobbins et al. [19] provided the grounds for the design of this type of cell in 2DSIL (Figure 5a):

$$\begin{aligned} R_{ES} &= \Phi[\mathbf{c_c}\phi(\mathbf{R_c}) - (\mathbf{c_{d1}}\phi(\mathbf{R_{d1}}) + \mathbf{c_{d2}}\phi(\mathbf{R_{d2}}))] \\ \Phi &= \frac{1 - e^{-R/\rho}}{1 + 1/\Gamma e^{-R/\rho}} \end{aligned} \tag{5}$$

c_c, c_{d1} and c_{d2} are the gains for the center and displaced cells. R_c, R_{d1} and R_{d2} are the responses of the center (a simple cell) and the two displaced complex cells. ϕ and Φ are rectification functions.

By varying the orientations of the inhibition zones (e.g., 45° and 135°) with respect to the center excitatory zone, we can obtain cells that respond to the sign of the curvature, that is, the direction to where it steers (one direction, let's call it positive, or the opposite, let's call it negative) (Fig. 5b).

$$\begin{aligned} R_+ &= \Phi[\mathbf{c_c}\phi(\mathbf{R_c}) - (\mathbf{c_{d1_{45}}}\phi(\mathbf{R_{d1_{45}}}) + \mathbf{c_{d2_{135}}}\phi(\mathbf{R_{d2_{135}}}))] \\ R_- &= \Phi[\mathbf{c_c}\phi(\mathbf{R_c}) - (\mathbf{c_{d1_{135}}}\phi(\mathbf{R_{d1_{135}}}) + \mathbf{c_{d2_{45}}}\phi(\mathbf{R_{d2_{45}}}))] \end{aligned} \tag{6}$$

where c_c, $c_{d1_\bullet}$ and $c_{d2_\bullet}$ are the gains for the center and displaced cells as before. R_c, $R_{d1_\bullet}$and $R_{d2_\bullet}$ are the responses of center and displaced cell at the specified orienations with respect to the center cell.

We can obtain curvature cells due to the neural convergence of these three types of endstopped cells. Let's call R_{curv_i} the response of an endstopped cell (Eq. 5) to the preferred direction (that at which $R_{sign} > R_{opposite\_sign}$, Eq. 6). The response of a model shape-selective cell follows on the works of [50, 51] in which the response to a shape would correspond to the response of the local curvatures of the shape with respect to its center (Fig. 5c):

$$\begin{aligned} R_{shape} &= \sum_{i=1}^{n} c_i R_{curv_i}(\lambda) \qquad \lambda = max_{j=1}^{m}(\lambda_j) \\ c_i &= \frac{1}{2\pi} e^{-(x^2+y^2)} \end{aligned} \tag{7}$$

where λ is the preferred curvature direction and c_i is a Gaussian weight that would account for partial excitation depending on the selective curvature in distance-angular position.

2DSIL outperformed or provided comparable results when compared to state of the art computer vision models as well as with Serre et al. [67] (without the need of a learning stage for intermediate layers) at a real-world image recognition [60]. The hypothesis that this may be the way neurons in area V4 represent shapes was tested in a later work [61], where the responses from model shape-selective cells were compared with the responses from real cells in area V4 to 370 stimuli. The model fitted real cells with an average accuracy of 83 %. An example of responses of a model shape cell to different stimuli is shown in Fig. 5d.

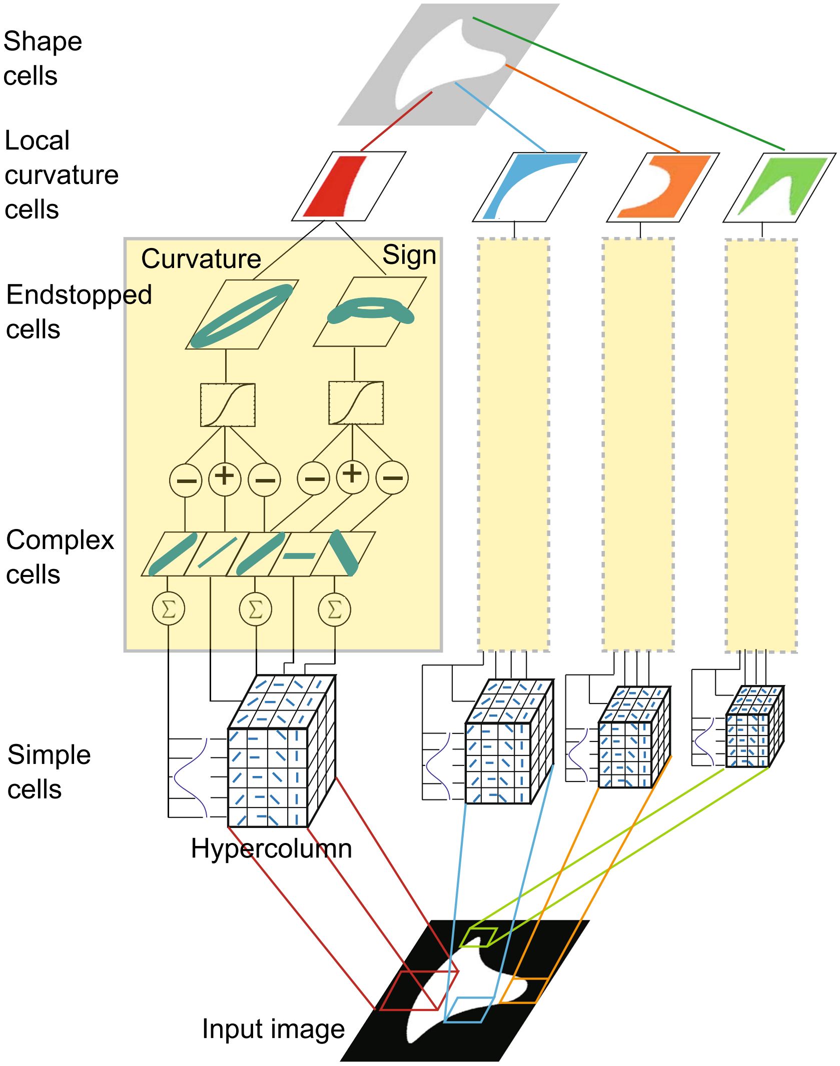

Fig. 4 Architecture of 2DSIL. Adapted from [61]

3 Discussion

The basic modeling here focuses on a mesoscopic level of scale in which layered representations of cells and their connections are considered (Figs. 3, 5). A critical novelty in comparison to static filters as suggested by the Hubel-Wiesel edge filter conception are the inclusion of direct higher-order computations (curvatures, contours, corners, etc.), lateral intra-cortical connections as well as top-down feedback cortico-cortical interactions, these last two form dynamic loops between processing units creating certain dynamics. Driving input to the layered architecture (that is composed of several hierarchically organized areas) is generated by feedforward signals. In the following we summarize what we think are next steps computer models need to take into account in order to advance towards such goal of building an advanced competence to process visual input.

The operations upon intermediate-level representations may be suited preferentially to access feature compositions in a deep cascade of stages along a feedforward, signal-driven process. In order to integrate context information at the various stages of such distributed, hierarchically-organized representations mechanisms of lateral signal integration and top-down feedback supplement the processing sketched here. We can then outline several characteristics of computations in the brain that seem to define a characteristic set of operational principles, namely: (1) a more complex bottom-up, feedforward filtering process, (2) lateral integration, and (3) feedback.

Regarding the first point, models usually define a filtering operation, each of which is characterized by a static kernel weighting function defined in space-feature domains. Several computer models have been successful at modeling the cortex this way, but there is still a long way to go to reach the point where we can say that we have built

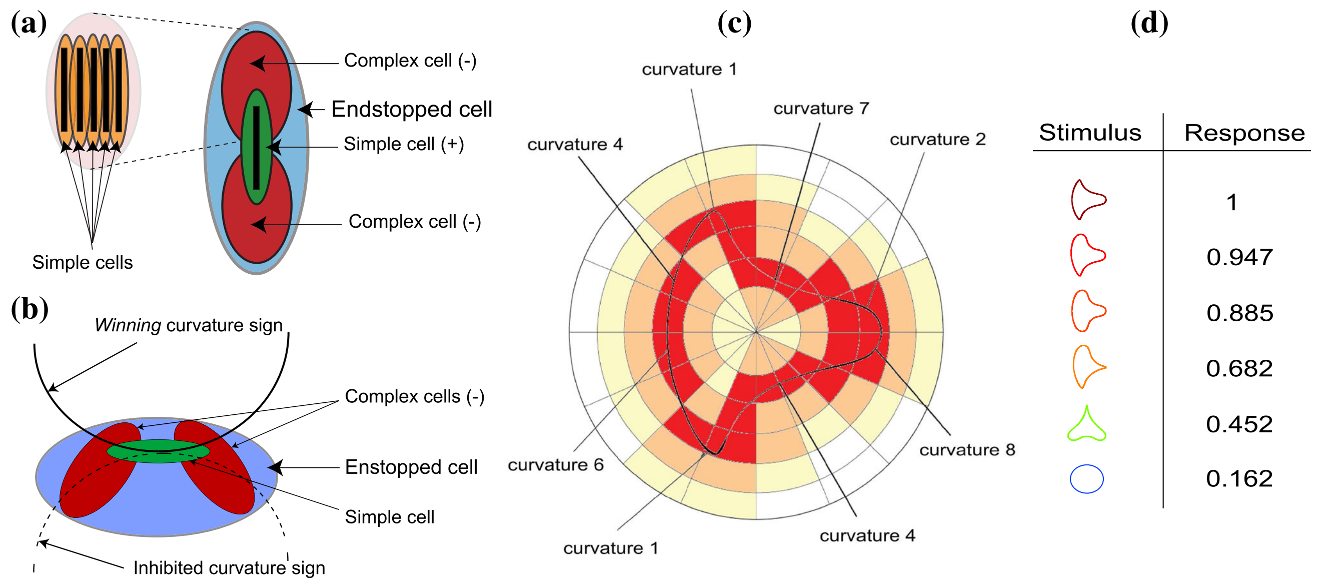

Fig. 5 Main cell types in 2DSIL (reproduced with permission from [61]). **a** Curvature-selective endstopped cell is composed of simple and complex cells. Note that a complex cell is the result of the addition of five simple cells at the same orientation. **b** Sign-selective endstopped cell. **c** Shape-selective cell **d** Response of a shape selective cell to different stimuli. This cell is selective to the stimulus on top, but provides high responses to other stimuli that are similiar (in a curvature-part fashion) to the preferred stimulus. These responses were compared with real V4 cell responses in [61]

an "artificial visual cortex". In chapter 2.2, we have summarized a model that reaches a higher level of abstraction than precedent models by considering the direct computation of intermediate-level representations. The differences between 2DSIL and other recent models, e.g. Serre et al. [67] and Cadieu et al. [10] are several. Whereas those previous models define their cell types as combinations of edge cenit responses successively over seven hierarchical layers, here our neurons in each layer compute quite different quantities. The goal was to include curvature computations directly, and not indirectly through the conjunctions of edges. How the visual cortex might accomplish this has been extensively investigated; endstopped cells play a major role. However, except for the notable exception of Dobbins et al. [19], they have not been adequately investigated computationally. This is where our approach diverges from others'. This is also what enables our faithful representation of curvature and 2D shape. The success of the approach in modeling the neural levels involved is evident in the matches to neural recordings which surpasses those shown by Cadieu et al. [10] as shown by Rodríguez-Sánchez and Tsotsos [61]. The results obtained by the model are very similar to those of the neurons in area V4, and are accomplished without any learning or classifier method.

Secondly, feedback and lateral signal integration influence the response of a target cell after its initial response to a driving input signal. We have focused our discussion in Sect. 2.1 on feedback which is generated at higher-level cortical stages or parallel processing pathways to provide contextual information to be re-entered at stages lower in the hierarchy influencing the evolution of each cell's responses [20]. The functional role feedback signals play remains a topic of intensive investigation to reveal how feedback signals interact and combine with the driving feedforward streams. Two main theoretical conceptions have been developed: (i) the reduction of the residual error between the feedforward signals and the predictive signals provided by the feedback from higher stages of processing [3, 74], and (ii) the modulatory enhancement of the gain for those cell responses where a matching top-down predictive signal template has been generated [17]. This feedback signal amplifies the sensory signal such that the subsequent competition between neurons yields a competitive advantage for the enhanced response patterns (biased competition) [18, 23, 63, 73]. The framework outlined above for boundary grouping and localized feature extraction focused on the latter which aims at amplifying activities that receive matching feedback while the predictive coding approach tends to drive the signal differences to zero. For example, boundary representations from grouping in model V2 and junction configurations in model V2/V3 send their output activations to curvature sensitive cells in model V4 where the activities are integrated. These cells, in turn, send their feedback to the input populations of neurons to further enhance their activation.

The temporal evolution of activations have demonstrated how a context-sensitive interpretation of features are derived in complex scenes and that the messaging processes require time to propagate their information. It remains to investigate a full scheme of boundary grouping in which multi-dimensional features such as disparity and border-ownership assignment are integrated. This would allow the selective reassignment of boundary fragments such that in case of mutual occlusions boundary segments can be completed amodally, as sketched in Fig. 1b, d. We should emphasize here that the grouping mechanisms discussed above are based on automatic and hardwired processes that utilize filters and operate in parallel over the visual scene.

It is now believed that feedback is mainly involved in the rapid computation of figural regions and assignment of ownership direction [16, 63, 78]. Following the analysis in Neumann et al. [48] neural mechanisms for such ownership should incorporate the following criteria: (i) long-range integration of real luminance contrasts and illusory boundaries (as discussed above), (ii) closed figural shape are composed of a majority of convex curvature or junction segments (which outnumber concave elements if those are present), and (iii) selective integration of (potential) spatial occlusion signals at T- and X-junctions as well as line endings. At T-junctions, for example, occlusions from opaque surface regions arise at the opposite side of the stem, while X-junctions often occur for overlapping configurations of transparent surfaces. So far, the proposed neural models that compute figure-ground direction and border-ownership utilize isotropic grouping of bottom-up input from prior boundary integration stages (e.g., [16]). Such models account for relatively simple stimulus configurations that have also been used in physiological experiments. It remains to demonstrate whether such mechanisms resolve ownership computation for more complex scenes in realistic camera images. Tschechne and Neumann [72] propose a recurrent computational network architecture that integrates the ideas outlined in this paper and embedded them in the recurrent feedback processing principles outlined above. In a nutshell, the model proposes hierarchical distributed representations of shape features to encode surface and object boundary over different scales of resolution corresponding to visual cortical areas V1–V4 and IT. Multiple low- and intermediate-level component representations interact by feedforward hierarchical processing which is combined with feedback signals driven by representations generated at higher stages. Global contextual configurations and local information is made available to represent contour details and to assign border ownership directions to eventually segregate figural shape from background. In addition, the integration of multi-dimensional features at the ownership grouping cells including boundaries as well as localized junctions features of T- and X-configurations remain to be simulated explicitly.

Finally, it is also worth including a few lines (due to their recent popularity) about deep learning and convolutional networks [30, 37]. These proposed network architectures utilize several principles suggested in this contribution. For example, deep networks rely on the stacked hierarchy of connected layers to accomplish learning representations of features and their combinations rich enough to detect structure in complex data. Our suggestion of intermediate-level representations and the processes operating upon them is related to this concept. We have emphasized how specific organizational principles support the establishment and integration of such mid-level can be utilized to derive richer shape related representations use in subsequent object recognition tasks. The processing principles involved, namely filtering, normalization and selective modulation via feedback, have been utilized in part in convolutional networks (see Sect. 2.1). In particular, variants of activity normalization have been proven useful in the past to stabilize recognition tasks that are organized in a layered hierarchical structure for object recognition. However, some criticisms remain regarding the biological plausibility of deep networks and convolutional nets in their current technical definition namely that they rely on large datasets and their status as greedy algorithms.

Acknowledgments H. N. has been supported by the Transregional Collaborative Research Centre SFB/TRR 62 “A Companion Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

References

1. Azzopardi G, Petkov N (2012) A corf computational model of a simple cell that relies on lgn input outperforms the gabor function model. Biol Cybernet 106(3):177–189
2. Azzopardi G, Petkov N (2013) Trainable cosfire filters for keypoint detection and pattern recognition. IEEE Trans Pattern Analy Mach Intell 35(2):490–503
3. Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76(4):695–711
4. Baylis G, Driver J (2001) Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. Nat Neurosci 4(9):937–942
5. Ben-Shahar O, Zucker S (2004) Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. Neural Comput 16(3):445–476
6. Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189
7. Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J Neurosci 17(6):2112–2127
8. Boynton G, Hegde J (2004) Visual cortex: the continuing puzzle of area V2. Curr Biol 14(13):523–524
9. Brosch T, Neumann H (2014) Computing with a canonical neural circuits model with pool normalization and modulating feedback

10. Cadieu C, Kouth K, Pasupathy A, Connor C, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. J Neurophysiol 98:1733–1750
11. Caplovitz GP, Tse PU (2007) V3a processes contour curvature as a trackable feature for the perception of rotational motion. Cereb Cortex 17(5):1179–1189
12. Carandini M, Ferster D (2000) Membrane potential and firing rate in cat primary visual cortex. J Neurosci 20(1):470–484
13. Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat R Neurosci 13(1):51–62
14. Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17(21):8621–8644
15. Connor C, Brincatt S, Pasupathy A (2007) Transformation of shape information in the ventral pathway. Curr Opin Neurobiol 17(2):140–147
16. Craft E, Schütze H, Niebur E, von der Heydt R (2007) A neural model of figure-ground organization. J Neurophysiol 97(6):4310–4326
17. De Pasquale R, Sherman SM (2013) A modulatory effect of the feedback from higher visual. J Neurophysiol 109:2618–2631
18. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Ann Rev Neurosci 18:193–222
19. Dobbins A, Zucker S, Cynader M (1987) Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature 329(6138):438–441
20. Edelman GM (1993) Neural darwinism: selection and reentrant signaling in higher brain function. Neuron 10(2):115–125
21. Fukushima K (1980) Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybernet 36(4):193–202
22. Gilbert CD, Li W (2013) Top-down influences on visual processing. Nat Rev Neurosci 14(5):350–363
23. Girard P, Bullier J (1989) Visual activity in area v2 during reversible inactivation of area 17 in the macaque monkey. J Neurophysiol 62(6):1287–302
24. Grossberg S (1968) Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. PNAS 2(59):368–372
25. Grossberg S (1970) Neural pattern discrimination. J Theoret Biol 2(27):291–337
26. Grossberg S, Mingolla E, Ross WD (1997) Visual brain and visual perception: how does the cortex do perceptual grouping? Trends Neurosci 20(3):106–111
27. Hansen T, Neumann H (2004) A simple cell model with dominating opponent inhibition for robust image processing. Neural Netw 17(5):647–662
28. Hawken M, Parker A (1987) Spatial properties of neurons in the monkey striate cortex. Proc R Soc Lon Ser B Biol Sci 231:251–288
29. von der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. Science 224(4654):1260–1262
30. Hinton G, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
31. Hubel D, Wiesel T (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol 28:229–289
32. Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195(1):215–243
33. Ito M, Komatsu H (2004) Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. J Neurosc 24(13):3313–3324
34. Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y (2009) What is the best multi-stage architecture for object recognition? In: Computer vision, 2009 IEEE 12th International Conference on, IEEE, pp 2146–2153
35. Kato H, Bishop P, Orban G (1978) Hypeercomplex and simple/complex cells classifications in cat striate cortex. J Neurophys, pp 1071–1095
36. Kellman PJ, Shipley TF (1991) A theory of visual interpolation in object perception. Cognit Psychol 23(2):141–221
37. LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw, 3361
38. Lecun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1(4):541–551
39. Logothetis N, Sheinberg D (1996) Visual object recognition. Ann Rev Neurosci 19:577–621
40. von der Malsburg C, Phillips WA, Singer W (2010) Dynamic coordination in the brain: from neurons to mind. MIT Press, Cambridge
41. Markov NT, Kennedy H (2013) The importance of being hierarchical. Curr Opin Neurobiol 23(2):187–194
42. Metzger W (1936) Gesetze des sehens. W. Kramer Frankfurt am Main.
43. Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. PNAS 99(23):15164–15169
44. Mutch J, Lowe DG (2008) Object class recognition and localization using sparse features with limited receptive fields. Int J Comput Vis 80(1):45–57
45. Nakayama K, Shimojo S, Ftamachandran VS (1990) Transparency: relation to depth, subjective contours, luminance, and neon color spreading. Perception 19(4):497–513
46. Neumann H, Mingolla E (2001) Computational neural models of spatial integration in perceptual grouping. In Fragments to Objects-Segmentation and Grouping in Vision, ch12 130:353–400
47. Neumann H, Sepp W (1999) Recurrent V1–V2 interaction in early visual boundary processing. Biol Cybernet 81(5–6):425–444
48. Neumann H, Yazdanbakhsh A, Mingolla E (2007) Seeing surfaces: the brain's vision of the world. Phys Life Rev 4(3):189–222
49. Parent P, Zucker S (1989) Trace inference, curvature consistency, and curve detection. IEEE Pattern Anal Mach Intell 11(8):823–839
50. Pasupathy A, Connor C (1999) Responses to contour features in macaque area V4. J Neurophysiol 82(5):2490–2502
51. Pasupathy A, Connor C (2002) Population coding of shape in area V4. Nat Neurosci 5(12):1332–1338
52. Pessoa L, Thompson E, Noë A (1998) Filling-in is for finding out. Behav Brain Sci 21(06):781–796
53. Piëch V, Li W, Reeke GN, Gilbert CD (2013) Network model of top-down influences on local gain and contextual interactions in visual cortex. PNAS 110(43):4108–4117
54. Qiu FT, von Der Heydt R (2005) Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. Neuron 47(1):155–166
55. Ramón y Cajal S (1888) Sobre las fibras nerviosas de la capa molecular del cerebelo. R Trim Histol Norm Patol 1:33–49
56. Rao R, Ballard D (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2(1):79–87
57. Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61(2):168–185
58. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2(11):1019–1025
59. Ringach DL, Shapley R (1996) Spatial and temporal properties of illusory contours and amodal boundary completion. Vis Res 36(19):3037–3050

60. Rodríguez-Sánchez A, Tsotsos J (2011) The importance of intermediate representations for the modeling of 2D shape detection: endstopping and curvature tuned computations. IEEE CVPR pp 4321–4326
61. Rodríguez-Sánchez A, Tsotsos J (2012) The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. PLOS ONE 7(8):1–13
62. Roelfsema PR (2006) Cortical algorithms for perceptual grouping. Ann Rev Neurosci 29:203–227
63. Roelfsema PR, Lamme VA, Spekreijse H, Bosch H (2002) Figureground segregation in a recurrent network architecture. J Cogn Neurosci 14(4):525–537
64. Rubin N (2001a) Figure and ground in the brain. Nat Neurosci 4(9):857–858
65. Rubin N (2001b) The role of junctions in surface completion and contour matching. Perception 30(3):339–366
66. Salin PA, Bullier J (1995) Corticocortical connections in the visual-system- structure and function. Physiol Rev 75(1):107–154
67. Serre T, Wolf L, Bileschi S, Riesenhuber M (2007) Robust object recognition with cortex-like mechanisms. IEEE T Pattern Anal Mach Intel 29(3):411–426
68. Sherman SM, Guillery R (1998) On the actions that one nerve cell can have on another: distinguishing drivers from modulators. PNAS 95(12):7121–7126
69. Tanaka K (1996) Inferotemporal cortex and object vision. Ann Rev Neurosci 19:109–139
70. Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66(1):170–189
71. Thielscher A, Neumann H (2008) Globally consistent depth sorting of overlapping 2d surfaces in a model using local recurrent interactions. Biol Cybernet 98(4):305–337
72. Tschechne S, Neumann H (2014) Hierarchical representation of shapes in visual cortexfrom localized features to figural shape segregation. Front Computat Neurosci 8
73. Tsotsos J, Culhane S, Winky W, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artif Intel 78(1–2):507–545
74. Ullman S (1995) Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. Cereb Cortex 5(1):1–11
75. Wallis G, Rolls E (1997) Invariant face and object recognition in the visual system. Prog Neurobiol 51(2):167–194
76. Weidenbacher U, Neumann H (2009) Extraction of surface-related features in a recurrent model of V1–V2 interactions. PLOS ONE 4(6):e5909
77. Williams LR, Jacobs DW (1997) Stochastic completion fields, a neural model of illusory contour shape and salience. Neural Computat 9(4):837–858
78. Zhou H, Friedman H, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594–6611
79. Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex. J Neurosci 16(22):7376–7389

Antonio J. Rodríguez-Sánchez is currently a senior research fellow in the department of Computer Science at the University of Innsbruck, Austria. He holds a MSc. degree from Universidade da Coruña, Spain. He completed his PhD. at York University, Canada on the subject of modeling attention and intermediate areas of the visual cortex. He is part of the Intelligent and Interactive Systems group and his main interests are computational neuroscience and computer vision.

Heiko Neumann is a professor of engineering and computer science at Ulm University, Germany. He received his Habilitation and PhD. degrees from the University of Hamburg, Germany. He leads the Vision and Perception Science lab at the Institute of Neural Information Processing. His main interests are mathematical and computation investigation of neural processes using empirical data derived from psychophysics, neurophysiology and imaging.

Justus Piater is a professor of computer science at the University of Innsbruck, Austria. He holds a MSc. degree from the University of Magdeburg, Germany, and MSc. and PhD. degrees from the University of Massachusetts Amherst, USA. He leads the Intelligent and Interactive Systems group that works on visual perception and inference in dynamic and interactive scenarios, including applications in autonomous robotics and video analysis.