



Hierarchical representation of shapes in visual cortex—from localized features to figural shape segregation

Stephan Tschechne* and Heiko Neumann

Faculty of Engineering and Computer Science (with Psychology and Education), Institute of Neural Information Processing, Ulm University, Ulm, Germany

Edited by:

Antonio J. Rodriguez-Sanchez,
University of Innsbruck, Austria

Reviewed by:

John K. Tsotsos, York University,
Canada

Jonathan R. Williford, Johns
Hopkins University, USA

***Correspondence:**

Stephan Tschechne, Institute of
Neural Information Processing, Ulm
University, Albert-Einstein-Allee 1,
89069 Ulm, Germany
e-mail: stephan.tschechne@
uni-ulm.de

Visual structures in the environment are segmented into image regions and those combined to a representation of surfaces and prototypical objects. Such a perceptual organization is performed by complex neural mechanisms in the visual cortex of primates. Multiple mutually connected areas in the ventral cortical pathway receive visual input and extract local form features that are subsequently grouped into increasingly complex, more meaningful image elements. Such a distributed network of processing must be capable to make accessible highly articulated changes in shape boundary as well as very subtle curvature changes that contribute to the perception of an object. We propose a recurrent computational network architecture that utilizes hierarchical distributed representations of shape features to encode surface and object boundary over different scales of resolution. Our model makes use of neural mechanisms that model the processing capabilities of early and intermediate stages in visual cortex, namely areas V1–V4 and IT. We suggest that multiple specialized component representations interact by feedforward hierarchical processing that is combined with feedback signals driven by representations generated at higher stages. Based on this, global configurational as well as local information is made available to distinguish changes in the object's contour. Once the outline of a shape has been established, contextual contour configurations are used to assign border ownership directions and thus achieve segregation of figure and ground. The model, thus, proposes how separate mechanisms contribute to distributed hierarchical cortical shape representation and combine with processes of figure-ground segregation. Our model is probed with a selection of stimuli to illustrate processing results at different processing stages. We especially highlight how modulatory feedback connections contribute to the processing of visual input at various stages in the processing hierarchy.

Keywords: ventral pathway, distributed representation, figure-ground segregation, modulatory feedback, computational model

1. INTRODUCTION

We visually perceive our environment as a stable and comprehensive combination of objects, where we can easily identify objects and persons and we efficiently analyse geometrical cues that allow a precise navigation and interaction. This happens so effortlessly and accurately that it is absolutely counterintuitive that this is an extraordinary achievement of our brain. The visual system of mammals achieves this result from input that is captured at the retinal level after light has been projected through the eye and hits light-sensitive neurons. The perception of our environment starts at this local level where our position, the direction of our gaze, the current illumination, an object's surface properties and its location relative to others causes a set of neurons in the retina to respond with increased activation that is a function of received light intensity. How the visual system transforms this concert of local visual inputs into a stable and informative perception of surfaces and objects is subject to intense research. Since the pioneering works on neural principles by Hubel and Wiesel (1959) many insights into cortical processing of visual input has been discovered. Neurophysiologists agree

that the processing in the mammalian brain is performed in a hierarchical way and processing is organized into various specialized brain areas (Felleman and Van Essen, 1991). Those brain areas receive connections from preceding processing stages, but also from regions later in the processing stream (Markov et al., 2013). Early areas in visual cortex are retinotopically arranged (Hubel and Wiesel, 1962), which means that juxtaposed retinal locations are mapped to juxtaposed locations in visual cortex, with foveal positions being represented at a higher resolution. Individual assemblies of neurons become activated when their preferred stimulus is presented in their receptive field (Hubel and Wiesel, 1962). With progression in the visual pathway, the size of those RFs increase from sizes smaller than one degree of visual angle to sizes covering a good part of the visual field. In parallel, the tuning toward the preferred stimulus changes from simple features like oriented contrasts (Hubel and Wiesel, 1962) to complex ones like image features, figure-ground-related cues, object categories or faces. Processing along the visual pathway is organized into two streams (Ungerleider and Haxby, 1994), the ventral stream that exhibits a tuning toward movement

and position, whereas the dorsal stream processes shapes and objects.

However, most of the achievements that the visual system exhibits, like the abilities to generalize and its robustness and adaptability, most probably stem from connections that connect higher cortical areas with lower ones (Hupé et al., 1998; Markov et al., 2013). Those feedback connections are believed to play an important role in visual processing, as they enrich local activations with contextual information that is represented at higher visual areas. We propose that on the way from generalizing early local features to higher meaningful representations, the role of object boundaries plays an essential part. Contrasts indicate spatial changes in local illumination which might coincide with object boundaries that allow segregation from background. However, contrasts indicating a real transition from one object to another or from the object to the background must be separated from those indicating an illumination change and those caused by textured regions. This must be accomplished using contextual information. The region delimited by such a boundary is a surface with locally constant parameters, and a set of surfaces forms objects, scenes and eventually our complete visual environment. We believe that the processing capabilities of early and intermediate stages of visual cortex are used to transform local representation into an intermediate, more meaningful representation of contours, shapes and surfaces. Following those ideas, we propose that a stable representation of shape may be established by interacting assemblies that are each devoted to specific features properties. We thus propose a hierarchical model of 2-dimensional shape representation that incorporates processing at low and intermediate areas of visual cortex. Each model area consists of a three-stage processing cascade of initial filtering, application of modulatory feedback effects and center-surround interactions leading to an activity normalization (Carandini and Heeger, 1994; Carandini et al., 1999; Kouh and Poggio, 2008; Carandini and Heeger, 2012). The functional effects of this columnar cascade can roughly be mapped onto compartments of cortical area subdivisions [as suggested in (Self et al., 2012)].

Our model combines the representation of visual shapes with mechanisms for figure-ground segregation on the basis of assigning border ownership and incorporates a distributed representation of local contour curvature over different cortical areas. In our model we emphasize the computational role of feed-forward and feedback mechanisms (Grossberg, 1980; Edelman, 1993) to generate a hierarchical distributed representation of shape information. The feedback amplifies the sensory signal such that the subsequent competition between neurons builds a competitive advantage (Tsotsos, 1988; Girard and Bullier, 1989; Desimone, 1998; Roelfsema et al., 2002; Reynolds and Heeger, 2009). Boundaries and their orientation are represented after initial processing in model area V1 and a grouping stage in model area V2. Contextual boundary configurations are also represented at a coarser spatial level at model V2 and V4 to achieve selectivities toward contour curvature. With the influence of feedback, those cells are enhanced at lower stages that contribute to a matching bottom-up signal.

The output of our model is a representation of shapes and shape segments where contextually compatible boundary

information benefits from recurrent feedback connections. Such a representation could provide input to subsequent processing stages for e.g., object classification tasks, which would clearly benefit from the enhanced representation.

This model extends previous own works (Neumann and Mingolla, 2001; Hansen and Neumann, 2004; Weidenbacher and Neumann, 2009) but introduces functional properties that have been inspired by the works of other groups. A model of curvature representation can also be found in Cadieu et al. (2007). The authors modeled physiological findings of the same group (Pasupathy and Connor, 1999; Connor et al., 2007) that has focussed on the dynamics of contour processing (Yau et al., 2013). Cell representations from early visual areas are combined to intermediate-level shape descriptors are used in a computational model by Rodríguez-Sánchez and Tsotsos (2012). Riesenhuber and Poggio (1999, 2000); Mutch and Lowe (2008) released very powerful models of object and object class categorization in a hierarchical modeling approach. The physiological (Zhou et al., 2000; O'Herron and von der Heydt, 2011) as well as the computational (Layton et al., 2012) aspects of border ownership are subject to intense research. Models of contour integration and perceptual grouping also exist from Zhaoping (1998) and Jehee et al. (2006); Roelfsema (2006). The role of feedback and physiological investigations are elaborated in Hupé et al. (1998); Markov et al. (2013) and very recently (De Pasquale and Murray Sherman, 2013) found evidence for the modulatory properties of feedback in the visual cortices of mice.

2. MODEL DEFINITION

We propose a biologically inspired model of two-dimensional shape representation that consists of a hierarchical structure of interconnected model areas (see **Figure 1**). These model areas resemble the mechanisms of early and intermediate stages of visual processing in the ventral pathway of visual cortex. Each of the model areas is represented by a staged columnar cascaded model (see **Figure 1**). This cascade consists of (i) initial filtering, (ii) activity modulation, and (iii) center-surround interaction.

2.1. NOMENCLATURE

The following list familiarizes the reader with the nomenclature that is used in our manuscript:

- Names of model areas are written in superscripts to indicate the affiliation to parameters or responses. A response for cells in area V1 thus would read R^{V1} .
- Greek symbols (like α, β, σ) are used for parameters of dynamic functions or shapes of receptive fields.
- N and M are constants indicated the number of orientations and directions used in our model.
- An instance of an orientation is indicated using the step width $\hat{\theta}$ between discretely sampled orientations. The variable i is used as index. A specific orientation in a population is indicated with $i\hat{\theta}$ with $i \in 0..N - 1$.
- \mathcal{N} stands for a normal (Gaussian) distribution that can be isotropic or anisotropic, rotated and spatially shifted, as defined by subscripted parameters.

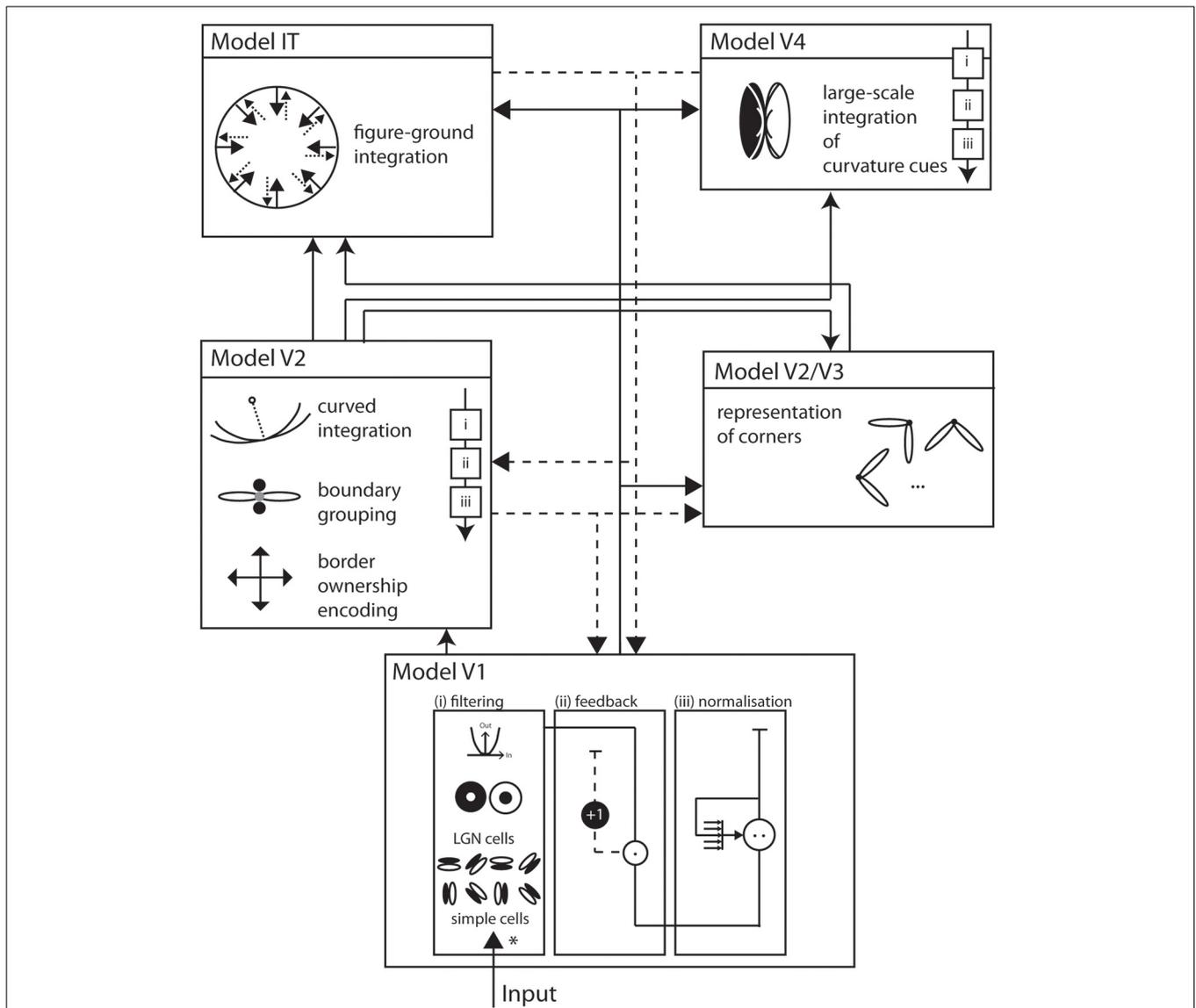


FIGURE 1 | Overall model architecture. Visual input enters the model at the bottom and is subsequently processed by interconnected functional areas with increasingly large receptive field sizes. Solid arrows indicate feedforward, dashed arrows indicate feedback, or modulatory, connections. Each area implements a generic architecture of building blocks that consists of (i) filtering (*) of the input, (ii) modulation by feedback, and (iii) response normalization. *Model V1* consists of image filters that resemble properties of early processing in LGN and V1, namely simple and complex cells that are tuned to circular or elongated image contrasts. *Model V2* integrates responses of model V1 with long-range integration cells. A multiplicative combination of subcells responds best to elongated contrasts of one dominant orientation. Also at V2, a population

of cells represents border ownership directions. At population of long-range curved integration cells help represent different boundary curvatures. The *Models V2, V3* complex hosts representations of corners by integrating V1 responses from orthogonal configurations over a small spatial surround. *Model V4* consists of cells that asymmetrically integrate responses from V1 and V2 to become curvature selective at an increased spatial scale. In *Model IT*, cells with large receptive fields integrate responses from V1, V2 and V4 at local figure convexities to achieve a contextual segregation into figure and ground. Area V4 allows a description of a shape by means of cues that are represented on distributed areas in the model. Those cues exist at different spatial scales and their mutual interaction generates dynamic processes in the model.

- Spatial positions are denoted in bold latin letters like **x** or **p**.
- To indicate an angle between two vectors we use $\angle(\mathbf{v}_1, \mathbf{v}_2)$.
- The convolution operation is abbreviated using the asterisk (*).
- A rectification operation is indicated by $[\dots]^+$.

2.2. PROCESSING CASCADE

In our model, neural activations or response levels are modeled using a scalar representation of the neural firing rate. For ease of

writing, we will in the following refer to the response of a cell, keeping in mind that this represents the activation level of a large number of real cells. The first model stage of the cascade is the initial filtering of available input *I*. To model the response for the preferred stimulus in the visual field, we employ a 2-dimensional convolution operation with the preferred stimulus as the convolution kernel K_{pref} . The response of model cells $R = I * K_{pref}$ is defined as

$$R(x, y) = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} K_{pref}(u, v)I(x - u, y - v) \quad \forall x, y \in D_I \quad (1)$$

A frequently used kernel in our model serves as elementary building block and is a 2-dimensional Gaussian distribution that is elongated along one axis and rotated around its center. We refer to this distribution by \mathcal{N} with parameters for orientation θ , deviation along the axes σ_1, σ_2 and the center of the distribution μ .

$$\mathcal{N}_{\theta, \sigma_1, \sigma_2, \mu}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\left(\frac{(\hat{x} - \mu_x)^2}{2\sigma_1^2} + \frac{(\hat{y} - \mu_y)^2}{2\sigma_2^2}\right)\right)$$

with

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (3)$$

If parameters are not specified they are considered having the following default values: $\theta = 0, \mu = (0, 0)^T, \sigma_1 = 1, \sigma_2 = \sigma_1$. In the following, functional filter kernels will often be designed as a combination of multiple such elementary components.

The coefficients of the kernel that models the preferred stimulus might incorporate negative weights to account for the inhibitory connections a cell may receive. This could lead to overall responses that are numerically negative. We thus use a rectification operator after convolution and feedback stages to ensure that numerically the response rate of a population is not negative:

$$[R]^+ = \max(0, R). \quad (4)$$

At the second stage of the cascade, response levels are modulated by recurring input from higher visual areas. We propose a feedback mechanism that exerts a purely modulatory gain control on the input. That means that feedback alone cannot generate activities without activation by the initial filtering step (see **Figure 2**). With R being the unmodulated driving signal and net_{FB} being the strength of the feedback, the modulated response is

$$R_{FB} \propto R \cdot (1 + net_{FB}). \quad (5)$$

Using this approach, given $R = 0$ no signal is generated as output irrespectible of the strength of the feedback net_{FB} . On the other hand, if no feedback signal is available, the right part of the equation leaves the input signal R unchanged (Salin and Bullier, 1995; Hupé et al., 1998; Eckhorn, 1999; Gilbert and Li, 2013).

Before normalization at the final stage of the cascade, we apply a non-linear transfer function to map the computed responses to a cell activation level. In our model, we use a function of type

$$f(R) = R^k \quad (6)$$

with k the non-linearity parameter. At the final stage, we incorporate a mechanism that keeps the response level limited by using a *shunting inhibition* that leads to a non-linear compression of high amplitude activities resembling the *Weber-Fechner-Law* of perceptual thresholds. In its dynamic formulation, the rate of change of

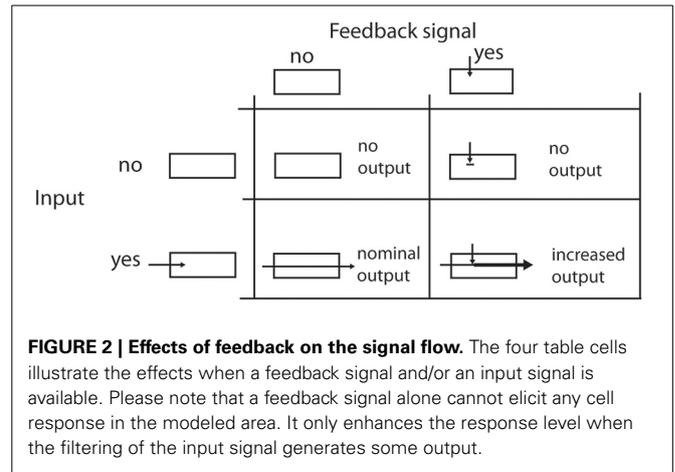


FIGURE 2 | Effects of feedback on the signal flow. The four table cells illustrate the effects when a feedback signal and/or an input signal is available. Please note that a feedback signal alone cannot elicit any cell response in the modeled area. It only enhances the response level when the filtering of the input signal generates some output.

the signal $\partial_t R_{i\hat{\theta}}^{norm}$ depends on the current activation level as well as the amount of input I_{net} :

$$\partial_t R_{i\hat{\theta}}^{norm} = -\alpha R_{i\hat{\theta}}^{norm} + \beta R_{i\hat{\theta}} - R_{i\hat{\theta}}^{norm} \cdot I_{net} \quad (7)$$

$$I_{net} = \frac{1}{N} \sum_{i=0}^{N-1} R_{i\hat{\theta}}. \quad (8)$$

With N the size of the used population, respective orientations. When this equation is solved at equilibrium, i.e., when $\partial_t R_{i\hat{\theta}} = 0$, the activation becomes

$$R_{i\hat{\theta}}^{norm} = \beta \frac{R_{i\hat{\theta}}}{\alpha + I_{net}} \quad (9)$$

The constants influence the steepness of the non-linearity (α) and the scale of the normalized signal (β). This model architecture has previously been used in various approaches touching different domains, such as the disambiguation of local motion (Bayerl and Neumann, 2004; Beck and Neumann, 2011), the processing of transparent motion (Raudies and Neumann, 2010) the detection of texture boundaries (Thielscher and Neumann, 2003), the extraction of object boundaries using texture compression (Weidenbacher and Neumann, 2009), and the analysis and representation of biological motion sequences (Layher et al., 2014).

In the following, we describe the forward sweep of our model, from early toward intermediate processing stages. After all areas have been described in detail, we will elaborate on the feedback connections that build the recurrent model structure.

2.3. MODEL AREA V1

The processing starts at early stages of visual cortex where we model the functionality of LGN and V1 cells where LGN cell responses provide feedforward input to V1 cells. Here, the visual input is initially processed to generate a representation of local image contrasts and local contrast orientations (Hubel and Wiesel, 1962).

In general, model cell responses follow first-order dynamics and represent the changes of membrane potentials. Such dynamics are influenced by excitatory and inhibitory inputs and a passive decay of activity. In order to simplify the computations in our large-scale simulations we use steady-state equations in calculations of feedforward filtering stages. Others are numerically integrated using a Euler one-step scheme. The response of LGN cells is calculated using the following linear equation,

$$\partial_t R^{LGN} = -R^{LGN} + (I * \mathcal{N}_\sigma) - (I * \mathcal{N}_\kappa) \quad (10)$$

As pointed above, we assume that such linear feedforward filtering operations quickly relax at their equilibrium state. Therefore, we utilize the steady-state equation

$$R^{LGN} = [I * (\mathcal{N}_\sigma - \mathcal{N}_\kappa)]^+, \quad (11)$$

with σ and κ denoting the width of center and surround kernel, respectively. To model cells that are tuned to oriented contrasts, we use elongated gaussian kernels that are combined into odd-symmetric simple cell profiles using anisotropic σ_1 and σ_2 and a radius ω_1 for the spatial shift of the integration kernels. The responses of such cells are denoted by the steady-state equation

$$R_{i\hat{\theta}}^{V1} = \left[R^{LGN} * (\mathcal{N}_{i\hat{\theta}, \sigma_1, \sigma_2, \mathbf{x}+\mathbf{p}} - \mathcal{N}_{i\hat{\theta}, \sigma_1, \sigma_2, \mathbf{x}-\mathbf{p}}) \right]^+ \quad (12)$$

with

$$\mathbf{p} = \omega_1 (\cos(\theta + \pi), \sin(\theta + \pi))^T \quad (13)$$

The filter kernel that is defined that way yields high response activations at positions with local luminance contrasts that match the layout of the filter kernel. To achieve insensitivity against the sign of contrast, pairs of equally oriented filters with opposite sensitivity to contrast polarity are used. Such filters populate a set with evenly distributed orientation tunings that represent possible contrast orientations. The locally dominant orientation can be derived by selecting the orientation channel with maximum response, $i_{max} = \text{argmax}_i R_{i\hat{\theta}}^{V1}$.

2.4. MODEL AREA V2/V3 COMPLEX

At the stage of V2 we model cells sensitive to contextual influences of contour segments that are arranged in larger spatial extent compared to V1 receptive fields. The integration of elongated contours in V2 makes use of a mechanism that links cells of like orientations over larger spatial distances. The filters are modeled using elongated Gaussian kernels positioned at \mathbf{p} with offset ω_{ex}^{V2} to the center of the cell. The parameters of the elongated Gaussian kernels are set to build a combined kernel of an elongated integration field, which reflects the highly significant anisotropies of long-range connections in visual cortex (Bosking et al., 1997). The subfields sample the activations generated by V1 complex cells (Grossberg and Mingolla, 1985; Neumann and Sepp, 1999).

The subfields are combined in a multiplicatively. This resembles a logical *and*-operation for the individual subfield activations. Modeled V2 cells only become activated when both subfields receive sufficient input. The response is thus able to bridge local gaps in contours. This is in line with physiological

findings, as V2 neurons respond to elongated luminance contrasts as well as to illusory contours (von der Heydt et al., 1984; Heitger et al., 1998) like in the Kanisza square.

This integration mechanism is enhanced by local inhibitory effects. Smaller and isotropic integration fields are positioned along an orthogonal axis from the cell's center with distance ω_{inh}^{V2} , building a cross-like zone of excitatory and inhibitory integration, compare (Piëch et al., 2013). At those positions \mathbf{p}_\perp , activity from all orientations is integrated and has an inhibitory effect on the total response. This has a strong suppressive effect on contour fragments that are positioned within a cluttered surround, while isolated boundary segments are not affected. The complete response for an elongated V2 cell is calculated by the steady state equation:

$$R_{i\hat{\theta}}^{V2} = \left[R^{V1} * \mathcal{N}_{i\hat{\theta}, \sigma_1, \sigma_2, \mathbf{x}+\mathbf{p}} \cdot R^{V1} * \mathcal{N}_{i\hat{\theta}, \sigma_1, \sigma_2, \mathbf{x}-\mathbf{p}} - \gamma \cdot R^{V1} * \mathcal{N}_{\sigma_3, \mathbf{x}+\mathbf{p}_\perp} - \gamma \cdot R^{V1} * \mathcal{N}_{\sigma_3, \mathbf{x}-\mathbf{p}_\perp} \right]_{i\hat{\theta}}^+ \quad (14)$$

with

$$\mathbf{p} = \omega_{ex}^{V2} (\cos(i\hat{\theta}), \sin(i\hat{\theta}))^T \quad (15)$$

$$\mathbf{p}_\perp = \omega_{inh}^{V2} (\cos(i\hat{\theta} + \pi), \sin(i\hat{\theta} + \pi))^T \quad (16)$$

We also model V2 neurons that respond to more complex stimuli like in curved or angular shape outlines. We propose a population of V2 cells tuned to curved contour outlines that allows integration of smooth and even fragmented boundary configurations (Field et al., 1993). We propose a population of V2 cells tuned to a curved contour outline, see **Figure 4**. They resemble the functionality of elongated V2 cells but their integration fields are designed such that they are curved. A curvature direction is defined either to the left or the right of the tangent orientation at the target location. the center of curvature defines an osculating circle with given curvature-radius. The integration weight is modeled by a function w^{dist} that decreases with distance from the cell's center. A second tuning function w^{ori} in the orientation domain specifies the weights for the orientation population. Here, the weight decreases with distance to the main tuning direction which is perpendicular to the dominant orientation. Basically, only those orientations are integrated with maximum that are tangential to the curvature trace at their relative positions. This yields a sharp tuning of the cell for a certain curvature level. The complete response for an curved V2 cell is calculated by the steady state equation:

$$R_{i\hat{\theta}}^{V2Curv} = \sum_{\mathbf{x}} w(\omega^C, \mathbf{x}) \cdot R_{\mathbf{x}, i\hat{\theta}}^{V1} \quad \text{with} \quad (17)$$

$$w = w^{dist} \cdot w^{ori} \quad (18)$$

$$w^{dist} = \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_0)^2}{\sigma_1^2}\right) \quad (19)$$

$$w^{ori} = \sin(\angle(\vec{\mathbf{x}}_0, \vec{\mathbf{x}})) \cdot \exp\left(-\frac{(\|\vec{\mathbf{x}}\| - \omega^c)^2}{\sigma_2^2}\right) \quad (20)$$

$$\mathbf{c} = \mathbf{x}_0 + \omega^c \left(\cos(i\hat{\theta}), \sin(i\hat{\theta}) \right)^T \quad (21)$$

In this equation w denotes a weighting function for responses in the currently integrated position \mathbf{x} . The reference point of the integrating cell is \mathbf{x}_0 . The weighting functions depends of the curvature radius being integrated w^c . w^{dist} produces a weight depending on the distance from the curvature cell's center \mathbf{x}_0 . w^{ori} returns a weight given the current angle between integrating position and center of curvature \mathbf{x}_c , depending on the reference position x_0 . In simple words, orientations orthogonal to the imaginary line between integration position and imaginary curvature center \mathbf{c} receive highest weight. w^{ori} is extended with a function that drops with increased distance of integrating position to imaginary center of curvature.

Cells in visual cortex V2 also show selectivity to the figure-ground arrangement of the scene in the visual field (Williford and von der Heydt, 2013). So-called border ownership cell responses are elicited when figure of arbitrary shape is presented on their preferred side with respect to the center of their receptive field. From the same group, O'Herron and von der Heydt (2013) have also shown that during visual motion caused by eye motion or object motion, these border ownership signals are remapped to different neurons. The visual system uses this information to resolve depth arrangements in the stimulus (Qiu and von der Heydt, 2005). The pointing of border ownership cells indicates the direction of the frontal surface at every image location. This reflects to commonly known Gestalt rule that a boundary is owned by the frontal figure.

We model border ownership cells by a retinotopically arranged population representing four potential directions where the figure can be positioned relative to the cell's center. Border ownership responses are initially isotropic and only occur together with local contrast activations. Cells indicating opponent border ownership direction are mutually rivaling in our model. The complete response for an border ownership V2 cell is calculated by the steady state equation:

$$R_{\lambda}^{V2Bown} = \begin{cases} f(R_{i\hat{\theta}}^{V2}) & \text{when } \lambda \perp i\hat{\theta} \\ 0 & \text{when } \lambda \parallel i\hat{\theta} \end{cases} \quad (22)$$

The mutual competition between activations indicating opposing border ownership directions R_a^{Bown} and R_b^{Bown} is calculated by

$$\partial_t R_a^{Bown} = -\alpha \cdot R_a^{Bown} + A(1 - R_a^{Bown}) - \beta \cdot R_b^{Bown} \quad (23)$$

$$\partial_t R_b^{Bown} = -\alpha \cdot R_b^{Bown} + A(1 - R_b^{Bown}) - \beta \cdot R_a^{Bown} \quad (24)$$

Based on empirical evidence of neural representations generated by cells selective to multiple orientations (Felleman and Van Essen, 1987; Ito and Komatsu, 2004; Anzai et al., 2007) we incorporate model representations of corners in a dedicated model area V2/V3 complex. We build upon the proposal developed in Weidenbacher and Neumann (2009) that corner and junction configurations can be made explicit by specific read-out mechanisms. Here, we employ a simplified version as of Hansen and Neumann (2004) to generate corner representations

by grouping V1 responses of orthogonal orientation fields. In a steady-state formalism the response reads

$$R_{i\hat{\theta}}^{V2/V3} = \left[R_{i\hat{\theta}}^{V1} \cdot R_{i\hat{\theta}+\pi}^{V1} \right]^+ \quad (25)$$

2.5. MODEL AREA V4

Inspired by experimental evidence cells in model V4 integrate responses of V1, V2, and V2/V3 to achieve a selectivity that considers large-scale boundary fragments as well as local variations in curvature and a selectivity for corners (Pasupathy and Connor, 1999; Yau et al., 2013). Curvature selective cells are modeled in a two-stage cascade of mechanisms. The first level integrates V2 contour responses and is selective to curvature directions, left or right (relative to the cell's orientation preference). The second level combines opposite curvature directions into one response, like in V1 complex cells. This model mechanism differs from the one proposed by Rodríguez-Sánchez and Tsotsos (2012), which utilizes single stage filter computations. In this approach specific subfield mechanisms sensitive to orientation, tangential contour outline and scale are combined in a non-linear fashion to selectively respond to contour fragments of different curvatures. We develop a mechanism that is distributed over different stages to first group responses to extended contour outlines in V1 and V2 suppressing non-contour clutter. In the case of sharply localized corners and junctions the dedicated representations of localized multi-orientation responses will be activated. Those responses of grouping cells (or the junction representations) are integrated at the subsequent stage. Here, curvature selectivity is made explicit that distinguishes left and right curvatures. Different integration scales generate selectivity to curvature. This distribution allows to associate regions of high contour curvature at an intermediate scale with localized outline details at the finer scale which enhances the selectivity of the model developed by Rodríguez-Sánchez and Tsotsos (2012).

The model cell responses in our model are described by the following equations:

$$\begin{aligned} \partial_t R_{i\hat{\theta}}^{V4, left} &= -\alpha_4 R_{i\hat{\theta}}^{V4, left} + \left(1 - R_{i\hat{\theta}}^{V4, left}\right) \cdot A_{i\hat{\theta}} \\ &\quad - \left(1 + R_{i\hat{\theta}}^{V4, left}\right) \cdot B_{i\hat{\theta}} \end{aligned} \quad (26)$$

$$\begin{aligned} \partial_t R_{i\hat{\theta}}^{V4, right} &= -\alpha_4 R_{i\hat{\theta}}^{V4, right} + \left(1 - R_{i\hat{\theta}}^{V4, right}\right) \cdot A_{i\hat{\theta}} \\ &\quad - \left(1 + R_{i\hat{\theta}}^{V4, right}\right) \cdot B_{i\hat{\theta}} \end{aligned} \quad (27)$$

with

$$A_{i\hat{\theta}} = \{R^{V2} * \mathcal{N}_{\sigma_4, \sigma_{4b}, x+\mathbf{p}}\}_{i\hat{\theta}} \quad (28)$$

$$B_{i\hat{\theta}} = \{R^{V2} * \mathcal{N}_{\sigma_4, \sigma_{4b}, x-\mathbf{p}}\}_{i\hat{\theta}} \quad (29)$$

$$\mathbf{p} = \omega^{V4} (\cos(i\hat{\theta}), \sin(i\hat{\theta}))^T \quad (30)$$

These responses are calculated at equilibrium and averaged subsequently, leading to the model V4 filter response

$$R_{i\hat{\theta}}^{V4} = \frac{1}{2} \frac{[(A_{i\hat{\theta}} - B_{i\hat{\theta}})]^+ + [B_{i\hat{\theta}} - A_{i\hat{\theta}}]^+}{\alpha_4 + A_{i\hat{\theta}} + B_{i\hat{\theta}}} \quad (31)$$

This integration mechanism yields a response for locally curved boundary segments at a larger spatial scale. For elongated contour segments that show no curvature, the response of individual cells will be equal and the combined response very low.

2.6. MODEL AREA IT

So far, we have described how our model integrated local features from model V1 into elongated, potentially curved boundaries at model V2–V4. Model area IT performs contextual integration that allows a segregation into figure and ground and a representation of prototypical objects at a large spatial scale. As discussed above, a population of V2 cells responds selectively to the direction of figure-ground direction. The local representation of border ownership at model V2 represents a set of available local hypotheses that cannot locally be resolved, as this step requires contextual influence from a larger spatial surround. Cells in IT cortex have been shown to be shape selective with properties generalizing over contrast polarity and mirror reversal (Baylis and Driver, 2001). The authors demonstrate that such cells do not, however, generalize over the assignments of figure-ground direction. The investigation supports the view that the population of probed IT cells is mainly driven by the sidedness of contours and less so by the contour itself. Given the rapidness of ownership selectivity observed in V2, we propose that ownership computation relies on a network of V2–V4–IT cell interaction. Our model uses local shape configuration in the outline of an object to collect confidence about the direction of figure and ground. We adopt an approach of Zhou et al. (2000) and model an integration cell at model IT that integrates border-ownership hypotheses from a larger spatial extent from model V2 input. For each location in the image, border ownership activations in a local neighborhood that point toward the inside of the respective receptive field contribute to the activation of an IT cell. This results in strong responses in model IT where local image regions are surrounded by contour convexities. Local activities of border ownership cells in model V2 then receive a positive enhancement if they contributed to such an integration process. This recurrent architecture resolves the initially ambiguous assignment of border ownership. Taken together, this makes the model belong to the class of feedback architectures according to the categorization in Williford and von der Heydt (2013). The response of cells and their interaction is denoted by the following equations:

$$R_{\mathbf{x}_0}^{IT} = \sum_i^{N-1} \sum_{\mathbf{p}} f(\mathbf{x}_0, \mathbf{p}, i\hat{\theta}) \cdot R_{i\hat{\theta}, \mathbf{x}_0 + \mathbf{p}}^{V2} \cdot \exp\left(-\frac{(\omega_{IT} - \|\overrightarrow{\mathbf{x}_0; \mathbf{x}_0 + \mathbf{p}}\|)^2}{\sigma_{IT}^2}\right) \quad (32)$$

with

$$f = \cos\left(\angle\left(\overrightarrow{\mathbf{x}_0; \mathbf{x}_0 + \mathbf{p}}, i\hat{\theta}\right)\right) \quad (33)$$

Such an IT cell at position \mathbf{x}_0 integrates responses of V2 cells in its proximity \mathbf{p} . The integration weight f depends on the angle between \mathbf{x}_0 and $\mathbf{x}_0 + \mathbf{p}$ and the currently integrated orientation $i\hat{\theta}$. This grants orientations parallel to an imaginary line toward x_0 high weights, while orthogonal orientations receive low weights.

This model area receives connections from the early as well as from the intermediate functional stages V1 and V2 where curvature is represented. This means that high-resolution local cues as well as contextual cues like corners from a larger region are available. A shape can thus be described as a set of contributing prototypical elements that contribute to the local configuration at every image location. Those elements are not solely generated through integration of lower areas, but exist as a distributed representation in all modeled areas and profit from mutual interaction through feedback and exhibit dynamic processes when a stimulus is presented.

2.7. FEEDBACK FOR CONTOUR ENHANCEMENT

The mechanisms so far presented contributed to the feedforward sweep of the model. We stated earlier that in visual cortex (and in neural processing in general), the input of cortical areas of higher stages highly contribute to the performance of individual earlier areas. By such recurring connections, contextual information is introduced in lower regions. We are thus now going to focus on the recurrent connections that are incorporated in our model.

Let's briefly recall that we model feedback connections that have a modulatory effect (Girard and Bullier, 1989) as outlined in Section 2, Equation 5. In **Figure 2** we illustrate how a feedback signal alone cannot elicit responses as long as no input activation is present. On the other hand, feedback that matches input configurations will increase those activations. We stick to this convention throughout our following elaborations.

V2 long-range and curved cells represent continuous straight or curved contours. Their multiplicative combination of receptive field subcomponents caused the cells to elicit responses whenever a contour of matching orientation was presented in their receptive fields. Now, those cells in V1 that contributed to the integration process will receive feedback and be thus increased in activity. The following non-linear transformation stage increases the difference in response strength with respect to other oriented contour cells that did not receive feedback. At the subsequent normalization stage, local response levels are now slightly increased by the recurrent input. Now, surrounding activations without feedback have a competitive disadvantage and receive a higher divisive normalization relative to their activation due to the increase response in their neighborhood that contributed to the sum. The dynamics of these interactions are denoted in formal terms. The enhancement of filter responses (Equation 12) via modulating feedback is defined by

$$\partial_t P_{i\hat{\theta}}^{V1} = -\alpha_1 P_{i\hat{\theta}}^{V1} + \beta_1 R_{i\hat{\theta}}^{V1} \cdot (1 + \lambda_1 \cdot \{R^{V2} * \mathcal{N}_{\sigma}^{FB}\}_{i\hat{\theta}}) - P_{i\hat{\theta}}^{V1} \cdot Q_{i\hat{\theta}}^{V1} \quad (34)$$

The subsequent competition to accomplish activity normalization is defined as

$$\partial_t Q_{i\hat{\theta}}^{V1} = -Q_{i\hat{\theta}}^{V1} + \{P^{V1} * \mathcal{N}_{\sigma}^{pool}\}_{i\hat{\theta}} \quad (35)$$

with TODO parameters. **Figure 3** shows an illustrated version of the mechanism with a small numerical model.

2.8. FEEDBACK FOR CURVATURE REPRESENTATION

As stated earlier, the modeled V4 cell do not at all or only marginally respond to straight elongated contours. Responses of V2 cells to curved boundaries are integrated in model V4, where integration cells sensitive to opposite sign of curvature mutually compete for equal orientations. These cells respond at positions with a local curved contour configuration, but are silent at elongated straight contours. Feedback is generated for those V2 cells that contribute to those curved boundary segments the corresponding model V4 cells respond to maximally. Regions with curved boundary segments thus elicit a strong response of V4 cells while regions with mostly straight contours do not elicit such a strong response. This signal can thus be used to differentiate regions of many straight contour segments from regions with many curved contours.

In formal terms, the V2–V4 cell interactions are defined by

$$R^{V2curv} = -\alpha R_{i\hat{\theta}}^{V2curv} + (1 - R_{i\hat{\theta}}^{V2curv}) \cdot A(1 + R_{i\hat{\theta}}) \quad (36)$$

$$A = \{R^{V2curv} * \mathcal{N}_\sigma\} \quad (37)$$

2.9. FEEDBACK FOR FIGURE-GROUND SEGREGATION

The contribution of feedback to figure-ground segregation is twofold in our model. First, local hypotheses of border ownership

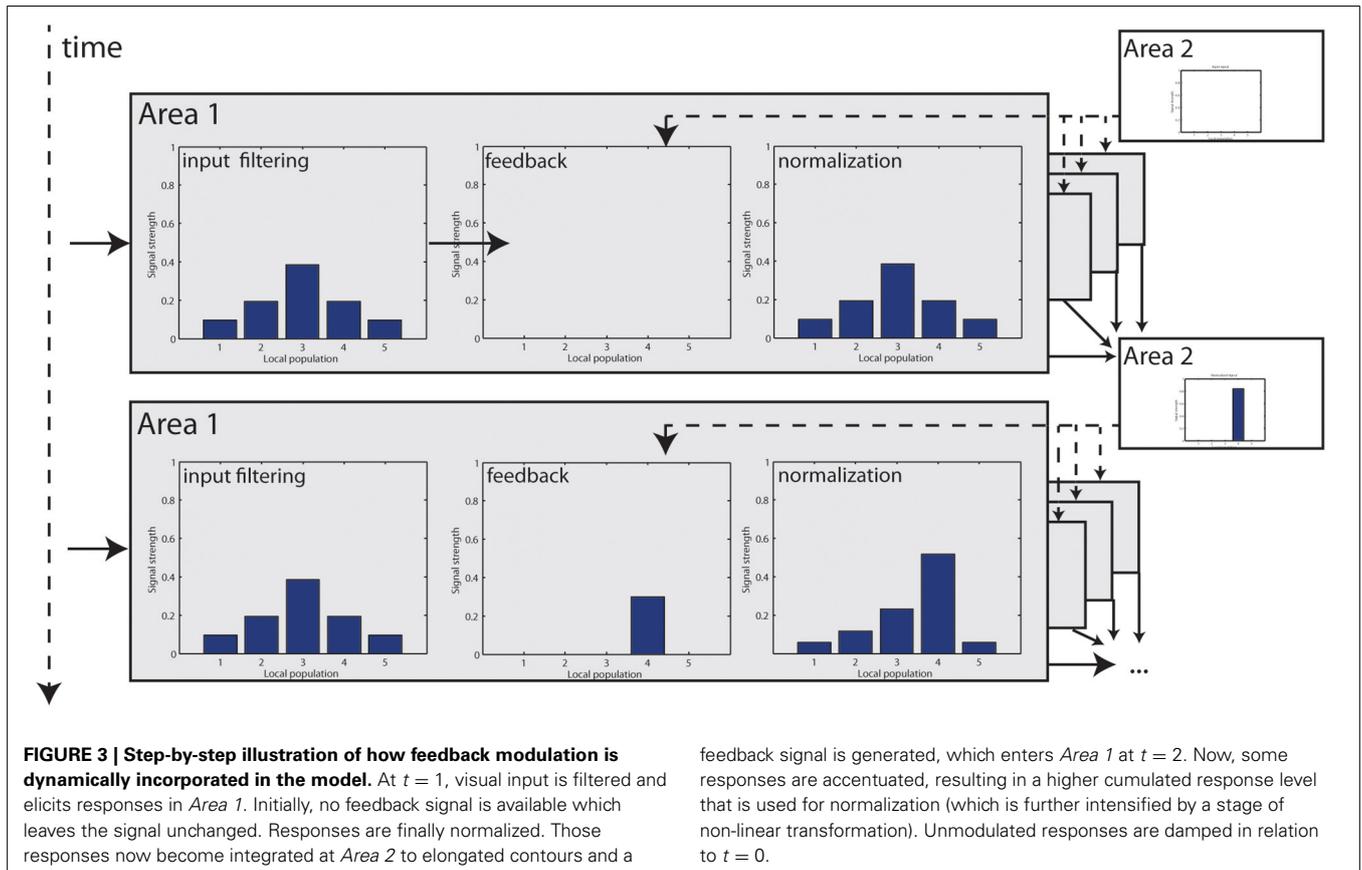
are generated by intra-area recurrent connections from long-range grouping cells. Contextual feedback from model IT resolves the remaining ambiguities. Initially, all directions of border ownership are equally likely at boundaries. With increasing confidence about local contrast orientations generated by V1 and V2, two options for border ownership directions are discarded and only two orthogonal border ownership directions remain. Activations of long-range V2 cells that indicate elongated surface boundaries and their orientation locally increase activities of those border ownership cells that are directed perpendicularly to the orientation of the boundary. Activity normalization for V2 border ownership cells then leads to a suppression of activities for ownership directions orthogonal to the boundary orientation. Formally, this is accomplished by the dynamics

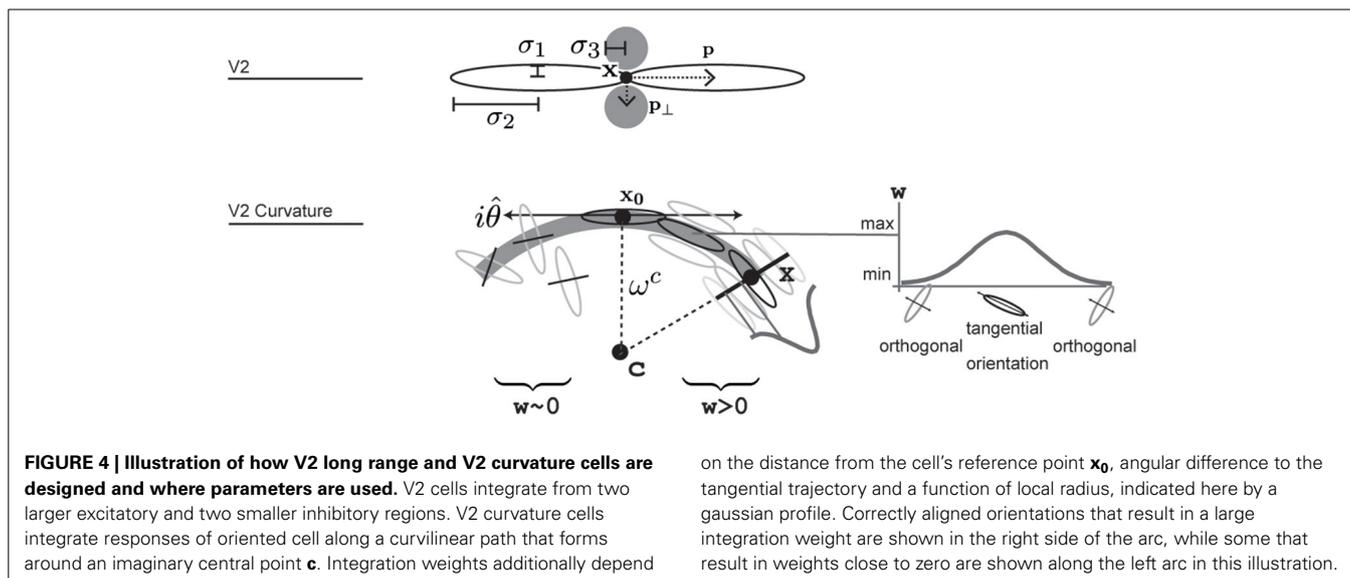
$$\begin{aligned} \partial_t R_{i\hat{\phi}}^{BOwn} = & -R_{i\hat{\phi}}^{BOwn} + \beta(R_{i\hat{\theta}}^{V2} + h_{tonic}) \\ & - R_{i\hat{\phi}}^{BOwn} \cdot \sum_{\gamma} R_{i\hat{\phi}}^{BOwn} \end{aligned} \quad (38)$$

with

$$\theta = \phi + \frac{1}{2} \text{ mod } \pi. \quad (39)$$

Second, V2 border ownership cells receive feedback from cells in model IT. Here, border ownership as well as figural cues, e.g., from local junctions, or curvature maxima, were integrated by





IT cells. For the correct inference of figure and ground, feedback from IT to V2 is essential. Figure-Ground cells at IT level integrate border ownership activations from V2 in a circular fashion to integrate the coherence of directions indicating a convex pattern of figure outline. In the feedback sweep, this contextual information is now fed back to these border ownership cells compatible with the configuration using recurrent connections. In formal terms, this extends the dynamics presented in Equation 38 above by incorporating a modulating feedback signal from model IT cells, namely

$$\begin{aligned} \partial_t R_{i\hat{\theta}}^{BOwn} = & -R_{i\hat{\theta}}^{BOwn} + \beta(R_{i\hat{\theta}}^{V2} + h_{tonic}) \cdot (1 + \lambda_2 \cdot R_{i\hat{\theta}}^{IT}) \\ & - R_{i\hat{\theta}}^{BOwn} \cdot \sum_{\gamma} R_{i\hat{\theta}}^{BOwn} \end{aligned} \quad (40)$$

This also concludes the feedback sweep of our recurrent model. In the following section, we will show the performance of the model and its individual areas in the Results Section.

3. RESULTS

In this section we illustrate the capabilities of our model in a number of simulations. To demonstrate how the model processes shapes, we use some artificial images to show working principles of various subcomponents of our model. These simple shapes were taken from the *Webdings* font freely available with a Microsoft® Windows™ 8.1 operating system. We also include also a depiction of a *Kanizsa* square (Kanizsa, 1955). This is a special stimulus because it elicits the perception of illusory contours at the outline of the occluding square, a sensation our model is also capable to represent.

To demonstrate the abilities of our model to process real world images we acquired the dataset of Fowlkes et al. (2007) and selected a few examples that we included in our Results Sections. These images have a resolution of 321×481 pixels in landscape or portrait orientation. They were converted to grayscale images using the Mathworks® Matlab® `rgb2gray`

function which performs a perceptually weighted combination of the red, green and blue channel. We used 8–12 iteration steps to allow recurrent feedback signals to build up. The angular resolution of cell populations is defined by selecting eight $\frac{\pi}{8}$ steps to encode orientation. Border ownership is represented by a population representing 4 directions. Model V2 curvature cells also used 8 orientations for tangential orientations, but due to two possible curvature directions, our model contains a population of 16 curvature cells. A list of parameters used is given in **Table 1**.

3.1. EARLY PROCESSING STAGES

To begin with, we show how the processing at early stages achieves a representation of the stimulus concerning contrasts and elongated contours. Local contrasts are represented in the early stages by model V1 and V2 cells. However, as can be seen in **Figure 5** the responses rapidly change in the first few iteration steps. The contained contour as well as the added noise signal both elicit responses at the V1 level (*second column*) and cause the shapes outline to be not clearly separated from the background. However, those responses are grouped into elongated contour representations in model V2 (*4th column*). Elongated contour segments are clearly emphasized. From these V2 activations, a recurrent feedback signal is generated that modulates V1 activations. After a few iterations, the representation at V1 dramatically changed, with the outline of the figure now clearly visible.

The effect of the feedback signal is also measurable in a quantitative way, see **Figure 5, right**. Along the boundary of an object we plotted the activation levels of the population of V1 neurons that represent the orientation. Initially, the neuron with preferred orientation responds best, but also those with orientation tunings close to the real contour (*first plot*). The situation changes when feedback is added (*second plot*). Now, representations of undesired orientations are attenuated and the activation of the cell representing the contextually valid orientation is highly increased.

Also in **Figure 5**, the representation of illusory contours at V2 stage is depicted. This is illustrated using an input depicting a *Kanizsa* square (*last row*). A complete square is highly salient

Table 1 | General model parameters used for simulations.

| Description | See equation | Value |
|--|--------------|------------------|
| Number of orientations used | | 8 |
| Number of feedback iterations | | 6 |
| Number of BOwn directions | | 4 |
| MODEL AREA V1 | | |
| Network size | | 321×481 |
| LGN σ | 10 | 1.00 |
| LGN κ | | 1.50 |
| LGN normalization α | (9 applied) | $2.17e-03$ |
| LGN normalization β | (9 applied) | $2.17e-03$ |
| V1 contrast σ_1 | 12 | 0.23 |
| V1 contrast σ_2 | | 0.12 |
| for p : Excentricity ω_1 | | 3.00 |
| V1 normalization α | (9 applied) | $2.16e-06$ |
| Non-linearity of V1 responses | (6 applied) | 4.00 |
| MODEL AREA V2/V3 | | |
| V1:V2 subsampling | 14 | 1:3 |
| Network size | | 107×161 |
| Filter size of V2 complex | | 41.00 |
| V2 complex cell $\sigma_1, \sigma_2, \sigma_3$ | | 0.21,0.02,0.10 |
| in p : ω_{ex}^{V2} | | 3.70 |
| in p : ω_{inh}^{V2} | | 2.50 |
| V2 inhibition strength γ | | 0.10 |
| V2 nonlinearity k | (6 applied) | 1.00 |
| Strength of V2-V1 feedback | | 0.11 |
| Non-linearity of BOwn | | 1.00 |
| In c : curvature radius ω^c | 19 | 15.00 |
| in w^{dist} : σ_1 | | 40.00 |
| in w_{ori} : σ_2 | | 2.00 |
| Strength of R^{V2Curv} feedback | | 0.15 |
| MODEL AREA V4 | | |
| V1:V4 subsampling | | 1:4 |
| Network size | | 81×121 |
| V4 filter size | | 31.00 |
| α_4 | 26 | 0.01 |
| σ_4, σ_{4b} | 28 | 0.43,1.35 |
| ω^{V4} | 28 | -1.00 |
| MODEL AREA IT | | |
| σ_{IT} | 32 | 0.43 |
| σ | | 29.00 |
| ω_{IT} | | 17.00 |
| IT Non-linearity | | 3.00 |
| α | (9 applied) | $5.49e-05$ |
| Strength of BOwn feedback | | 50.00 |

for human observers despite the fact that only a series of circles with cut-out corners are depicted. This is reflected in the grouping responses of V2 neurons, they also show activity in the gap between the real contour fragments. **Figure 5** shows V2 responses for the same parameter set and for a parameter set with changed receptive field sizes, to illustrate the effect even stronger (*framed part*). **Figure 6** shows a result of the corner representation in the model.

3.2. CURVATURE TUNING

Figure 7 illustrates the tuning functions we defined for model V2 curved cells. A curved cell with distinct radius tuning was selected and we presented arcs of different curvature to this cell and simulated the response. We performed this for four cells with curvature tuning to 10, 15, 20, and 25 pixels radius. This curvature definition happens in V2, where the initial resolution of the image had been subsampled. For this reason, the value here correspond to values 40, 60, 80, and 100 in V1 resolution. In each plot, the peak response occurred when the stimulus with the matching radius was presented. In this simulation, subsampling artifacts cause the first two plots to elicit some discontinuities.

3.3. SHAPE REPRESENTATION

In the final setup, we show how our model independently represents different elements of a shape, and how this depends on the recurrent feedback connections. **Figure 8** illustrates the results we achieved for an artificial image. Initially, we configured the model to only use feedforward connections from V1 to V2. The model only achieves an representation at model V1 and a representation at V2 where the elongated boundaries are visible, but surrounded by many spurious activations. When recurrent feedback from V2 is added, the representation at V1 improves in the first few iterations before a steady representation is reached. In parallel, elongated boundaries at V2 are integrated and noise is highly reduced.

To represent prototypical objects at an intermediate level of detail, we stated that the model needs to represent different contour properties. In the second row of **Figure 8** we show how the model achieves to emphasize V1 responses when they contribute to a certain contour fragment with desired properties. We deliberately exaggerated the effect and chose a very narrow tuning so that all other responses become almost completely suppressed. On the *left* side, we let the model emphasize contour parts that are oriented almost vertical but in a curved context of a matching radius. As can be seen, the model highlights that parts on the left side of the stimulus that matches and leaves others suppressed, even if their local orientation would match. On the *right* side of **Figure 8**, we perform the same for a different part of the shape outline.

In **Figure 9** we perform the same selection for a realistic photograph depicting an elephant. On the *left* side, we show interaction of model V1-V2 causes an appealing representation of the animal at stages V1-V2. On the *right* side, we configured the model using model area V4 to emphasize parts of the outline of the animal that match a certain context and configuration, here, a part of the outline.

3.4. BORDER OWNERSHIP AND FIGURE-GROUND ASSIGNMENT

In the segregation of a scene into figure and ground the modeled *border ownership* cells participate by indicating the direction where the frontal surface is positioned at a boundary (Zhou et al., 2000). Our model incorporates a mechanism using such border-ownership cells to resolve the direction of a frontal surface from local boundary cues (Zhou et al., 2000). We performed such a assignment for our sample images, see **Figure 10** for an

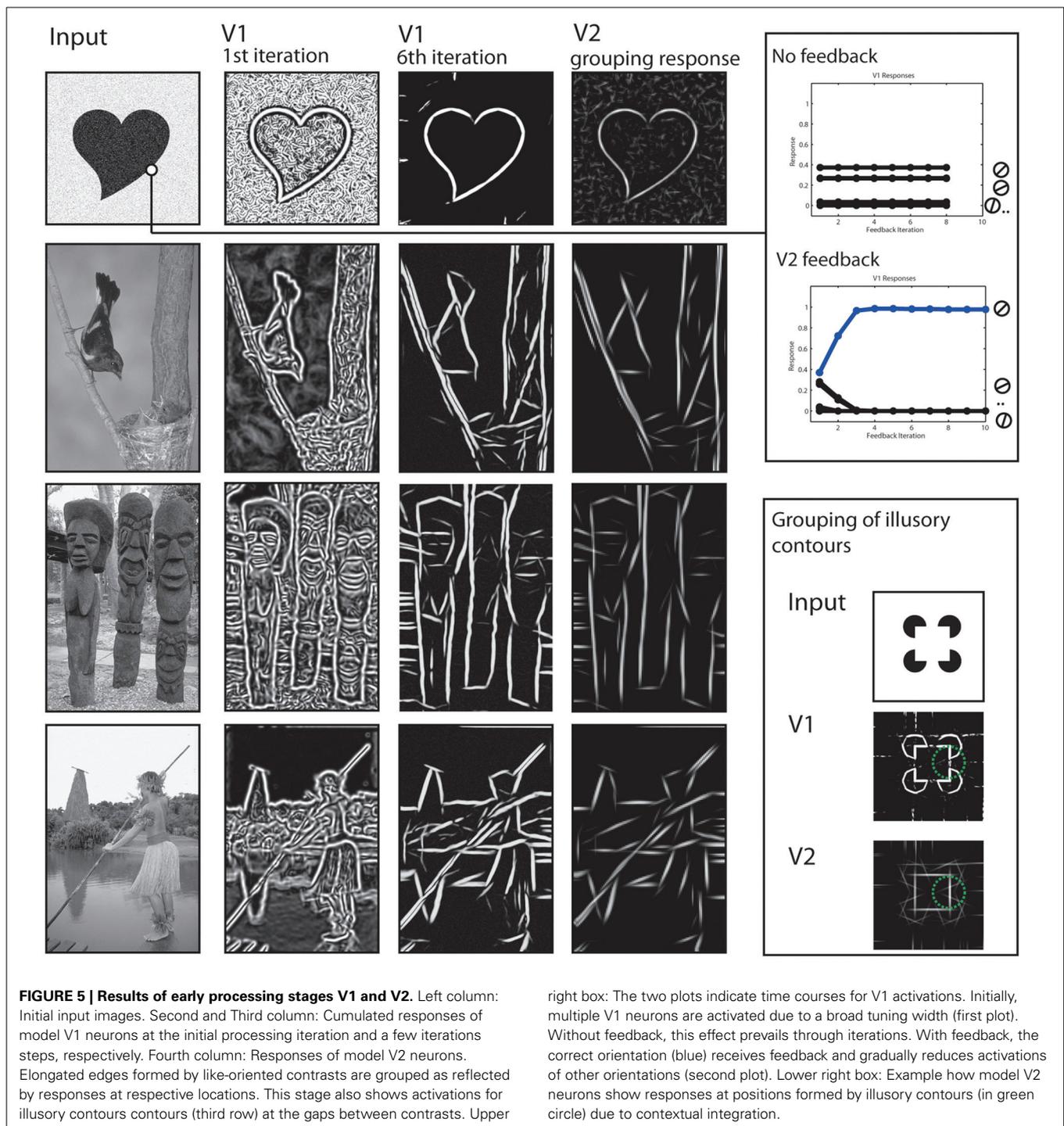
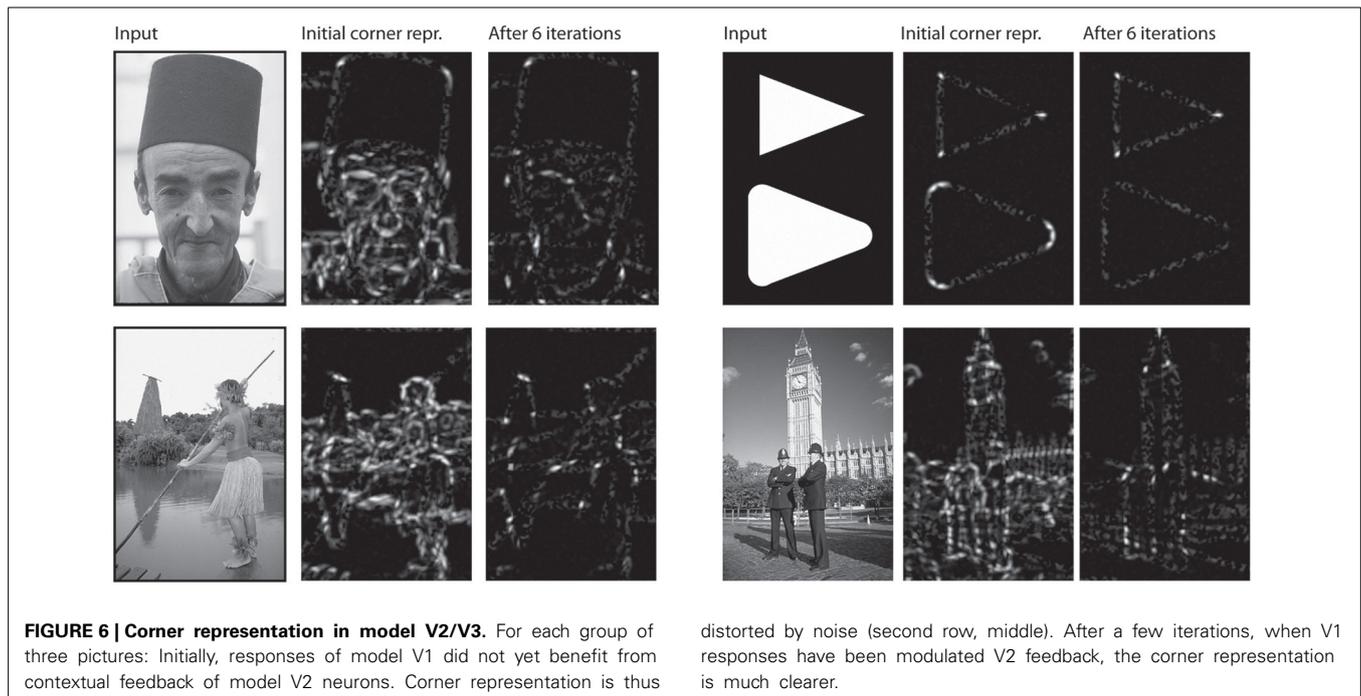


illustration of the result. The output of model area V1 and of V2 long-range integration cells are acquired to generate initial hypotheses of border ownership direction at image regions where local contrasts are situated. Initially, all four border ownership directions show equal responses at a boundary location. After stimulus onset, three dynamic effects occur and their contribution to the resolution of border ownership is reflected in the time course of cell activation, see **Figure 10** for an illustration.

First, local feedback from V2 cells enhances two hypotheses of border ownership for the directions orthogonal to the local boundary orientations.

A local normalization causes an attenuation of the other two representations **Figure 10**, *second row*; Timestep 0 and 1). Second, shape-level integration at model area IT contributes positive feedback to those border ownership cells that are directed toward the inside of the figural depiction. Again, normalization leaves the



net response of the cells constant (timesteps 2–4). Finally, mutual inhibition among border ownership cells with opposite direction selectivity causes the dominant direction to gain all available net energy (timestep 5–8). At this point, a stable point is reached and the local ambiguity for border ownership direction is resolved using feedback from higher cortical areas. The interpretation of the final representation would be that the frontal surface is to the inside of the curved boundary.

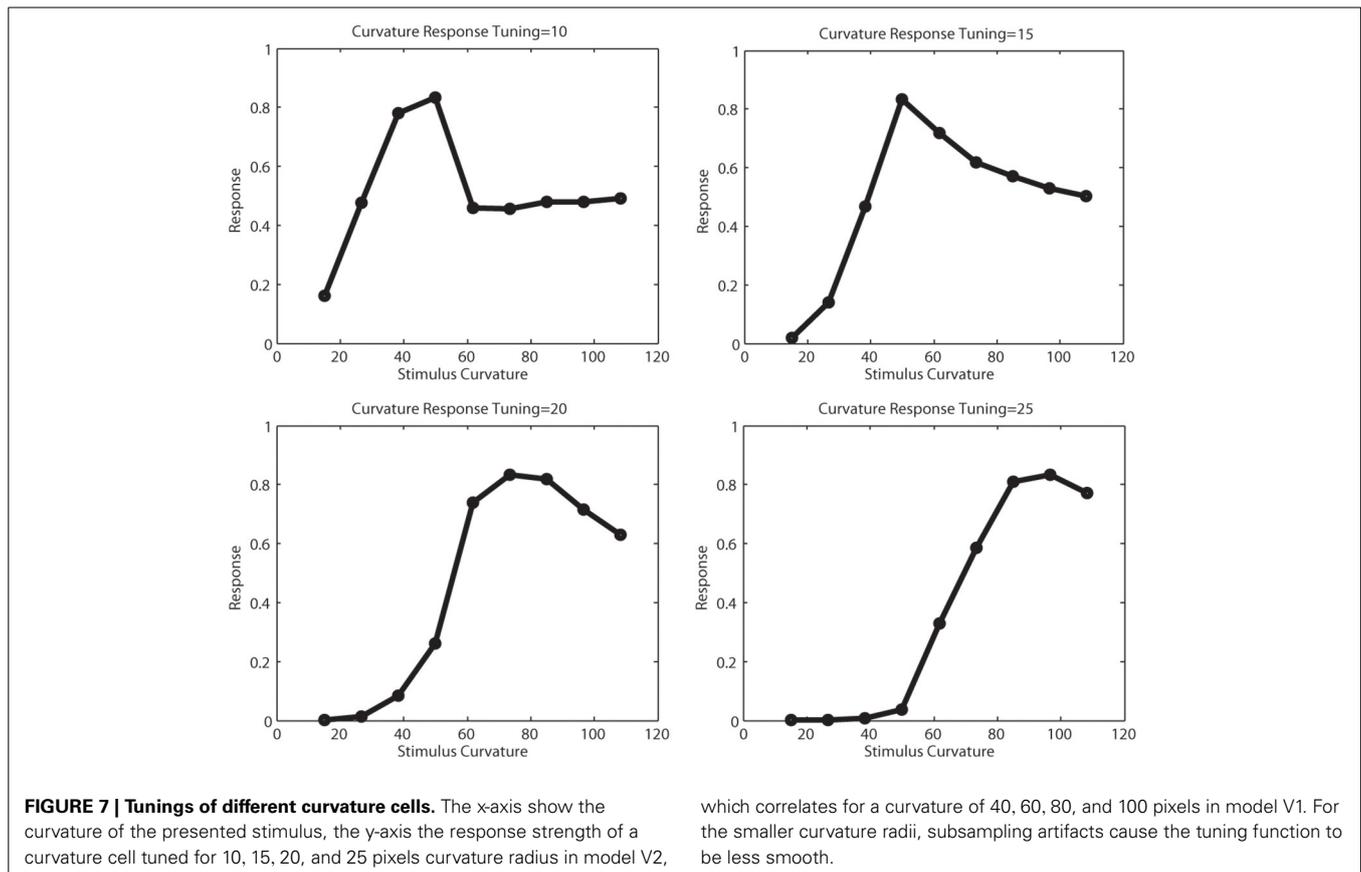
4. DISCUSSION

4.1. SUMMARY OF CONTRIBUTIONS

In this contribution we emphasized the role hierarchical representations have in the organization of shape features and their combinations into a coherent form. Like some previous model developments (Cadiou et al., 2007; Hatori and Sakai, 2012; Rodríguez-Sánchez and Tsotsos, 2012) our model is based on low and intermediate representations of shape features. These proposals are all based on a strictly hierarchical feedforward processing sequence. We propose here that such shape encoding mechanisms may be based on distributed representations that are established by interacting assemblies each devoted to specific feature properties. Such interactions in the model are organized by recurrent interactions of feedforward and feedback signals. The underlying structural principles are based on the cortical architecture of the ventral pathway with mutual interactions between such distributed representations (Markov et al., 2013). The model architecture incorporates principles that have been predicted to minimize the computational efforts of visual systems to successfully deal with the complexity problem of perception (Tsotsos, 1988) [compare also (Tsotsos, 2005)]. Among those, the hierarchical organization of representations in model areas, the specific receptive field properties of model columnar

mechanisms, hierarchical pooling of spatially separated input representations, and top-down (modulatory) feedback are proposed here to account for the functional properties of cortical shape processing. We did not discuss complexity advantages in this contribution. However, given the theoretical predictions by such earlier work our proposal of a model architecture provides a evidence how distributed intermediate-level mechanisms may help to shape our understanding of modeling complex visual machinery that captures key cortical principles.

The main contributions of the work presented in the manuscript are twofold. First, we propose a computational network architecture that utilizes a hierarchical distributed representation of shape features. Contour features play a major role to track moving shape in which their strength parametrically change as a function of their saliency (Caplovitz and Tse, 2007). This necessitates global configurational as well as local information to distinguish rather tiny differences in the outline of a 2-dimensional form [such as curved boundaries vs. localized corners (Pasupathy and Connor, 1999; Ito and Komatsu, 2004)]. In order to generate a representation with sufficient spatial resolution combined with spatial context we suggest that multiple specialized component representations interact by feedforward hierarchical processing that is combined with feedback from representations generated at higher stages in the hierarchy. Second, we incorporate grouping mechanisms to integrate like-oriented contour responses that are integrated if they form a smooth outline fragment of a surface boundary (e.g., Grossberg and Mingolla, 1985; Neumann and Sepp, 1999; Ben-Shahar and Zucker, 2004). Such grouping mechanisms operate at the stage of area V2 and are, thus, involved in the hierarchical processing of shape. Given the hierarchical processing and representation of boundary information in the ventral pathway (see the overview

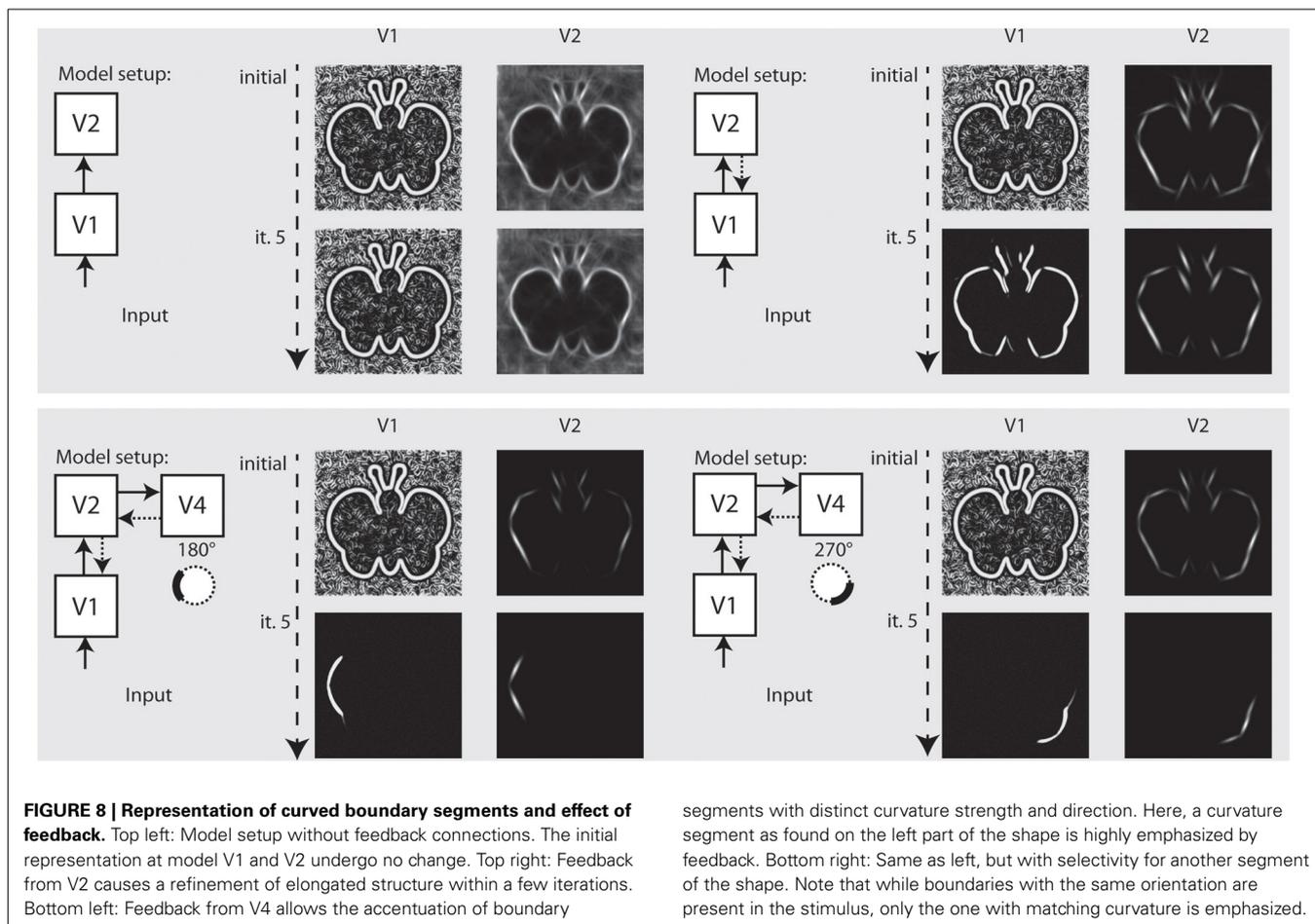


in Neumann et al., 2007) the shape processing observed in area V4 is mainly driven by the output of grouping responses. It may be supplemented by input from simple/complex cells in V1, a principle of convergent signal streams also used in the models described in Thielscher and Neumann (2003); Rodríguez-Sánchez and Tsotsos (2012). In addition, we suggest that the shape representation built at the stages of V4 and IT influences the assignment of border ownership in surface representation (Zhou et al., 2000) (see overview in Neumann et al., 2007). Model IT cells send modulatory feedback to those V4 cells that provide relevant input (in V4 and V2) such that the net sum of convex corners/curvatures determines the ownership direction. The proposed model thus combines separate findings about the generation of cortical shape representation with figure-ground segregation mechanisms by assigning border ownership.

4.2. RELATION TO PREVIOUS MODELS OF SHAPE REPRESENTATIONS IN CORTEX

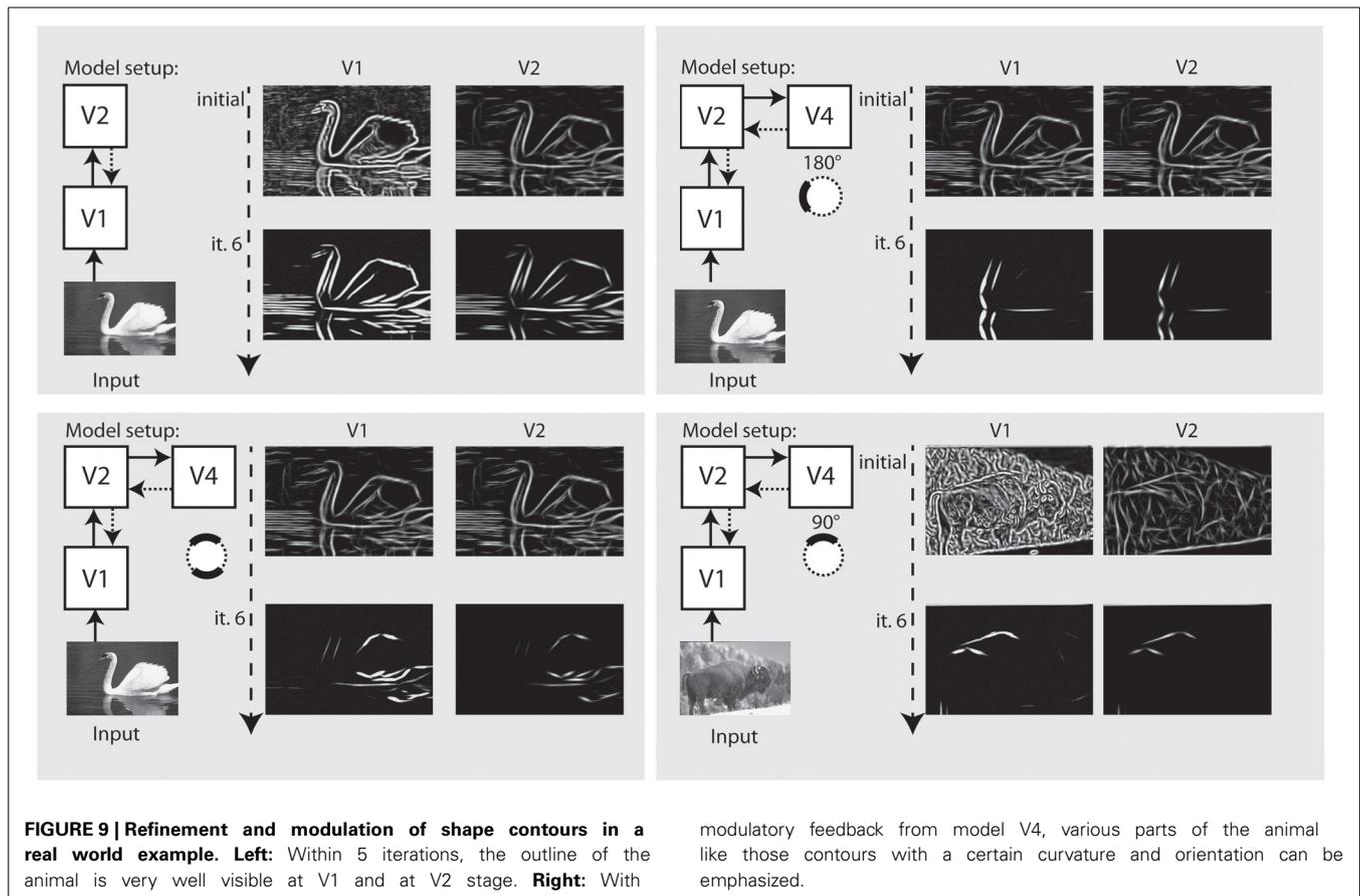
Visual shape recognition has already been investigated intensively by considering the 3-dimensional (3D) surface appearance for object recognition (Riesenhuber and Poggio, 1999; Serre et al., 2007; Mutch and Lowe, 2008; Yamane et al., 2008; Serre and Poggio, 2010) as well as 2-dimensional (2D) shape recognition (Schwartz et al., 1983; Mokhtarian and Mackworth, 1986; Mokhtarian, 1995; Rodríguez-Sánchez and Tsotsos, 2012).

In the context of view-based models of object recognition stable views (Logothetis et al., 1995) are associated with 2D shapes so that their analysis can be considered as an intermediate stage of object processing (Cadieu et al., 2007). The computational model approaches of 2D shape representation can be subdivided into flat and hierarchical schemes. Examples of flat processing schemes, e.g., utilize Fourier descriptors (Schwartz et al., 1983), multi-scale representations of curvature features in the shape outline (Mokhtarian and Mackworth, 1986; Mokhtarian, 1995), or global schemes for integrating oriented line features (Wilson and Wilkinson, 1998). Hierarchical multi-layer processing schemes are based on different stages to generate an increasingly coarse-grained representation of shape features utilizing repetitive application of local filtering operations (Riesenhuber and Poggio, 1999; Cadieu et al., 2007; Rodríguez-Sánchez and Tsotsos, 2012). In order to resemble the feature selectivity of V4 cells in monkey cortex such cells build coarse-grained orientation-curvature representation of the shape under inspection. The hierarchical organization of a sequence of processing stages follows the idea of the Neocognitron (Fukushima, 1980, 1988) by developing low and intermediate representations of richer shape feature compositions (LeCun et al., 1998; Riesenhuber and Poggio, 1999; Mutch and Lowe, 2008; Tabernik et al., 2014). The orientation-curvature representation of V4 cells reported by Pasupathy and Connor (1999); Connor et al. (2007) has been investigated in the models reported in Cadieu et al.



(2007); Rodríguez-Sánchez and Tsotsos (2012); Hatori and Sakai (2012). We share the principles of the hierarchical organization of processing and the emergence of rich orientation-curvature sensitivity in our proposal. Initial processing utilizes orientation sensitive filters to extract local oriented contrast. Unlike the previous models we incorporate a stage of boundary grouping at the interface between low and intermediate levels of representation. Such grouping operations integrate oriented contrast responses that are arranged in the local neighborhood of a target location. The local responses are enhanced by evaluating a support function that measures feature compatibility [(Neumann and Mingolla, 2001) for an overview and taxonomy of grouping schemes]. The measure of compatibility, or reliability, depends on the lateral integration that utilizes oriented weighting functions for contrast features arranged along a model shape outline, e.g., circular arcs with different radii (Parent and Zucker, 1989). Such a scheme thus implicitly incorporates curvature as a local contour feature. In order to make this explicit, different contour radii and signs of curvature (for individual orientations) have been considered in Rodríguez-Sánchez and Tsotsos (2012). Rather than implementing this curvature selectivity in a hard-wired scheme of local oriented filter conjunctions, we propose that this selectivity is generated via bottom-up and top-down filter mechanisms organized in a hierarchy. In this architecture

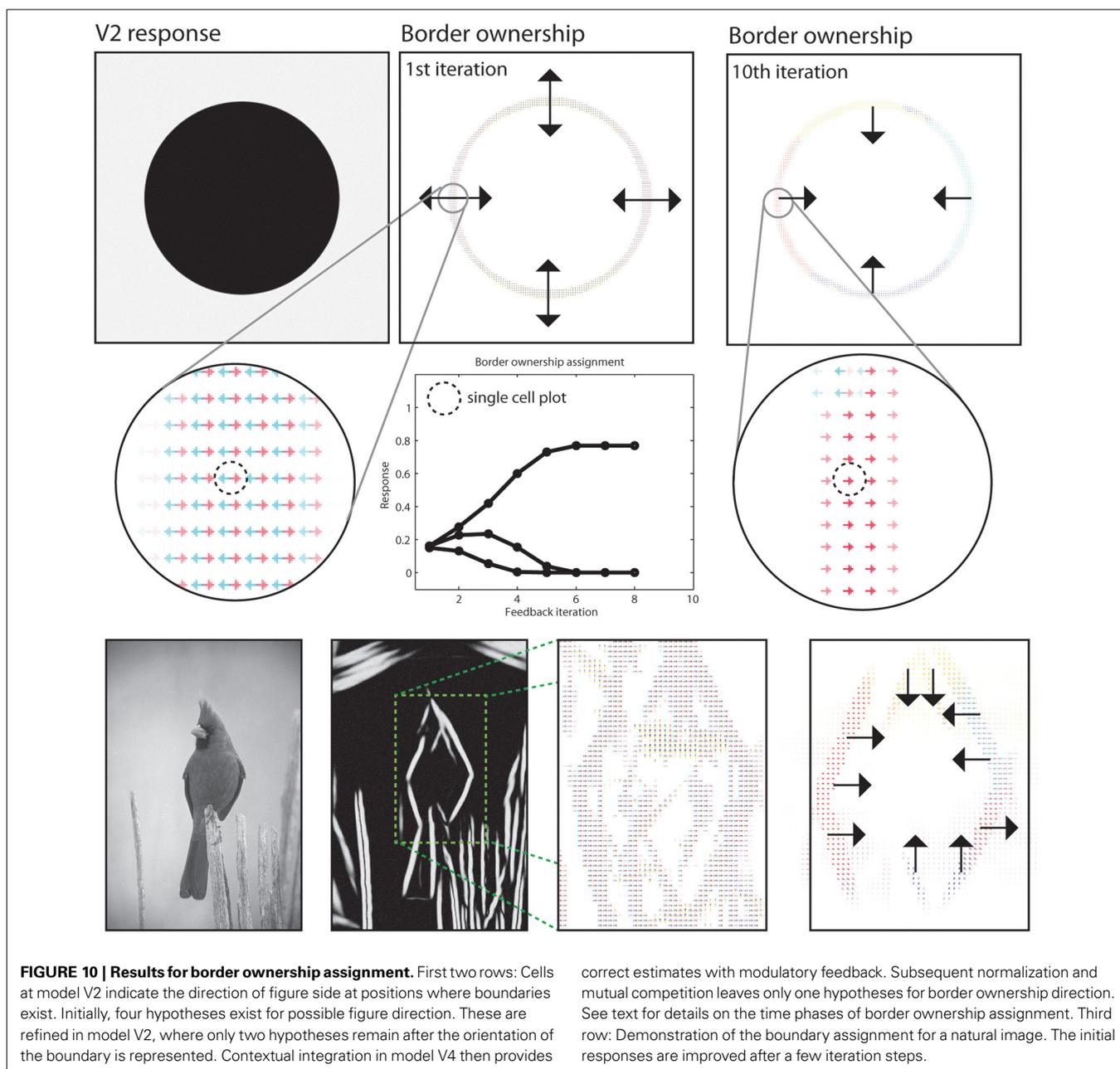
the responses from model V2 contour groupings (based on different radii) are integrated by model V4 curvature sensitive cells with coarse bipartite odd-symmetric receptive fields (similar to simple cell profiles, but at much larger spatial scale). The sign of curvature is distinguished by cells of opposite polarity that mutually compete for each orientation. As a consequence responses are generated preferentially in cases where a single dominant curvature is present while responses are suppressed for straight contours which feed curvature cells symmetrically. The curvature radius is represented through a family of differently scaled integration sizes of such model V4 cells. Each of these cells have a specific peak selectivity. In the simulations we used three different sizes for each curvature sign. In order to make those cell responses selective to the feature specificity but mainly invariant to luminance contrast we suggested that each V4 cell response competes against the responses of other curvature selective cells in a local pool that interact via a mechanism of shunting inhibition. This leads to normalization of responses just like in those mechanisms proposed to account for various non-linearities at different stages in cortical processing, e.g., for context related contour responses in V1 (Carandini and Heeger, 1994; Carandini et al., 1999), attention selection (Carandini and Heeger, 2012), and higher level cognitive functions (Louie et al., 2011). Since the curvature sensitive model V4 cells, in turn, send feedback



to their input contour representations in model V2 and filter response in model V1 those corresponding input activations will be enhanced. The amplitude of responses in distributed boundary representations will be amplified as an emergent net effect such that local salient curvature features in a shape outline will be amplified to yield distributed component feature representations of figural shapes.

These local boundary and curvature representations also feed mechanisms of border ownership assignment at the level of the model V4/IT complex. Such mechanisms have been investigated before in e.g., Zhou et al. (2000); O'Herron and von der Heydt (2011). Our computational framework belongs to the group of feedback models for border ownership encoding (see the overview of the current state in Williford and von der Heydt, 2013, see discussion below). We adopted this generic scheme by integrating responses from curvature selective cells with the compatible sign of curvature. In such a way the ownership configuration favors contributions from coarsely presented convex components. If a shape with multiple convex and concave segments is present then the ownership cells with opponent direction selectivities compete in order to arrive at a disambiguated assignment of surface belongingness. This makes the testable prediction that bumpy outlines should lead to slightly longer ownership disambiguation than for smooth convex shapes since the disambiguation will take more time when initially opposite assignment hypotheses coexist.

An additional investigation was argued to be of importance in the work proposed here. Several experimental investigations have reported that cells in extra-striate cortex selectively respond to corner junctions. For example, Ito and Komatsu (2004) (compare also Hegdé and Van Essen, 2000) reported that cells in area V2 selectively respond as to generate representations of sharp corners, or angles, selective for a particular opening angle. Similarly, Pasupathy and Connor (1999); Yau et al. (2013) show that area V4 cells respond to sharp shape corners with a sub-population of cells preferring sharp corners with different orientation and opening angles while another sub-population prefers smooth rounded corners. While the previous hierarchical models can account for the response selectivity for any of these generic corner types the perceptual representation of sharp localized features that allow, e.g., to distinguish between sharp and rounded corners remain unanswered. Sharp corners of any opening angle would be indistinguishable from the smooth variants of these corners given the increasing smoothing and subsampling of the visual representation while proceeding in the hierarchy. Our model argues in favor of a distributed representation: While shape sensitive cells at an intermediate level represent the salient shape protrusions (as in V4) the localized detail of an outline is represented at a higher spatial resolution in lower-level representations, e.g., in V1, V2, V3. In our model we suggest representations of smooth boundaries with different curvatures represented by groupings in model V2 while



sharp corners are implicitly represented by convergent V1 input in local representations in model V2/V3. We assume that responses of cells in the model V2/V3 complex mutually compete such that their energy provides a measure to normalize individual responses. These provide convergent input to curvature selective contour cells in model V4 which, in turn, send feedback signals to their input sites at preceding stages. Since they are driven by either smooth or sharp contour arrangements the interaction of bottom-up sensory and top-down context-driven signals leads to selective enhancement of the particular corner configuration in the present stimulus. The specific details of the interaction between such counter-stream signal flows are discussed below.

4.3. FEEDBACK AS PREDICTION MECHANISM TO LINK SHAPE COMPONENTS

The hierarchical model architecture proposed here is composed of multiple model areas each of which is represented by a three-stage columnar cascade model. In a nutshell, the model cascade consists of (i) an initial stage of input filtering, (ii) a stage of activity modulation of filter outputs by top-down or lateral re-entrant signals, and (iii) a stage of center-surround interaction of target cells against an inhibitory pool of cells leading to activity normalization to generate the net output response of the model area. These three stages can be roughly mapped onto compartments of cortical area subdivisions (as suggested in Self et al., 2012). The filtering stage of the driving feedforward input signals is specific

to the particular (model) area under consideration. At the output stage, the activity normalization is computed by a mechanism of shunting inhibition, like the non-linear divisive mechanisms proposed in Carandini and Heeger (1994); Carandini et al. (1999); Kouh and Poggio (2008); Carandini and Heeger (2012). The feedback signal is generated at higher-level cortical stages or parallel processing pathways and is thought to provide context information that is re-entered at the stage earlier in the processing hierarchy (Grossberg, 1980; Edelman, 1993).

The functional role feedback signals play still remains controversial. Different proposals how feedback signals interact and combine with the driving feedforward stream have been discussed in the literature which have received different support from the experimental literature (Markov et al., 2013). One such framework proposes that the goal of computation is to reduce the residual error between the different signal streams in order to approach the sensory prediction generated by higher stages of processing (Ullman, 1995; Bastos et al., 2012). This idea is rooted in the Bayesian theory of predictor-corrector mechanisms which yields to the Kalman optimal filter realization under some restricting assumptions (Rao and Ballard, 1999). We follow an alternative route in which the feedback mechanism is modulatory in nature. Unlike predictive coding which tried to drive the difference between driving signals and the prediction to zero bottom-up input signals are amplified by matching feedback signals. This leads to a gain enhancement for those cell responses where a matching top-down predictive signal template has been generated. 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; (Girard and Bullier, 1989; Desimone, 1998; Roelfsema et al., 2002; Reynolds and Heeger, 2009)]. The modulation mechanism is reminiscent of the linking mechanism suggested by Eckhorn et al. (1990); Eckhorn (1999) to account for activity synchronization in networks of spiking neurons. We have recently demonstrated (Brosch and Neumann, 2014) that such mechanism of convergent bottom-up feedforward and top-down feedback signal correlation accounts for the signal amplification as measured at the level of cortical pyramidal cells (Larkum, 2013).

In the shape processing architecture described here the modulatory feedback serves the role of a predictor (Spratling, 2008). For example, bottom-up input in oriented contrast is integrated by mechanisms of contour grouping and integration to generate continuous boundary representations. This is similar in spirit as the recent investigation of Piëch et al. (2013) who emphasized how context information at higher cortical stages influence more local feature representation at lower levels. Here, the same principle is replicated over different stages of model cortical processing. Contour representations after grouping in model V2 and junction configurations in model 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 that have generated their input. The computational logic is that the curvature responses provide a template of context-related information about the local presence of oriented shape features. The modulatory feedback amplifies those inputs

that are consistent with the curvature feature representation. The mutual competition of responses in a pool of cells at the lower level leads to a suppression of inputs that do not contribute to the present curvature feature. In all, a distributed representation of shape information is created that contains coarse-grained configurational information about stimulus shape and, at the same time, the spatially localized detail needed to distinguish between sharp and smooth corners. Similarly, the action of feedback sent from ownership sensitive cells (in the V4/IT complex of the model) to curvature sensitive and grouping cells in model V2 and V4 also provides context information for the assignment of configurational information. Here, the ownership assignment is based on the consolidation of evidence which convex shape elements make to establish a closed shape region in the visual field. This context is delivered via feedback to their input that represents fragments of shape components (irrespective of the sign of curvature) and also to the grouping representations. Those shapes that finally receive assigned direction of border ownership, and thus figure-ground direction, will enhance the associated inputs at the intermediate level orientation-curvature representations.

In all, the hierarchical processing scheme proposed here relies on extensive bidirectional flow of information in which the feedback signals that represent context-sensitive templates are gated by feedforward driving input signals. Such a modulating feedback driven gain control mechanism relates to mechanisms proposed by Roelfsema and colleagues (Lamme and Roelfsema, 2000; Roelfsema et al., 2002; Roelfsema, 2006) in which spatial detail is generated by feature-driven low-level processes and representations and subsequently associated with coarse-grained context information provided by intermediate and higher-levels of cortical computation. The mechanisms implemented in the proposed model are consistent with theoretical predictions from computational constraints visual perception imposes on the underlying architecture (Tsotsos, 1988). The advantages in computational complexity have been calculated for principles such as hierarchical organization, localized receptive field computations, and dedicated (distributed) maps of feature representation and their combination. Feedback has been suggested to steer an attentional beam by selecting a spatial region and their computational resources (Tsotsos, 2005). In the proposed architecture feedback also selectively enhances representations of features by increasing their gain which are coherent with the predictions generated at higher-level stages with more condensed coding of shape and figural properties. Also we emphasize that this provides a key to enhance (and make accessible) localized shape features, such as sharp edges, as part of a shape configuration that is represented on a coarser scale.

4.4. MODEL LIMITATIONS AND FURTHER EXTENSIONS

The proposed model architecture emphasized the computational role of feedforward and feedback mechanisms in order to generate a hierarchical distributed representation of shape information. For that reason, we focused on the representational aspects as steady-state solutions of an otherwise dynamic interaction between neuronal populations and representations distributed over several model areas. We did not, so far, investigate the temporal response phases observed for shape sensitive cells in V4

(Yau et al., 2013). The work of Roelfsema and colleagues has shown that different response phases exist that can be reliably assigned to different mechanisms in processing, namely for feature detection, figure-ground segregation, and attention (Roelfsema et al., 2007). We have demonstrated that such separate but temporally overlapping phases can be accounted for by a recurrent network of mutually interacting neuronal sites. The network model has been composed of the same components like the present model architecture (Raudies and Neumann, 2010). It would thus be interesting to reveal whether similar temporal phases can be identified for model V4 cells that may give rise to identify different signatures indicative of contributions from delayed neuronal mechanisms that are involved in the computation of figural shape information.

Different signal streams (particularly in the feedforward sweep of feature processing) operate on different temporal scales. Several lines of evidence suggest that the dorsal and the ventral streams of processing do not operate entirely in isolation but mutually interact at different levels (Felleman and Van Essen, 1991; Markov et al., 2013). Also different response characteristics of cells may define different temporal routes of fast and slow processing (Born, 2001) that may help fusing information from different pathways. Here, we did not take into account such interactions based on different temporal effectivenesses. However, other model investigations capitalized on combining information from different channels to improve the selectivity of representation. For example, edge detection and grouping (in the ventral pathway) could be enhanced through mutually inhibitory gain control (which is similar as the normalization stage described here) generated by representations in the dorsal pathway. Since the dorsal representation is created by magno-cellular responses, such inhibition arrives already early to shape the selectivity of shape representations in the ventral path that is mainly driven by parvo-cellular responses (Shi et al., 2013). Similarly, interactions between the motion and form pathway have been suggested to help disambiguating localized features that give rise to occlusion cues which, in turn, support the disambiguation of object representation in the motion representation (Bayerl and Neumann, 2007; Beck and Neumann, 2010). Such detailed mechanisms would further enhance the proposed model architecture in refining the selectivities at different levels of low and intermediate representation.

As already pointed out above, the focus here is on the processing of 2D shape representations. In Cadieu et al. (2007) the authors have highlighted that their specific model investigation on shape representation in V4 is part of a larger hierarchically organized architecture for object recognition (Riesenhuber and Poggio, 1999; Serre and Poggio, 2010). Since their model principles relied on purely feedforward processing the insights provided in the work presented here might also shed some light on the mutual interactions between different processes on an even larger scale of object recognition processes. In addition, it would be interesting to find out how the representation of 3D surface patches (Yamane et al., 2008) seamlessly fit into a model computational architecture of recurrent shape computation.

In the presented coverage our model does not respond to contours elicited by contrasts of spatial luminance statistics caused

by differently textured regions. However, the core mechanisms, including initial filtering, modulatory feedback and competitive interaction for normalization, are like those proposed in the current contribution. A model that focuses on the processing of such boundaries has been developed in Thielscher and Neumann (2003). It is thus very likely that the recent model architecture proposed here can be extended with processing stages capable to process texture define boundaries as well without changing the basic architecture and computational principles. Also not considered in the current version is a multi-scale approach. We acknowledge the theoretical justification of hierarchical multi-stage processing to build up a pyramid-like structure (Tsotsos, 2005). Incorporating this representational diversity would allow the processing of a wider range of curvature configurations in shape outlines. In addition, this would support a more robust segregation of border ownership on the basis of convexities in the figural outline. We have focused our efforts on the specification of a hierarchically organized network architecture that utilized bottom-up and top-down convergent processing flows. In order to keep the computational efforts and the simulation times within reasonable bounds we restricted our description to single scale components at the different model stages within the hierarchy. A more extended realization of components is certainly desired but left for future investigations.

Intermediate level representations involve cells with receptive fields that recruit multiple sub-field components (Mineault et al., 2012; Yau et al., 2013). The model of Cadieu et al. (2007) accounts for this by sequentially fitting the subunits of intermediate level receptive field models to match the response profiles of V4 responses measured experimentally. This yields a sampling structure of statistically significant inputs in a feature space that contributes a significant amount of feature input to generate the final response of a shape selective cell. So far, in our modeling we sampled the spatial and the feature domains regularly. This of course demands high representational as well as computational resources. Consequently, it would be of interest to see how an irregularly sampled 4D space-feature domain (with orientation and curvature features) can be embedded into the scheme of shape representation proposed here.

ACKNOWLEDGMENTS

We thank our two anonymous reviewers for their valuable comments and constructive criticism that helped to improve the manuscript. This work has been supported by a grant from the Transregional Collaborative Research Center SFB/TRR62 “A Companion Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

REFERENCES

- Anzai, A., Peng, X., and Van Essen, D. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nat. Neurosci.* 10, 1313–1321. doi: 10.1038/nn1975
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bayerl, P., and Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Comput.* 16, 2041–2066. doi: 10.1162/0899766041732404

- Bayerl, P., and Neumann, H. (2007). Disambiguating visual motion by form-motion interaction – a computational model. *Int. J. Comp. Vis.* 72, 27–45. doi: 10.1007/s11263-006-8891-8
- Baylis, G. C., and Driver, J. (2001). Shape coding in IT cells generalizes over contrast and mirror reversal but not figureground reversal. *Nat. Neurosci.* 4, 937–942. doi: 10.1038/nn0901-937
- Beck, C., and Neumann, H. (2010). Interactions of motion and form in visual cortex – a Neura model. *J. Physiol. Paris* 104, 61–70. doi: 10.1016/j.jphysparis.2009.11.005
- Beck, C., and Neumann, H. (2011). Combining feature selection and integration. A neural model for mt motion selectivity. *PLoS ONE* 6:e21254. doi: 10.1371/journal.pone.0021254
- Ben-Shahar, O., and Zucker, S. (2004). Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Comput.* 16, 445–476. doi: 10.1162/089976604772744866
- Born, R. T. (2001). Visual processing: parallel–er and Parallel–er. *Curr. Biol.* 11, R566–R568. doi: 10.1016/S0960-9822(01)00345-1
- Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* 17, 2112–2127.
- Brosch, T., and Neumann, H. (2014). Interaction of feedforward and feedback streams in visual cortex in a firing-rate model of columnar computations. *Neural Netw.* 54, 11–16. doi: 10.1016/j.neunet.2014.02.005
- Cadiou, C., Kouh, M., Pasupathy, A., Connor, C., Riesenhuber, M., and Poggio, T. (2007). A model of V4 shape selectivity and invariance. *J. Neurophysiol.* 98, 1733–1750. doi: 10.1152/jn.01265.2006
- Caplovitz, G., and Tse, P. (2007). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cereb. Cortex* 17, 1179–1189. doi: 10.1093/cercor/bhl029
- Carandini, M., and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science* 264, 1333–1336. doi: 10.1126/science.8191289
- Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136
- Carandini, M., Heeger, D. J., and Movshon, J. A. (1999). “Linearity and gain control in V1 simple cells,” in *Cerebral Cortex Volume 13 Models of Cortical Circuits*, eds P. S. Ulinski, E. G. Jones, and A. Peters (New York, NY: Kluwer Academic; Plenum Publishers), 401–443.
- Connor, C. E., Brincat, S. L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Curr. Opin. Neurobiol.* 17, 140–147. doi: 10.1016/j.conb.2007.03.002
- De Pasquale, R., and Murray Sherman, S. (2013). A modulatory effect of the feedback from higher visual areas to V1 in the mouse. *J. Neurophysiol.* 109, 2618–2631. doi: 10.1152/jn.01083.2012
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. B* 353, 1245–1255. doi: 10.1098/rstb.1998.0280
- Eckhorn, R. (1999). Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Vis. Cogn.* 6, 231–265. doi: 10.1080/135062899394975
- Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex. *Neural Comput.* 2, 293–307. doi: 10.1162/neco.1990.2.3.293
- Edelman, G. (1993). Neural darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. doi: 10.1016/0896-6273(93)90304-A
- Felleman, D. J., and Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *J. Neurophysiol.* 57, 889–920.
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Field, D. J., Hayes, A., and Hass, R. F. (1993). Contour integration by the human visual system: evidence for a local association field. *Vision Res.* 33, 173–193. doi: 10.1016/0042-6989(93)90156-Q
- Fowlkes, C. C., Martin, D. R., and Malik, J. (2007). Local figure-ground cues are valid for natural images. *J. Vis.* 7, 1–9. doi: 10.1167/7.8.2
- Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202. doi: 10.1007/BF00344251
- Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Netw.* 1, 119–130. doi: 10.1016/0893-6080(88)90014-7
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Girard, P., and Bullier, J. (1989). Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *J. Neurophysiol.* 62, 1287–1302.
- Grossberg, S. (1980). How does the brain build a cognitive code? *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1
- Grossberg, S., and Mingolla, E. (1985). Neural dynamics of perceptual grouping: textual, boundaries, and emergent segmentations. *Percept. Psychophys.* 38, 141–171. doi: 10.3758/BF03198851
- Hansen, T., and Neumann, H. (2004). Neural mechanisms for the robust representation of junctions. *Neural Comput.* 16, 1013–1037. doi: 10.1162/089976604773135087
- Hatori, Y., and Sakai, K. (2012). “Surface-based construction of curvature selectivity from the integration of local orientations,” in *19th International Conference on Neural Information Processing (ICONIP)*, Vol. 7665, eds H. Tingwen, Z. Zhigang, L. Chuandong, and L. ChiSing (Berlin; Heidelberg: Springer), 425–433. doi: 10.1007/978-3-642-34487-9_52
- Hegde, J., and Van Essen, D. (2000). Selectivity for complex shapes in primate visual area V2. *J. Neurosci.* 20, 1–6.
- Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: representing anomalous contours. *Image Vis. Comput.* 16, 407–421. doi: 10.1016/S0262-8856(97)00083-8
- Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *J. Physiol.* 148, 574–591.
- Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.* 160, 106–154.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature* 394, 784–787. doi: 10.1038/29537
- Ito, M., and Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J. Neurosci.* 24, 3313–3324. doi: 10.1523/JNEUROSCI.4364-03.2004
- Jehee, J. F. M., Lamme, V. A. F., and Roelfsema, P. R. (2006). Boundary assignment in a recurrent network architecture. *Vision Res.* 47, 1153–1165. doi: 10.1016/j.visres.2006.12.018
- Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista Psicologia* 49, 7–30.
- Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* 20, 1427–1451. doi: 10.1162/neco.2008.02-07-466
- Lamme, V., and Roelfsema, P. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Layher, G., Giese, M., and Neumann, H. (2014). Learning representations of animated motion sequences – a neural model. *Top. Cogn. Sci.* 6, 170–182. doi: 10.1111/tops.12075
- Layton, O. W., Mingolla, E., and Yazdanbakhsh, A. (2012). Dynamic coding of border-ownership in visual cortex. *J. Vis.* 12, 1–21. doi: 10.1167/12.13.8
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791
- Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563. doi: 10.1016/S0960-9822(95)00108-4
- Louie, K., Grattan, L. E., and Glimcher, P. W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *J. Neurosci.* 31, 10627–10639. doi: 10.1523/JNEUROSCI.1237-11.2011
- Markov, N., Ercsey-Ravasz, M., Van Essen, D., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Science* 342, 1238406. doi: 10.1126/science.1238406
- Mineault, P., Khawaja, F., Butts, D., and Pack, C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proc. Natl. Acad. Sci. U.S.A.* 109, E972–E980. doi: 10.1073/pnas.1115685109

- Mokhtarian, F. (1995). Silhouette-based isolated object recognition through curvature scale space. *IEEE Trans. PAMI* 17, 539–544. doi: 10.1109/34.391387
- Mokhtarian, F., and Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. PAMI* 8, 34–43. doi: 10.1109/TPAMI.1986.4767750
- Mutch, J., and Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comp. Vis.* 80, 45–57. doi: 10.1007/s11263-007-0118-0
- Neumann, H., and Mingolla, E. (2001). “Computational neural models of spatial integration in perceptual grouping,” in *From Fragments to Objects: Grouping and Segmentation in Vision*, Chapter 12, eds T. Shipley and P. Kellman (Amsterdam: Elsevier), 353–400.
- Neumann, H., and Sepp, W. (1999). Recurrent v1-v2 interaction in early visual boundary processing. *Biol. Cybern.* 81, 425–444. doi: 10.1007/s004220050573
- Neumann, H., Yazdanbakhsh, A., and Mingolla, E. (2007). Seeing surfaces: the brain’s vision of the world. *Phys. Life Rev.* 4, 189–222. doi: 10.1016/j.plev.2007.09.001
- O’Herron, P., and von der Heydt, R. (2011). Representation of object continuity in the visual cortex. *J. Vis.* 11, 1–9. doi: 10.1167/11.2.12
- O’Herron, P., and von der Heydt, R. (2013). Remapping of border ownership in the visual cortex. *J. Neurosci.* 33, 1964–1974. doi: 10.1523/JNEUROSCI.2797-12.2013
- Parent, P., and Zucker, S. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Trans. PAMI* 11, 823–839. doi: 10.1109/34.31445
- Pasupathy, A., and Connor, C. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* 82, 2490–2502.
- Piëch, V., Li, W., Reeke, G. N., and Gilbert, C. D. (2013). Network model of top-down influences on local gain and contextual interactions in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4108–4117. doi: 10.1073/pnas.1317019110
- Qiu, F. T., and von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron* 47, 155–166. doi: 10.1016/j.neuron.2005.05.028
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Raudies, F., and Neumann, H. (2010). A model of neural mechanisms in monocular transparent motion perception. *J. Physiol. Paris* 104, 71–83. doi: 10.1016/j.jphysparis.2009.11.010
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819
- Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci.* 3, 1199–1204. doi: 10.1038/81479
- Rodríguez-Sánchez, A. J., and Tsotsos, J. K. (2012). The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS ONE* 7:e42058. doi: 10.1371/journal.pone.0042058
- Roelfsema, P., Lamme, V., Spekreijse, H., and Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *J. Cogn. Neurosci.* 14, 525–537. doi: 10.1162/0899290260045756
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Ann. Rev. Neurosci.* 29, 203–227. doi: 10.1146/annurev.neuro.29.051605.112939
- Roelfsema, P. R., Tolboom, M., and Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron* 56, 785–792. doi: 10.1016/j.neuron.2007.10.006
- Salin, P.-A., and Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol. Rev.* 75, 107–154.
- Schwartz, E., Desimone, R., Albright, T., and Gross, C. (1983). Shape Recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci. U.S.A.* 80, 5776–5778. doi: 10.1073/pnas.80.18.5776
- Self, M. W., Kooijmans, R. N., Supér, H., Lamme, V. A., and Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11031–11036. doi: 10.1073/pnas.1119527109
- Serre, T., and Poggio, T. (2010). A neuromorphic approach to computer vision. *Commun. ACM* 53, 54. doi: 10.1145/1831407.1831425
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *TPAMI* 29, 411–426. doi: 10.1109/TPAMI.2007.56
- Shi, X., Wang, B., and Tsotsos, J. K. (2013). “Early recurrence improves edge detection,” in *Proceedings of the 24th British Machine Vision Conference* (Bristol, UK).
- Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Front. Comput. Neurosci.* 2:4. doi: 10.3389/fncom.2008.10.004.2008
- Tabernik, D., Kristan, M., Boben, M., and Leonardis, A. (2014). “Using discriminative analysis for improving hierarchical compositional models,” in *Proc. 19th Computer Vision Winter Workshop*, ed Z. Kukulova (Krtiny).
- Thielscher, A., and Neumann, H. (2003). Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach. *Neuroscience* 122, 921–939. doi: 10.1016/j.neuroscience.2003.08.050
- Tsotsos, J. K. (1988). “Computational processes in human vision: an interdisciplinary perspective,” in *How Does Human Vision Beat the Computational Complexity of Visual Perception?* ed Z. Pylyshyn (Norwood, NJ: Ablex Press), 286–338.
- Tsotsos, J. K. (2005). “Neurobiology of attention,” in *Computational Foundations for Attentive Processes*, eds L. Itti, G. Rees, and J. K. Tsotsos (Amsterdam, NL: Elsevier), 3–7.
- Ullman, S. (1995). Sequence-seeking and counter streams: a computational model for bi-directional information flow in the visual cortex. *Cereb. Cortex* 5, 1–11. doi: 10.1093/cercor/5.1.1
- Ungerleider, L. G., and Haxby, J. V. (1994). ‘what’ and ‘where’ in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165. doi: 10.1016/0959-4388(94)90066-3
- von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science* 224, 1260–1262. doi: 10.1126/science.6539501
- Weidenbacher, U., and Neumann, H. (2009). Extraction of surface-related features in a recurrent model of V1–V2 interactions. *PLoS ONE* 4:e5909. doi: 10.1371/journal.pone.0005909
- Williford, J., and von der Heydt, R. (2013). Border-ownership coding. *Scholarpedia* 8, 30040. doi: 10.4249/scholarpedia.30040
- Wilson, H., and Wilkinson, F. (1998). Detection of global structure in Glass patterns: implications for form vision. *Vision Res.* 38, 2933–2947. doi: 10.1016/S0042-6989(98)00109-6
- Yamane, Y., Carlson, E., Bowman, K., Wang, Z., and Connor, C. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360. doi: 10.1038/nn.2202
- Yau, J., Pasupathy, A., Brincat, S., and Connor, C. (2013). Curvature processing dynamics in macaque area V4. *Cereb. Cortex* 23, 198–209.
- Zhaoping, L. (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput.* 10, 903–940. doi: 10.1162/089976698300017557
- Zhou, H., Friedman, H. S., and Von Der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611. doi: 10.1113/jphysiol.2002.033555

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 March 2014; accepted: 22 July 2014; published online: 11 August 2014.

Citation: Tschechne S and Neumann H (2014) Hierarchical representation of shapes in visual cortex—from localized features to figural shape segregation. *Front. Comput. Neurosci.* 8:93. doi: 10.3389/fncom.2014.00093

This article was submitted to the journal *Frontiers in Computational Neuroscience*. Copyright © 2014 Tschechne and Neumann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.