Biol Cybern (2008) 98:305–337
DOI 10.1007/s00422-008-0211-7

ORIGINAL PAPER

Biological Cybernetics

Globally consistent depth sorting of overlapping 2D surfaces in a model using local recurrent interactions

Axel Thielscher · Heiko Neumann

Received: 19 June 2007 / Accepted: 16 January 2008 / Published online: 4 March 2008

Abstract The human visual system utilizes depth information as a major cue to group together visual items constituting an object and to segregate them from items belonging to other objects in the visual scene. Depth information can be inferred from a variety of different visual cues, such as disparity, occlusions and perspective. Many of these cues provide only local and relative information about the depth of objects. For example, at occlusions, T-junctions indicate the local relative depth precedence of surface patches. However, in order to obtain a globally consistent interpretation of the depth relations between the surfaces and objects in a visual scene, a mechanism is necessary that globally propagates such local and relative information. We present a computational framework in which depth information derived from T-junctions is propagated along surface contours using local recurrent interactions between neighboring neurons. We demonstrate that within this framework a globally consistent depth sorting of overlapping surfaces can be obtained on the basis of local interactions. Unlike previous approaches in which locally restricted cell interactions could merely distinguish between two depths (figure and ground), our model can also represent several intermediate depth positions. Our approach is an extension of a previous model of recurrent $V1$–$V2$ interaction for contour processing and illusory contour formation. Based on the contour representation created by this model, a recursive scheme of local interactions subsequently achieves a globally consistent depth sorting of several overlapping surfaces. Within this framework, the induction of illusory contours by the model of recurrent $V1$–$V2$ interaction gives rise to the figure-ground segmentation of illusory figures such as a Kanizsa square.

1 Introduction

Robust recognition of objects in complex and cluttered environments crucially relies upon two concurring mechanisms, namely grouping and segregation (Grossberg and Mingolla 1985; Sajda and Finkel 1995). Grouping denotes the problem of binding together distinct visual items and attributes belonging to the same object embedded in a visual scene containing, e.g., partially occluded and mutually overlapping objects. In contrast, segregation addresses the task to separate those items and attributes from each other that belong to different objects. No trivial solution exists for these tasks. For example, it is likely to assume that neighboring positions in the visual scene contain items belonging together (as addressed by the Gestalt rule of proximity). However, when the overlapping object is transparent, some positions in the visual scene simultaneously contain information from two objects, which subsequently have to be segregated from each other (Adelson and Anandan 1990; Anderson 1997). In contrast, when an object is partially occluded by another opaque one such that belonging parts appear spatially split, the distinct items have to be grouped together in order to allow for an unified percept (the task termed as amodal completion; Kanizsa 1979; Kellman and Shipley 1991).

The human visual system utilizes depth information as a major cue to robustly solve the above depicted problems in the process of grouping and segmentation (Baumann et al. 1997; Kovacs et al. 1995; Nakayama et al. 1989). The depth

A. Thielscher (✉)
High-Field Magnetic Resonance Center,
Max Planck Institute for Biological Cybernetics,
Spemannstraße 38, 72076 Tübingen, Germany
e-mail: axel.thielscher@tuebingen.mpg.de

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

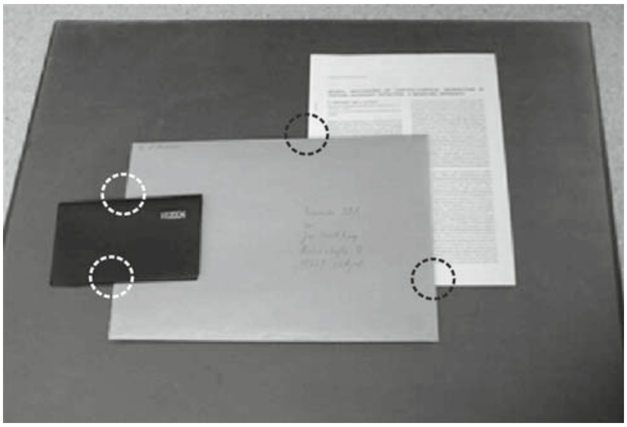


Fig. 1 An example of overlapping real-world objects (diary, envelope, paper). T-junctions occur at those positions where the contours of two objects overlap (indicated as *dotted circles*). The contour corresponding to the cross-bar of the T belongs to the object being locally in the foreground and the stem of the T belongs to a surface contour that continues behind the occluding object

relations between the items in the visual scene are utilized to obtain a globally consistent depth sorting, in turn allowing for a segregation of those items that belong to distinct objects and enabling the segmentation of figure from ground. Here, we present a neural model in which mechanisms of grouping and depth processing interact in order to segregate the contours of overlapping objects according to their position in depth.

Depth information can be inferred from a variety of different visual cues dividing mainly into binocular (i.e., disparity) and monocular (e.g., occlusions, perspective, relative size, etc.) cues (e.g., Howard 2003; Kellman and Shipley 1991; Poggio et al. 1988). Most cues are locally restricted, i.e., their information is only available at sparse locations in the visual scene. For example, while disparity information may be unambiguously measured only at surface *boundaries*, it also has to be available on closed *regions* in order to be able to distinguish flat from curved surfaces. Likewise, contour intersections between occluding and occluded objects (denoted as T-junctions; see dashed circles in Fig. 1) are hints allowing one to determine the *local* figure-ground direction (Rubin 2001a): In the case of occlusion, the top of the T intrinsically belongs to the object in the foreground, while the stem refers to the background, respectively. Moreover, most cues do not allow one to directly determine the absolute position of an object in depth. Instead, they are relative cues to depth for one object in relation to another one (i.e., relative disparity, depth ordering at T- and X-junctions). Consequently, in order to obtain a globally consistent interpretation of the depth relations between all objects in a visual scene, a mechanism is required that allows one to globally propagate the local and relative information of the depth cues.

The necessity of a global mechanism seems to argue in favor of depth processing being implemented in rather high levels of the hierarchy of cortical visual areas (Felleman and van Essen 1991). This view is supported by electrophysiological studies indicating that cell activities in IT are sensitive to figure-ground reversal but are invariant to partial occlusions of object shape (Baylis and Driver 2001; Kovacs et al. 1995). However, IT neurons pool information over wide parts in the visual scene and therefore lack the sensitivity to specifically react to small local cues such as T-junctions. Also, the electrophysiological findings might result from IT neurons that were driven by depth-selective bottom–up input. In this case, low- and midlevel visual processes would succeed in determining a depth sorting of the objects in the visual scene, which is then passed on to the higher visual processes of object recognition. Indeed, electrophysiological studies demonstrating the ability of neurons in $V2$ to use occlusion cues for figure-ground segregation (Baumann et al. 1997; Zhou et al. 2000) indicate that mechanisms for depth processing might already be integrated at stages of early visual processing. However, in order to fulfill the demands of spatially high-resolution processing, cortical neurons in early visual areas have rather small receptive fields and they interact only in a restricted spatial neighborhood (Gilbert and Wiesel 1989; Peterhans 1997; Smith et al. 2001). Therefore, the question arises how the framework of neuronal interaction has to look like in order to allow these neurons to exchange and promote their local information to achieve a globally consistent interpretation of the depth relations in the visual scene.

In the following, we present a computational framework in which depth information from local relative cues (namely T-junctions) is propagated along surface contours using local recurrent interactions between neighboring neurons. We demonstrate that within this framework a globally consistent depth sorting of overlapping surfaces can be obtained on the basis of local interactions. Our approach is an extension of a biologically plausible model of recurrent $V1$–$V2$ interaction for contour processing and illusory contour formation (Grossberg and Mingolla 1985; Neumann and Sepp 1999; Thielscher and Neumann 2003). The contour representation created by this model is subsequently used in a recursive scheme of local interactions to determine a globally consistent depth sorting. We start our description with an overview of the overall model architecture (Sect. 2.1). Then the model stages for contour processing (Sects. 2.2 and 2.3) and for the detection of corners and T-junctions (Sect. 2.4) are introduced. The model stage of recurrent depth processing is depicted in Sect. 3, starting with the presentation of the general scheme of depth sorting (Sect. 3.1) and continuing with the neural model used to implement that scheme (Sects. 3.2–3.6). Results of the simulations performed with the model are shown in Sect. 4. The presentation ends with a discussion of the model and the results in Sect. 5.

2 Initial contour processing and detection of T-junctions and corners

2.1 Overview of the overall model architecture

The overall model framework is structured in three main processing stages (Fig. 2). Initially, the stage of recurrent $V1$–$V2$ interaction processes the surface contours in the input image. Oriented contrast is measured by a cascade of model LGN-On/Off cells followed by $V1$ simple and complex cells. Model $V1$ complex cell activity is passed on to $V2$ bipole cells having elongated receptive fields for line groupings and illusory contour formation. Recurrent interaction between $V1$ complex and $V2$ bipole cells stabilizes the initial activation pattern and results in a robust representation of the contours in the input image. Within this hierarchy, $V1$ complex cell activity signals the amount and orientation of physical luminance contrast in the input image, whereas $V2$ bipole cell activity represents continuous and completed surface outlines. Details on the first model stage are depicted in Sects. 2.2 and 2.3.

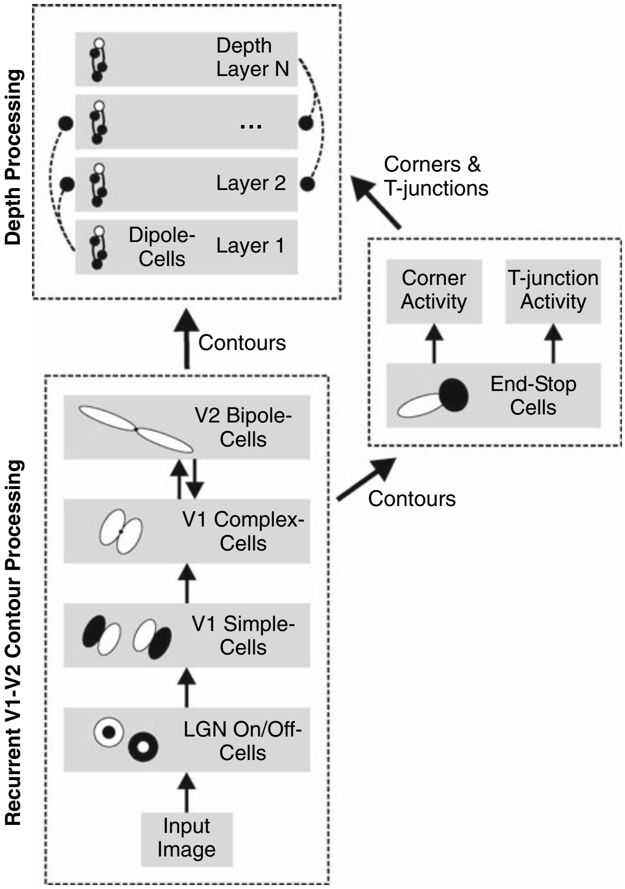


Fig. 2 Overall system diagram. The model is structured in three main processing blocks. (1) Initially, a feedforward cascade of model LGN On/Off cells, $V1$ simple and $V1$ complex cells measures oriented contrast along contours in the input image. The contrast information signaled by $V1$ complex cells is passed on to model $V2$ bipole cells for long-range contour groupings, and recurrent interaction between these two cell types stabilizes the overall contour representation. (2) The contour information is passed on to a feed-forward scheme which uses end-stop cells to extract the position of contour corners and T-junctions in the image. (3) The information on contours, corners and T-junctions is used by a recurrent model stage to solve the depth relations between the surface patches in the input image. The receptive field kernels and the channels of the gated dipole channels are shown in white (excitatory subfields and ON-channels) and black (inhibitory subfields and OFF-channels). The field sizes were note drawn to scale and the positions of cells do not imply any pattern of spatial interactions

The second model stage consists of a feed-forward scheme for the detection of contour corners and T-junctions. $V1$ complex cell activity is filtered by end-stop cells to detect the position and orientation of putative line endings. The combination of end-stop cell activities representing roughly perpendicular line orientations is used to determine the position and orientation of T-junctions and contour corners in the input image (for details see Sect. 2.4).

The third model stage uses the information on contours, corners and T-junctions delivered by the first and second stages to resolve the depth relations between overlapping surfaces in the input image. It consists of several layers representing different positions in depth, ranging from foreground to background. Each depth layer contains a topographic map of gated dipole cells (throughout the paper, "dipole" refers to gated dipole cells in the third model stage of depth processing, and "bipole" refers to $V2$ bipole cells in the first model stage of contour processing). Within a depth layer, the dipole cells propagate the local depth information delivered by T-junctions along contours. By that propagation, the local information from all T-junctions belonging to a surface contour is combined. The combined evidence is used to determine if or if not a contour has the position in depth that is represented by the layer. Contours being at a different depth position are passed on to other layers for further processing via connections between dipole cells corresponding to the same topographical position in the visual field, but being allocated to different depth layers. The recurrent interactions outlined above, acting *within* and *between* depth layers, finally determine globally consistent depth positions for all surface contours in the input image. A comprehensive description of the model stage of recurrent depth processing can be found in Sect. 3.

2.2 Recurrent $V1$–$V2$ interaction for contour processing: model architecture

The first model stage represents early visual mechanisms in areas LGN, $V1$ and $V2$ for contour processing based on luminance contrast. It consists of a hierarchy of areas containing topographical maps of single-compartment cells with gradual activation dynamics. Each model cell represents the average response (or firing-rate) of groups of neurons with similar selectivities. The size of the cells' receptive fields increase within the hierarchy of model areas (Smith et al. 2001). The following paragraph outlines the functionality and the receptive field properties of the model cells. The corresponding mathematical equations are presented in Appendix A.1. A more detailed description of the $V1$–$V2$ model stage of contour processing with respect to, e.g., line grouping and illusory boundary formation can be found in (Thielscher and Neumann 2003) or (Neumann and Sepp 1999).

Initially, *Model LGN ON and OFF cells* with concentric center-surround receptive fields (Hubel and Wiesel 1962) detect local luminance contrast in the input image, based on a subtractive and half-wave rectified interaction between Gaussian weighted input intensities (Fig. 2). The output of appropriately aligned LGN cells is subsequently pooled by *model $V1$ simple cells* having elongated juxtaposed ON and OFF subfields. They respond to local luminance transitions along a given orientation preference and are selective to contrast polarity (dark–light and light–dark in 8 discrete orientations). These first two processing steps emulate roughly the functionality seen in the parvocellular layers of LGN and simple cells in $V1$. They were incorporated in the model for preprocessing the initial luminance distribution and their activation level is determined in a purely feed-forward fashion. Their activity constitutes the input to the scheme of recurrent contour processing implemented by bi-directionally linked *model $V1$ complex* and *$V2$ bipole cells* that is depicted in Fig. 3a:

- *Model $V1$ complex cells* form the first level of recurrent processing for surface boundary computation (lower part of Fig. 3a). They pool the activity of two equally oriented $V1$ simple cells of opposite polarity at each position. In combination, the computation performed by model LGN, simple and complex cells result in complex cell activity that is sensitive to orientation but insensitive to the direction of contrast. The output of the model $V1$ complex cells thus resembles that of real cortical complex cells. This output activation is subsequently modulated by excitatory top–down interaction from model $V2$ bipole cells and intra-areal center-surround competition utilizing two sequential computational steps. Details on these two computational steps are presented in Sect. 2.3.
- *Model $V2$ bipole cells* use two prolated subfields aligned along the axis of the cell's orientation preference to pool the input delivered by appropriately aligned $V1$ and $V2$ cells (upper part of Fig. 3a; Grossberg and Mingolla 1985; Neumann and Mingolla 2001). Model $V2$ bipole cells respond to luminance contrasts as well as to illusory contours, thus resembling the functional properties of contour neurons in $V2$ (Heitger et al. 1998; v. d. Heydt et al. 1984, 1993). The bottom–up activity delivered by $V1$ complex cells is pooled by the two subfields and subsequently combined by a soft-AND-gate mechanism that only generates significant responses when both fields are excited simultaneously (Neumann and Sepp 1999; Thielscher and Neumann 2003). This mechanism enables the cells to complete fragmented contours and to respond to illusory contours induced by two or more contrast fragments or line ends exciting both subfields. The initial $V2$ bipole cell response to the driving $V1$ bottom–up input is subsequently modulated by long-range activity from neighboring $V2$ cells. More specifically, $V2$ cells pool the output of neighboring bipole cells using two elongated subfields. The architecture of these subfields is identical to those pooling the bottom–up input from $V1$. The pooled long-range activity modulates the initial cell response to the $V1$ input in a multiplicative fashion, thereby helping to normalize the $V2$ cell activation strength along contours. This $V2$ horizontal long-range interaction stabilizes the initial contour representation in particular for noisy contours, e.g., at surface boundaries with varying luminance levels in the background. Finally, the $V2$ activity that is determined by driving $V1$ bottom–up input and modulatory $V2$ long-range interaction undergoes center-surround competition.

2.3 Recurrent $V1$–$V2$ interaction for contour processing: model cell dynamics

In model V1 complex and $V2$ bipole cells, the integration of bottom–up activity via a cell's receptive field is followed by two successive computational steps to determine its final activation level (Fig. 3a; Appendix A.2):

- The initial $V1$ complex cell activity is modulated via feedback interaction from model $V2$ bipole cell. Correspondingly, the initial $V2$ bipole activation level is modulated by horizontal $V2$ long-range feedback. The excitatory feedback activity that is delivered either by the descending cortical pathway (in case of $V1$ complex cells) or by horizontal anisotropic long-range connections (in case of $V2$ bipole cells) *multiplicatively* enhances the initial activity at the earlier processing stage. This type of feedback interaction is only effective at positions with non-zero initial activation, which prevents unspecific activity to spread unintentionally within the topographical maps. Several physiological studies indicate that, e.g., feedback from higher visual areas is not capable of driving cells in lower areas, but *modulates* their activity (Hupe et al. 1998; Mignard and Malpeli 1991; Przybyszewski et al. 2000; Salin and Bullier 1995; Sandell and Schiller 1982). We use multiplicative instead of, e.g., additive excitatory feedback as one possible implementation of such a modulatory interaction. In the "no-strong-loops hypothesis" of

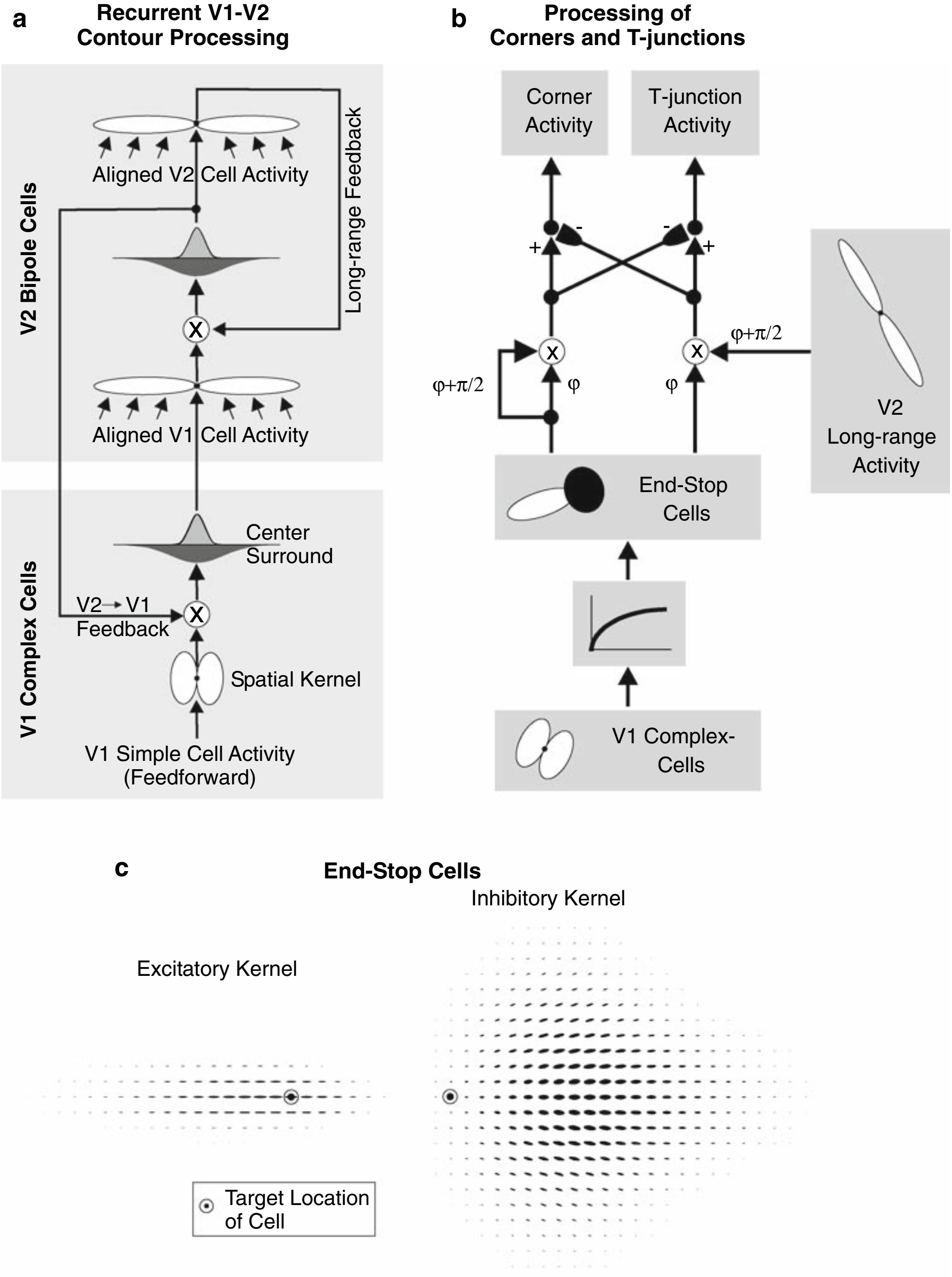


Fig. 3 **a** Diagram of the model stage of recurrent $V1$–$V2$ contour processing. *Lower half* Model $V1$ complex cells pool the input of appropriately aligned LGN-cells delivered via the $V1$ simple cells. This bottom–up input is modulated by feedback from model $V2$ bipole cells at corresponding topographical positions. Finally, the top–down modulated activity undergoes a stage of ON-center/OFF-surround competition to suppress spurious activations and to normalize cell activity. *Upper half* Model $V2$ bipole cells integrate the input of appropriately aligned $V1$ and $V2$ cells using two identical, elongated spatial kernels. The initial $V2$ bipole cell activity is determined by the driving $V1$ bottom–up input pooled by the first spatial kernel. The cell activity is then modulated by excitatory $V2$ intra-areal feedback pooled by the second kernel. This activity that is determined by driving $V1$ input and modulated by long-range $V2$ activity finally undergoes a stage of center-surround competition. **b** Feed-forward scheme of the model stage used to detect the position of corners and T-junctions. $V1$ complex cell activity is normalized and filtered by end-stop cells. Subsequently, the output of end-stop cells sensitive to approx. perpendicular orientations is multiplied at each spatial position to detect the likely position of contour corners. Likewise, the output activity of end-stop cells and $V2$ long-range activity of approximately perpendicular orientations is multiplied to detect the likely position of T-junctions. Finally, the initial corner and T-junction activities compete with each other to suppress ambiguous activations. **c** Spatial kernels of the end-stop cells used to filter $V1$ complex cell activity. The relative weight and the orientation preference at a specific position are represented by the size and orientation of the ellipse

Crick and Koch (1998) it is argued that the existence of one driving and one modulatory connection, instead of two driving connections in a directed loop between two cortical areas, avoids uncontrolled oscillations of the overall system and limits the amount of inhibition necessary to achieve a stable network behavior. The same argument holds when considering recurrent horizontal connections between cells of the same hierarchy level.

- The activity modulated by feedback undergoes a stage of shunting ON-center/OFF-surround competition between cells in a spatial and orientational neighborhood. This competitive interaction suppresses spurious and perceptually irrelevant activities. It normalizes the final activity level and enhances it by contrast.

Together, the two stages of computation realize a soft-gating mechanism: $V1$ activities which match the activity pattern in model area $V2$ are further enhanced via excitatory $V2 \rightarrow V1$ feedback and inhibit cells in their neighborhood. Likewise, initial $V2$ activities which correlate with other $V2$ activations representing continuous contours are enhanced and inhibit non-matching activations in the neighborhood. Thus, salient contour arrangements are enhanced and stabilized while at the same time spurious and perceptually irrelevant activities are suppressed. These computational mechanisms were motivated by previously proposed principles of recurrent interaction for response integration and cortical prediction (Grossberg 1980; Mumford 1994; Neumann and Sepp 1999) as well as reentry processes for integration and disambiguation of localized feature measurements (Sporns et al. 1991). In particular, the computations in our model stage of recurrent $V1$–$V2$ interaction resemble the contour groupings performed by the Boundary Contour System (BCS) suggested by Grossberg and Mingolla (1985). In contrast to the BCS, feedback activity in our model cannot induce new activity, but only modulate the cell responses driven by bottom–up input. The combination of intra-areal horizontal feedback in $V2$ with recurrent interaction between model areas $V1$ and $V2$ resembles to some extent the computational mechanisms captured by the LAMINART model (Raizada and Grossberg 2003). LAMINART was developed to investigate the patterns of interaction between feedforward, feedback and horizontal activity. It demonstrated that the range of possible computations is greatly extended by allowing top–down and lateral signals to reciprocally interact. Likewise, in our model, the horizontal interactions between model $V2$ bipole cells enhance the overall model robustness and help to further stabilize the model activation pattern signaling salient contours. In contrast to our model, LAMINART makes more explicit and detailed predictions about the functionalities and interactions between the cells in the different layers of the cerebral cortex. Our model, instead, emphasizes the processing of more complex visual features in early visual areas by utilizing a more abstract description of the neural computations performed in that areas based on a three-level cascade of processing stages.

Prior to the actual simulations, the neural connection strengths and the constants of the model cell dynamics (as listed in Table A.1 in the Appendix) were empirically determined in such a way that the whole network could reach a stable activation pattern quickly after onset of input pattern presentation. In order to speed up processing, the differential equations of the model cell dynamics (Appendix A.2) were solved at equilibrium in response to a constant input. Initially, the activities of all model layers were set to zero. The input image was clamped and the activities of the model areas were sequentially updated. The final activation patterns were achieved after 3–4 iterative cycles. Each simulation was continued until iteration 7 in order to visually demonstrate the stability of the solution. A comparison with results obtained by numerical integration of the model equations revealed that the use of equilibrium responses did not affect the results of the final activation patterns.

2.4 Detection of corners and T-junctions

The second model stage (Fig. 2) determines the positions and orientations of corners and T-junctions using a feed-forward processing scheme, based on the steady-state activation patterns of model $V1$ complex cells and model $V2$ contour cell activity from long-range integration. Details of the feed-forward scheme are depicted in Fig. 3b.

Model end-stop cells signal the position and orientation of contour endings. In the context of our model, a contour ending represents the spatial position at which either one contour is ended by another, occluding contour, or a contour of a certain orientation abruptly "ends" and is continued by another contour of different orientation (e.g., at L-junctions). For three-dimensional surface shapes, more complex ending patterns might occur depending on the viewpoint (Koenderink and v. Doorn 1982), which are not captured by the simple feed-forward processing scheme outlined here. End-stop cells pool the model $V1$ complex cell activity using excitatory and inhibitory subfields (Fig. 3c). At spatial positions corresponding to contours, both the excitatory and inhibitory subfields are activated, resulting in a suppressed final cell activity. At contour endings, the activity integrated by the excitatory subfield significantly exceeds that of the inhibitory kernel and, consequently, the cell responds. Detailed mathematical equations can be found in Appendix B.1.

Contour corners can be seen as two intersection contour endings that are roughly perpendicular to each other. Accordingly, in our feed-forward scheme, the likely positions of contour corners are signaled by the product of the end-stop activities at orientation θ with the activities at roughly perpendicular orientations $\theta + \pi/2$. The stem and hat of a

T-junction is formed by the intersection of a contour ending with a continuous line at a roughly perpendicular orientation. In our feed-forward scheme, candidate positions for these intersections are determined by the product of the end-stop activities at orientation θ and the $V2$ long-range activity at orientation $\theta + \pi/2$. The initial corner and T-junction activities undergo a stage of subtractive inhibition. This results, e.g., in the suppression of initial corner-related activity at the inducing pacmans of a Kanizsa square due to T-junction activity evoked by $V2$ long-range groupings between the inducers. To summarize, T-junction activity is not solely based on the local physical stimulus properties, but is also driven by $V2$ bipole cell activity signalling illusory contours. In addition, initial corner and T-junction activities compete to signal which type of junction is more likely given the information integrated from a medium-scale neighbourhood around the junction. This framework is supported by two studies from McDermott and Adelson (2004a,b) showing that (i) interaction between T-junctions and contour information occurs "at an intermediate semilocal scale", and that (ii) "what matters is not junctions per se, but whether illusory contours are introduced when junction category is changed (from L- to T-junctions)". The model equations are depicted in Appendix B.2.

3 Recurrent depth processing

The following section outlines the third model stage (Fig. 2) that computes a globally consistent depth sorting of the 2D surfaces in the input image, based on the information on contours, contour corners and T-junctions delivered by the first and second stage. The section starts with a theoretical description how overlapping contours can be arranged in depth in a globally consistent way using local junction information. Subsequently, the mechanisms necessary to integrate the proposed scheme into a neural model architecture are discussed. Mathematical details of the implementation of these mechanisms are described in the remaining sections of this chapter.

3.1 Depth sorting of surface contours based on local relative depth cues

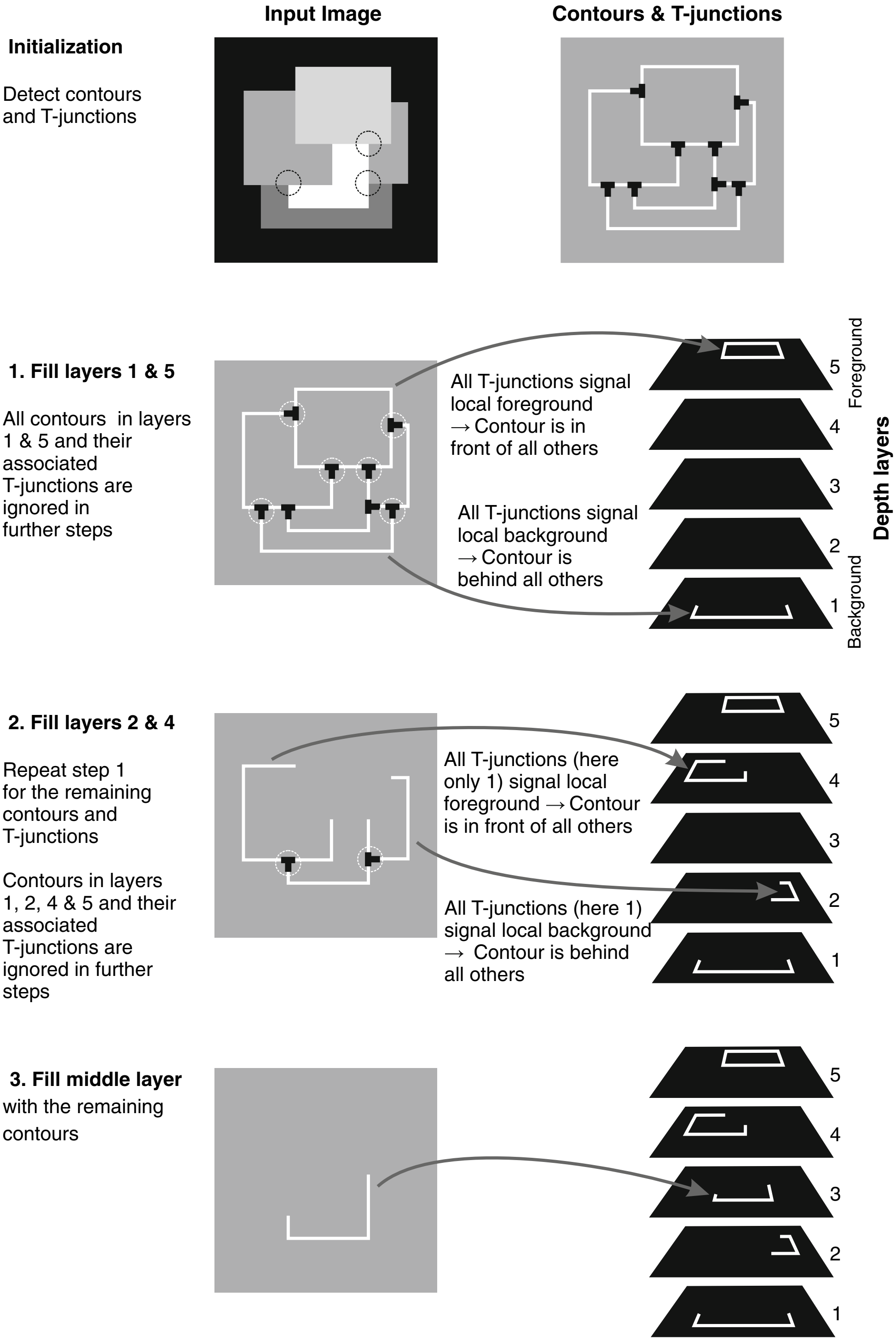
In general, the information given by local cues, such as T- or X- junctions, is not sufficient to unambiguously determine the position of a specific surface in depth. This limitation holds true even when simultaneously considering the pooled information of all local depth cues related to a surface. For example, the number of local T-junctions which indicate a contour is in the front of or behind other contours does not allow conclusions about its exact depth position, but merely depends on the contour shapes and on the layout of the overall visual scene. This can be demonstrated by considering the white surface patch in the center of the stimulus in Fig. 4 (left side, top row). On the right side, the corresponding surface contours and the positions of the T-junctions are depicted. The white surface patch has two T-junctions which indicate that the surface boundaries are behind other contours and one showing that the patch is in front of another contour, as indicated by the black dashed circles. Based on this information, we can conclude that the white patch is at some intermediate depth, but we cannot determine its depth position more precisely.

Importantly, there are two exceptions to that rule, namely the contours belonging to surfaces that are either in front of or behind *all* other surfaces. These contours can easily be identified as they possess either *all* of the tops or *all* of the stems of their T-junctions. Consequently, when trying to determine the depth relations between overlapping surface contours, one can start by "earmarking" these contours. In Fig. 4 (second row), this was done by transferring them to depth layers representing the foreground (layer 5) and background (layer 1), respectively. They and their T-junctions can subsequently be ignored when determining the depth positions of the remaining contours. Accordingly, in Fig. 4 (third row), they were deleted from the contour representation. When considering the remaining surfaces, there will again be contours that are either in the front of or behind all other remaining contours, which is now unambiguously indicated by their T-junctions. In Fig. 4, these contours are stored in depth layers 2 and 4 (third row) and deleted from the contour representation (forth row). Each time when contours with unambiguous depth positions have been determined and are deleted from the representation, there will be new contours which are now either in front of or behind all other remaining contours. Consequently, the scheme outlined above can be recursively continued until the depth positions of all contours are determined, given that we initially provided a sufficient number of depth layers to store the intermediate results (as shown on the right side of Fig. 4).

In the general case of N depth layers with #1 as the background and #N as the foreground layer (#2 to #$N - 1$ are intermediate layers), the recursive scheme can be summarized as follows:

- First, the outer layers #1 and #N are considered and the contours being either in the background or in the foreground of all others (as indicated by their T-junctions) are arranged in these layers. Contours owned by an isolated surface not overlapping with others will be simultaneously assigned to layers #1 and #N per convention.
- Second, the contours assigned to layers #1 and #N as well as the local depth cues belonging to them are no longer taken into account. Now layers #2 and #$(N-1)$ are

Fig. 4 Recursive scheme to determine a globally consistent interpretation of the depth of surface contours using relative depth cues (see Sect. 3.1)


Input Image
Contours & T-junctions
Initialization
Detect contours and T-junctions
1. Fill layers 1 & 5
All contours in layers 1 & 5 and their associated T-junctions are ignored in further steps
All T-junctions signal local foreground → Contour is in front of all others
All T-junctions signal local background → Contour is behind all others
5
4
3
2
1
Foreground
Background
Depth layers
2. Fill layers 2 & 4
Repeat step 1 for the remaining contours and T-junctions
Contours in layers 1, 2, 4 & 5 and their associated T-junctions are ignored in further steps
All T-junctions (here only 1) signal local foreground → Contour is in front of all others
All T-junctions (here 1) signal local background → Contour is behind all others
3. Fill middle layer with the remaining contours

considered and the first step is repeated for the remaining contours and depth cues: Remaining contours that are behind or in the front of or all other remaining contours are assigned to layers #2 and #$(N-1)$, respectively.

- Third, the contours assigned to layers #1, #2, #$(N-1)$ and #N and the depth cues belonging to them are no longer taken into account and layers #3 and #$(N-2)$ are considered, etc.

This scheme continues until all contours have been assigned to a depth layer or until all layers have been utilized. If the number of necessary layers is unknown in the beginning, N

should be an odd number. In this case, all remaining contours can be stored in the middle layer to indicate their depth in relation to the contours assigned to the outer layers (when the number of contours stored in the middle layer is K, $N + K - 1$ is the maximal number of layers necessary to assign all contours to a depth layer). When a contour is assigned to layers #1 *and* #N, then it is an isolated surface contour not overlapping with others. The above depicted recursive scheme successively resolves the correct depth positions of the surface patches in a visual scene without having to consider any global relationships between them. Instead, only local T-junction information is pooled along each surface contour, which is repeated at each step of the recursion until all contours have been assigned to a depth layer.

3.2 Outline of the overall architecture and key processing mechanisms of the third model stage

The above depicted scheme allows for a globally consistent depth sorting of overlapping surface contours based on locally restricted junction information. In the following, we will highlight the key structures and neural mechanisms of the third model stage that integrate this computational scheme into a neural architecture. As indicated in Fig. 2, processing in the third stage builds upon the activation pattern of the stage of recurrent $V1$–$V2$ interaction as the underlying representation of the contours in the image. The T-junction detectors of the second stage signal topographical positions at which activation patterns that represent different surface contours intersect and allow one to determine local figure-ground relationships.

Based on physiological findings of cell pools in the visual cortex that exhibit different disparity profiles (Poggio et al. 1988), the third model stage employs a stack of depth layers (Figs. 2, 5b) containing topographical 2D maps of model cells. The depth of a contour is signaled by the activity of neurons at corresponding spatial positions within the corresponding depth map. Contours are recursively assigned to depth layers, whereby the activation patterns of the two outermost depth layers initially represent *all* contours. This indicates that the depth of the contours is unknown at the beginning of the process. Subsequently, activation patterns that signal contours with ambiguous depth positions are continuously passed on from the outer to the more medial (or inner) depth layers by means of two recursive model mechanisms. One mechanism acts *within* a depth layer, and propagates T-junction information along contours (Fig. 5a). The other mechanism acts *between* layers and transfers model cell activity signaling ambiguous contours from the outer to more medial layers (Fig. 5b).

The neural mechanism acting *within* a layer is based on dipole cell dynamics (Fig. 5a; for details on dipole dynamics see, e.g., Grossberg 1991). A dipole cell consists of two antagonistic channels (ON and OFF) that continuously compete with each other in order to signal the channel receiving the stronger input activation (the output of the other channel is suppressed). The activities delivered by the ON- and OFF-channels of neighboring cells are pooled by the cell's receptive fields and constitute the two inputs to the dipole. A central functionality of dipoles is that of antagonistic rebound: When the relative strength between the two input channels changes, the dipole resets the formerly active output channel and the formerly inhibited channel responds at a high initial activity level. Consequently, when a sufficient number of dipoles in the neighborhood of a cell reset, the input to one of the channels becomes very strong, in turn resetting the cell. By this, a "wave" of automatic activity resets is triggered in the two-dimensional map of model cells. Without applying any spatial restrictions, the activity resets would propagate in a circular manner, originating from the point where the first reset occurred. In order to prevent this type of unspecific spread of activity, the $V2$ bipole cells from the first model stage control the input gain of the dipole channels, thereby guiding the propagation of activity along surface contours. Dipole cells have initially their ON-channels activated, while the OFF-channels are inhibited. The activity of T-junction detectors locally resets cells that represent contours not belonging to the current depth layer. By this, waves of OFF-channel activity are induced that propagate autonomously within the depth layer and inhibit the ON-channels of all cells along contours having a different depth position. For example, in the foreground layer, T-junction activities locally reset dipoles that correspond to the *stems* of the Ts, so that the resulting activity waves inhibit the ON-channels of cells that signal contours being behind at least one other contour. In the background layer, the T-junction activities reset dipoles signaling the *hats* of the Ts, so that the activity waves suppress the ON-channel activities of all cells representing contours that are in front of other contours.

After outlining how dipole activity propagates *within* the topographical map of a depth layer via locally restricted interactions between neighboring cells, we now turn to the mechanism which distributes the dipole activity patterns *between* depth layers by means of stereotyped inhibitory connections (Fig. 5b). Dipole cells with active ON-channels inhibit all neurons that are located at the corresponding topographical positions in more medial depth layers. For example, the active ON-channel of a cell in the background layer #1 suppresses all dipoles at the corresponding positions in the intermediate layers #2 to #$(N-1)$. The same pattern of suppression occurs between cells in the foreground layer #N and the corresponding cells in the intermediate layers. As outlined above, a wave of OFF-channel activity traveling *within* a depth layer inhibits the ON-channels of all cells along a contour having a depth position that differs from the one represented by the depth layer. This results in a *release of inter-layer inhibition* that

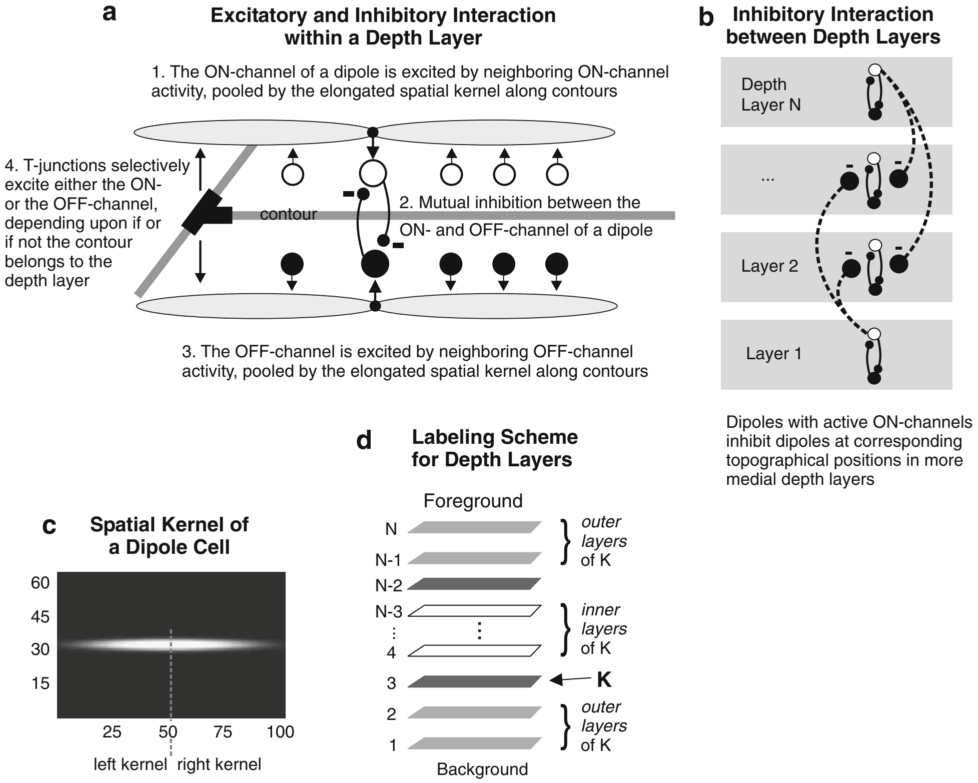

a
Excitatory and Inhibitory Interaction within a Depth Layer
1. The ON-channel of a dipole is excited by neighboring ON-channel activity, pooled by the elongated spatial kernel along contours
4. T-junctions selectively excite either the ON- or the OFF-channel, depending upon if or if not the contour belongs to the depth layer
contour
2. Mutual inhibition between the ON- and OFF-channel of a dipole
3. The OFF-channel is excited by neighboring OFF-channel activity, pooled by the elongated spatial kernel along contours
b
Inhibitory Interaction between Depth Layers
Depth Layer N
...
Layer 2
Layer 1
Dipoles with active ON-channels inhibit dipoles at corresponding topographical positions in more medial depth layers
c
Spatial Kernel of a Dipole Cell
60
45
30
15
25
50
75
100
left kernel
right kernel
d
Labeling Scheme for Depth Layers
Foreground
N
N-1
N-2
N-3
4
3
2
1
outer layers of K
inner layers of K
K
outer layers of K
Background

Fig. 5 Model stage of recurrent depth processing. **a** Interactions within a depth layer. A dipole cell pools the ON- and OFF-channel activities in its neighborhood using elongated receptive fields. The ON- and OFF-channels of the dipole continuously compete against each other to signal the channel receiving the stronger input activation. Antagonistic rebounds of the dipole are triggered when the difference between the inputs to the ON- and OFF-channels changes its sign. The high output activity of the dipole directly after the rebound can cause the neighboring dipoles to also reset their channels, thereby triggering "waves" of dipole activity propagating within the depth layers. T-junction detectors trigger waves of OFF-channel activity to reset those dipoles corresponding to contours having a different position in depth. **b** Interaction between depth layers. Active ON-channels of dipoles in the outer layers inhibit the dipoles in more medial depth layers. By this mechanism, all contours are initially represented in the outer depth layers. Subsequent release of inhibition transfers those activity patterns that represent contours at intermediate depth positions from the outer to more medial depth layers. **c** Layout of the spatial dipole kernel used to integrate activity of neighboring dipoles. The kernel is cut into a left and a right half. **d** Labeling scheme used to indicate the relation between depth layer #K and the other layers. *Outer* layers depict those layers (foreground and background) that have a larger distance to the middle depth layer compared to layer #K. For example, for $K = 3$, the outer layers are #1, 2, $N - 1$ and N. The *inner* layers are closer to the middle depth layer than layer #K is. For $K = 3$, the inner layers range from #4 to #$N - 3$

transfers the activity pattern of that contour from the outer to more medial depth layers.

To summarize, waves of OFF-channel activity triggered by T-junction detectors initially reset the neurons in the two outer layers which signal contours being not in the fore- or background, respectively. The cells lose their inhibitory impact on the cells in the more medial layers, which, in turn, start to represent these contours. Since depth layers #2 and #$(N-1)$ again suppress the more medial layers via inhibitory connections, these two layers start to represent *all* contours which have intermediate depth positions. T-junction activity again triggers waves of activity inhibiting those contour activities not belonging to the two layers. These contours are again transferred to more medial depth layers via the release of inter-layer inhibition, and so on. Taken together, the combination of inhibitory mechanisms acting within and between layers enables the model to continuously pass on those activities from the outer to the inner depth layers that represent contours having more intermediate, but still ambiguous positions in depth.

In the following subsections, we will have a closer view on the key components of the third model stage. We start with the mathematical equations that describe how a dipole cell determines its output by continuously comparing the activation in two input channels (Sect. 3.3). Next, the terms determining the spatial interactions of neighboring dipoles *within* a depth layer are introduced (Sect. 3.4). The equations used to model the inhibitory interactions *between* depth layers are discussed in Sect. 3.5. As last step to complement the description of the mechanisms of the third model stage, Sect. 3.6 outlines how the T-junction detectors exert influence on the neighboring dipoles.

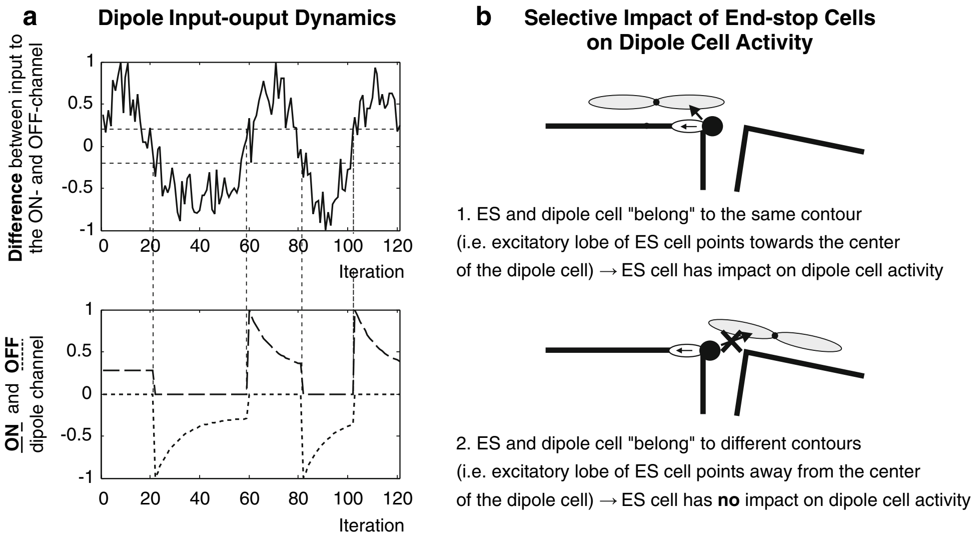


Fig. 6 Model stage of recurrent depth processing. **a** Example time course of the dipole input–output dynamics. The difference between the two input channels is continuously calculated and the dipole resets when the difference either exceeds or falls below a threshold value (indicated by the *two dashed lines in the upper plot*). The formerly active output channel is then set to zero, and the formerly inactive channels becomes active (the activities of both output channels are ⩾0; here, the time course of the OFF channel is mirrored at the x-axis for purposes of visualization). As long as the difference between the input channels does not change its sign again, the output of the active channel slowly decays to asymptotically reach a baseline level. **b** A dipole cell selectively integrates output activity at contour corners only if the excitatory kernels of the corresponding end-stop cells point approximately in the direction of the dipole cell

3.3 Dynamics of model dipole cells

A model dipole cell continuously calculates the difference between the input to its ON- and OFF-channels and determines those points in time at which the sign of this difference reverses (Fig. 6a). When this is the case, an antagonistic rebound is triggered: The output channel corresponding to the stronger input is set to the maximal output activity and the output channel of the weaker input is inhibited (this only happens if the difference exceeds a given threshold value in order to gain noise robustness). After the rebound, the output corresponding to the weaker input remains inhibited while the output activity corresponding to the stronger input slowly decays and asymptotically approaches a base activity level. The dipole remains in this condition until the difference between the inputs to the two channels reverses the sign again to trigger a rebound, and so on.

The functionality of the dipoles as described above is captured by the following equations. First, at time point t, the thresholded and half-wave rectified difference between the input channels g_t^{ON} and g_t^{OFF} is determined. This difference is then multiplied by the dipole output activities $d_{t-1}^{\mathrm{ON/OFF}}$ stemming from the prior time step t-1, and the result is represented as a_t^{ON} and a_t^{OFF}:

$$\begin{aligned} a_t^{\mathrm{ON}} &= \left[g_t^{\mathrm{ON}} - g_t^{\mathrm{OFF}} - C_{\mathrm{thres}}\right]^+ \cdot d_{t-1}^{\mathrm{OFF}} \\ a_t^{\mathrm{OFF}} &= \left[g_t^{\mathrm{OFF}} - g_t^{\mathrm{ON}} - C_{\mathrm{thres}}\right]^+ \cdot d_{t-1}^{\mathrm{ON}} \end{aligned} \tag{3.1}$$

$[x]^+ := \max\{x, 0\}$ stands for half-wave rectification. The constant C_{thres} is included to prevent that antagonistic rebounds of the dipole are triggered by spurious fluctuations of the input activities. The numerical value of the constant as used in the simulation study as well as the values of all model parameters described in the following equations are listed in Table C.1 in the Appendix. In Eq. (3.1), a non-zero value of, e.g., a_t^{ON} indicates that at time point t the ON-channel receives stronger input than the OFF-channel, while at the same time the output of the OFF-channel is still active. Consequently, non-zero values of $a_t^{\mathrm{ON/OFF}}$ signal those points in time at which an antagonistic rebound should be triggered. As long as one input remains stronger than the other one, both a_t stay at zero. The a_t are subsequently used to update the internal activity states $b_t^{\mathrm{ON/OFF}}$ of both channels:

$$\begin{aligned} b_t^{\mathrm{ON}} &= d_{t-1}^{\mathrm{ON}} - \beta\left[d_{t-1}^{\mathrm{ON}} - C_{\mathrm{base}}\right]^+ + C_{\max}\frac{a_t^{\mathrm{ON}}}{a_t^{\mathrm{ON}}+\alpha} \\ b_t^{\mathrm{OFF}} &= d_{t-1}^{\mathrm{OFF}} - \beta\left[d_{t-1}^{\mathrm{OFF}} - C_{\mathrm{base}}\right]^+ + C_{\max}\frac{a_t^{\mathrm{OFF}}}{a_t^{\mathrm{OFF}}+\alpha} \end{aligned} \tag{3.2}$$

As long as both a_t are zero, the b_t-activation corresponding to the active channel slowly decays and asymptotically approaches the base activity level C_{base} (β determines the decay rate). The b_t-activation corresponding to the weaker input is zero in this case (see Eq. 3.3). At time points at which the sign of the difference between the inputs changes, the a_t-activity of the newly dominating input exceeds zero and resets the corresponding b_t to its maximal value $C_{\max}$.

Mutual subtractive inhibition between the internal activity states $b_t^{\mathrm{ON/OFF}}$ is used to realize the competition between the two channels of a dipole:

$$\begin{aligned} d_t^{\mathrm{ON}} &= b_t^{\mathrm{ON}} \cdot \left[C_{\mathrm{gain}} \left[b_t^{\mathrm{ON}} - b_t^{\mathrm{OFF}} \right]^+ \right]^{\leqslant 1} \\ d_t^{\mathrm{OFF}} &= b_t^{\mathrm{OFF}} \cdot \left[C_{\mathrm{gain}} \left[b_t^{\mathrm{OFF}} - b_t^{\mathrm{ON}} \right]^+ \right]^{\leqslant 1} \end{aligned} \tag{3.3}$$

The activities d_t represent the output of the dipole cell. $[x]^{\leqslant 1} := \min\{x, 1\}$ denotes a bounded linear transfer function. The competition in Eq. (3.3) results in an ongoing suppression of the weaker channel. After resetting a b_t to its maximal value $C_{\max}$, it will dominate the b_t of the previously active channel, which is subsequently suppressed by the competition in Eq. (3.3). To summarize, antagonistic rebounds are triggered by non-zero values of a_t that reset the previously non-active channel to its maximal value and suppress the output of the previously active one. The output activities d_t are pooled by the receptive fields of the neighboring dipoles (as outlined in the next subsection) to constitute their new input activities g_t, thereby creating a recurrent flow of activity in each depth layer.

3.4 Interactions between neighboring dipole cells

Within a depth layer, T-junction information is propagated along contours by means of dipole cell resets (Figs. 5a, 6a). When a sufficient number of antagonistic rebounds occurs in the neighborhood of a dipole cell, the input to one of the channels becomes very strong and in turn resets the cell. The recurrent interaction between model cells can, therefore, result in a "wave" of activity resets traveling automatically in the two-dimensional map. In the following, the mechanisms are presented that restrict this wave to travel along a contour and prevent an unspecific spread of activity.

A key element is the usage of *anisotropic* receptive fields (or kernels) to specifically pool the activity delivered by the ON- and OFF-channels of those neighboring dipoles that correspond to contours (Fig. 5c). As dipole cells are not orientation selective *per se*, the shapes of their receptive fields are biased by the activity of the model $V2$ bipole cells of the first model stage at the corresponding topographical positions. For example, a horizontal contour maximally activates horizontally oriented $V2$ bipole cells, which in turn bias the dipole receptive fields to have a horizontally elongated shape. The adaptive shaping of the dipole kernels can be seen as adaptive filtering approach in which the input activation pattern is initially filtered at each spatial location by several elongated kernels having orientations ranging from 0 to π. Subsequently, the activity of the kernel that best matches the orientation of the underlying contour is selected. The adaptive filtering proceeds in several subsequent steps. First, the dipole activities corresponding to contours are extracted from the overall activation pattern in a depth layer by multiplying the dipole activities d_t with the normalized model $V2$ bipole activity l^{V2_Norm}:

$$\begin{aligned} e_{ti\theta}^{\mathrm{ON}} &= d_{ti}^{\mathrm{ON}} \cdot l_{i\theta}^{V2\_\mathrm{Norm}} \\ e_{ti\theta}^{\mathrm{OFF}} &= d_{ti}^{\mathrm{OFF}} \cdot l_{i\theta}^{V2\_\mathrm{Norm}} \end{aligned} \tag{3.4}$$

Subscripts t, i and θ represent time, the spatial location and orientation, respectively. In the computational implementation of the model, activities l^{V2_Norm} and $e_t^{\mathrm{ON/OFF}}$ were represented by 3D matrices: 2D—space, 1D—orientation. Eight discrete orientations θ were used, ranging from 0 to $7/8\pi$ in steps of $\pi/8$. The normalized activity l^{V2_Norm} is derived from the steady-state $V2$ bipole activity l^{V2} using subtractive inhibition in the orientation domain combined with shunting self-inhibition to scale the activity range from 0 to 1 (details in Appendix C; Eq. C.1). The maps of modulated dipole activities $e_{ti\theta}^{\mathrm{ON/OFF}}$ selectively signal the dipole activities at contour positions. As activity $e_{ti\theta}^{\mathrm{ON/OFF}}$ results from multiplying the dipole activities with the orientation-selective activity l^{V2_Norm}, the orientation of the underlying contours is also encoded in the activity pattern of $e_{ti\theta}^{\mathrm{ON/OFF}}$. Subsequently, the dipoles integrate the activities $e_{ti\theta}^{\mathrm{ON/OFF}}$ in their neighborhood using elongated receptive field kernels V (Fig. 5c):

$$\begin{aligned} f_{ti\theta}^{\mathrm{ON}} &= \left\{ e_t^{\mathrm{ON}} * \Psi_f * V \right\}_{i\theta} \\ f_{ti\theta}^{\mathrm{OFF}} &= \left\{ e_t^{\mathrm{OFF}} * \Psi_f * V \right\}_{i\theta} \end{aligned} \tag{3.5}$$

Ψ_f denotes a Gaussian weighting function in the orientation domain, V is the kernel in the spatial domain, and $*$ is the convolution operator. V is modeled as modified anisotropic Gaussian kernel which is normalized to yield steeper flanks at the boundaries of the kernel (Appendix C; Eq. C.2). At each topographical position, eight kernels were used with the main axes oriented between 0 and $7/8\pi$ in steps of $\pi/8$. Each kernel selectively pools the dipole activities $e_{ti\theta}^{\mathrm{ON/OFF}}$ from the orientation channel which corresponds to its own kernel orientation. At the intersections of overlapping surfaces, this prevents crosstalk between dipoles corresponding to (differently oriented) contours of different surfaces. Activities $f_{ti\theta}^{\mathrm{ON/OFF}}$ are finally weighted with the normalized $V2$ bipole activity l^{V2_Norm} to adaptively shape the dipole kernels: The multiplication with l^{V2_Norm} selects those activities f which were pooled by anisotropic kernels V that correspond best to the orientation of the underlying contour. The weighted activities are summed over all orientations, resulting in the new input activities $g_{(t+1)}^{\mathrm{ON/OFF}}$ (see Eq. 3.1) of the dipole at time $t+1$:

$$\begin{aligned} g_{(t+1)i}^{\mathrm{ON}} &= \sum_{k=1}^{N\_\mathrm{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\mathrm{orient}}^{\mathrm{ON}} \cdot l_{i(k-1)\cdot\pi/N\_\mathrm{orient}}^{V2\_\mathrm{Norm}} \right) \\ g_{(t+1)i}^{\mathrm{OFF}} &= \sum_{k=1}^{N\_\mathrm{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\mathrm{orient}}^{\mathrm{OFF}} \cdot l_{i(k-1)\cdot\pi/N\_\mathrm{orient}}^{V2\_\mathrm{Norm}} \right) \end{aligned} \tag{3.6}$$

The kernel orientation is coded by $(k-1)\cdot\pi/N\_\text{orient}$ that ranges from 0 and $7/8\pi$ ($k=1,\ldots,8$ as $N\_\text{orient}=8$). The elongated layout of the kernels V in Eq. (3.5) allows the dipoles to selectively pool the activities of those dipoles in their neighborhood which correspond to the same underlying contour. This mechanism helps the overall model to gain noise robustness as it prevents crosstalk between dipoles which signal contours of neighboring surfaces patches and in consequence are likely to have different orientations. At the same time it allows for a faster propagation of the waves of activities along contours, as the dipoles can integrate activity in a wider spatial range. However, problems might occur at surface corners at which the contours segments are differently oriented but nevertheless belong to the same surface patch. Consequently, the wave of activities might be disrupted at corners as the elongated kernels cannot integrate the dipole activity from the part of the contour that is "around the corner". It might even happen that they pool the activity of an appropriately aligned contour belonging to a neighboring surface patch (as depicted in Fig. 6b). In order (i) to allow the waves of activity to flow along contour corners and (ii) to prevent crosstalk between aligned contours of neighboring surface patches, the end-stop cell activities of the second model stage amplify the impact of those dipoles situated at contour corners. Additionally, the kernels selectively pool only the dipole activities amplified by those end-stop cells pointing in the direction of the contour (see Fig. 6b; the mathematical details are given in Appendix C).

The six equations outlined in this and the previous subsection describe the basic mechanisms of dipole cell dynamics and dipole interactions that are necessary for a recurrent flow of activity *within* the depth layers, in turn enabling the waves of activity resets to autonomously propagate along contours. In the following sections, the last two key components of the model are outlined that distribute the model activation patterns across the model layers according to the depth positions of the corresponding contours. The above-depicted equations are adapted to incorporate (i) the inhibitory interactions between the model layers and (ii) the mechanisms by which the T-junction detectors trigger the waves of activity within the model layers.

3.5 Interaction between depth layers

As already outlined above, inhibitory mechanisms acting between the depth layers transfer those activity patterns from the outer to the inner layers that represent contours having medial depth positions (Sect. 3.2; Fig. 5b/d): Dipole cells with active ON-channels inhibit all neurons that are at the corresponding topographical positions in the inner depth layers. Once a T-junction detector triggers a wave of OFF-channel activity that travels along a contour, the neurons representing that contour release their inhibition on the dipoles in more medial depth layers. By this release of interlayer inhibition, the contour is subsequently signaled by the activities of dipoles in the more medial depth layers. For example, in a model with N depth layers as depicted in Fig. 5d, the activity in layer $K=3$ is suppressed by ON-channel activity in layers 1, 2 as well as $N-1$ and N. Formally speaking, a dipole in layer #K is inhibited by active ON-channels of the corresponding dipoles in the *outer layers* of K, whereby the *outer layers* are defined as:

$$\text{outerlayers} = \{1,2,\ldots,(\kappa-1)\}\cup\{(N-\kappa+2),\ldots,N\}$$
$$\text{with } \kappa=\min\{K,|N-K+1|\} \qquad (3.7)$$

Using this formal definition, one can incorporate the inhibition of dipole output activity in layer #K into the model using a modified version of Eq. (3.4)

$$e_{ti\theta}^{\text{ON}} = \left[d_{ti}^{\text{ON}} - C_{ol}\cdot\sum_{j\in\text{outerlayers}} d_{ti}^{\text{ON\_layer(j)}}\right]^{+}\cdot l_{i\theta}^{V2\_\text{Norm}}$$
$$e_{ti\theta}^{\text{OFF}} = \left[d_{ti}^{\text{OFF}} - C_{ol}\cdot\sum_{j\in\text{outerlayers}} d_{ti}^{\text{ON\_layer(j)}}\right]^{+}\cdot l_{i\theta}^{V2\_\text{Norm}} \qquad (3.8)$$

whereby the superscript $ON\_layer(j)$ is used to indicate that the dipole activity $d_t^{\text{ON\_layer(j)}}$ stems from the outer layer j. The ON-channel activities of all dipoles in the outer layers are added to suppress the outputs of both the ON-and OFF-channel via subtractive inhibition (constant C_{ol} determines the strength of inhibition). By this mechanism, waves of OFF-channel activity in the outer layers release the inhibition of dipoles in more medial layers.

3.6 The impact of T-junctions on dipole cell activity

The last key mechanism of the model specifies the impact of the model T-junction detectors on the dipole activities in order to enable the overall model to recursively determine the depth of overlapping contours. The T-junction detectors selectively trigger waves of OFF-channel activity to reset those dipoles that signal contours not belonging to the current depth layer. In the background layer, for example, the triggering of the waves is achieved by a local increase of the OFF-channel input g^{OFF} to those dipoles that correspond to the *hats* of the Ts. Likewise, in the foreground layer, the OFF-channel input to the dipoles that represent the *stems* of Ts is increased. In order to prevent the resulting waves of OFF-channel activity to unintentionally spread along the intersecting contours, the input to the ON-channels of the remaining dipoles in the neighborhood of the T-junction detectors is also increased.

The specific impact of a T-junction detector on dipole positions corresponding either to the hat or to the stem of a T is achieved by applying a stage of normalization and Gaussian filtering in the spatial and orientation domain to the initial junction activity (for details please refer to Appendix D). This results in the activity distribution q_{ti}^{foregr} that signals positions at which a contour element is locally in the foreground of another contour (i.e., at the hat of a T), and in q_{ti}^{backgr} that represents the positions of contours being locally in the background (i.e., at the stem of a T). In the background layer #1, activity q_{ti}^{foregr} increases the input g^{OFF} to the OFF-channels of dipoles corresponding to the hats of the Ts. At the same time, the dipoles representing the stems are stabilized by q_{ti}^{backgr} that increases the input to their ON-channels. Accordingly, the modified Eq. (3.6) reads:

$$\begin{aligned} g_{(t+1)i}^{\text{ON}} &= \sum_{k=1}^{N\_\text{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\text{orient}}^{\text{ON}} \cdot l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \right) \\ &\quad + q_{ti}^{\text{backgr}} \\ g_{(t+1)i}^{\text{OFF}} &= \sum_{k=1}^{N\_\text{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\text{orient}}^{\text{OFF}} \cdot l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \right) \\ &\quad + q_{ti}^{\text{foregr}} \end{aligned} \tag{3.9}$$

In the foreground layer $\#N$, activity q_{ti}^{backgr} excites the OFF-channels of the dipoles representing the stems of the Ts, and activity q_{ti}^{foregr} stabilizes the input to the ON-channels of the dipoles representing the hats:

$$\begin{aligned} g_{(t+1)i}^{\text{ON}} &= \sum_{k=1}^{N\_\text{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\text{orient}}^{\text{ON}} \cdot l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \right) \\ &\quad + q_{ti}^{\text{foregr}} \\ g_{(t+1)i}^{\text{OFF}} &= \sum_{k=1}^{N\_\text{orient}} \left(f_{ti(k-1)\cdot\pi/N\_\text{orient}}^{\text{OFF}} \cdot l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \right) \\ &\quad + q_{ti}^{\text{backgr}} \end{aligned} \tag{3.10}$$

In case of intermediate layers, it depends upon the depth position of a layer whether its OFF-channels are triggered either by q_{ti}^{foregr} or by q_{ti}^{backgr} activity. For layers #1 to $\lfloor N/2 \rfloor$ that represent rather distant depth positions, Eq. (3.9) describes the impact of activity q_{ti}^{foregr} to reset those dipoles corresponding to contours being locally in the foreground of other contours. For layers $\#\lceil N/2 \rceil + 1$ to N representing more frontal positions, Eq. (3.10) is used to specify how activity q_{ti}^{backgr} resets the dipoles corresponding to contours being locally in the background.

In the theoretical scheme of depth sorting outlined in Sect. 3.1, a T-junction is no longer considered in the further processing steps once one of the intersecting contours has been assigned to a depth layer (Fig. 4). Accordingly, T-junction information in the third model stage that, e.g., stems from the intersection of the contour in the foreground with other contours should not exert influence on the activation patterns in the intermediate depth layers. Only T-junction activity resulting from the intersection of contours both having intermediate depth positions should exert influence on the dipole activities in the intermediate layers. This is achieved by the inhibition of activities q_{ti}^{foregr} and q_{ti}^{backgr} in the intermediate layers by active ON-channels of dipoles in the outer layers (analogous to the inhibition between the dipoles across depth layers, as described in Sect. 3.5; please refer to Appendix D for details). By release of this inhibition, the T-junction information corresponding to contours having intermediate depth positions is transferred from the outer to the inner depth layers.

The computational stages and mechanisms of the third model stage were motivated by the physiological finding that cells tuned to different disparity profiles can already be found in early visual areas (Poggio et al. 1988). Furthermore, neurons in visual areas $V2$ and $V4$ have been shown to signal border ownership relations and the relative disparity between objects (e.g., Baumann et al. 1997; Heider et al. 2000; Qiu and v. d. Heydt 2005; Zhou et al. 2000). Consequently, information about the depth relations between objects seems to be represented already at early processing stages of the ventral visual stream. Determining globally consistent depth relations also requires an exchange of information between remote positions in the visual scene. However, the range of neural interaction is limited in early visual areas even when considering horizontal long-range connections spanning several cortical hypercolumns (Gilbert and Wiesel 1989; Hirsch and Gilbert 1991). So the question arises if depth information that is in accordance with the global arrangement of the scene can already be processed by cells in early visual areas, or if it necessarily requires the involvement of higher visual areas with cells having large receptive fields (Smith et al. 2001). In this modelling approach, we explore the components and mechanisms which are necessary to achieve a globally consistent depth representation of overlapping 2D surface patches based on locally restricted cell interactions.

In this section, we developed the equations specifying the three key components of third model stage that

- allow for a propagation of activity *within* model layers via recurrent interactions between cells in a locally restricted neighborhood (Sects. 3.3 and 3.4),
- distribute the model activity *between* the depth layers via local inhibitory interactions between cells in different layers (Sect. 3.5), and
- enable the T-junction detectors to trigger waves of activity resetting those dipole activations corresponding to contours having different depth positions (Sect. 3.6).

These mechanisms are captured in a set of six equations used in the subsequent computational simulations: Eq. (3.1) compares the input and output of a dipole cell to signal time points at which an antagonistic rebound should be triggered. Equation (3.2) updates the internal activity states of a dipole, and Eq. (3.3) determines its output. Equation (3.8) determines the dipole outputs that correspond to contours and incorporates the inter-layer inhibition into the model. Equation (C.5) is based on Eq. (3.5) and describes how cells integrate the neighboring dipole activity, but additionally contains mechanisms to selectively enhance activity at contour corners to ensure a smooth flow of activity "around corners". Finally, Eqs. (3.9) and (3.10) determine the new input to a dipole cell and describe the impact of T-junction activities on the dipole input. In the following results section, it will be demonstrated that the combination of these mechanisms allows for a globally consistent depth sorting based on the locally restricted dipole interactions.

4 Simulation results

In the following, simulation results are used to demonstrate the computational capabilities of the model stages of contour processing (Sect. 2) and recurrent depth processing (Sect. 3) in determining a globally consistent representation of the depth of surface contours. In the first two subsections, the presentation is focused to a specific example stimulus in order to demonstrate the overall functionality of the models in a step-by-step fashion. Subsequently, additional input stimuli were utilized to sketch the full computational properties of the models.

4.1 First and second model stages: contour processing and T-junction detection

In the following, the key properties of the first model stage of recurrent $V1$–$V2$ contour processing followed by the second stage for T-junction and corner detection (Sect. 2) are outlined using the stimulus depicted in Fig. 7. The final equilibrated $V1$ and $V2$ model activation patterns after seven iterations are shown. For purposes of visualization the activities are summed up over all orientations at each spatial position. The resulting two-dimensional activity distributions are illustrated as gray-scale images, with the maximal activity of each model area coded as white. The stimulus consists of four mutually overlapping gray and white rectangles. Model $V1$ complex cells signal the positions and orientations of the visible outlines of the rectangles. The $V1$ activity pattern serves as bottom–up input to model $V2$ bipole cells that group commonly aligned contour fragments to form continuous activity patterns. Recurrent feedback interaction within model area $V2$ and between model areas $V2$ and $V1$ helps to stabilize and enhance the initial $V2$ responses to fragmented contours, resulting in smooth and continuous $V2$ activation patterns.

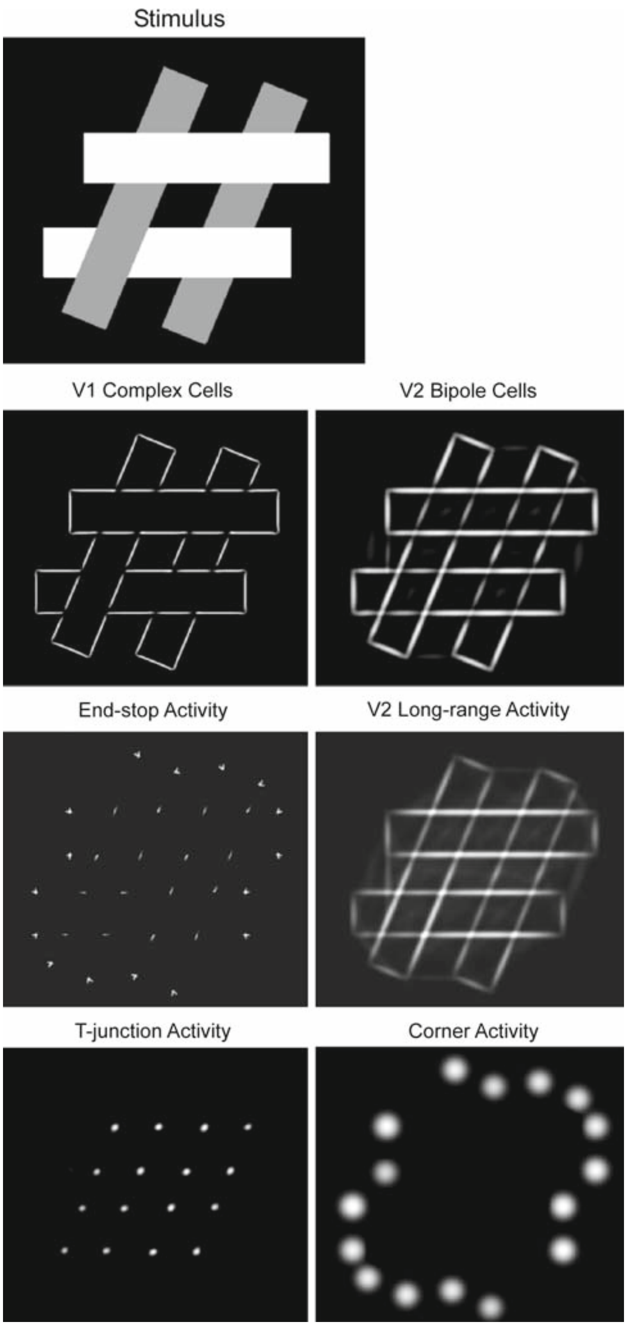


Fig. 7 Example for the processing of contours by the model of recurrent $V1$–$V2$ interaction. The physical outlines of the *four gray or white rectangles* are signaled by $V1$ complex cells. $V2$ bipole cells subsequently insert illusory contours to bridge gaps created by overlapping surfaces. The model $V1$–$V2$ activation patterns constitute the input to subsequent feed-forward stages for T-junction and corner processing. $V1$ complex cell activity is linearly filtered by end-stop cells to detect contour endings. Finally, end-stop activity and $V2$ long-range activity is combined to detect T-junctions and corners (see Sect. 2.3)

Taken together, the overall model stage of recurrent $V1$–$V2$ interaction achieves a robust processing of surface contours capable of bridging gaps caused by noisy illumination or due to overlaps (please refer to Neumann and Sepp 1999 for a comprehensive description of the key model properties).

The activity pattern of the model stage of recurrent $V1$–$V2$ contour processing constitutes the input to the feed-forward stage for T-junction and corner detection. $V1$ complex cell activity is linearly filtered by model end-stop cells to determine contour endings (Fig. 7: third row left). End-stop activity of approximately perpendicular orientation but same topographical position is combined to indicate corners and endings (L-junctions) in the input image (Fig. 7: last row right). Likewise, combination of end-stop activity and $V2$ long-range activity of approximately perpendicular orientation results in an activity pattern signaling the position and orientation of T-junctions in the input image (Fig. 7: last row left). In combination, the model $V2$ bipole activities that signal the (completed) contours in the input image as well as the T-junction and corner activity patterns constitute the input to the model of recurrent depth processing (Fig. 2).

4.2 Third model stage of recurrent depth processing

The functionality of the neural mechanisms of the model stage of recurrent depth processing (Sect. 3) is demonstrated step-by-step starting with the dipole dynamics which enables local T-junction activity to be propagated along surface contours. The interaction between depth layers via inhibitory connections resulting in a globally consistent representation of the depth of surface contours is subsequently outlined.

Propagation of T-junction information within a depth layer

T-junction activity signals local foreground-background relations at crossings of surface contours. In the third model stage, this local information is distributed along contours, as depicted exemplarily for the background layer in Fig. 8. For purposes of visualization, only the dipole activities at topographical positions corresponding to surface contours are shown. This is achieved by multiplying the ON and OFF-channel activities with a mask constructed by thresholding the $V2$ bipole activation pattern. Model $V2$ bipole activity is summed over all orientations at each spatial position (as shown in Fig. 7) and subsequently normalized to yield a contour mask being approximately 1 at contour positions and 0 otherwise. Here and in the following subsections, we will use snapshots of the model activity at selected points in time to demonstrate the dynamics of the development of the model activation pattern. The time points shown were chosen so that they show a sequence of activation states which best illustrates the overall model functionality. The overall time which the model needs to reach a final stable activation pattern does not only depend on the model parameters (determining, e.g., the spatial range of dipole interaction), but also, e.g., on the number of overlapping objects in the image or the length of the contours. In consequence, the selected time points shown in the figures vary from stimulus to stimulus.

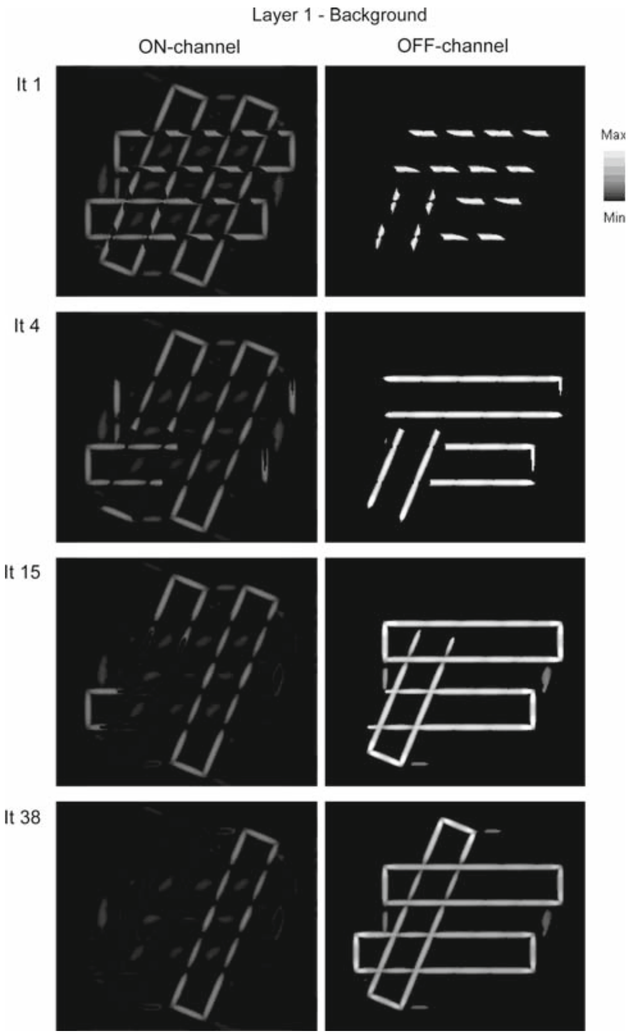


Fig. 8 Recurrent model of depth processing: development of the dipole activity pattern in the background model layer in response to the stimulus of Fig. 7. Only the dipoles positions corresponding to surface contours are shown by applying a mask based on thresholded model $V2$ bipole activity (see Sect. 4.2). Iteration 1: initially, all contours are represented by active ON-channels and T-junction activity locally resets dipoles at contours being locally in the foreground of other contours. Iterations 4 and 15: T-junction information is distributed by waves of OFF-channel activity running along contours. Iteration 38: finally, only the ON-channels of those dipoles corresponding to the contour being in the absolute background remain active

Initially, the ON-channels of all dipoles of a layer are active at the (low) base activity level. Subsequently, T-junction activity triggers the rebounds of dipoles in a local neighborhood, resulting in topographical regions of high OFF-channel activity at contours not belonging to the actual depth layer. For example, in the background layer, T-junction activity locally resets the dipoles corresponding to contours which are locally in the foreground of another contour (Fig. 8

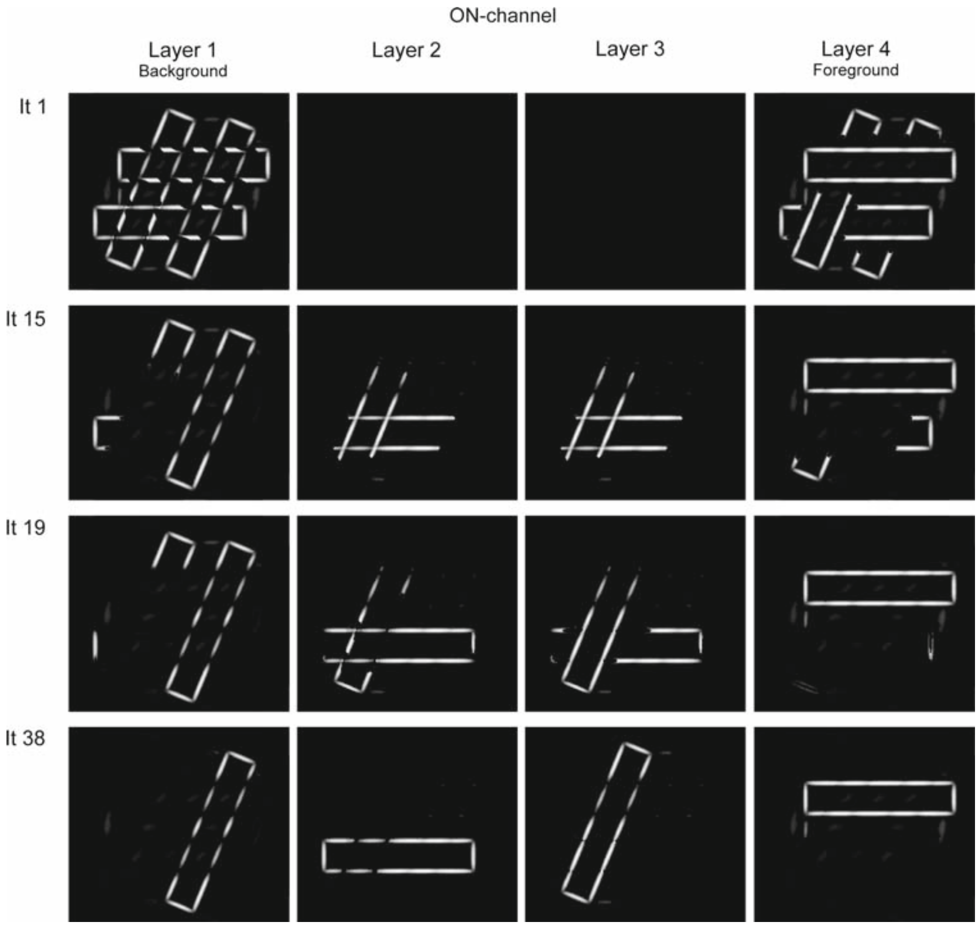


Fig. 9 Development of the activation pattern of the model of recurrent depth processing in response to the stimulus of Fig. 9. Iteration 1: only the dipoles in layers 1 (background) and 4 (foreground) are active and suppress the dipole activity in the intermediate layers. Iteration 15: T-junction activity has triggered waves of OFF-channel activity resetting those dipoles in layers 1 and 4 which correspond to contours not belonging to these depth layers. At topographical positions at which neither the dipoles in layer 1 nor those in layer 4 have activated ON-channels, dipole activity in the intermediate layers is released. Iteration 19: T-junction activity triggers waves of activity in the intermediate layers 2 and 3 resetting dipoles corresponding to contours belonging to another depth layer. Iteration 38: the final model activation pattern is reached, signaling the correct depth sorting of the rectangles in the stimulus

top row: iteration 1). The high OFF-channel activities of the reset dipoles are pooled by neighboring dipole cells that in turn reset themselves (Fig. 8 second row: iteration 4). This results in "waves" of OFF-channel activity propagating along the contours not belonging to the actual depth layer. Finally, dipoles with active ON-channels signal the positions of contours having a depth corresponding to the actual layer (Fig. 8 bottom row: iteration 38). Successful completion of fragmented contour elements by the first model stage of recurrent $V1$–$V2$ interaction is crucial to prevent the waves of dipole activity of being stopped at gaps caused by overlapping surfaces. Furthermore, the elongated receptive fields of the dipole cells enable the waves of dipole activity to bridge small contour gaps, thereby helping the overall model to gain noise robustness. Otherwise, ON-channel activity would persist in wrong depth layers, resulting in an ambiguous representation of a contour being distributed over several depth layers. The anisotropic receptive fields integrate dipole activity related to a specific surface contour. This increases the spatial range of information pooling by preventing influences of dipoles which correspond to contours of neighboring or crossing surfaces. However, the anisotropic spatial layout of the receptive fields would normally result in a disruption of the waves of activity at contour corners. This effect is avoided by selectively enhancing the impact of dipole activity at contour corners (see end of Sect. 3.4; Fig. 6b).

The flow of dipole activity in the foreground layer (Fig. 9 right column) is contrary to that in the background layer: Initially, T-junction activity locally resets the dipoles corresponding to contours that are in the background of another contour (Fig. 9 right column: iteration 1). This information is distributed along the contours by waves of OFF-channel dipole activity (Fig. 9 right column: iteration 15 and 19). Finally, ON-channel activities remain only at positions corresponding to contours being in the foreground of all others (Fig. 9 right column: iteration 38).

Interaction between depth layers

Inhibitory connections from the outer to the inner depth layers enable interaction between the layers to develop a consistent depth representation (Fig. 9). Dipole cells in outer layers with active ON-channels inhibit the dipole outputs in the inner layers at corresponding spatial positions (Sect. 3.5). Consequently, the ON- and OFF-channel activities of the dipoles in depth layers 2 and 3 are initially suppressed (Fig. 9 top row: iteration 1). The dipoles in these layers receive no activity from their neighboring dipoles and remain in their actual state with the ON-channels active at the (low) base activity level, which is subsequently suppressed by the inhibitory interaction from the outer model layers.

After some iterations the output activity of dipoles in layers 2 and 3 is released at those topographical positions at which the dipoles in layers 1 and 4 have been reset and have inactive ON-channels (Fig. 9 second row: iteration 15). At this point in time, activation patterns in layers 2 and 3 are equal to each other, i.e., the mutual depth relations between the contours represented in these layers are still unresolved. Consequently, the representation of contour depth by the model activation patterns starts to differentiate between contours being globally in foreground and background as well as at some intermediate position.

The waves of OFF-channel activity in the outer depth layers stop the suppression of T-junction activity in the intermediate layers. In layer 2, the released T-junction activity results in a reset of dipoles at topographical positions corresponding to contours being locally in the foreground of other contours (Fig. 9 third row: iteration 19). Likewise, the dipoles of layer 3 which correspond to contours being in the background of another contour are reset in a local neighborhood of the T-junctions. Consequently, the release of dipole activity in the intermediate layers is immediately followed by waves of OFF-channel activity (triggered by T-junction activity) to resolve the depth relations between the contours having intermediate depth positions. Finally, the model reaches a stable final activation pattern signaling a globally consistent depth sorting of the surface contours (Fig. 9 bottom row: iteration 38).

Taken together, the overall model capabilities in depth sorting emerge from two key properties: Within each layer, local T-junction activity is propagated along contours by waves of dipole activity, thereby resetting all dipoles corresponding to contours not belonging to the actual depth layer. Between two layers, inhibition of cell activity in the inner layer by active ON-channels in the outer layer results in a recursive scheme for the assignment of surface contours to the depth layers. In this recursive scheme, the determination of the surface contours assigned to the overall foreground and background is fastest. The remaining surface contours are assigned to an intermediate depth level and the exact determination of their absolute depths takes more time to develop.

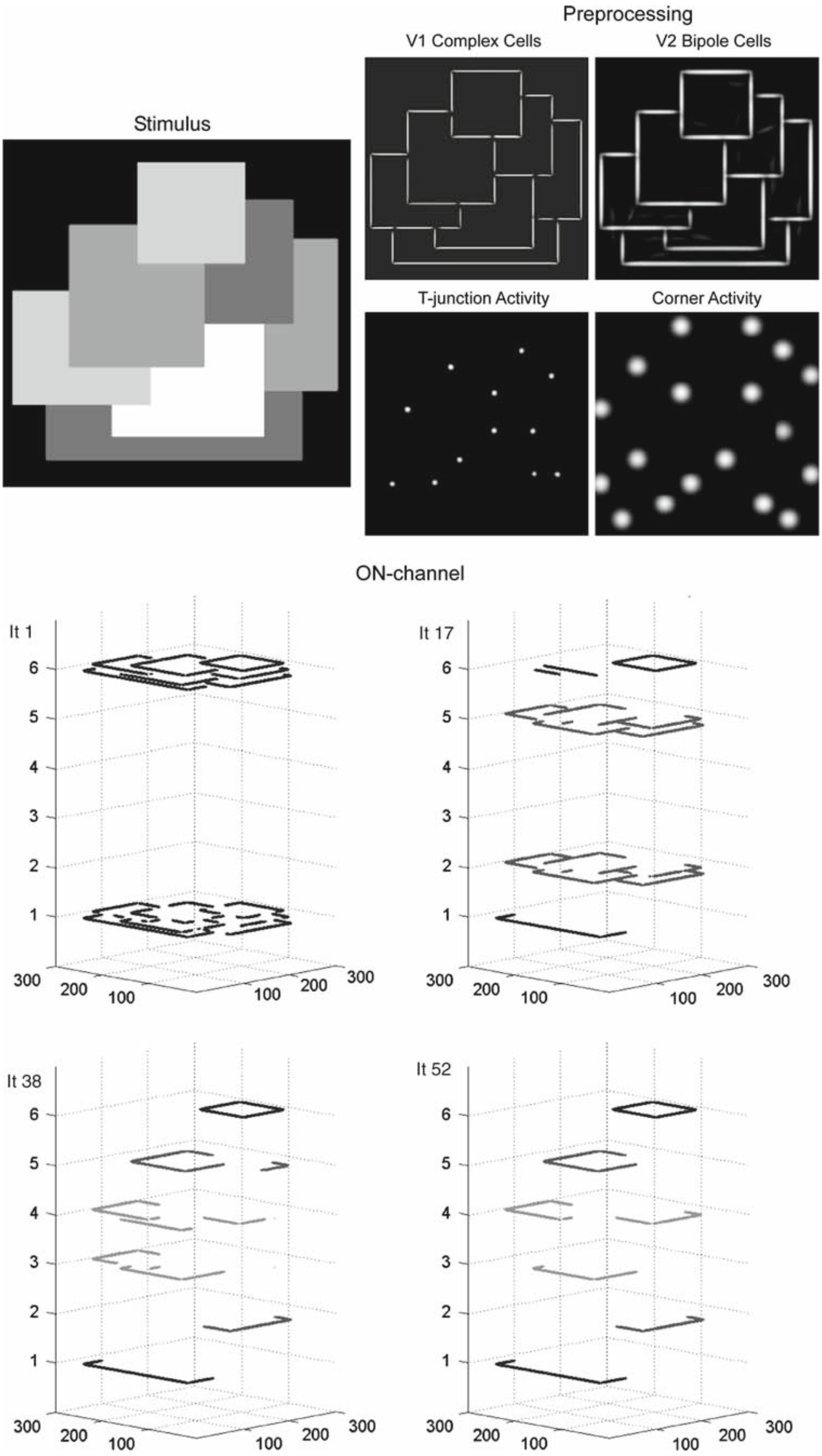
4.3 Further simulation results

In the previous subsections, the presentation was focused to a specific stimulus in order to allow for a demonstration of the general model functionality. In the following, additional simulation results help to highlight the model capabilities as well as its limitations in more detail. First, a stimulus containing overlapping rectangles arranged in six different relative depths is considered (Fig. 10). Human observers are unable to instantaneously determine the exact depths of all surface patches at the same time. This indicates that early (i.e., pre-attentive) visual processes in isolation do not succeed in completely segmenting the stimulus. It seems that only the surfaces being in the absolute foreground and background are easily detected and active scanning of the scene is required to determine the exact depth sorting of the intermediate surface patches. In contrast, no theoretical limit of the maximal number of overlapping patches exists for the model. Instead, only the number of dipole layers has to be high enough (see Sect. 3.1).

The stimulus is preprocessed by the model stage of recurrent $V1$–$V2$ interaction, as depicted in the upper right area of Fig. 10. The contours of the surface patches are signaled by model $V2$ bipole cells. The subsequent feed-forward stage filters the $V1$ and $V2$ model activation patterns to determine the position and orientation of T-junctions and corners in the stimulus. The resulting activity patterns constitute the input to the model stage of recurrent depth processing. In Fig. 10, positions at which the ON-channel activities exceed a certain threshold are shown. Additionally, a mask obtained by thresholding the $V2$ bipole activity is applied to isolate dipole activity corresponding to contours (see Sect. 4.2). We utilized displays in which the depth layers were arranged along the z-direction, starting from layer 1 (background) to layer 6 (foreground). Additionally, different gray scale values were utilized to discriminate between the different depth layers. Dark and light grays indicate outer and inner depth layers, respectively.

Despite the higher number of overlapping objects and model depth layers utilized, the development of the model activation pattern in time is similar to the one in response to the previous stimulus: Initially, only the dipoles in the outer layers are active (iteration 1) and waves of activity triggered by T-junctions start to reset all dipoles corresponding to contours being not in the foreground (layer 6) and background (layer 1), respectively. After some iterations, most parts of the intermediate contours are no longer represented by ON-channel activity in layers 1 or 6, and dipole activity is released at the corresponding topographical positions in next inner layers 2 and 5 (iteration 17). T-junctions again trigger waves of dipole activity to reset the contours not belonging to these layers. Finally, in iteration 38, depth layers 1, 2, 5 and 6 have reached their final activation patterns. Dipole activity corresponding to contours at the remaining intermediate depth positions is released in layers 3 and 4 and the depth relations are subsequently resolved in these depth layers triggered by the T-junction activity. The final model activation pattern is reached in iteration 52. This scheme for the assignment of contours to depth positions implemented by the inhibitory connections between the model layers could

Fig. 10 Model responses to a stimulus containing overlapping surface patches arranged in six depth levels. Preprocessing: the model of recurrent $V1$–$V2$ interaction signals the surface contours, followed by the feed-forward stages detecting the positions of T-junctions and corners in the input image. Model stage of recurrent depth processing: dipoles with active ON channels are shown at several points in time (see Sect. 4.3). The depth layers are arranged in z-direction, ranging from 1 (background) to 6 (foreground). Iteration 1: only the dipoles in the background and foreground layers are active and suppress the dipole activity in the intermediate layers. In layers 1 and 6, T-junction activity triggers waves of activity to reset dipoles corresponding to contours which are neither in the background or foreground, respectively. Iteration 17: at topographical positions at which the dipoles in layers 1 and 6 do not have activated ON channels, dipole activity in the intermediate layers 2 and 5 is released. Iteration 38: at positions without ON-channel activity in layers 1, 2, 5 and 6, dipole activity in layers 3 and 4 is released. Iteration 52: the final model activation pattern is reached, signaling the correct depth sorting of the surface patches



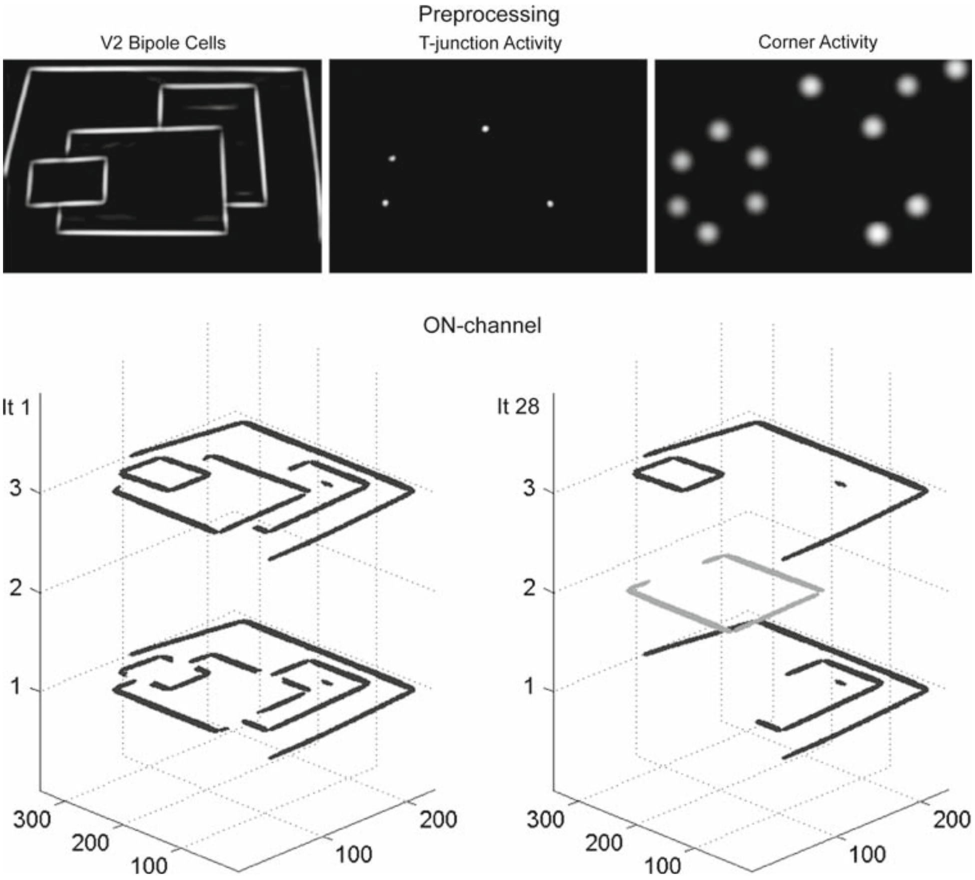


Fig. 11 Model responses to the real-world stimulus of Fig. 1a. The preprocessing network signals the positions of contours, corners and T-junctions in the input image. The model stage of recurrent depth processing successfully achieves the correct depth sorting of the three central overlapping objects. However, the depth relation between these objects and the surrounding contour of the *dark gray* underlay remains unresolved, as no common contour crossings exist. This demonstrates a general limitation of object representations which are based only on contour information

be recursively continued and is (in principle) only limited by the number of depth layers.

In Fig. 11, the model reactions to the real-world picture of Fig. 1 are shown. $V2$ bipole cells signal the outlines of the three objects and of the underlay. The activation patterns of the model stage of $V1$–$V2$ interaction constitutes the input to the feed-forward processing stage determining the positions of T-junctions and corners in the image. Subsequently, the depth relations of the three overlapping central objects are successfully resolved by the model stage of recurrent depth processing that reaches its final activation pattern in iteration 28. The three central objects are located on a common underlay. Parts of the outline of this underlay are visible in the image and are signaled by the $V2$ bipole cells. These outlines also result in ON-channel activities in layers 1 and 3 of the model of recurrent depth processing at corresponding spatial positions. Furthermore, some spurious $V2$ activations caused by the text on the paper also result in ON-channel activities in layers 1 and 3. However, as these activities do not overlap with others and consequently no corresponding T-junctions exist, the model stage of recurrent depth processing is unable to resolve their depth relations. This demonstrates a general limitation of object representations that are based on contour information alone. In order to enable the processing of depth relations of objects with non-overlapping contours, surface-based mechanisms have to be involved.

The final example demonstrates that interaction of the model stage of recurrent $V1$–$V2$ contour processing and the model stage of recurrent depth processing can achieve successful figure-ground segmentation of stimuli based on illusory contours, such as a Kanizsa square (Fig. 12): $V1$ complex cell activity signals the physical outlines of the pac-man inducers. $V2$ bipole cells subsequently bridge the gaps between the inducers by reacting to the co-linearly aligned contour fragments. Consequently, $V2$ activation patterns additionally signal the illusory contours of the Kanizsa square (see also Grossberg and Mingolla 1985; Neumann and Sepp 1999). The straight parts of the physical outlines of the inducers in combination with the illusory contours signaled by $V2$ bipole cells represent the complete outline contour of the Kanizsa square. The overlaps between the contour of the Kanizsa square and the remaining circular physical outlines of the inducers create “illusory” T-junctions, which are detected by the second model stage. Finally, the depth relations between the contours are resolved by the model stage of recurrent depth processing using the “illusory” T-junction information as cue. To conclude, in our model architecture, illusory contours are able to transform

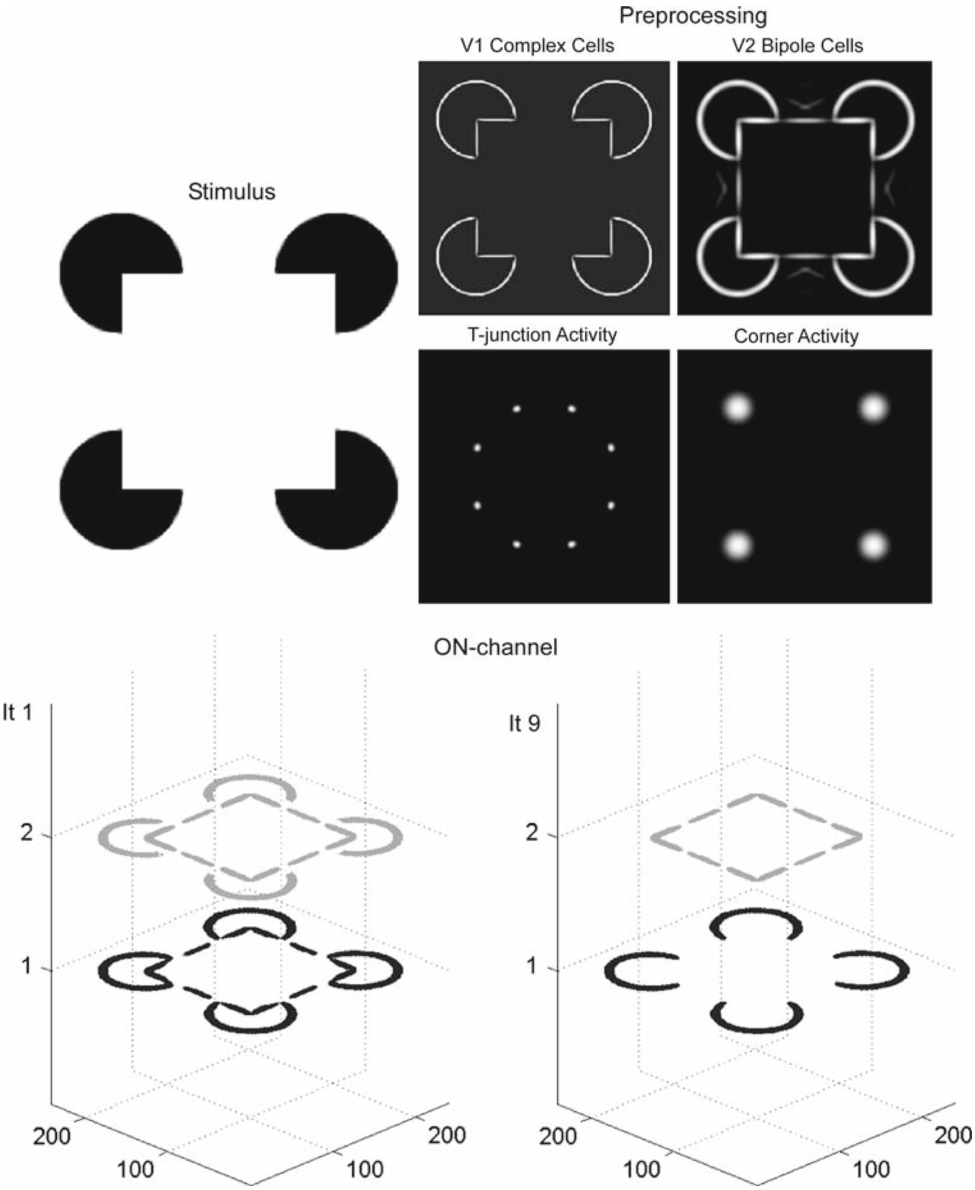


Fig. 12 Model responses to a Kanizsa square. Model $V2$ bipole cells respond to the illusory contours constituting the square. The subsequent feed-forward stages detect the "illusory" T-junctions created by the overlap of the outline of the Kanizsa square (as signaled by $V2$ bipole cells) and the remaining circular contour elements of the inducers. Finally, the model of recurrent depth processing successfully segregates the contours of the square from the contours of the inducers

physical corners to "illusory" T-junctions, thereby enabling a depth segmentation of the resulting contour representation.

5 Discussion

5.1 General model framework: surface boundary processing and depth propagation

Our overall computational framework divides into two successive steps: First, surface boundaries are processed in the model stage of recurrent $V1$–$V2$ interaction, thereby interpolating contour gaps via $V2$ long-range grouping mechanisms. Second, the completed boundary representation is used to recursively determine the depth of the contours. A globally consistent depth representation is obtained by the recurrent propagation of local T-junction information along the contours. This hierarchy of processing steps is supported by a series of psychophysical studies of Shipley and colleagues (Kellman and Shipley 1991; Kellman et al. 1998; Shipley and Kellman 1992). The authors propose a formal concept of relatability which defines those spatial arrangements which result in a common grouping of two contour elements to form smooth continuous curves by the human visual system. The psychophysical results demonstrate that relatability is identical for modal and amodal completions, indicating that the human visual system uses common grouping mechanisms independent from depth relations. Likewise, in our model, grouping by recurrent $V1$–$V2$ interaction occurs prior to depth processing and depth cues are subsequently utilized to determine whether gaps were completed modally (i.e., in front of other objects) or amodally (i.e., behind other objects). In Neumann and Sepp (1999) the relatability constraint was implemented in a model of $V1$–$V2$ interaction by utilizing excitatory interactions between co-circular tangents while inhibiting continuations that result in inflections. Here, we utilize a simplified version of these $V2$ grouping mechanisms by using only excitatory relations for collinear contour configurations which speeds up processing.

Baumann et al. (1997) demonstrated that neurons in area $V2$ of the macaque are selective to border ownership. A significant amount of $V2$ cells exhibited activation patterns signaling the figure-ground direction of surface contours independent of contrast polarity. Furthermore, subsequent studies revealed that illusory contours defined either by disparity (Heider et al. 2002) or occlusions cues (Heider et al. 2000; Zhou et al. 2000) evoke similar $V2$ cell response patterns. Our model dipole cells replicate the $V2$ response properties observed in the physiological studies. In particular, multiplication of model $V2$ bipole activity and model dipole activity is used to isolate the responses of those dipole cells lying at physical or illusory surface contours (see Sect. 3.4). The resulting activity pattern signals the depth order of contours independent of contour type (i.e., illusory or physical) or contrast polarity.

The short latencies of border ownership-related response differences ($<$25 ms) after response onset observed in the study of Zhou et al. (2000) led the authors to propose that figure-ground relations are computed within the visual cortex, either by horizontal activity propagation within $V2$ or by feedback from $V4$, rather than projected down from higher levels that were concerned with object recognition. In our model, $V2$ bipole activity is used to adaptively shape the dipole receptive fields according to the orientation of the surface contours, allowing a model cell to selectively pool contour-related dipole activity in an extended topographical region. This speeds up the propagation of local T-junction information, in turn helping to reduce the number of iterations necessary to reach a globally consistent interpretation of the depth ordering present in the visual scene.

In the electrophysiological studies discussed above, two overlapping surface patches or one isolated patch presented on a uniform background were used as stimuli. Based on our modeling investigations, we suggest that it might be useful to study cell responses for three or more overlapping surfaces in future investigations. In particular, in our model, dipole cells can signal more than two depth positions, depending on the number of model dipole layers. Consequently, it would be interesting to investigate whether cortical $V2$ cells exist which selectively signal intermediate depth positions. In case such cells exist a straight-forward question arising from our model would be if they can distinguish between several intermediate depth positions. Exemplarily, we demonstrated the capabilities of our model to sort surface contours into 6 depth layers (Fig. 10). The number of surfaces that can be automatically, or pre-attentively, distinguished by the visual system is most likely limited to a lower number by time and resolution constraints. Due to our recursive scheme of depth sorting, the depth of surface contours is successively determined starting with the contours being in the fore- and background. Further electrophysiological studies could, therefore, address the question of the dependence of the latencies of border ownership-related response differences on the depth position. However, the time that is necessary to determine the exact depth of surfaces at medial depth positions increases with the overall number of surfaces in the visual scene. The amount of time available for pre-attentive processing is highly restricted so that the maximal number of surfaces that can be sorted by such a mechanism is likely to be limited to a relatively low number. Also, the spatial resolution at which contours and T-junctions are represented decreases for peripheral positions in the visual field, so that the sizes of the objects in terms of visual field angle might affect the maximal number of surfaces that can be automatically processed. When the visual information on more peripheral objects is too coarse and blurred to determine their *local* depth relations (as indicated by the T-junctions), these relations are, in turn, not available in a scheme of *global* depth sorting. In this case, depth sorting will depend on eye movements and attentional shifts. Tyler and Kontsevich (1995) presented psychophysical evidence for a 3D-attentional mechanism that focuses visual processing onto a certain depth range to gain noise robustness in case of ambiguous or complex scenes. According to their suggestion, in this case pre-attentive processing is restricted to the depth relations between the attended object and its neighbors rather than the global relations between all objects.

5.2 Depth layers and dipole dynamics for activity propagation

In our model, several layers of dipole cells were used to distinguish between distinct positions in depth. Within each depth layer, dipole dynamics is used to successively propagate local T-junction information along contours. The assumption of several depth layers has been motivated by psychophysical and physiological findings in stereopsis which indicate that cell pools of different disparity profiles exist in the visual cortex (Poggio et al. 1988; Regan et al. 1986). Furthermore, a recent psychophysical study investigated the period of time after which depth information originating from unambiguous depth cues at specific topographical positions is available at other positions (Nishina et al. 2003). In the study of Nishina et al. the length of the period increased monotonically with increasing distance between the positions, indicating that depth information is propagated over an object using a time-consuming process.

The integration of dipole dynamics into our model allows for an isolated propagation of information in layers of model cells sensitive to the same position in depth. The concept of dipole fields was introduced by Grossberg (1980, 1991) to be a basic building block of processing in the nervous system. In his theoretical framework, feedforward dipole fields act as major tool to reset an error and to search for a correct code. In particular, the circuit proposes that mutual inhibition between the ON- and OFF-dipole channels creates a balance between mutually exclusive categories or features (e.g., perpendicular motion directions). In cases of abruptly changed input, stimulation of an antagonistic rebound of dipole cell activation is elicited which, in turn, enables the overall system to quickly reset itself (Francis et al. 1994). In our model, the different depths act as mutually exclusive features that were coded by the dipole cells. Due to the recursive scheme of depth sorting, the exact depth order of intermediate surface patches is ambiguous during the first iterations of the model after stimulus onset and is resolved later on (see Sects. 3.1 and 4.2). The use of dipole fields enables us to specifically exclude intermediate surface patches from the outer depth layers without influencing activity in the inner depth layers. This effect reliably maintains the depth ambiguity by preventing a specific intermediate depth layer from unintentionally obtaining a biasing advantage due to cross-talk between the depth layers and results in an overall robust model behavior. In case of “impossible figures” based on mutually overlapping surfaces, the depth positions of the contours are ambiguous per se. The model would, therefore, transfer the contour representations successively to more medial depth layers until they are commonly represented in the middle layer. This is consistent with the interpretation that none of the contours is completely in front of or behind all other contours.

5.3 T-junctions as depth cues

The impact of local occlusion cues such as T- or X-junctions on the global interpretation of the visual scene by the human visual system was demonstrated by several psychophysical studies (Boselie 1994; Howard 2003; Nakayama et al. 1990; Nakayama et al. 1989; Pianta and Gillam 2003; Rubin 2001b; Shimojo and Nakayama 1990; Shipley and Kellman 1990). It was demonstrated that monocular depth cues arising, e.g., from occlusion can be equally effective in generating quantitative depth as disparity information (Howard 2003; Pianta and Gillam 2003). Furthermore, it was shown that elimination or manipulation of T-junctions notably affected depth perception, despite the presence of other, more global cues in the image (Rubin 2001b). Shimojo and Nakayama tested depth perception of partially occluded and only monocularly visible image regions (Shimojo and Nakayama 1990). Based on their results, they propose that occlusion-related constraints must be embodied at early levels of visual processing, as such perception necessitates eye-of-origin information which is lost relatively early in the hierarchy of cortical visual processing. Although psychophysical studies clearly demonstrate the importance of 2D-junctions in depth perception, it is unclear if neurons specialized to such image features exist in the visual cortex. In our model, the orientation and position of T-junctions is processed by combining $V2$ long-range activity and end-stop activity. The existence of $V1$ and $V2$ neurons reacting to end-stop configurations has been confirmed by several electrophysiological studies (Heider et al. 2000; Hubel and Wiesel 1965; Peterhans 1997). Filtering of model $V1$ complex cell activity is used in our modeling approach to achieve activity patterns resembling those of cortical single-stopped cells. This activity is combined with the output of model $V2$ long-range grouping mechanisms which have functional properties resembling those seen in the studies of Peterhans and colleagues (Baumann et al. 1997; Heider et al. 2000). It is not necessary to assume that combination of both activities necessitates the existence of specialized cells in order to achieve T-junction specificity. The same result can be obtained when, e.g., $V2$ long-range activity gates the input of end-stop cells to the model $V2$ bipole cells (Spratling and Johnson 2001). This is clearly distinct from specialized filters working directly on the luminance distribution of the input image. Furthermore, our approach is capable of creating ‘illusory’ T-junctions at the corners of the pacman inducers which leads to the percept of a Kanizsa square (Sect. 4.3).

5.4 Other models of depth processing

Most computational approaches determining the depth of objects from occlusion cues roughly divide into two classes, depending on the kind of representation used for object primitives such as contours, corners and junctions: *Graph representations* and *contour representations*. The first class, based on graph representations, uses the vertices of the graph to encode contour intersections such as contour corners, T-or X-junctions and the corners of the graph to represent contour elements which connect the contour intersections (Liu and Wang 1999; Singh and Huang 2003; Williams and Hanson 1996; Williamson 1996). Based on the graph representations of contours and occlusion cues, algorithms were proposed that enable figure-ground segmentation of the objects in the input image. Some of the approaches use hand-labeling or semi-automatic strategies to obtain the graph representation from the stimulus (Liu and Wang 1999; Singh and Huang 2003; Williams and Hanson 1996). In contrast, in the model of Williamson (1996), competitive interactions between cells at the highest model level result in a dynamically allocated (neural) graph representation which is obtained from the activation patterns of the cells at lower model levels. Based on T-junction information, subsequent cooperative and competitive interactions of the cells at the highest model level achieve a depth ordering of the objects in the input image.

Graphs are rather abstract representations of the objects in the input image. As a consequence, all approaches discussed above argue in favor of depth segmentation as a high-level visual process. However, as discussed in Sect. 5.1, electrophysiological studies of visual areas $V2$ and $V4$ indicate that neurons in these areas already signal relative disparity as well as border ownership, which is in contrast to approaches modeling depth perception to be an isolated high-level process (e.g., Baumann et al. 1997; Heider et al. 2000).

Other approaches extend previously developed models of early contour processing to achieve a depth sorting of the objects in the visual scene. The functionality of cells signaling the position and orientation of contours in the input image is complemented so that their response profile is additionally selective to border ownership or relative depth. In most approaches, additional areas are introduced which contain topographical maps of cells signaling the position of line endings or T-junctions. The output of these model areas is used to guide the process of boundary formation in the contour processing stages, resulting in a final activation pattern of the overall model which signals the depth relations in the input image. For example, in the model of Peterhans and Heitger (2001) the illusory contours of Kanizsa squares and triangles are detected by $V2$ contour cells. A contour cell is sensitive to direction of contrast and selectively pools the activity of end-stop cells signaling line endings which have appropriate contrast polarity and are arranged perpendicular to the main axis of the contour cell. This results in $V2$ contour cell activity sensitive to the figure-ground direction of the Kanizsa square or triangle. In accordance with our approach, surface borders are processed and completed by $V2$ cells by utilizing elongated receptive fields to mediate long-range groupings. In contrast to our model, no specific T-junction detectors are necessary as the $V2$ contour cells are modeled to be sensitive to the direction of contrast. However, the model of Peterhans and Heitger (2001) can only distinguish between two positions in depth, namely background and foreground. In contrast, our model is capable to signal intermediate positions in depth, depending on the number of depth layers. The computational capabilities of the model of Peterhans and Heitger are limited compared to our approach. Their approach, on the other hand, is more closely linked to electrophysiological data (Heider et al. 2000) compared to our preprocessing stages that consists of recurrent $V1$–$V2$ contour processing followed by the feedforward stage for T-junction and corner-processing.

Li (2005) proposes a model in which border ownership is determined by recurrent processing within a layer of orientation selective cells resembling those in visual area $V2$. Like the model of Peterhans and Heitger (2001), it can only distinguish between figure and ground, but cannot signal intermediate depth positions. For example, Li presents the model response pattern to two overlapping surface patches on a uniform background (Li, 2005: Fig. 4a). The neurons which signal the surface contours succeed in resolving the direction of the figure for each of the two surfaces, but do not signal the depth relation between the two patches.

Sajda and Finkel propose a model in which the contour-based mechanisms for figure-ground segmentation are complemented by surface-based mechanisms (Finkel and Sajda 1992; Sajda and Finkel 1995). This enables their model to determine the depth of even those objects that show no contour intersections with other objects. For example, black dots lying at the surface of a Kanizsa square are grouped together with the (illusory) outline of the square, enabling one to determine their depth to be in the foreground. In their model, the grouping of contours which belong to an object and the representation of depth is signaled by two independent mechanisms. Contour grouping is achieved using a temporal binding value which is unique to all contours belonging to a common object. Depth is encoded by the firing rates of neurons in two model layers representing background and foreground. The existence and functional role of temporal binding via spike train correlations is a matter of an ongoing and highly controversial debate (e.g., Shadlen and Movshon 1999; Engel et al. 2001). For example, in the model of Sajda and Finkel (1995) each new object in the visual scene necessitates the use of an additional, unique temporal binding value. However, it is unclear how many different temporal patterns of spike trains may be coded by cortical neurons without having cross-talk between these patterns. Furthermore, the maximal spatial range of binding via spike train correlations seems to be rather limited (Eckhorn 2000; Frien and Eckhorn 2000). In contrast, in our model, grouping and depth are coded by a common mechanism, namely the firing rate of neurons in several depth layers in which the binding is augmented by the mechanism of long-range integration and the filling-in of relative depth activation.

In Grossberg's FACADE theory the depth of objects is determined by recurrent interaction of $V2$ bipole cells for contour grouping and filling-in domains for surface reconstruction (Grossberg 1993; Kelly and Grossberg 2000). In FACADE, information from relative depth cues at contour intersections as well as stereoscopic and color information is processed in a common model framework, resulting in a depth-sensitive representation of contours and surfaces. The scope of FACADE is beyond our model, which is restricted to the use of relative depth cues such as T-junctions. In FACADE feedback from the filling-in domains for surface reconstruction to the stages of contour grouping is crucial in order to obtain a clear-cut and unambiguous representation of depth. In contrast, our modeling approach demonstrates that the same result can be achieved by purely contour-based mechanisms.

Kumaran et al. (1996) propose a model using the positions of corners, T- and Y-junctions and end-stoppings as sparse data which is fed into a two-dimensional diffusion process to recover the surfaces of the objects in the image. The model is capable of using several diffusion layers and can, e.g., reconstruct the illusory surface of a Kanizsa square in the first layer and the (completed) surfaces of the inducers in the second layer, thereby enabling figure-ground segmentation. This model is an interesting alternative to our approach and the approaches discussed above in that the initial processing and completion of surface contours is substituted by surface-based mechanisms. However, diffusion-based approaches often suffer from rather slow convergence towards the final solution (Rubin 2001a).

Fukushima (2001) proposes an extension of his neocognitron model which can recognize partly occluded patterns. Occluders are detected using differences in brightness and are subsequently represented in a special masking layer. Information in this layer is excluded from further processing and, consequently, the occluders do not influence pattern recognition in the remaining network. This in turn enables the overall model to recognize the occluded patterns. The model extensions proposed by Fukushima allow it to distinguish only between two depth levels (background and foreground) and recognition is restricted to the patterns in the background. Furthermore, figure-ground segmentation based solely on differences in brightness is likely to work merely on a restricted sample of images.

Taken together, our approach fits into the class of models which integrate the mechanisms for figure-ground segmentation and depth sorting in the overall process of early boundary formation observed in low and mid-level visual areas (Grossberg 1993; Peterhans and Heitger 2001; Sajda and Finkel 1995). Unlike previous approaches, we demonstrate that contour-based mechanisms are sufficient to recover the depth order of *several* overlapping surface patches. For fast activity propagation along contours, we propose model cells which have elongated receptive fields to pool the neighboring activity. The cells exhibit dipole dynamics to allow for isolated activity propagation in specific depth layers which, in turn, guarantee robust model behavior.

5.5 Model properties, limitations and future extensions

The aim of this study was to demonstrate that monocular depth processing can directly build upon a representation of surface contours, as created by early visual processing, and that it can be successfully achieved within a neural architecture in which cells interact only in a locally restricted neighborhood. In order to limit the complexity of the overall approach, minimalist solutions were used for some computational details of the model which were outside the central scope of the study. For example, the most obvious limitation is the use of T-junctions as depth cues only. A straightforward extension may be the integration of X-junction information as well. Adelson and Anandan (1990) as well as Anderson (1997) proposed rules for the classification of X-junctions that allow one to determine which junctions are consistent with a transparency interpretation. This classification scheme could be integrated into our model by additional mechanisms based on $V1$ simple cell activities which signal local contrast polarity at junctions. The usage of disparity information would necessitate the integration of neural mechanisms for binocular image matching (Banks et al. 2004; Dev 1975; Grossberg and Grunewald 2002; Marr and Poggio 1979; Ohzawa et al. 1997). In particular, in our model hierarchy, binocular matching would be used to create a common representation of surface boundaries independent of their depth. Based on this border representation, relative disparity information (Thomas et al. 2002) might then be used to trigger waves of dipole activity in order to obtain a globally consistent representation of depth.

The assumption of having no cross-talk between model cells sensitive to different positions in depth may be too strict in order to be biologically plausible. However, a more plausible model would necessitate the usage of a higher number of model layers each optimally tuned to a slightly different depth position. We claim that in such a continuum of depth positions the effect of cross-talk between adjacent model layers could be counteracted by mechanisms of ON-center/OFF-surround interaction acting along the depth axis, in turn keeping activity focused to a few adjacent depth layers. Furthermore, simultaneous gradual activation of several adjacent depth layers would allow for the smooth representation of slanted surfaces. Finally, while the concept of dipole cells was introduced to describe certain aspects of biological vision (Grossberg 1991), alternative mechanisms could be used in our model to propagate information along contours when having a high number of depth layers. For example, in the modeling approach of Bayerl and Neumann (2004), inhibition between cell layers tuned to different motion preferences was used to propagate motion information along contours.

Tse (1999) used mutually overlapping 3D volumes to demonstrate cases of amodal completion that neither contour- nor surface-based approaches can fully explain, arguing in favor of a higher-level completion process based on "mergeable" volume fragments. However, the selectivity of $V2$ cells to border ownership and the short latencies of their border ownership-related response differences strongly suggest that depth relations are already partly processed in early visual areas (Heider et al. 2000; Zhou et al. 2000), as also explored in our modeling study. The combined evidence clearly hints towards depth processing being distributed across several stages of the visual system. To our view, additional experimental and modeling work is needed to shed light on the specific contributions of all the involved areas to this process, as well as on their interactions. Our modeling work helps to clarify the putative role of early and mid-level areas by demonstrating the amount of depth processing that can be achieved in a recurrent network of locally interacting cells.

Taken together, in our model, a robust contour representation is established by recurrent $V1$–$V2$ interaction capable of completing fragmented contours based on a relatability measure. The contour representation is taken as basis for further depth processing implemented in a recursive scheme of depth sorting. In this scheme, contours being at an unambiguous position in depth are assigned to the according depth layers and this information is then used to recursively resolve ambiguity of the remaining contours. More specifically, based on local occlusion cues, contours being in the fore- and background are first assigned to the according depth layers, followed by a successive depth sorting of the contours at intermediate depth positions. The recursive scheme allows us to achieve a globally consistent depth sorting of overlapping surfaces using a neural model restricted to local mechanisms of recurrent interactions to propagate local and relative depth cues.

Acknowledgments We thank the two reviewers for their insightful comments that helped to substantially improve the manuscript. HN was supported in part by the European Union (EU FP6 IST Cognitive Systems Integrated project: Neural Decision-Making in Motion; project number 027198).

Appendix A: First model stage of recurrent $V1$–$V2$ contour processing

A.1 Receptive field properties: model equations

In the following, the model equations describing the receptive field properties are depicted. The initial activation determined by pooling of bottom–up activity by the cell's receptive field is denoted by c. The letters l^{V1} and l^{V2} denote the final activation of $V1$ and $V2$ cells, respectively, after top–down modulation and center-surround competition. Capital letters A, B, C, D or E denote constants. Anisotropic Gaussians in the spatial domain are described by $\Lambda_{\sigma x,\sigma y,\tau x,\tau y,\theta}$. Their size (in pixel) is defined by standard deviations σ_x and σ_y in the x- and y-direction, respectively. They are shifted by τ_x and τ_y pixel in the x- and y-direction and finally rotated by angle θ. Spatial locations in the topographical maps are expressed by the index i. Ψ denotes isotropic Gaussians in the orientation domain, $*$ is the convolution operator and $[x]^+ := \max\{x, 0\}$ stands for half-wave rectification.

A.1.1 LGN ON/OFF cells

LGN cell signal local luminance transitions in the input image. Their activity is determined by convolution of the input image with a difference-of-Gaussians operator followed by a half-wave rectification:

$$\begin{aligned} x &= I * (\Lambda_{\text{Center}} - \Lambda_{\text{Surround}}) \\ x^{\text{on}} &= [x]^+ \\ x^{\text{off}} &= [-x]^+ \end{aligned} \tag{A.1}$$

Constants $\sigma_{\text{Center}} = 0.8$ and $\sigma_{\text{Surround}} = 3\sigma_{\text{Center}}$ denote the standard deviations of the center and surround isotropic Gaussian kernels, respectively.

A.1.2 $V1$ simple cells

Simple cells are sensitive to local luminance transitions along a given orientation preference and are selective to contrast. They have elongated subfields which pool the input of appropriately aligned LGN responses. The subfields are modeled as anisotropic Gaussian kernels that are shifted perpendicular to their main axis and then rotated by θ. The activity of the subfields is denoted by p. It is fed into a softAND-circuit to determine the response s of the simple cell. Simple cells exist for two polarities (dark–light, dl; light–dark, ld) and eight orientations. The softAND-circuit that determines, e.g., the activation $s_{i\theta}^{ld}$ of a simple cell sensitive for polarity light–dark is determined by

$$s_{i\theta}^{ld} = \frac{A_{\text{s}}(p_{i\theta}^{\text{on\_left}} + p_{i\theta}^{\text{off\_right}}) + 2B_{\text{s}}\, p_{i\theta}^{\text{on\_left}}\, p_{i\theta}^{\text{off\_right}}}{A_{\text{s}} D_{\text{s}} + E_{\text{s}}(p_{i\theta}^{\text{on\_left}} + p_{i\theta}^{\text{off\_right}})} \tag{A.2}$$

with $\begin{aligned} p_{\theta}^{\text{on\_left}} &= x^{\text{on}} * \Lambda_{\sigma_x,\sigma_y,0,-\tau_y/2,\theta} \\ p_{\theta}^{\text{off\_right}} &= x^{\text{off}} * \Lambda_{\sigma_x,\sigma_y,0,\tau_y/2,\theta} \end{aligned}$ representing the two subfield activities given by the convolution of LGN-activity $x^{\text{on/off}}$ with shifted and rotated anisotropic Gaussian kernels. Their standard deviations are $\sigma_y = 0.8$ and $\sigma_x = 3.0\sigma_y$, and they are shifted by $\tau_y = 0.8\sigma_y$ in y-direction. Their orientation θ is given by $\theta = 0, \pi/N\_\text{orient}, \ldots, (N\_\text{orient} - 1)\pi/N\_\text{orient}$ (with $N\_\text{orient} = 8$). Constants $A_{\text{s}} = 1.0$ and $B_{\text{s}} = 10000.0$ determine the relative strength of the additive and multiplicative component, respectively, of the softAND-circuit. The amount of self-normalization is controlled by the ratio between constants $D_{\text{s}} = 0.05$ and $E_{\text{s}} = 100.0$.

A.1.3 $V1$ complex cells

Complex cells are sensitive to the orientation of luminance contrast, but insensitive to its direction. The initial $V1$ complex cell response $c_{i\theta}$ is determined by the sum of half-wave rectified differences of the two simple cell activities of opposite polarity (dark–light and light–dark) at each position:

$$c_{i\theta}^{V1} = A_{\text{c}}([s_{i\theta}^{ld} - s_{i\theta}^{dl}]^+ + [s_{i\theta}^{dl} - s_{i\theta}^{ld}]^+) \tag{A.3}$$

with $A_{\text{c}} = 0.1$.

A.1.4 $V2$ bipole cells

Model bipole cells consist of two prolated subfields which are shifted parallel to the main axis of the cell. They pool the activity l^{V1} of $V1$ complex cells after top–down modulation and center-surround competition, which are approximately sensitive to the same orientation as the bipole cell. The activities k of the subfields are combined using a softAND-gate to give the initial activation $c_{i\theta}$ of a $V2$ bipole cell:

$$c_{i\theta}^{V2} = \frac{A_{\text{t}}(k_{i\theta}^{\text{left}} + k_{i\theta}^{\text{right}}) + 2B_{\text{t}}k_{i\theta}^{\text{left}}k_{i\theta}^{\text{right}}}{A_{\text{t}}D_{\text{t}} + E_{\text{t}}(k_{i\theta}^{\text{left}} + k_{i\theta}^{\text{right}})} \tag{A.4}$$

Please refer to Eq. (A.2) for an explanation of constants $A_{\text{t}} = 1.0$, $B_{\text{t}} = 50{,}000.0$, $D_{\text{t}} = 0.15$ and $E_{\text{t}} = 100.0$. The subkernel activities

$$\begin{aligned} k^{\text{left}} &= l^{V1} * \Psi_f * K^{\text{left}} \\ k^{\text{right}} &= l^{V1} * \Psi_f * K^{\text{right}} \end{aligned} \tag{A.5}$$

are determined by the convolution of $V1$ cell activity l^{V1} with an isotropic Gaussian Ψ_f in the orientation domain (standard deviation $\sigma_{f\_\text{orient}} = 0.25$) and with spatial kernels $K^{\text{left/right}}$. The spatial kernels are modeled as anisotropic Gaussians, which are cut off in the central part of the cell by means of a sigmoid function:

$$\begin{aligned} K_{i\theta}^{\text{left\_unnorm}} &= \Lambda_{\sigma_{k\_x},\sigma_{k\_y},\tau_{k\_x},0,\theta}(\vec{x}_i) \\ &\quad \bullet \frac{1}{1+\exp\left(-A_{\text{k}}\vec{x}_i \bullet \begin{pmatrix}\cos\theta \\ \sin\theta\end{pmatrix} - B_{\text{k}}\right)} \\ K_{i\theta}^{\text{right\_unnorm}} &= \Lambda_{\sigma_{k\_x},\sigma_{k\_y},-\tau_{k\_x},0,\theta}(\vec{x}_i) \\ &\quad \bullet \frac{1}{1+\exp\left(+A_{\text{k}}\vec{x}_i \bullet \begin{pmatrix}\cos\theta \\ \sin\theta\end{pmatrix} + B_{\text{k}}\right)} \end{aligned} \tag{A.6}$$

Vector $\vec{x}_i$ denotes the Cartesian coordinates of the spatial position with index i, and symbol "$\bullet$" is used to express the dot vector product. Constants $A_{\text{k}} = 0.5$ and $B_{\text{k}} = 2.0$ determine the slope and threshold of the sigmoidal weighting function. The standard deviations of the anisotropic Gaussians are given by $\sigma_{k\_x} = 18.0$ and $\sigma_{k\_y} = 1.25$, and $\tau_{k\_x} = 16.0$ determines their spatial offsets in x-direction. The partial overlap of the subfields in the center of the cell defines the classical receptive field. The resulting kernels are normalized at each spatial position and the areas under the kernels are subsequently scaled to 1:

$$\begin{aligned} K_{i\theta}^{\text{left}} &= \frac{K_{i\theta}^{\text{left\_unnorm}}}{\alpha_K + K_{i\theta}^{\text{left\_unnorm}}} \cdot \left(\sum_j \frac{K_{j\theta}^{\text{left\_unnorm}}}{\alpha_K + K_{j\theta}^{\text{left\_unnorm}}}\right)^{-1} \\ K_{i\theta}^{\text{right}} &= \ldots \end{aligned} \tag{A.7}$$

with $\alpha_K = 0.0004$.

A.1.5 $V2$ long-range interaction

The $V2$ long-range activity l^{V2_LR} is based on pooling the final bipole cell activity l^{V2} by two prolated subfields that have identical shapes as the kernels previously used to integrate the $V1$ bottom–up input. The resulting activities $r^{\text{left/right}}$ of the subfields are combined using a softAND-gate to yield the long-range activity l^{V2_LR}:

$$l_{i\theta}^{V2\_LR} = \frac{A_{\text{t}}(r_{i\theta}^{\text{left}} + r_{i\theta}^{\text{right}}) + 2B_{\text{t}}r_{i\theta}^{\text{left}}r_{i\theta}^{\text{right}}}{A_{\text{t}}D_{\text{t}} + E_{\text{t}}(r_{i\theta}^{\text{left}} + r_{i\theta}^{\text{right}})} \tag{A.8}$$

$$\text{with} \quad \begin{aligned} r^{\text{left}} &= l^{V2} * \Psi_f * K^{\text{left}} \\ r^{\text{right}} &= l^{V2} * \Psi_f * K^{\text{right}} \end{aligned}$$

The kernel specifications of Ψ_f and $K^{\text{left/right}}$ as well as the constants of the softAND-gate are identical to those used in Eqs. (A.4)–(A.7). Long-range activity l^{V2_LR} enhances the initial $V2$ cell activation in a multiplicative way (see next section) and helps to stabilize the initial $V2$ cell responses to contour elements.

A.2 Model dynamics

After the integration of bottom–up activity via their receptive fields, model $V1$ complex and $V2$ bipole cells determine their final activation level using two successive computational steps (Fig. 3a; Thielscher and Neumann 2003).

- First, the initial activation is modulated via shunting feedback from model area $V2$ (in case of $V1$ complex cells) or shunting intra-areal interaction from neighboring $V2$ cells. The dynamics of this step is denoted by the equation

$$\frac{\partial}{\partial t} m_{i\theta} = -\alpha_m m_{i\theta} + (\beta_m - \gamma_m m_{i\theta}) c_{i\theta} \left[1 + C h_{i\theta}\right] \tag{A.9}$$

$$\text{with} \quad \begin{aligned} h_{i\theta} &= l_{i\theta}^{V2} (V1 \text{ complex cells}) \\ h_{i\theta} &= l_{i\theta}^{V2\_LR} (V2 \text{ bipole cells}) \end{aligned}$$

The input activation $c_{i\theta}$ depends on the bottom–up input weighted by the receptive field kernel (first processing stage as described in Sect. A.1). Thus, $c_{i\theta}$ is a specific function of the cortical area the cell belongs to and is sensitive to the spatial location i and to orientation θ. Top–down activity $h_{i\theta}$ is delivered by descending cortical pathways ($V1$ complex cells) or horizontal long-range connections ($V2$ bipole cells), multiplicatively enhancing the initial activation. Feedback is sensitive to spatial location and orientation. The constants α_m, β_m and γ_m define the parameters of cell interaction dynamics (see Table A.1). α_m determines the activity decay, β_m and γ_m denote the saturation level of the modulated activity $m_{i\theta}$.

Table A.1 Constants of the two-stage cell dynamics and standard deviations σ of the Gaussians pooling the center and surround activity in the spatial and orientational domains (see Eq. A.9, A.10)

	Parameters for m				Parameters for l							
	α_m	β_m	γ_m	C	α_l	β_l	δ_l	ζ_l	σ of			
									ψ^+	Λ^+	ψ^-	Λ^-
$V1$	12.0	0.73	3.7	10.0	1.0	11.2	20.0	500.0	0.3	1.0	0.8	3.0
$V2$	12.0	0.34	5.9	0.088	1.0	6.0	5.6	300.0	0.5	1.6	0.8	6.0

The constant C represents the gain factor of top–down modulation.

- Second, the top–down or intra-areal modulated activity undergoes a stage of shunting ON-center/OFF-surround competition between neighboring cells resembling a "Mexican Hat" shape for a spatial and orientational information. This is expressed by

$$\frac{\partial}{\partial t} l_{i\theta} = -\alpha_l l_{i\theta} + \beta_l \left\{ m * \Psi^+ * \Lambda^+ \right\}_{i\theta} - (\delta_l + \zeta_l l_{i\theta}) \left\{ m * \Psi^- * \Lambda^- \right\}_{i\theta} \qquad \text{(A.10)}$$

The excitatory and inhibitory weighting function, both in space and orientation, are denoted by Λ^+, ψ^+, Λ^- and ψ^-, respectively; $\alpha_l, \beta_l, \delta_l, \zeta_l$ are scalar constants (see Table A.1). Subtractive inhibition of cell activity by the OFF-surround is incorporated using the term $\delta_l \left\{ m * \Psi^- * \Lambda^- \right\}_{i\theta}$. Additionally, divisive inhibition by the term $\zeta_l l_{i\theta} \left\{ m * \Psi^- * \Lambda^- \right\}_{i\theta}$ enables shunting cell interaction. With this scheme of competitive interaction, the initial top–down modulated activities are enhanced by contrast and normalized. The normalization achieves an activity dependent tuning of the cells' responsiveness.

Appendix B: Feed-forward scheme of the second model stage for the detection of corners and T-junctions

B.1 End-stop cells: model equations

A cell's end-stop behavior is modeled using a cascade of feed-forward processing steps, as described by the following equations. Initially, the $V1$ complex cell activity is normalized in the orientation domain at each spatial position.

$$l_{i\theta}^{V1\_\text{Norm}} = \frac{l_{i\theta}^{V1}}{\alpha + \sum_{k=1}^{N\_\text{orient}} l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V1}} \qquad \text{(B.1)}$$

Constant $\alpha = 0.01$ determines the steepness of the saturation curve. Subsequently, activity $l_{i\theta}^{V1\_\text{Norm}}$ is filtered by the excitatory and inhibitory subfields of the end-stop cell (Fig. 3c). The activity that is pooled by the *excitatory* subfield is denoted by

$$\begin{aligned} \text{Act}_{i\theta}^{ex} &= \left\{ l^{V1\_\text{Norm}} * \Psi_{es} * \Lambda^{ex} \right\}_{i\theta} \\ &\quad \text{for } 0 \leqslant \theta < \pi \\ \text{and } \text{Act}_{i\theta}^{ex} &= \left\{ l^{V1\_\text{Norm}} * \Psi_{es} * \Lambda^{ex} \right\}_{i(\theta-\pi)} \\ &\quad \text{for } \pi \leqslant \theta < 2\pi \end{aligned} \qquad \text{(B.2)}$$

End-stop cells use an extended range of orientations given by $\theta = 0, \pi/N\_\text{orient}, \ldots, (2 \cdot N\_\text{orient} - 1) \cdot \pi/N\_\text{orient}$ (with $N\_\text{orient} = 8$). Usage of the two sectors $[0, \pi)$ and $[\pi, 2\pi)$ stems from the fact that $V1$ complex cells signal the orientation of a contour element in the range from 0 to π, but the direction of a contour ending has the full range of $[0, 2\pi)$. The subfield is modelled by a combination of an isotropic Gaussian Ψ_{es} (with $\sigma_{es\_\text{orient}} = 0.35$) in the orientation domain and an anisotropic spatial Gaussian kernel Λ^{ex} (with $\sigma_x = 8.0$ and $\sigma_y = 1.5$) that is shifted by $\tau_x = -3.0$ in x-direction and rotated by θ.

The *inhibitory* subfield consists of three anisotropic and shifted Gaussians, each pooling activity from slightly different, but neighboring input orientations:

$$\begin{aligned} 0 \leqslant \theta < \pi : \text{Act}_{i\theta}^{\text{inh}} &= C_{\text{cen}} \left\{ l^{V1\_\text{Norm}} * \Psi_{es} * \Lambda^{\text{inh\_cen}} \right\}_{i\theta} \\ &+ C_{\text{lat}} \left\{ l^{V1\_\text{Norm}} * \Psi_{es} * \Lambda^{\text{inh\_lat1}} \right\}_{i(\theta+1)} \\ &+ C_{\text{lat}} \left\{ l^{V1\_\text{Norm}} * \Psi_{es} * \Lambda^{\text{inh\_lat2}} \right\}_{i(\theta-1)} \\ \pi \leqslant \theta < 2\pi : &\ldots \end{aligned} \qquad \text{(B.3)}$$

The activities pooled by the central and the two lateral kernels is weighted by constants $C_{\text{cen}} = 2.0$ and $C_{\text{lat}} = 1.2$, respectively, and subsequently summed up to yield the output $\text{Act}_{i\theta}^{\text{inh}}$ of the inhibitory subfield. The standard deviations and spatial offsets are $\sigma_x = 6.0$, $\sigma_y = 3.0$ and $\tau_x = 10.0$ for the central kernel $\Lambda^{\text{inh\_cen}}$, as well as $\sigma_x = 4.0$, $\sigma_y = 4.0$, $\tau_x = 6.0$ and $\tau_y = \pm 4.0$ for the two lateral kernels $\Lambda^{\text{inh\_lat1/inh\_lat2}}$.

The half-wave rectified difference between the excitatory and inhibitory subfield activities is weighted with l^{V1_Norm} to restrict the activations to spatial positions corresponding to contours. Subsequently, the final end-stop cell activity is determined by collapsing the full orientation range of $[0, 2\pi)$

to the range $[0, \pi)$ by pooling the activities corresponding to orientations θ and $(\theta + \pi)$:

$$\begin{aligned} es^{\mathrm{left}}_{i\theta} &= \left[\mathrm{Act}^{ex}_{i\theta} - \mathrm{Act}^{\mathrm{inh}}_{i\theta}\right]^{+} \cdot l^{V1\_\mathrm{Norm}}_{i\theta} \\ es^{\mathrm{right}}_{i\theta} &= \left[\mathrm{Act}^{ex}_{i(\theta+\pi)} - \mathrm{Act}^{\mathrm{inh}}_{i(\theta+\pi)}\right]^{+} \cdot l^{V1\_\mathrm{Norm}}_{i\theta} \\ es_{i\theta} &= es^{\mathrm{left}}_{i\theta} + es^{\mathrm{right}}_{i\theta} \end{aligned} \tag{B.4}$$

with $\theta = 0, \pi/N\_\mathrm{orient}, \ldots, (N\_\mathrm{orient} - 1)\pi/N\_\mathrm{orient}$.

B.2 Detection of corners and T-junctions: model equations

The sequence of feed-forward processing steps employed to determine the T-junction and corner activities is sketched in Fig. 3b. Initially, end-stop activities corresponding to roughly perpendicular orientations θ and $\theta + \pi/2$ are multiplied and subsequently summed over all orientations to signal the likely positions of contour corners in the input image:

$$\mathrm{Act}^{\mathrm{corner}}_{i} = \sum_{k=1}^{N\_\mathrm{orient}/2} es^{sm}_{i(k-1)\cdot\pi/N\_\mathrm{orient}} \cdot es^{sm}_{i(k-1)\cdot\pi/N\_\mathrm{orient}+\pi/2} \tag{B.5}$$

Activity $es^{sm}_{i\theta} = \{es * \Psi_{\mathrm{corner}} * \Lambda^{\mathrm{corner}}\}_{i\theta}$ represents a smoothed version of the original end-stop activity given by Eq. (B.4) that is blurred in the spatial and orientation domains using isotropic Gaussian kernels $\Lambda^{\mathrm{corner}}(\sigma_{\mathrm{corner}} = 1.5)$ and $\Psi_{\mathrm{corner}}(\sigma_{\mathrm{corner\_orient}} = 0.1)$, respectively. Candidate positions of T-junctions are determined by multiplying the end-stop activities at orientation θ with the $V2$ long-range activity at the perpendicular orientation $\theta + \pi/2$ and subsequently summing up across all orientations:

$$\mathrm{Act}^{T}_{i} = \sum_{k=1}^{N\_\mathrm{orient}} es^{sm}_{i(k-1)\cdot\pi/N\_\mathrm{orient}} \cdot l^{V2\_sm}_{i(k-1)\cdot\pi/N\_\mathrm{orient}+\pi/2} \tag{B.6}$$

Activity $l^{V2\_sm}_{i\theta} = \left\{l^{V2\_LR} * \Psi_{T} * \Lambda^{T}\right\}_{i\theta}$ represents a smoothed version of $V2$ long-range activity, blurred by isotropic Gaussians Λ^{T} ($\sigma_{\mathrm{T}} = 0.1$) and $\Psi_{T}(\sigma_{\mathrm{T\_orient}} = 0.3)$ in the spatial and orientation domain, respectively.

The initial activities $\mathrm{Act}^{\mathrm{corner}}$ and Act^{T} undergo a stage of mutual subtractive inhibition. The final corner activity Cor_i is determined by subtracting Act_T from $\mathrm{Act}_{\mathrm{corner}}$. The resulting difference map is half-wave rectified, spatially blurred and subsequently normalized:

$$\mathrm{Cor}_i = \frac{\left[\left\{(\mathrm{Act}^{\mathrm{corner}} - C_T \mathrm{Act}^{T}) * \Lambda^{sm}\right\}_i\right]^{+}}{\alpha_{\mathrm{Cor}} + \left[\left\{(\mathrm{Act}^{\mathrm{corner}} - C_T \mathrm{Act}^{T}) * \Lambda^{sm}\right\}_i\right]^{+}} \tag{B.7}$$

Constant $C_T = 0.1$ is used to roughly match the activation levels of $\mathrm{Act}_{\mathrm{corner}}$ and Act_T. The spatial isotropic Gaussian Λ^{sm} ($\sigma = 6.0$) is used for blurring, and constant $\alpha_{\mathrm{Cor}} = 0.0035$ determines how quickly Cor_i saturates. T-junction activity $T^{\mathrm{unorient}}_{i}$ is determined by subtracting $\mathrm{Act}_{\mathrm{Corner}}$ from Act_T, followed by a half-wave rectification, spatial blurring and normalization:

$$T^{\mathrm{unorient}}_{i} = \frac{\left[\left\{(\mathrm{Act}^{T} - C_{\mathrm{Cor}} \mathrm{Act}^{\mathrm{corner}}) * \Lambda^{sm}\right\}_i\right]^{+}}{\alpha_T + \left[\left\{(\mathrm{Act}^{T} - C_{\mathrm{Cor}} \mathrm{Act}^{\mathrm{corner}}) * \Lambda^{sm}\right\}_i\right]^{+}} \tag{B.8}$$

Constant $C_{\mathrm{cor}} = 8.0$ is used to match the activation strength of Act_T with that of $\mathrm{Act}_{\mathrm{corner}}$, and constant $\alpha_T = 0.03$ determines the steepness of the saturation curve.

Activities Cor_i and $T^{\mathrm{unorient}}_{i}$ signal the spatial positions of corners and T-junctions, respectively, but do not contain information about the orientations of the underlying contours. In order to enable the T-junction activities to exert selective impact on the contours corresponding either to the hats or to the stems of Ts (depending on the depth layer), orientations are assigned to T-junctions by weighting $T^{\mathrm{unorient}}_{i}$ with end-stop activities es^{sm}:

$$T_{i\theta} = \frac{T^{\mathrm{unorient}}_{i} \cdot es^{sm}_{i\theta}}{\alpha_{T2} + T^{\mathrm{unorient}}_{i} \cdot \sum_{k=1}^{N\_\mathrm{orient}} es^{sm}_{i(k-1)\cdot\pi/N\_\mathrm{orient}}} \tag{B.9}$$

The relative impact of a T-junction on a certain orientation channel is determined by the strength of the end-stop activity in that channel, normalized to the overall activation strength across all channels (constant $\alpha_{T2} = 0.09$ determines the steepness of the saturation curve). The end-stop activities that signal contour corners are isolated by weighting es^{left} and es^{right} (Eq. (B.4)) with corner activity Cor_i:

$$\begin{aligned} es^{\mathrm{Corner\_left}}_{i\theta} &= es^{\mathrm{left}}_{i\theta} \cdot \mathrm{Cor}_i \\ es^{\mathrm{Corner\_right}}_{i\theta} &= es^{\mathrm{right}}_{i\theta} \cdot \mathrm{Cor}_i \end{aligned} \tag{B.10}$$

Appendix C: Third model stage of recurrent depth processing: supplements

C.1 Determining the normalized $V2$ bipole activity l^{V2_Norm}

The $V2$ bipole activity l^{V2} undergoes a stage of subtractive inhibition in the orientation domain in order to suppress spurious activations. Subsequently, it is normalized at each spatial position i:

$$l^{V2\_\mathrm{Norm}}_{i\theta} = \frac{C_{Eq}\left[l^{V2}_{i\theta} - \beta_{Eq} \sum_{k=1}^{N\_\mathrm{orient}} l^{V2}_{i(k-1)\cdot\pi/N\_\mathrm{orient}}\right]^{+}}{\alpha_{Eq} + \left[l^{V2}_{i\theta} - \beta_{Eq} \sum_{k=1}^{N\_\mathrm{orient}} l^{V2}_{i(k-1)\cdot\pi/N\_\mathrm{orient}}\right]^{+}} \tag{C.1}$$

Table C.1 Constants of the equations delineating the model of recurrent depth sorting in Sect. 3

Equation	Parameter	Value
3.1	C_{thres}	0.022
3.2	C_{base}	0.28
	$C_{\max}$	see Eq. (C.4)
	α	0.00001
	β	0.02
3.3	C_{gain}	100,000
3.5	SD of Gaussian Ψ_f: σ_{f_orient} =	0.25
3.8	C_{ol}	7.2

Constant $\beta_{Eq} = 0.1$ determines the impact of the suppressive activity summed across all orientations, constant $\alpha_{Eq} = 0.015$ controls the steepness of the saturation curve, and $C_{Eq} = 1.6$ determines the maximal activation strength of l^{V2_Norm}. $N\_\text{orient} = 8$ depicts the number of orientations used in the model, that are given by $\theta = 0, \pi/N\_\text{orient}, \ldots, (N\_\text{orient} - 1)\pi/N\_\text{orient}$.

C.2 Spatial layout of kernels V

Kernel V is modelled using an anisotropic Gaussian kernel which undergoes divisive self-normalization to yield steeper flanks at the boundaries of the kernel. The area under the kernel is subsequently scaled to 1:

$$V_{i\theta} = \frac{\Lambda_{\sigma\_x,\sigma\_y,0,0,\theta}(i)}{\alpha_K + \Lambda_{\sigma\_x,\sigma\_y,0,0,\theta}(i)} \cdot \left(\sum_j \frac{\Lambda_{\sigma\_x,\sigma\_y,0,0,\theta}(j)}{\alpha_K + \Lambda_{\sigma\_x,\sigma\_y,0,0,\theta}(j)} \right)^{-1} \tag{C.2}$$

The standard deviations of the anisotropic Gaussian Λ are given by $\sigma_x = 13.0$ and $\sigma_y = 0.7$. The steepness of the flanks is controlled by constant $\alpha_K = 0.004$. The kernel is cut into the two halves V^{left} and V^{right}. The two halves enable the dipoles to selectively integrate those dipole activities at contour corners which correspond to end-stop cells pointing in the direction of the contour (see Fig. 6b; details are depicted below):

$$\begin{aligned} V^{\text{left}}_{i\theta} &= \begin{cases} V_{i\theta}\ \forall \vec{x}_i \bullet \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} < 0 \\ 0 \forall \vec{x}_i \bullet \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \geqslant 0 \end{cases} \\ V^{\text{right}}_{i\theta} &= \begin{cases} V_{i\theta}\ \forall \vec{x}_i \bullet \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \geqslant 0 \\ 0 \forall \vec{x}_i \bullet \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} < 0 \end{cases} \end{aligned} \tag{C.3}$$

Vector $\vec{x}_i$ denotes the Cartesian coordinates of spatial position i, and symbol "$\bullet$" is used to express the dot vector product.

C.3 Integration of dipole activity at contour corners

As outlined in Sect. 3.2, the recurrent interactions between dipoles result in waves of activity propagating along contours. However, at contour corners, the dipoles cannot pool activity from the dipoles corresponding to the part of the contour "around the corner", which might disrupt the waves. This is prevented by three complementary mechanisms. First, corner activity Cor_i (see Eq. B.7) enhances the maximal value $C_{\max}$ of the internal activity b_t in Eq. (3.2) at spatial positions corresponding to corners:

$$C_{\max}(i) = 1.0 + C_{\text{corner\_amp}} \cdot \text{Cor}_i \tag{C.4}$$

Index i depicts the spatial location, and constant $C_{\text{corner\_amp}} = 2.4$ determines the maximal impact of Cor_i on $C_{\max}$. Second, the dipole activities $e_t^{\text{ON/OFF}}$ in Eq. (3.5) are enhanced by end-stop activities at contour corners. Third, the spatial kernels V use two sub-kernels $V^{\text{left/right}}$ to selectively pool those dipole activities which correspond to end-stop cells pointing in the direction of the contour (see Fig. 6b). The modified version of Eq. (3.5) then reads:

$$\begin{aligned} f^{\text{ON/OFF}}_{ti\theta} = {} & \left\{ \left[(1 + C_{es} es^{\text{Corner\_left}}) \cdot e_t^{\text{ON/OFF}} \right] \right. \\ & \left. * \Psi_f * V^{\text{left}} \right\}_{i\theta} + \left\{ \left[(1 + C_{es} es^{\text{Corner\_right}}) \right. \right. \\ & \left. \left. \cdot e_t^{\text{ON/OFF}} \right] * \Psi_f * V^{\text{right}} \right\}_{i\theta} \end{aligned} \tag{C.5}$$

The terms $es^{\text{Corner\_left}}$ and $es^{\text{Corner\_right}}$ denote the activities of end-stop cells at contour corners (Eq. B.10). The impact of these activities on dipole output strength is controlled by the constant $C_{es} = 10.0$. The terms $(1 + C_{es} es^{\text{Corner\_left/Corner\_right}})$ selectively enhance the dipole output activities $e_t^{\text{ON/OFF}}$ at contour corners. Kernel V^{left} selectively pools dipole activity e_t modulated by $es^{\text{Corner\_left}}$, while V^{right} selectively pools dipole activity e_t modulated by $es^{\text{Corner\_right}}$. This restricts the impact of $es^{\text{Corner\_left}}$ and $es^{\text{Corner\_right}}$ to the contour to which the actual corner belongs to and prevents a wave of activity to jump over to neighboring contours (Fig. 6b).

Appendix D: Third model stage of recurrent depth processing: impact of the model T-junction detectors on model dipole activity

In Sect. 3.6 it is depicted how activities q_{ti}^{foregr} and q_{ti}^{backgr} trigger waves of OFF-channel activity to reset those dipoles signaling contours not belonging to the actual depth layer. Here, we will outline (i) how activations $q_{ti}^{\text{foregr/backgr}}$ are derived from the activity distribution of the T-junction detectors and (ii) how the T-junction information corresponding to contours having intermediate depth positions is transferred from the outer to the inner depth layers.

As outlined in Sect. 3.6, activations $q_{ti}^{\text{foregr/backgr}}$ locally increase the OFF-channel input to dipoles corresponding to contours *not* belonging to the actual depth layer and thereby induce the waves of OFF-activity propagating along that contours. They also increase the input to the ON-channels of the remaining dipoles in the neighborhood to stabilize the dipole activities corresponding to the intersecting contours. The increased input to the ON-channels might result in a disruption of a wave of activity triggered by other T-junction detectors, which is prevented by limiting the impact of the T-junction detectors on model dipole activity in time. For the fore- and background layers #1 and #N, this time limit is denoted by the following equation:

$$T_{ti\theta}^{\text{mod}} = T_{i\theta} \cdot \frac{1}{\left(1 + \left[t - \beta_q\right]^+\right)^2} \tag{D.1}$$

with the time constant $\beta_q = 9.0$. $T_{i\theta}$ denotes the activity of the T-junction detectors (Eq. B.9). T^{mod} is maximal until time point β_q and quickly decays afterwards. In an intermediate layer #K, the impact of T-junction activity on dipole input is inhibited by activated ON-channels of dipoles at corresponding positions in the *outer layers* of K:

$$u_{ti} = \left[\frac{1 - \beta_u \cdot \sum\limits_{j \in \text{outerlayers}} d_{ti}^{\text{ON\_layer(j)}}}{\alpha_u + \sum\limits_{j \in \text{outerlayers}} d_{ti}^{\text{ON\_layer(j)}}}\right]^+ \tag{D.2}$$

with decay $\alpha_u = 0.01$ and scaling constant $\beta_u = 33.0$. This inhibition prevents T-junctions that stem from intersections of those contours which are represented in the outer depth layers with other contours to exert influence on the dipole activation patterns in the more mediate depth layers. The time limit of the impact of T-junction information on the dipole activities in the intermediate depth layers is captured in the following equations. First, changes of activity u_{ti} caused by the release of inhibition are detected and signalled by

$$v_{ti} = [u_{ti} - \beta_v u_{(t-1)i}]^+ \tag{D.3}$$

with scaling constant $\beta_v = 1.08$. Subsequently, activity v_{ti} is low-pass filtered in time and the resulting activity w_{ti} is used to modulate the original T-junction detector activity T

$$\begin{aligned} w_{ti} &= ([\beta_w w_{(t-1)i} + C_v v_{ti}]^{\leqslant 1})^2 \\ T_{ti\theta}^{\text{mod}} &= T_{i\theta} \cdot [C_w w_{ti}]^{\leqslant 1} \end{aligned} \tag{D.4}$$

with decay $\beta_w = 0.97$ and scaling constants $C_v = 2.4$ and $C_w = 5.0$. The time-limited T-junction activity T^{mod} signals the position of contour intersections or, more specifically, the position and orientation of those contour elements being locally in the background of other contours. Activity T^{mod} is blurred in the orientation domain (using an isotropic Gaussian Ψ_p) and convolved with anisotropic Gaussians Λ in the spatial domain. For activity p^{foregr}, which locally signals the position of the contour being in the foreground, it is rotated by $\pi/2$ for that purpose. The resulting activity is weighted with the normalized $V2$ bipole activity l^{V2_Norm} and summed over all orientations:

$$\begin{aligned} p_{ti}^{\text{foregr}} &= \sum_{k=1}^{N\_\text{orient}} \Bigg(l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \\ &\quad \cdot \left\{ T_t^{\text{mod}} * \Psi_p * \Lambda \right\}_{i((k-1)\cdot\pi/N\_\text{orient}+\pi/2)} \Bigg) \\ p_{ti}^{\text{backgr}} &= \sum_{k=1}^{N\_\text{orient}} \Bigg(l_{i(k-1)\cdot\pi/N\_\text{orient}}^{V2\_\text{Norm}} \\ &\quad \cdot \left\{ T_t^{\text{mod}} * \Psi_p * \Lambda \right\}_{i(k-1)\cdot\pi/N\_\text{orient}} \Bigg) \end{aligned} \tag{D.5}$$

The standard deviations are $\sigma_{\text{orient}} = 0.25$ for the Gaussian in the orientation domain and $\sigma_x = 11.0$ and $\sigma_y = 1.0$ for the anisotropic Gaussian spatial kernel. Finally, $p^{\text{foregr/backgr}}$ undergo self-inhibition to yield the normalized activities $q^{\text{foregr/backgr}}$

$$q_{ti}^{\text{foregr/backgr}} = \frac{C_T\, p_{ti}^{\text{foregr/backgr}}}{\alpha_q + p_{ti}^{\text{foregr/backgr}}} \tag{D.6}$$

with decay $\alpha_q = 0.1$ and scaling constant $C_T = 2.1$.

References

Adelson EH, Anandan P (1990) Ordinal characteristics of transparency. AAAI workshop on qualitative vision, pp 77–81

Anderson BL (1997) A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. Perception 26(4):419–453

Banks MS, Gepshtein S, Landy MS (2004) Why is spatial stereoresolution so low?. J Neurosci 24(9):2077–2089

Baumann R, v. d. Zwan R, Peterhans E (1997) Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. Eur J Neurosci 9(6):1290–1303

Baumann R, Zwan Rvan der , Peterhans E (1997) Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. Eur J Neurosci 9:1290–1303

Bayerl P, Neumann H (2004) Disambiguating visual motion through contextual feedback modulation. Neu Comput 16:2041–2066

Baylis GC, Driver J (2001) Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. Nat Neurosci 4(9):937–942

Boselie F (1994) Local and global factors in visual occlusion. Perception 23:517–528

Crick F, Koch C (1998) Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. Nature 391:245–250

Dev P (1975) Perception of depth surfaces in random-dot stereograms: a neural model. Int J Man-Machine Stud 7:511–528

Eckhorn R (2000) Neural mechanisms of visual feature grouping. Neurol Neurochir Pol 34:27–42

Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top–down processing. Nat Rev Neurosci 2(10):704–716

Felleman DJ, Essen DCvan (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47

Finkel LH, Sajda P (1992) Object discrimination based on depth-from-occulsion. Neural Comput 4:901–921

Francis G, Grossberg S, Mingolla E (1994) Cortical dynamics of feature binding and reset: control of visual persistence. Vis Res 34(8):1089–1104

Frien A, Eckhorn R (2000) Functional coupling shows stronger stimulus dependency for fast oscillations than for low-frequency components in striate cortex of awake monkey. Eur J Neurosci 12(4):1466–1478

Fukushima K (2001) Recognition of partly occluded patterns: a neural network model. Biol Cybern 84:251–259

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinisic horizontal and corticocortical connections in cat visual cortex. J Neurosci 9(7):2432–2442

Grossberg S (1980) How does a brain build a cognitive code. Psychol Rev 87(1):1–51

Grossberg S (1991) Why do parallel cortical systems exist forthe perception of static form and moving form. Percept Psychophys 49(2):117–141

Grossberg S (1993) A solution of the figure-ground problem for biological vision. Neural Netw 6:463–483

Grossberg S, Grunewald A (2002) Temporal dynamics of binocular disparity processing with corticogeniculate interactions. Neural Netw 15:181–200

Grossberg S, Mingolla E (1985) Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. Percept Psychophys 38(2):141–171

Heider B, Meskenaite V, Peterhans E (2000) Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours. Eur J Neurosci 12:4117–4130

Heider B, Spillmann L, Peterhans E (2002) Stereoscopic illusory contours—cortical neuron responses and human perception. J Cogn Neurosci 14(7):1018–1029

Heitger F, v.d. Heydt R, Peterhans E, Rosenthaler L, Kübler O (1998) Simulation of neural contour mechanisms: representing anomalous contours. Image Vis Comput 6:407–421

Hirsch JA, Gilbert CD (1991) Synaptic physiology of horizontal connections in the cat's visual cortex. J Neurosci 11:1800–1809

Howard IP (2003) Neurons that respond to more than one depth cue. Trends Neurosci 26(10):515–517

Howard IP, Duke PA (2003) Monocular transparency generates quantitative depth. Vis Res 43:2615–2621

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154

Hubel DH, Wiesel TN (1965) Receptive fields, and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol 28:229–289

Hupe JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by $V1$, $V2$ and $V3$ neurons. Nature 394(6695): 784–787

Kanizsa G (1979) Organization in vision: essays on Gestalt perception. Praeger, New York

Kellman PJ, Shipley TF (1991) A theory of visual interpolation in object perception. Cogn Psychol 23(2):141–221

Kellman PJ, Yin C, Shipley TF (1998) A common mechanism for illusory and occluded object completion. J Exp Psychol Hum Percept Perform 24(3):859–869

Kelly F, Grossberg S (2000) Neural dynamics of 3D surface perception: figure-ground separation and lightness perception. Percept Psychophys 62(8):1596–1618

Koenderink JJ, v. Doorn A (1982) The shape of smooth objects and the way contours end. Percept Psychophys 11:129–137

Kovacs G, Vogels R, Orban GA (1995) Selectivity of macaque inferior temporal neurons for partially occluded shapes. J Neurosci 15(3):1984–1997

Kumaran K, Geiger D, Gurvits L (1996) Illusory surface perception and visual organization. Netw Comput Neural Syst 7:33–60

Liu X, Wang DL (1999) Perceptual organization based on temporal dynamics. In: Paper presented at the IJCNN'99, Washington DC, USA

Marr D, Poggio T (1979) A computational theory of human stereo vision. Proc R Soc Lond B 204:301–328

McDermott J, Adelson EH (2004a) The geometry of the occluding contour and its effect on motion interpretation. J Vis 4(10):944–954

McDermott J, Adelson EH (2004b) Junctions and cost functions in motion interpretation. J Vis 4(7):552–563

Mignard M, Malpeli JG (1991) Paths of information flow through visual cortex. Science 251(4998):1249–1251

Mumford DB (1994) Neuronal architectures for pattern-theoretic problems. In: Koch C, Davis J (eds) Large-scale neuronal theories of the brain. MIT Press, Cambridge, pp 125–152

Nakayama K, Shimojo S, Ramachandran VS (1990) Transparency: relation to depth, subjective contours, luminance, and neon color spreading. Perception 19(4):497–513

Nakayama K, Shimojo S, Silverman GH (1989) Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. Perception 18(1):55–68

Neumann H, Mingolla E (2001) Computational neural models of spatial integration in perceptual grouping. In: Shipley TF, Kellman PJ (eds) From fragments to objects—segmentation and grouping in vision.. Elsevier, Amsterdam pp 353–400

Neumann H, Sepp W (1999) Recurrent $V1$–$V2$ interaction in early visual boundary processing. Biol Cybern 81:425–444

Nishina S, Okada M, Kawato M (2003) Spatio-temporal dynamics of depth propagation on uniform region. Vis Res 42:2493–2503

Ohzawa I, DeAngelis GC, Freeman RD (1997) The neural coding of stereoscopic depth. Neuroreport 8(3):iii–xiii

Peterhans E (1997) Functional organization of area $V2$ in the awake monkey. In: Rockland KS, Kaas JH, Peters A (eds) Extrastriate cortex in primates, vol 12. Plenum Press, New York

Peterhans E, Heitger F (2001) Simulation of neuronal responses definining depth order and contrast polarity at illusory contours in monkey area $V2$. J Comput Neurosci 10(2):195–211

Pianta MJ, Gillam BJ (2003) Paired and unpaired features can be equally effective in human depth perception. Vis Res 43:1–6

Poggio GF, Gonzalez F, Krause F (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. J Neurosci 8(12):4531–4550

Przybyszewski AW, Gaska JP, Foote W, Pollen DA (2000) Striate cortex increases contrast gain of macaque LGN neurons. Vis Neurosci 17(4):485–494

Qiu FT, v.d. Heydt R (2005) Figure and ground in the visual cortex: $V2$ combines stereoscopic cues with Gestalt rules. Neuron 47(1): 155–166

Regan D, Erkelens CJ, Collewijn H (1986) Visual field defects for vergence eye movements and for stereomotion perception. Invest Ophthalmol Vis Sci 27:806–819

Raizada RD, Grossberg S (2003) Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. Cereb Cortex 13:100–113

Rubin N (2001a) Figure and ground in the brain. Nat Neurosci 4(9): 857–858

Rubin N (2001b) The role of junctions in surface completion and contour matching. Perception 30:339–366

Sajda P, Finkel LH (1995) Intermediate-level visual representations and the construction of surface perception. J Cogn Neurosci 7(2): 267–291

Salin PA, Bullier J (1995) Corticocortical connections in the visual system: structure and function. Physiol Rev 75(1):107–154

Sandell JH, Schiller PH (1982) Effect of cooling area 18 on striate cortex cells in the squirrel monkey. J Neurophysiol 48(1):38–48

Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77

Shimojo S, Nakayama K (1990) Real world occlusion constraints and binocular rivalry. Vis Res 30(1):69–80

Shipley TF, Kellman PJ (1990) The role of discontinuities in the perception of subjective figures. Percept Psychophys 48(3):259–270

Shipley TF, Kellman PJ (1992) Strength of visual interpolation depends on the ratio of physically specified to total edge length. Percept Psychophys 52(1):97–106

Singh M, Huang X (2003) Computing layered surface representations: an algorithm for detecting and separating transparent overlays. In: Paper presented at the IEEE CVPR'03, Wisconsin USA

Smith AT, Singh KD, Williams AL, Greenlee MW (2001) Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. Cereb Cortex 11:1182–1190

Sporns O, Tononi G, Edelman GM (1991) Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. PNAS 88:129–133

Spratling MW, Johnson MH (2001) Dendritic inhibition enhances neural coding properties. Cereb Cortex 11:1144–1149

Thielscher A, Neumann H (2003) Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach. Neuroscience 122:921–939

Thomas OM, Cumming BG, Parker AJ (2002) A specialization for relative disparity in $V2$. Nat Neurosci 5(5):472–478

Tse PU (1999) Volume completion. Cogn Psychol 39:37–68

Tyler CW, Kontsevich LL (1995) Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. Perception 24:127–153

v. d. Heydt R, Heitger F, Peterhans E (1993) Perception of occluding contours: neural mechanisms and a computational model. Biomed Res 14:1–6

v. d. Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. Science 224(4654):1260–1262

Williams LR, Hanson AR (1996) Perceptual completion of occluded surfaces. Comput Vis Image Understand 64(1):1–20

Williamson JR (1996) Neural network for dynamic binding with graph representation: from, linking, and depth-from-occlusion. Neural Comput 8:1203–1225

Zhou H, Friedman HS, v.d. Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20(17):6594–6611