

ulm university universität
uulm

Nonlinear Recurrent Mechanisms for the Processing and Representation of Surface Boundaries Based on Luminance and Texture Gradients

DISSERTATION

zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

(DR.-ING.)

der Fakultät für Ingenieurwissenschaften
und Informatik der Universität Ulm

von

Dr. biol. hum. Axel Thielscher

aus Kösching

Betreuer:	Prof. Dr.-Ing. Jürgen Lindner Prof. Dr. rer. nat. Heiko Neumann
Amtierender Dekan:	Prof. Dr. rer. nat. Helmuth Partsch

Ulm, 30.04.08

Acknowledgements

First, I would like to thank Prof. Dr. J. Lindner for the interest he showed in the topic of this PhD thesis (which is not in the mainstream of his research) and for kindly undertaking the responsibilities as supervisor. Furthermore, I would like to thank Prof. Dr. M. Spitzer for giving me the opportunity to realize this thesis when I was member of the Department of Psychiatry (University of Ulm).

Prof. Dr. M.A. Cohen (CNS, Boston University) provided helpful comments on the stability analysis in chapter 2.

My former colleagues PD Dr. G. Grön and M. Koelle offered substantial support and steady motivation that enabled me to perform two fMRI-experiments to test some of the predictions of the neural model.

I am grateful for Prof. Dr. H. Neumann's steady support during the last years, his motivating and enlightening remarks and the frequent discussions which proved very helpful for my work. I would like to thank Heiko for establishing a pleasant and at the same time productive working atmosphere. In particular, I enjoyed the journal club meetings of his vision science group, with its current and former members Dr. T. Hansen, Dr. M.S. Keil, Dr. P. Bayerl and Dr. S. Corchs (among others).

Last but not least I would like to thank my wife Susanne for all her love and encouragement.

Table of Contents

1 Introduction **5**

1.1 Motivation and Overview 5

1.2 Architecture of the Human Visual System 8

1.3 How Texture- and Depth-Processing Relates to the Overall Task of Object Recognition in the Ventral Visual Pathway 14

1.4 Shunting Cell Dynamics: Motivation and Functional Properties 20

2 A Model for the Recurrent Processing of Texture Boundaries **25**

2.1 General Model Architecture: Physiological and Psychophysical Foundations 26

2.2 Model Areas and Receptive Field Organization 31

2.3 Model Cell Dynamics: Overview and Stability Analysis 33

2.3.1 Boundedness 39

2.3.2 Test for the Applicability of the Cohen-Grossberg Theorem 40

2.3.3 Test for Global Stability Using Hirsch's Theorem 42

2.3.4 Stability Analysis of a Hierarchy of Model Areas 48

2.4 General Model Behavior 51

2.5 Systematic Variation of Background Orientation Noise: Effect of Feedback 53

2.5.1 Stimuli & Analysis of Model Activation Patterns 54

2.5.2 Results 55

2.5.3 Relation to Psychophysics 60

2.6 Functional Roles of Specific Model Connections 62

2.7 Functional Role of Model Area V2 64

2.7.1 Example: Lesioning Feedback from Model Area V4 to V2 65

2.7.2 Relation to Psychophysics 66

2.8 Variations of Texture Density 69

2.8.1 Simulation Results 70

2.8.2 Relation to Psychophysics 72

2.9 Modulation of V1 Responses by Feedback Activity 73

2.10 Segmentation of Real-world Images: An Example 75

2.11 Main Discussion 77
2.11.1 Key Model Features 77
2.11.2 Relation to Psychophysics 78
2.11.3 Relation to Physiological and Neuroimaging Data: The Functional Roles of Areas V1 and V4 81
2.11.4 Other Models 88
2.11.5 Model Limitations and Putative Future Developments 91
2.12 Conclusion 92

3 Globally Consistent Depth Sorting of Overlapping 2D Surfaces in a Model of Local Recurrent Interactions 95
3.1 Recurrent V1-V2 Contour Processing 98
3.2 Recurrent Depth Processing 102
3.2.1 A Recursive Scheme for Depth Sorting of Surface Contours Using Local Relative Depth Cues 102
3.2.2 Outline of the Overall Model Architecture and Key Processing Mechanisms 105
3.2.3 Dipole Dynamics: A Close-up 109
3.2.4 Interactions between Neighboring Dipoles 111
3.2.5 Interaction between Model Layers 114
3.2.6 The Impact of T-junctions on Dipole Activity 115
3.3 Simulation Results 117
3.3.1 Pre-Processing 118
3.3.2 Recurrent Depth Processing 120
3.3.3 Further Simulation Results 125
3.4 Discussion 130
3.4.1 General Model Framework: Surface Boundary Processing and Depth Propagation 130
3.4.2 Depth Layers and Dipole Dynamics for Activity Propagation 132
3.4.3 General Model Framework 133
3.4.4 Other Models of Depth Processing 134
3.4.5 Model Properties, Limitations and Future Extensions 138

4 Summary 141

A Model of Texture Border Processing: Supplements **149**

A.1 Receptive Field Equations 149

A.2 Model Cell Dynamics: Reformulation of Equations for Stability Analysis 154

A.3 Boundedness 155

A.4 Stability Analysis of a Hierarchy of Model Areas: Supplements 157

B Model Determining the Depth of Surface Contours: Supplements **159**

B.1 Recurrent V1-V2 Contour Processing 159

B.2 Feed-forward Scheme for the Detection of Corners and T-Junctions 161

B.3 Interactions between Neighboring Dipoles: Supplements 163

B.4 Impact of the Model T-junction Detectors on Model Dipole Activity 166

Bibliography **169**

1 Introduction

1.1 Motivation & Overview

Recognition or identification of objects on the basis of two-dimensional images (such as pictures taken by cameras or the retinal images of the human eyes) has to cope with several non-trivial problems (Liter and Bülthoff, 1996): Subsequent images of the same object can differ substantially from each other, depending on variations of viewpoint, distance, illumination, etc. Furthermore, in real world situations, objects are often partially occluded by others, in turn complicating a clear-cut recognition. The development of computational frameworks and mechanisms enabling robust object recognition is in the focus of ongoing and vivid research activities. However, it is a very challenging problem and no general solution exists up to now. In contrast, humans and mammals achieve apparently effortless and robust object recognition even in cluttered and noisy real-world environments. So one way to improve the computational architectures and mechanisms used for image processing might be to try to understand how the same tasks have been solved by Nature. Clearly, a technical solution will never be a strict one-to-one implementation of biological vision. However, the hope is to identify general principles, constraints, mechanisms and design frameworks which can be used to guide the development of technical systems.

In the last few decades our knowledge about the structure and functionality of living beings has dramatically increased. This increase has led to a distribution and parcellation of this knowledge over various scientific disciplines and, at the moment, new scientific disciplines emerge as an attempt to integrate research on the basis of life into a common framework. Biological vision is a good example for that development. It is a traditional topic of Biology (focusing on the physiology of the vision system) and Psychology (focusing on the overall process of perception or *Wahrnehmung*). More lately, it has become one of the multi-disciplinary research focuses of Neurosciences, a prominent member of the newly emerging

Life Sciences. Research on biological vision is demanding and requires the integration of data obtained by several classical research disciplines such as Biology, Physics, Psychology, Medicine and Mathematics into a common concept. In this work, I use the tools developed in the field of Neuroinformatics to capture specific aspects of biological vision (namely the processing of surface textures and depth information) into a coherent theoretical framework, in turn allowing for a rigid testing of the underlying assumptions. The aim of my work is two-fold: On the one hand the results presented here allow for a better understanding of the neural mechanisms underlying biological object recognition, and on the other hand they might help to improve future technical approaches to image processing.

Although our understanding of biological vision is still rather at the beginning, many important key properties and principles have already been discovered. One central principle is the initial processing of surface boundaries by the mammalian visual system as necessary pre-requisite for object recognition. The visual system is faced with the problem that the same surface can look very differently when changing illumination. Even different parts of a uniform surface can have very different local brightness levels and colors due to reflections and shadows originating from neighboring objects. However, when we look at such a surface we perceive it as uniform and contiguous, i.e. our visual system has build-in mechanisms to *discount the illuminant*. A central strategy is the initial processing and representation of salient luminance discontinuities representing the *borders* of a surface, instead of the processing of the surface per se. For example, as demonstrated by the famous research of Hubel and Wiesel (1959, 1962, 1965) on the cat's visual system, a significant proportion of the neurons in the primary visual cortex (i.e. the brain area receiving visual input from the retina) selectively reacts to luminance steps having a specific position and orientation in the visual stimulus. This indicates that these neurons signal the *outline* or *shape* of a surface rather than the surface itself. More recent studies revealed that not only the representation of a surface having a common luminance value depends upon the initial processing of salient discontinuities, but that this basic principle also holds for surfaces defined by other properties such as color, texture or stereopsis (Lennie et al., 1990, Nothdurft, 1991, Lamme et al., 1999,

Conway et al., 2002). However, relatively little is known about the neural mechanisms underlying the processing of boundaries defined by changes of surface properties other than luminance. This served as motivation for the first part of this work in which I concentrate on the processing of texture borders. Texture processing has been intensively studied by many authors using psychophysical, physiological and functional imaging methods that have led to a vast amount of experimental data. However, up to now, all these details have not been put together into a common, coherent framework describing how texture processing is implemented in the visual system. I present a neural model to integrate the experimental data on texture processing into a common computational framework, in turn allowing to identify the probable neural origins of human texture processing. One focus of the work was to integrate previous modeling approaches into my model, which simulate early visual processes of boundary formation on the basis of luminance information (Grossberg and Mingolla, 1985, Neumann and Sepp, 1999). The computational capabilities of these approaches are incorporated and preserved, thereby integrating the processing of luminance and texture information for surface boundary finding into a unified computational framework.

Creating a stable representation of surface boundaries and thereby getting rid of spurious changes in, e.g. luminance, color or texture is only one necessary prerequisite for robust object recognition. It is likewise important to *group* (or *bind*) together distinct visual items and attributes belonging to the same object and to reliably *segregate* (or separate) those items from others belonging to different objects or the background. In complex and cluttered real-world environments containing, e.g. partially occluded and mutually overlapping objects, the tasks of *grouping* and *segregation* are far from trivial. The human visual system utilizes depth information as a major cue to robustly solve these problems (Nakayama et al., 1989, Kovacs et al., 1995, Baumann et al., 1997). The depth relations between items in the visual scene can be utilized to obtain a globally consistent depth sorting, in turn allowing for a segregation of those items belonging to distinct objects and enabling the segmentation of figure from ground. In the second part of this thesis, I 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. The model builds upon contour representations of objects as obtained by the model depicted in the first part of my work.

Depth information can stem from a variety of different visual cues dividing mainly into binocular (i.e. disparity) and monocular (e.g. occlusions, perspective, relative size ...) cues (e.g. Poggio et al., 1988, Kellman and Shipley, 1991, Howard, 2003). Most cues are locally restricted, i.e. their information is only available at some positions in the visual scene. For example, disparity information stems from the vertical displacement between the two projections of an object on both retinas. A continuous horizontal contour has no vertical displacement and, consequently, no disparity information can be derived from it. Likewise, contour intersections between occluding and occluded objects (denoted as T-junctions) are hints allowing to determine the local figure-ground direction (Rubin, 2001, Fig. 1.6A): The contour representing the top of the T belongs to the object in the foreground and the stem to the one in the background, respectively. T-junctions can thus be used as highly localized depth cues. As most depth information is locally restricted, a mechanism to globally propagate this information is necessary in order to obtain a consistent interpretation of the depth relations between all objects in a visual scene. In this thesis, I use T-junctions as depth cues and focus on a mechanism to propagate this information along object contours. The goal is to identify a (minimal) set of assumptions and mechanisms needed for a functioning model of activity propagation for depth sorting.

1.2 Architecture of the Human Visual System

Light which is projected on the background of the eye is converted into electrical signals by the photoreceptors (the *rods* and *cones*) of the retina (Kandel et al., 2000). The signals are passed on by nerve cells (the *ganglion cells*) to the optic nerve and further on to the brain. A ganglion cell reacts to light in a circumscribed and approximately circular area of the retina. This area is called the *receptive field* of the ganglion cell and typically possesses a circular *center-surround* organization: The cell reacts to light in the center of the receptive field, but

this reaction is inhibited when the surrounding area is simultaneously illuminated. Consequently, a ganglion cell reacts best to a differential illumination of the receptive field and is almost unresponsive to diffuse and uniform light. Most ganglion cells belong to one of two different classes, the *parvo-* and *magno-cellular* class. Magno cells have large receptive fields and are *not* selective to specific wavelength. Their responses to changes in illumination are very short and they are thus able to react to, e.g. rapidly moving stimuli. Parvo cells have small receptive fields (i.e. their spatial resolution is higher than that of the magno cells), their responses are longer in time and they respond selectively to specific wavelengths. They are thus the first processing stage in the perception of color. The optic nerves deliver the visual information from the eyes to the two hemispheres of the brain (Fig. 1.1). The two nerve bundles intersect in the *optic chiasm*, where the visual information is reorganized depending

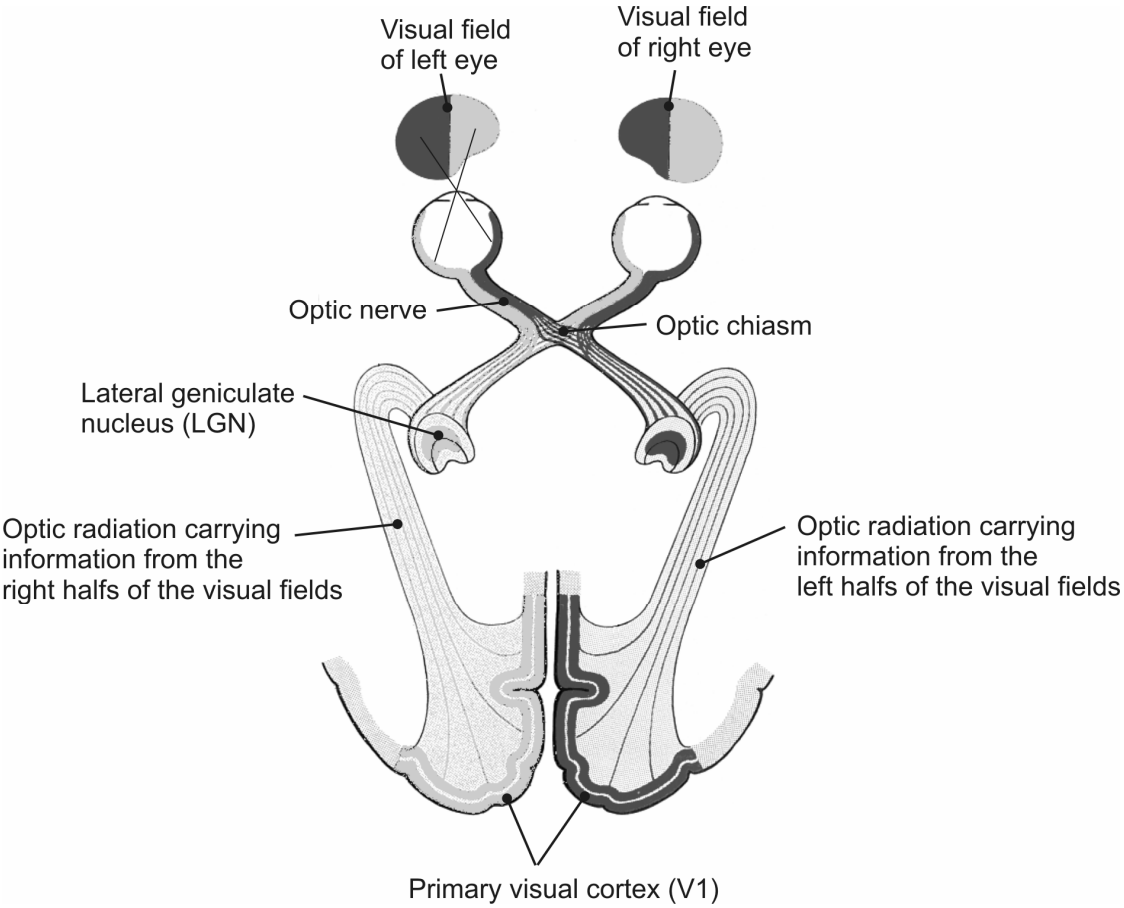


Figure 1.1: The visual pathway from the eyes to the primary visual cortex (adapted from Zeki, 1993)

on its spatial origin in the visual field: Nerve fibers originating, e.g. from the right halves of the retinas are combined in the optic chiasm and commonly projected to the right hemisphere of the brain. As the right halves of the eyes both receive light from the left half of the visual field, the right hemisphere exclusively gets information from this part of the visual field. Likewise, the left hemisphere of the brain exclusively receives information from the right half of the visual field. The next processing stage are the *lateral geniculate nuclei (LGN),* where the nerve fibers are re-sorted depending on if they carry magno- or parvo-cellular information. This information is then transferred via the optic radiations to the primary visual areas (also called visual area 1, V1, striate cortex or Brodman area 17) of the two hemispheres.

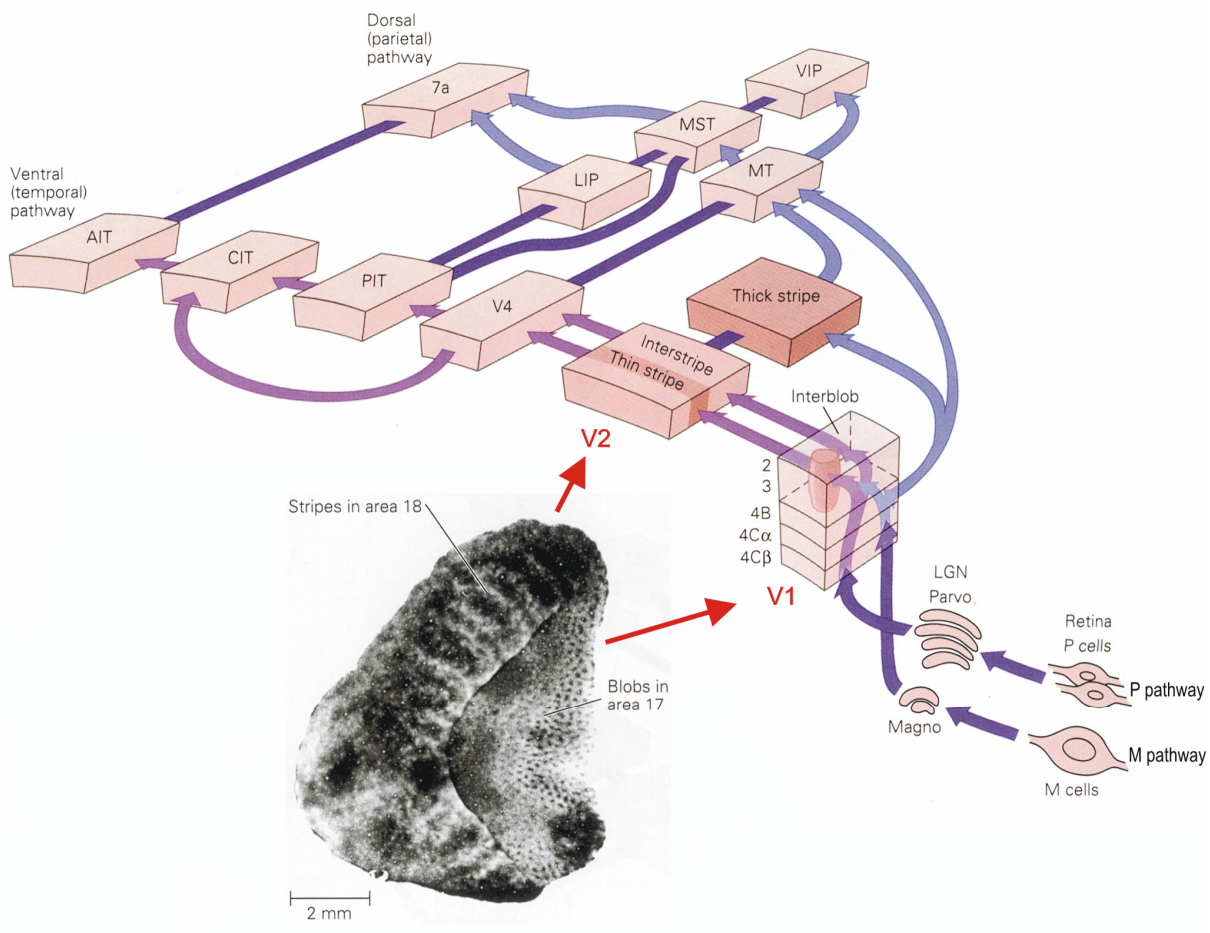


Figure 1.2: The parvo- (P) and magno- (M) cellular pathways from the retina to the primary visual cortex are reorganized in V1. Cells in V1 project via area V2 to higher visual areas in the temporal and parietal parts of the brain (from Kandel et al., 2000).

Staining of the visual cortex (in case of Fig. 1.2 with cytochrome oxidase) reveals a regular pattern in the primary visual area V1 and the second visual area V2: V1 can be divided into blob- and interblob-regions. V2 contains thin-, thick-, and interstripe regions. The primary visual cortex is approx. 2 mm high and can be divided into several distinct vertical layers based on results from histological staining (as schematically depicted for area V1 in Fig 1.2; the organization into altogether six vertical layers is a general feature of the cerebral cortex). The magno- and parvo-cellular pathways terminate in different layers of the primary visual area, so that the separation between these two functionally different pathways is maintained within V1. For example, most cells in layer 4B of V1 are sensitive to motion, but insensitive to color. They project either directly to the higher visual area MT (middle temporal area, or V5; see Fig. 1.2) or first to the thick stripes of V2 and the cells there further on to MT. In contrast, cells in the blob- and interblob regions of layers 2 and 3 receive mainly input from the parvo-cellular pathway. Most cells in the blobs are color-sensitive and have a concentric center-surround organization of their receptive fields. They project to the thin stripes of V2, and then further on to the higher visual area V4. Cells in the interblob regions are the first cells of the visual processing hierarchy which have a receptive field layout *different* from the concentric center-surround organization of the cells in the retina and LGN. In contrast, they possess elongated receptive fields (see model area V1 in Fig. 2.2) consisting of several non-circular subfields. They are hardly sensitive to circular light dots, but prefer local luminance transitions along a given orientation preference such as light bars or surface edges. The two major subgroups of cells in the interblob regions are *simple*- and *complex-cells*. Simple-cells are selective to the orientation of a bar or surface border *and* are sensitive the contrast polarity (i.e. dark-light and light-dark). Complex-cells are also orientation selective but insensitive to the direction of contrast. They project to the interstripe region of visual area V2 and further on to area V4.

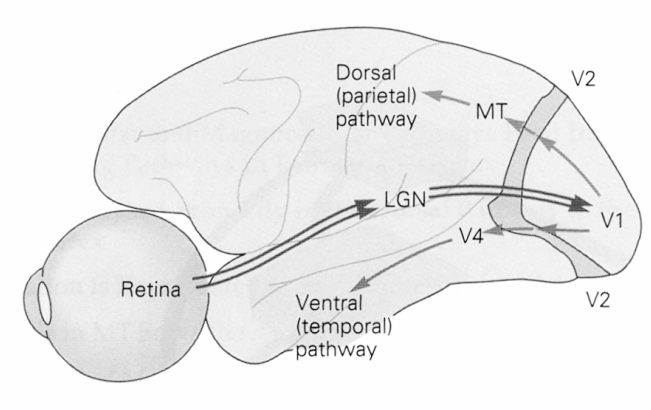


Figure 1.3: The two major visual pathways: The ventral pathway of color and form processing, and the dorsal pathway of motion processing and representation of visual space (from Kandel et al., 2000).

Taken together, two major pathways can be distinguished within the visual system, one carrying motion information and the other one carrying color and form information (Fig. 1.3). The motion pathway is fast, not color-sensitive and has a lower spatial resolution. It starts with the magno cells of the retina and includes layer 4B of V1, the thick stripes of V2 and area MT, and projects further on to higher visual areas in upper rear part (or: dorsal posterior part) of the brain. Most commonly it is called the *dorsal visual pathway*. The color and form pathway is slower and has a higher spatial resolution. It starts with the parvo cells of the retina, includes the blobs and interblob regions of V1, the thin stripes and interstripes of V2, and area V4. From there, it projects to higher visual areas in the lower rear part (or: ventral posterior part) of the brain. It is mostly referred to as the *ventral visual pathway*. The neural models of texture and depth processing presented in this thesis are meant to imitate a part of the information processing performed in the ventral visual pathway (Fig. 1.3).

While it is sensible and important to distinguish between the two functionally different visual pathways, it is likewise important to point out that a lot of anatomical connections exist between them (as visible from Fig. 1.2). An interesting and open question is how and which kind of information is exchanged between them in order to mutually support each other in the overall tasks of object recognition and construction and representation of visual space.

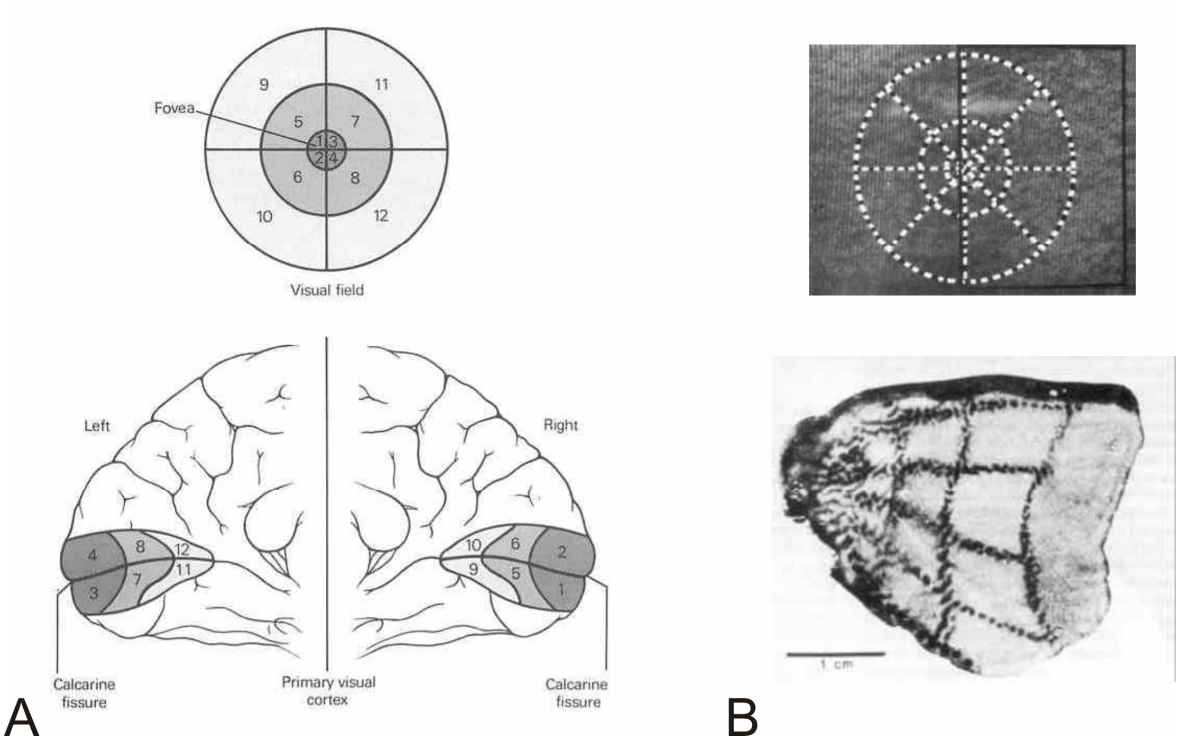


Figure 1.4: (A) Schematic illustration of the representation of the visual field in the retina and in the primary visual cortex. The topographic relationship between cells in the retina is maintained in V1, i.e. V1 is retinotopically organized. The foveal part of the retina has the largest representation in V1. (B) Experiment determining the topographic organization of V1: The circular light pattern shown in the upper part is projected on the retina and dyes which are sensitive to metabolic activity are simultaneously injected in the visual cortex. The dyes tag the cells with high activity and thereby reveal the retinotopic organization in V1 (from Tootell et al., 1982, Kandel et al., 2000).

The cells in many visual areas such as V1, V2 or V4 are arranged in a topographic fashion, i.e. they maintain the neighborhood relationships of the corresponding cells in the retina (Fig. 1.4). In other words, neighboring cells in V1 receive input from neighboring cells in the retina. This *retinotopic representation* of the visual field gets more crude in higher visual areas such as V4 or MT, and is finally almost lost in areas such as AIT or 7a (Fig. 1.2). The cells' mean receptive field sizes increases monotonically from the center of the visual field to the periphery. Consequently, the area on the cortex necessary to represent the periphery of the visual field is equal or even smaller than the area representing the central (or foveal) part (Fig 1.4A). The average receptive field size increases from V1 to V2, V4, etc., i.e. the receptive fields of cells in higher visual areas cover larger parts of the visual field. Furthermore, the higher the area within the hierarchy of the visual system, the more complex

the stimuli its cells optimally respond to. For example, while most cells in V1 prefer quite simple visual patterns such as light dots or oriented bars, cells in inferior temporal cortex (areas PIT, CIT and AIT in Fig. 1.2) selectively respond to, e.g. monkey faces and smilies.

1.3 How Texture- and Depth-Processing Relates to the Overall Task of Object Recognition in the Ventral Visual Pathway

The overall task of object recognition can be divided into a hierarchy of several functionally distinct sub-entities (Fig. 1.5): In a first step, a representation of a variety of informative features in the input image such as edges, texture elements, line crossings, etc. is build. For that purpose, a bank of feature detectors (such as linear filters) each working in a restricted local neighborhood is typically used. Subsequently, the features are grouped in larger spatial neighborhoods to construct coherent representations of the surfaces and volumes in the image. This information is then used by the higher hierarchy levels to recognize or classify the objects in the visual scene. Due to the complexity of the task of object recognition, the analysis first starts with the detection of locally restricted features, and subsequently recovers a global interpretation of the complete scene in several successive steps. It is questionable if the human visual system utilizes a fixed and clearly defined set of non-intersecting and distinct features (e.g., Beck, 1982, Julesz, 1986) and the above depicted distinct functional tasks are surely intermingled and distributed across several visual areas. However, it is likely to assume that the human visual system *generally* follows the above depicted architecture: As mentioned in the previous section, the cells at the first levels of visual processing react to rather simple and locally restricted stimulus arrangements. The higher a neuron in the hierarchy of ventral visual areas, the larger its field of spatial integration and the more complex its preferred stimulus arrangements.

Our visual system has to achieve robust and stable object recognition that is invariant to changes in illumination, rotation, scaling, etc. It has been proposed that the selectivity of cells in V1 for oriented luminance steps is the first step to build a contour representation of the

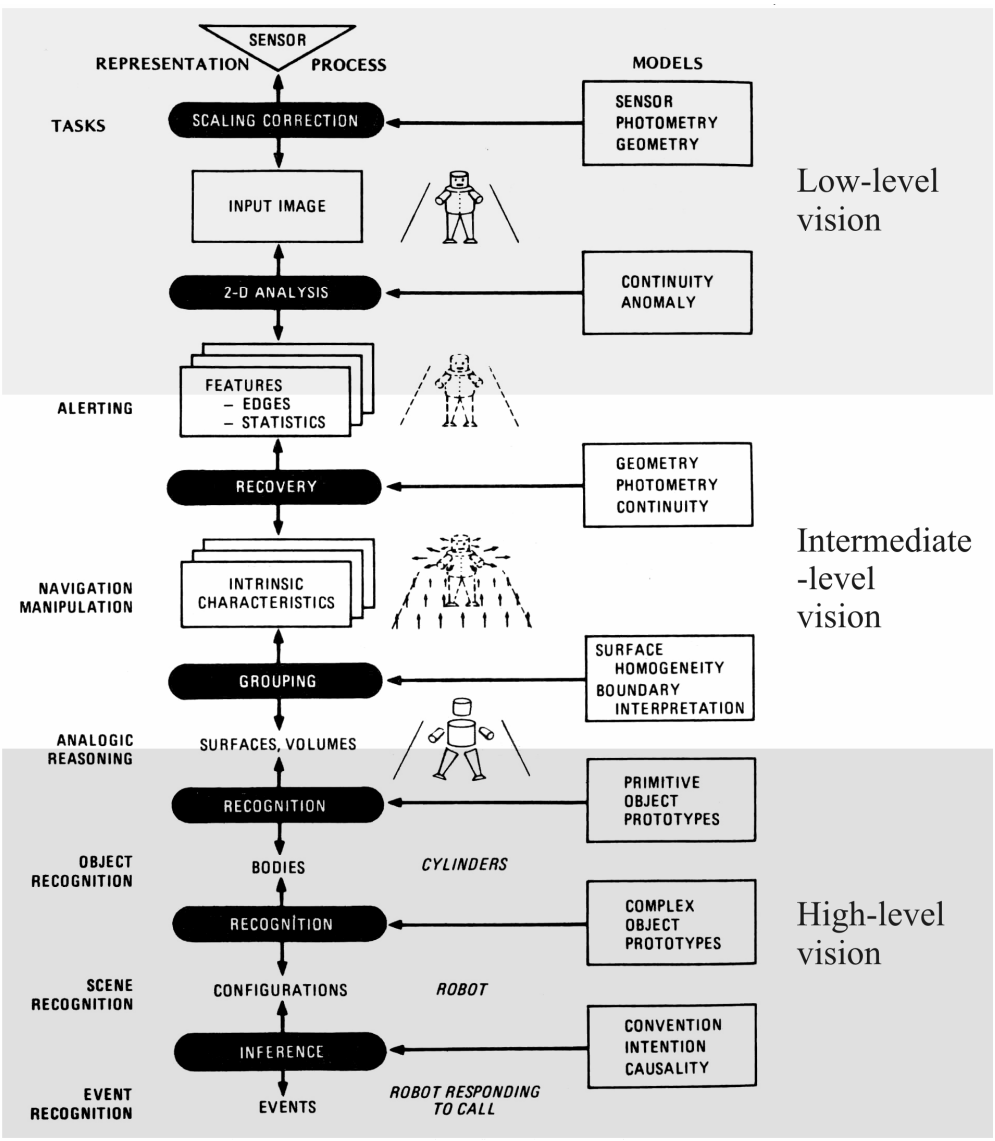


Figure 1.5: Architecture of a general-purpose system for object recognition. The overall task of object recognition can be divided into low-, intermediate- and high-level processes basing on successively more abstract representations of the visual input (from Witkin and Tenenbaum, 1983).

objects in the visual scene which is invariant to changes in illumination. However, the problems arising from such a representation based on contour information are manifold. For example, contours in real-world images are often not perfectly continuous, but are interrupted due to occlusions and variations in illumination (see Fig. 1.6A). Consequently, a mechanism to complete interrupted contours is necessary. Indeed, cells in visual area V2 are capable of signaling the existence of putative surface contours even at image positions where no

physical contour information is available (v. d. Heydt et al., 1984). These cells also respond to so called *illusory contours* as visible in Fig. 1.6B. For example, they might use the "knowledge" that one or several line endings aligned perpendicular to main axis of their receptive fields probably represent an occluded object, and that consequently an invisible contour exists which is perpendicular to the line endings. Less is known about the question how the surface information itself is "*filled in*" again after first constructing a contour representation of the objects in the visual scene. Clearly, from a subjective point of view, we don't see contours but complete objects consisting of several surfaces. So the surface information has to be recovered again after first discarding it and concentrating on the contours. Some authors use the word "*filling-in*" of surface information just as metaphor having no direct physiological counterpart. On the other hand, neural models have been suggested which are one-to-one implements of "*filling-in*" algorithms (Pessoa et al., 1995). These models base on contour representations and are capable of, e.g. filling in the corrected brightness value of a surface after eliminating luminance changes caused by shadows and reflections of neighboring objects.

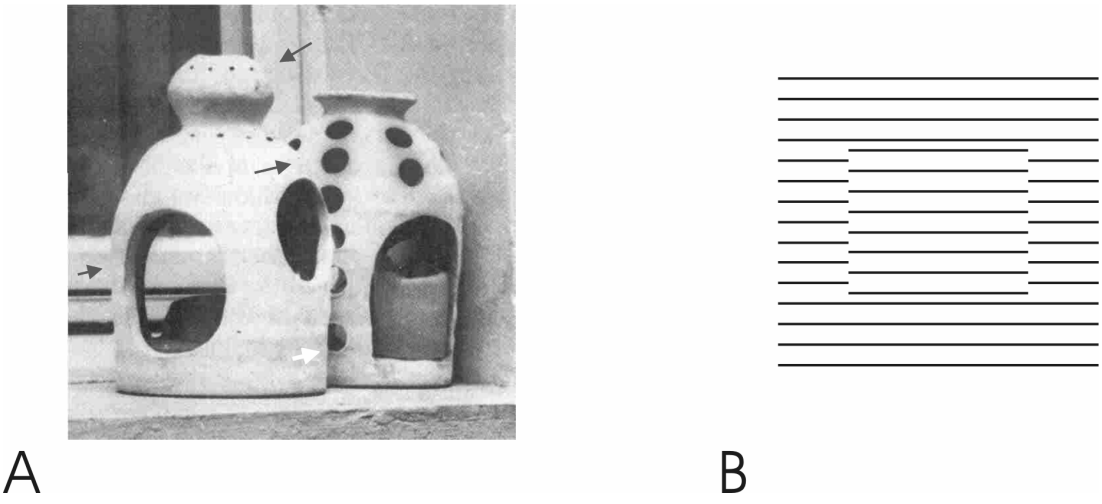


Figure 1.6: (A) A real-world example of occlusions. The dark arrows indicate positions at which the gray-scale values of the objects in the fore- and background are approx. equal, and consequently no contour information is available. The white arrow depicts a position at which the contours of the fore- and background objects form a T-junction (from Peterhans and v. d. Heydt, 1991). (B) An artificial image containing an illusory rectangle.

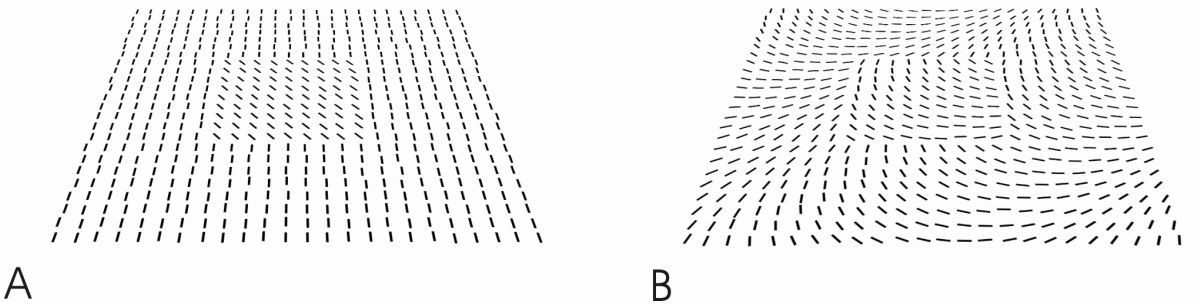


Figure 1.7: Examples for different sources of texture gradients. (A) In the center of the planar surface, a square pops out, created by an angle of 37° between neighboring line elements of the square and the background texture. The angle between the lines in the outermost left and right columns of the overall image is 40°, creating the impression of a tilted plane. The regular texture gradient of the background is informative when trying to determine the tilt of the surface, but has to be discarded when segmenting the inner pop-out square from the background texture. (B) The same situation as in (A), but with an additional orientation shift between neighboring lines. This orientation shift symbolizes the texture noise in natural images, and has to be discarded (as the regular texture gradient caused by the tilt, too) when segregating the inner pop-out square.

Above, it is argued that cells in V1 and V2 build a contour-based representation of the objects in the visual scene using oriented luminance steps as feature. How is this representation combined with other representations using other stimulus features such as color, texture, or motion? Are the latter representations also based on surface borders? In the first part of this thesis, I focus on human texture processing. It is known that neurons in area V1 signal the *borders* of texture surfaces (Lamme et al., 1999) and it is currently an open question if these neurons are capable of detecting these borders on their own, or if their activity reflects feedback information which they receive from higher visual areas. It is also known that an intact area V4 is necessary for successful texture segregation (de Weerd et al., 1996, Merigan, 1996, 2000). On the first view, the processing of surfaces based on luminance- and texture-information seem to be quite different from each other. However, both visual features share one critical property, namely that they hardly ever stay constant across a surface. The local luminance value changes, e.g. due to shadows and reflections. Likewise, a natural surface normally possesses a varying texture structure and the slant of the surface additionally adds systematic deviations to the texture (Fig. 1.7). Consequently, in both cases it is favorably to start processing of the surfaces by detecting *salient* changes of

the features which most likely occur at surface borders. The enhanced V1 cell activity at texture borders indicates that our visual system follows that strategy in case of texture processing. However, the rather low-frequent spatial changes of texture patterns across a surface requires a comparison of surface properties over wide areas of the visual field to detect the salient texture gradients or discontinuities representing a border. As a consequence, it is questionable that cells in V1 with their small receptive fields can process these borders on their own. The processing of texture boundaries most likely involves higher visual areas such as V4 as well.

In this thesis, I tackle the problem of how to process texture boundaries within a hierarchy of ventral visual areas from V1 to V4, and how to integrate the processing of luminance boundaries within visual areas V1 and V2 into that hierarchy. This is a contribution to the more general problem of how our visual system can cope with noisy features which define the surfaces when building a robust representation of the objects in a visual scene. Furthermore, I investigate the neural mechanisms which are necessary to separate different, partially overlapping 2D surface patches from each other based on monocular depth information, and to group together different parts of an occluded surface. The goal is to create a neural representation in which the contours of different objects are disentangled from each other. This helps to prevent cross-talk between parts belonging to different objects in the successive steps of object recognition. Clearly, the model of depth processing is minimal in terms of the kind of stimuli (2D patches) it can handle. However, up to now, little is known about the problem how the human visual system represents more complex 3D objects and the spatial relationships between several of such objects. To my knowledge, no biologically plausible neural model of 3D object segmentation exists up to now. Once a contour representation has been created, the next task is to *fill in* or reconstruct the surfaces itself, which is beyond the scope of this study

Taken together, the visual processes depicted above create a representation of the objects in the visual scene which is invariant to spurious variations in illumination and texture and in which the visual information of different objects is separated from each other. These are

critical prerequisites for subsequent higher-level vision tasks solving, e.g. problems like rotation- and scaling-invariance. For example, rotation-invariant object recognition can be achieved by learning several contour representations (or “snapshots”) of the same object from different points of view (Liter and Bülthoff, 1996, Riesenhuber and Poggio, 1999). It was shown by simulation studies that objects can be reliably recognized when a sufficient number of views has been learned, as new views can be interpolated from the learned ones. It is likely that these kinds of tasks mainly involve higher ventral visual areas such as PIT, CIT and AIT.



Figure 1.8: A famous optical illusion illustrating the impact of top-down information on object recognition. When seeing it for the first time, most people hardly recognize more than a clutter of black and white patches in the image. However, it is easy to detect the Dalmatian in the middle of the image when its existence has been pointed out to you. Furthermore, once the dog has been recognized, it is impossible to *not* seeing it again when looking at the picture for a second time.

Finally, it is important to note that the above outlined serial and feed-forward hierarchy of visual processes is a gross simplification of the processes of object recognition in the ventral visual pathway. For example, the anatomical connections between two successive areas mostly include dense feedback projections, enabling a bi-directional flow of information. It is likely to assume that efficient and powerful object recognition is hardly possible in a strict feed-forward structure, but crucially relies on a recurrent flow of information. For example, the problem of determining which surface patches belong to which objects in a cluttered

visual scene may be unambiguously solved only when knowledge about the shapes of these objects is available (the importance of top-down information for segmentation can be nicely seen in the famous optical illusion illustrated in Fig. 1.8). However, the exact functional role of the information relayed via the feedback connections is still a matter of intense debate.

1.4 Shunting Cell Dynamics: Motivation and Functional Properties

The question of the functional role of the feedback projections from higher to lower visual areas also includes the problem of how bottom-up and top-down information is combined in the cortex. As mentioned in section 1.2, the cerebral cortex can be divided into 6 distinct vertical layers. A first hint that bottom-up and top-down information is processed differently comes from the finding that feed-forward projections mainly terminate in layer 4, whereas feedback projections mostly avoid this layer and concentrate on layers 1/2 and 6 (Felleman and van Essen, 1991). Furthermore, several physiological studies show that feedback from higher visual areas is not capable of driving cells in lower areas, but *modulates* their activity (Sandell and Schiller, 1982, Mignard and Malpeli, 1991, Salin and Bullier, 1995, Hupe et al., 1998, Przybyszewski et al., 2000). For example, feedback activation alone is not sufficient to drive V1 neurons if they are not stimulated by a visual feeding input (Sandell and Schiller, 1982). Likewise, projections from the striate cortex to the LGN multiplicatively enhance responses of parvocellular neurons to grating stimuli (Przybyszewski et al., 2000). Hupé et al. (1998) demonstrated that feedback from area V5 (or MT) to areas V1, V2 and V3 amplifies and facilitates neural responses in these areas. Based on these findings, Crick and Koch (1998) proposed in their "*no-strong-loops hypothesis*" that a directed loop between two cortical areas will *not* consist of two driving connections, but will use one driving and one modulatory connection in order to avoid uncontrolled oscillations of the overall system and to limit the amount of inhibition necessary to achieve a stable network behavior. Using multiplicative instead of, e.g. additive excitatory feedback is one possible implementation of such a modulatory connection.

Another general feature of the cerebral cortex is a circular ON-center/OFF-surround architecture: Cells in a close spatial neighborhood excite each other and inhibit cells which are further away (Grossberg and Williamson, 2001, Kastner et al., 2001). It has been argued that in a hierarchy of feed-forward connected cortical areas lateral inhibition is necessary to keep the activity in higher areas focused and narrow despite the larger receptive field sizes in these areas (Kastner et al., 2001, Fig. 1.9). To our view, the same holds for the integration of activity from higher areas in lower areas via feedback. Consider, for example, the case that the activity distribution in a higher model area is somewhat “blurred” or broadened due to the larger receptive field sizes of the cells in that area. In a first step, feedback from this area will enhance the cell activity in a lower area at topographical positions corresponding to the same blurred region. Without center-surround competition following the feedback interaction, this in

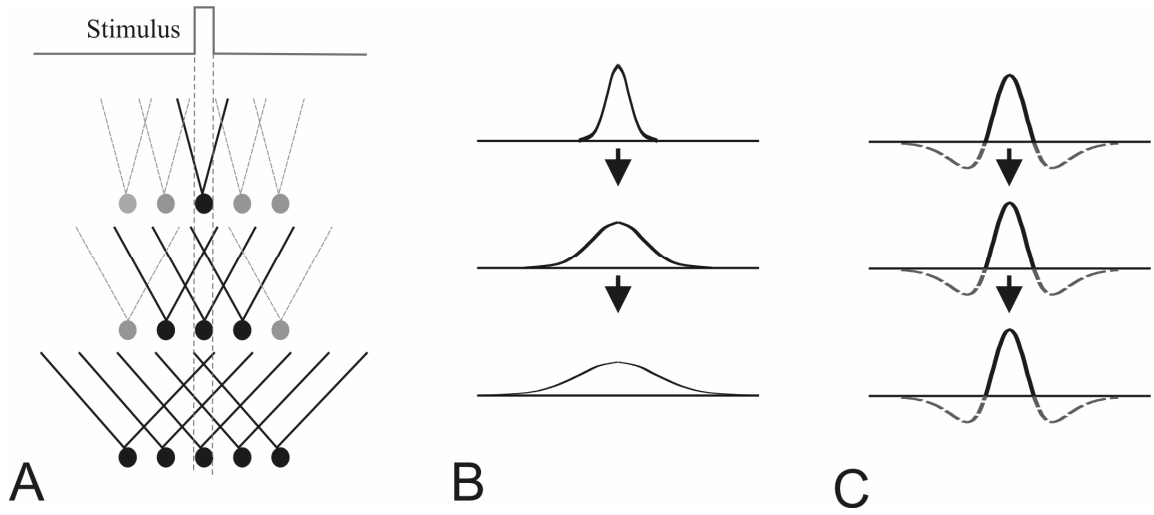


Figure 1.9: (A) The receptive field sizes increase monotonically within the hierarchy of visual areas. Consequently, the number of neurons receiving input from a spatially focused stimulus increases from lower to higher visual areas. (B) This results in a successive broadening or blurring of the activity distribution in higher visual areas. (C) Center-surround competition counteracts this effect by inhibiting the weak responses at the borders of the activity distribution.

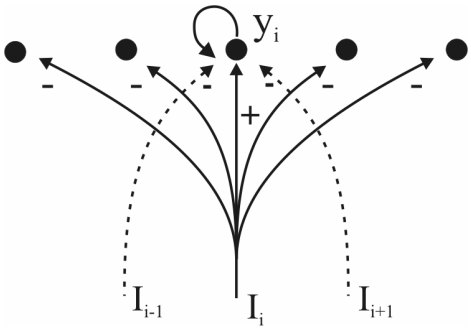


Figure 1.10: A very simple ON-center/OFF-surround network. Neuron i is only excited by input I_i, and inhibited by all other inputs. The activity of the neuron (denoted by y_i) has an multiplicatory impact on both the excitatory and inhibitory input.

turn will lead to an activity pattern in the higher area which tends to be even more unfocused, and so on. However, that effect is prevented by center-surround competition in the lower area narrowing the zone of activity enhancement via feedback and, in consequence, helping to spatially focus the responses to an input pattern in both the lower and the higher model area.

The functional properties of multiplicatory feedback interaction in combination with center-surround interaction have been intensely studied by Grossberg (Grossberg 1973, Cohen and Grossberg, 1983, 1988). He demonstrated that multiplicatory or shunting interaction combined with a center-surround architecture possesses properties such as *automatic gain control* and *normalization* of the total neuronal activity. These properties can be easiest demonstrated using a simple shunting feed-forward network (Fig. 1.10) in which neuron i is excited only by its own input I_i, and inhibited by the remaining input:

$$\frac{\partial}{\partial t} y_i = -A y_i + (B - y_i) I_i - y_i \sum_{k \neq i} I_k \qquad (1.1)$$

The impact of the input on the cell activation level is controlled (or *shunted*) by the cell activity itself: The term $(B - y_i)\, I_i$ ensures that input I_i excites the cell only until its activity y_i asymptotically reaches the upper bound given by B. The next term $- y_i \sum_{k \neq i} I_k$ gradually

decreases the impact of the inhibitory surround input when cell activity y_i reaches 0. Finally, -Ay_i denotes the passive activity decay. For a constant input, y_i approaches an equilibrium value (i.e. $\partial/\partial t\, y_i = 0$), which is given by

$$y_i = \theta_i \frac{BI}{A+I} \tag{1.2}$$

$I = \sum_{k=1}^{N} I_k$ represents the overall input to the network, and θ_i denotes the relative proportion of I_i among I (i.e. $I_i = \theta_i I$). The overall activity in the network given by

$$y = \sum_{k=1}^{N} y_k = \frac{BI}{A+I} \tag{1.3}$$

is *independent* of the number of active cells. In other words, the activities y_k of the single cells are *normalized* so that the overall activity y is constant for a specific input strength I. The relative activity of a cell is given by $Y_i = y_i/y = \theta_i$, i.e. it always equals the relative strength of the input I_i to that cell. In other words, the network doesn't suffer from saturation effects even for strong overall inputs I. This is due to the fact that the *gain* of the cells (i.e. the sensitivity of cell i to changes of I_i) is *automatically controlled* by the multiplication of the overall input I and the cell activation level y_i in equation (1.1): The variations of y_i in dependence on I_i are large for low levels of I and decrease monotonically with increasing I. Grossberg demonstrated that recurrent versions of the above depicted shunting ON-center/OFF-surround possess the same properties of automatic gain control and normalization (Grossberg, 1973), and have absolutely stable responses to constant inputs (Cohen and Grossberg, 1983). As a consequence, they can, e.g. be used as content-addressable memory.

2 A Model for the Recurrent Processing of Texture Boundaries

In this chapter, a neural model for the processing of texture boundaries is developed and tested. As emphasis is placed on the biologically plausibility of the model, the chapter starts with an outline of the physiological and psychophysical key findings underlying the model architecture and model cell dynamics in section 2.1. In the next section 2.2, the hierarchy of model areas is briefly described together with the receptive field properties of the model cells employed in the successive areas. Section 2.3 introduces the model cell dynamics and highlights its role as *context-selective gain enhancement* mechanism. Furthermore, its global stability properties are discussed and exemplarily demonstrated using a simplified model network. The subsequent results sections 2.4 to 2.10 highlight several key findings of the simulation studies performed with the model. In particular, the performance of the overall model as well as the functionality of the feedback connections are systematically investigated using texture stimuli adapted from psychophysical studies (section 2.5). Subsequently, the functional roles of specific areas and connections are studied using artificial lesions of selected model components (section 2.6). In sections 2.7 and 2.8, the model behavior is related to two key psychophysical results on human texture processing, allowing to trace back these findings to the putatively underlying neural mechanisms. The two final result sections 2.9 and 2.10 link the model behavior with electrophysiological data on the response characteristics of cells in the primary visual cortex and exemplarily demonstrate the model's ability to segment real-world texture images. In section 2.11, the physiological and psychophysical literature on human texture processing is discussed in the light of the model architecture and the simulation results demonstrated in the preceding sections. Finally, the current modeling approach is compared to previous models of texture processing and the

limits and putative future developments of the model are outlined. Chapter 2 concludes with a short summary.

2.1 General Model Architecture: Physiological and Psychophysical Foundations

The use of texture information for segmentation is compelling, as it requires a sophisticated analysis and comparison of surface properties over wide areas of the visual field. For example, the detection of the Dalmatian in Fig. 2.1 clearly relies on texture, since no other information can be utilized for object segregation, such as, e.g. different luminance levels or continuous contrast contours. Texture processing has been intensively studied by many authors using psychophysical, physiological and functional imaging methods. Based on the

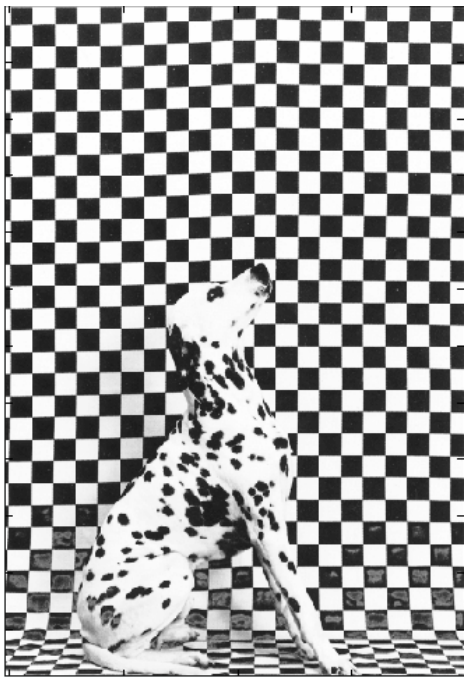


Figure 2.1: Example of figure-ground-segmentation relying on texture information. The silhouette of the dog is defined by the change between the regular background texture containing black and white squares and the randomly patterned skin of the dog (Original figure reprinted with permission by Gert Weigelt 1985 / Art Unlimited Amsterdam; image size 400x590).

early results of the experiments by Beck (1966) and Julesz (1965), respectively, psychophysical studies have focused on the question which features characterize two juxtaposed textures so that they can be automatically, or pre-attentively, separated from each other (Julesz, 1981, Beck, 1983, Bergen and Adelson, 1988, Landy and Bergen, 1991). Common to all these studies is the finding that orientation and size are key features for the segregation of otherwise homogeneous texture regions. The influence of other features like crossings or terminations on automatic segmentation has been identified to be substantially weaker. Furthermore, psychophysical results indicate that texture segmentation depends on the detection of local first-order feature dissimilarities at region borders (Nothdurft, 1985, Landy and Bergen, 1991). Segmentation based on feature differences leading to, e.g. orientation contrast between bar items, was investigated psychophysically in detail by Nothdurft (1985, 1991). Having stimuli composed of lines that form continuous flow patterns (see Fig. 1.7B), both a central pop-out bar and the background may contain lines of same orientation or statistics of oriented items. The central bar nevertheless pops out pre-attentively when the orientation contrast along the border is strong enough, that is the slope in the gradient of texture orientation is sufficiently large. These findings indicate that the visual system does not achieve texture segmentation by grouping together similar feature elements of a region (according to a homogeneity criterion), but rather by detecting boundaries defined by sufficient contrast along a particular feature dimension. This view is further supported by physiological studies using single- and multi-unit recordings in monkey V1 which reveal an enhanced activity of orientation selective cells when texture borders defined by an orientation contrast between neighboring line elements fall on their receptive fields (Gallant et al., 1995, Lamme et al., 1998, Lamme et al., 1999, Nothdurft et al., 1999, Nothdurft et al., 2000).

Robust segmentation of stimuli embedded in background orientation noise necessitates the pooling of information over extended areas of the retinal image in order to selectively identify meaningful orientation contrasts which signal a texture border. In the visual system, receptive field sizes increase monotonously within the hierarchy of cortical areas (Smith et al., 2001).

Recent functional imaging studies demonstrated a significant contribution of higher visual areas, including V4, in texture segmentation as well as in illusory contour detection (Mendola et al., 1999, Kastner et al., 2000). Likewise, lesion studies in monkeys showed that the detection of illusory contours and texture borders as well as the discrimination of shapes is severely impaired after ablation of V4 (de Weerd et al., 1996, Merigan, 1996, 2000). In consequence, it is likely to assume that cells in higher visual areas, such as V4, provide the necessary substrate of detecting texture boundaries in complex and noisy scenes.

Taken together, the available experimental data suggests that a hierarchically organized stream of several visual areas is concerned with texture processing. Within this hierarchy, segmentation of texture regions is achieved by detection of salient pattern arrangements signaling region boundaries. In accordance with the above depicted results, I suggest a model of recurrent texture boundary processing in the ventral visual stream of static form processing. It builds upon the two key properties that (i) texture segmentation is based on boundary detection and that (ii) texture border detection is mainly achieved by cells in higher model areas. According to the known anatomical data, it consists of a hierarchy of bi-directionally linked model areas V1, V2 and V4 (Felleman and van Essen, 1991, Ungerleider and Haxby, 1994, Fig. 2.2). Each area contains a retinotopically organized map (Sereno et al., 1995, Peterhans, 1997, Tootell et al., 1997) of model cells with gradual activation dynamics representing the average response (firing-rate) of groups of cortical neurons having similar selectivities. The model cells are orientation selective (Zilles and Clarke, 1997) and their receptive field sizes increase monotonically within the hierarchy of model areas (Smith et al., 2001). Their relative sizes have been parameterized in a ratio V1:V2:V4 = 1:2.5:8. The response selectivity of a model cell is determined by the spatial layout and organization of its receptive field pooling the bottom-up cell activities of the previous model area. Model cells belonging to a particular area exhibit stereotyped receptive field organizations and response selectivities. However, receptive field organization differs between the model areas, thereby determining the specific functional role of an area in texture boundary processing.

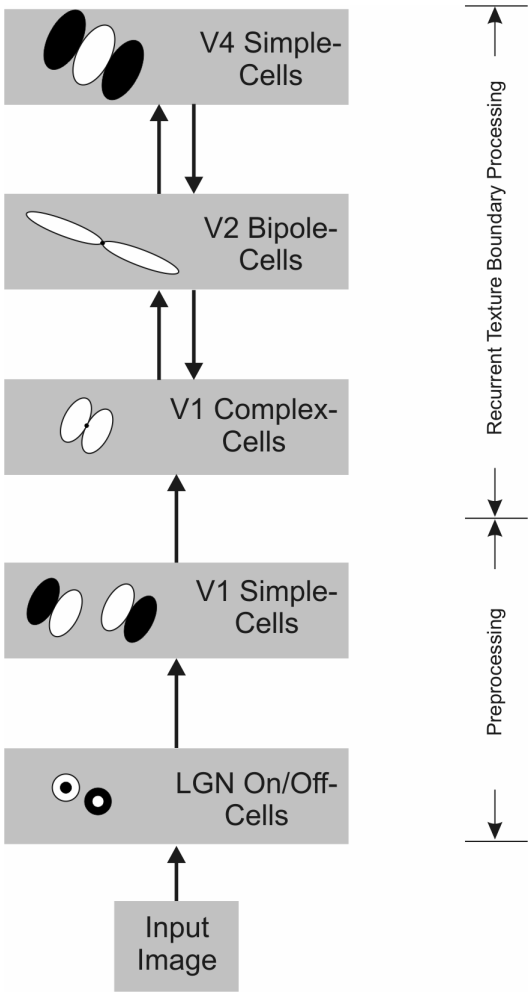


Figure 2.2: Outline of the hierarchy of model areas and model cells. Receptive field kernels are sketched in the left column (white: excitatory subfields; black: inhibitory subfields). The field sizes are not drawn to scale.

As outlined in sections 1.3 and 1.4, dense feedback projections exist between two successive areas in the hierarchy of the ventral visual pathway. Attempts to describe the functional impact of theses connections have led to the principles of recurrent interaction for response integration and cortical prediction as proposed by Grossberg (1980) and Mumford (1994), respectively, or to the reentry processes for integration and disambiguation of localized feature measurements described by Sporns et al. (1991). These principles served as motivation for the computational mechanisms utilized for the model of recurrent texture boundary processing. In particular, the model shares the key processing principles of the model suggested in (Neumann and Sepp, 1999), namely (i) the bi-directional information flow

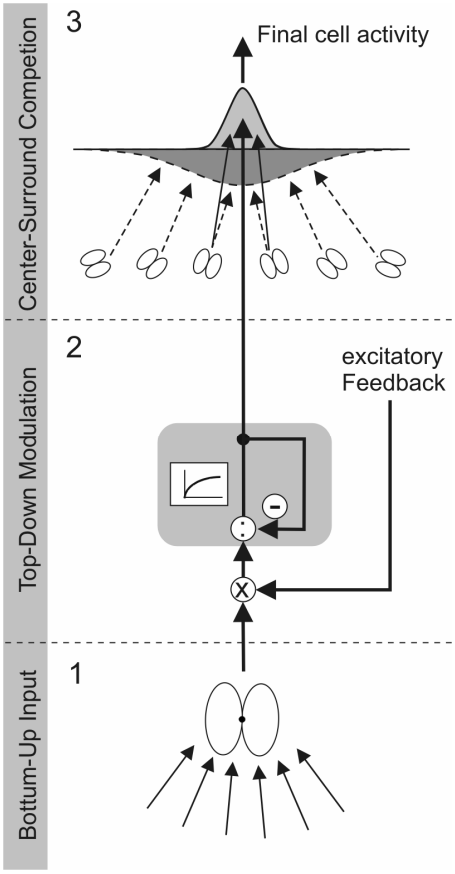


Figure 2.3: Three-stage dynamics used to determine the activation level of a cell. (1) The bottom-up input is pooled by the cell's receptive field to generate its initial activation. (2) Multiplicative feedback interaction from higher model areas modulates the initial activation. (3) The top-down modulated activity finally undergoes a stage of shunting ON-center/OFF-surround competition in a spatial and orientational neighborhood.

between cortical areas, enabling the modulation of cell responses via feedback from higher visual areas, and (ii) the intra-areal normalization of cell responses by means of competitive interaction in a spatial and orientational neighborhood.

More specifically, at each point in time, the activity level of a model cell is determined in three successive computational stages (Fig. 2.3): First, the bottom-up activity pooled within the receptive field of a cell determines its initial activation. Second, this initial activation level is modulated by excitatory feedback from the following model area. Feedback is specific to topographical position and orientation, enhancing those cell activities in the lower model area which fit to the activation pattern in the higher area. At the third computational stage, the top-down modulated cell activities undergo intra-areal center-surround competition for contrast

enhancement and normalization of activation levels. After onset of stimulus presentation, the three-stage cell dynamics is used to successively update the cell activation levels in all model areas until the overall model reaches its stable final activation pattern.

2.2 Model Areas and Receptive Field Organization

The model consists of a feed-forward preprocessing stage followed by the recurrent stage of texture border processing (Fig. 2.2). The preprocessing stage is constituted by model LGN and V1 simple cells. The combination of LGN and V1 simple cell processing results in V1 simple cell activity which signals the spatial position and local orientation of luminance discontinuities in the input image. This activity is passed on to the recurrent stage of texture border processing which is implemented by bi-directionally linked model areas V1, V2 and V4. In the following, the receptive field properties of the model cells incorporated in these areas are outlined. Please refer to Appendix A.1 for a complete description of the mathematical equations used.

- *Model LGN cells* and *model V1 simple cells* constitute the first two stages used to initially filter the luminance distribution of the input image in a feed-forward process. Model LGN ON and OFF cells detect local luminance transitions using concentric center-surround receptive fields, which are modeled by a subtractive and half-wave rectified interaction between Gaussian weighted input intensities. Model V1 simple cells have elongated ON and OFF subfields (modeled by shifted and rotated anisotropic Gaussian weighting functions) to pool the input delivered by appropriately aligned LGN cells. 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). Taken together, the first two model stages emulate roughly the functionality seen in the parvocellular layer of LGN and simple cells in V1 (Hubel and Wiesel, 1959, Hubel and Wiesel, 1962).
- *Model V1 complex cells* form the lowest level of recurrent texture border processing in our model. Model complex cells pool the activity of two 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 which 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 (Hubel and Wiesel, 1962). This output activation is subsequently modulated by excitatory top-down feedback from model area V2 and intra-areal center-surround competition, according to the three-stage model cell dynamics motivated above (please refer also to chapter 2.3 and Fig 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 complex cells. The subfield activations are combined using a soft-AND-gate, in turn only generating significant initial V2 cell responses when both fields are excited simultaneously. In accordance to the three-stage model cell dynamics, the initial V2 cell activities subsequently undergo feedback modulation by V4 activity and intra-areal center-surround competition. Model V2 bipole cells respond to luminance contrasts as well as to illusory contours, thus resembling the functional properties of contour neurons in V2 (v. d. Heydt et al., 1984, v. d. Heydt et al., 1993, Peterhans, 1997). The processing principles implemented by the model bipole cells are motivated by previous models of recurrent V1-V2 interaction for robust contour formation in early vision (Grossberg and Mingolla, 1985, Neumann and Sepp, 1999, Ross et al., 2000, Neumann and Mingolla, 2001). The computational capabilities of these approaches are incorporated and enhanced in our model.
- *Model V4 cells* represent the highest level in the model of texture border processing. The orientation of texture elements changes abruptly at region borders, causing activity transitions at corresponding topographical positions in the 2D maps of orientation selective model V1 and V2 cells. Model V4 cells measure the differences between the V2 bipole cell activities pooled by an excitatory center field and left and right inhibitory subfields (modeled as anisotropic Gaussian kernels). Consequently, they respond to gradients of orientation activity at borders while being insensitive to homogeneous

regions of same oriented items. Discontinuities in the activity distribution can occur in each input orientation field (depending on the orientation of the elements that define the texture stimulus) and can themselves have different orientations (depending on the orientation of the texture boundaries). Accordingly, a range of model V4 cells with differently oriented receptive fields was defined for each input orientation, i.e. for each of the eight V1 complex and V2 bipole cell orientations V4 cells in eight orientations exist. This results in a matrix of 8x8 = 64 V4 cells in total. All V4 cells that are selective to the same input orientation undergo center-surround competition. The activities are subsequently pooled at each position for each separate input orientation field and finally fed back to corresponding orientation channels in the preceding model areas.

2.3 Model Cell Dynamics: Overview and Stability Analysis

Cells were modeled by single-compartment units of gradual activation dynamics. Such model cells represent groups, or assemblies, of individual cortical neurons. The activation of a model cell in response to an input stimulation is determined in three successive computational stages (Fig. 2.3), namely (i) the pooling of bottom-up activity followed by (ii) activity modulation via feedback interaction and (iii) intra-areal center-surround competition. The pooling of bottom-up activity is implemented as a feed-forward process that is determined by the receptive field organization of a cell, as described in the previous section. The second and third stages of the cell dynamics are denoted by first-order differential equations, which are outlined in the following.

- The dynamics of modulatory top-down interaction (second stage) is denoted by the equation:

$$\frac{\partial}{\partial t} x_{i\theta} = -\alpha_1 \, x_{i\theta} + (\beta_1 - \delta_1 \, x_{i\theta}) \, I_{i\theta} \left[\, 1 + C \, h_{i\theta} \, \right] \tag{2.1}$$

  The input activity $I_{i\theta}$ represents the bottom-up activity from previous model areas weighted by the cells' receptive field kernels (first processing stage). Spatial locations are coded by index i, the orientation is given by θ. The input activity is multiplicatively

combined with the excitatory feedback activity $h_{i\theta}$ from higher model areas. The equation denotes a *compression function* following a first-order Naka-Rushton nonlinearity (Naka and Rushton, 1966), which can be easily observed when considering the equilibrated responses ($\partial / \partial t \, x_{i\theta} = 0$):

$$x_{i\theta} = \frac{\beta_1 \, I_{i\theta}[1 + C \, h_{i\theta}]}{\alpha_1 + \delta_1 \, I_{i\theta}[1 + C \, h_{i\theta}]} \tag{2.2}$$

The strength of the excitatory feedback activity $h_{i\theta}$ is controlled by the constant β_1 together with the gain factor C. Feedback is sensitive to spatial location and orientation. As feedback activity $h_{i\theta}$ multiplicatively enhances the initial activation $I_{i\theta}$, it is only effective at positions with non-zero initial activation. This prevents unspecific activity to spread unintentionally within the topographical map. The term $\delta_1 \, I_{i\theta}[1+C \, h_{i\theta}]$ accounts for a gradual saturation of activity $x_{i\theta}$, which is given by the ratio β_1/δ_1. The constant α_1 determines the rate of activity decay (the numerical values of the constants as utilized in the simulation studies are listed in Table A.1).

- In the third computational stage, the top-down modulated activity undergoes a process of ON-center/OFF-surround competition within the model area in order to determine the cell output activation level:

$$\frac{\partial}{\partial t} y_{i\theta} = -\alpha_2 \, y_{i\theta} \; + \beta_2 \left\{ x * \Psi^+ * \Lambda^+ \right\}_{i\theta} - \left(\delta_2 y_{i\theta} + \zeta_2 \, \varepsilon(y_{i\theta}) \right) \cdot \left\{ x * \Psi^- * \Lambda^- \right\}_{i\theta} \tag{2.3}$$

The excitatory strength of the ON-center is determined by parameter β_2. Constants ζ_2 as well as δ_2 are used to control the subtractive and divisive impact of the inhibitory OFF-surround. The term $\varepsilon(y_{i\theta})$ represents a step function which switches off the subtractive inhibition once $y_{i\theta}$ reaches 0. This prevents activity $y_{i\theta}$ from being negative. The overall structure of equation (2.3) is more easily seen in the equilibrated version:

$$y_{i\theta} = \frac{[\beta_2 \; \{ x * \Psi^+ * \Lambda^+ \}_{i\theta} \; - \; \zeta_2 \; \{ x * \Psi^- * \Lambda^- \}_{i\theta}]^+}{\alpha_2 \; + \; \delta_2 \; \{ x * \Psi^- * \Lambda^- \}_{i\theta}} \tag{2.4}$$

The equation is based on a “Mexican Hat” shape for spatial as well as orientational information. The letters Λ^+, ψ^+, Λ^- and ψ^- denote excitatory (+) and inhibitory (-) Gaussian weighting functions in space and orientation, used to integrate the top-down modulated activity $x_{i\theta}$ ($*$ is the convolution operator). The constant ζ_2 controls the strength of subtractive inhibition by surround activity given by the term $\zeta_2 \{ x * \psi^- * \Lambda^- \}_{i\theta}$. The have-wave rectification $[x]^+ = \max(x,0)$ in equation (2.4) achieves the same effect as the step function $\varepsilon(y_{i\theta})$ in equation (2.3), i.e. it prevents $y_{i\theta}$ from being negative. Shunting interaction is employed to incorporate divisive inhibition by the term $\delta_2 \{ x * \psi^- * \Lambda^- \}_{i\theta}$. Shunting interaction achieves a contrast-enhancement and normalization of the initial top-down modulated activities, resulting in an activity dependent tuning of the cells’ responsiveness.

Taken together, the three stages of computation were motivated by the physiological studies and theoretical considerations outlined in section 1.4: Feedback activity is only capable of modulating the cell activities in previous areas, but cannot drive cells without bottom-up input. Center-surround competition following the activity enhancement via feedback guarantees that the activity distribution stays focused despite the recurrent flow of activity. The model cell dynamics *normalize* the cell activities in a spatial and orientational neighborhood and the surround activity of a cell *automatically controls its gain* via shunting inhibition. Furthermore, the three stages of model cell dynamics realize a *context-selective gain enhancement* or *soft-gating mechanism*: Initial cell activities, which match the activation pattern in the next model area, are enhanced via excitatory feedback connections and in turn inhibit cells in their neighborhood via center-surround competition. This results in a facilitation of bottom-up activity matching the “expectation” of the cells in the higher model area (Grossberg, 1980, Mumford, 1994). As cells in higher areas integrate information over wide parts of the input image, the overall process of recurrent activity interaction enables a context-selective enhancement of salient texture arrangements, while at the same time spurious and perceptually irrelevant activities are suppressed.

Human visual perception exhibits a high degree of *stability* and *selectivity*, as can be seen from our own subjective experience as well as from objective results obtained by psychophysical experiments. The stability of the recognition process is, e.g. demonstrated by the fact that we will easily re-recognize the same object over and over again when it is presented several times. Furthermore, we are able to selectively detect a specific target object and distinguish it from a variety of distractor objects in a visual scene. However, when comparing the firing patterns of a single cortical neuron in response to the same stimulus which is presented repeatedly, then these patterns often exhibit a high inter-trial variability. Furthermore, single cortical neurons do not selectively and unambiguously respond to a specific input stimulus, but rather react to a broad range or class of different stimuli sharing some common properties (Kobatake and Tanaka, 1994). Consequently, a specific visual object or configuration of objects cannot be coded by the response of a single neuron, but has to be represented by the distributed response pattern of a group (or assembly) of several cortical neurons. This distributed response pattern (rather than one of the single cell responses constituting the pattern) exhibits the selectivity and stability necessary for a reliable and stable perceptual process. In the model of texture border processing, artificial neurons are utilized which represent groups, or assemblies, of individual cortical neurons having similar response selectivities. The model cell responses follow a gradual activation dynamics to signal the average firing-rate of the cortical neurons constituting these groups. The simulation results in the following sections demonstrate that the response pattern of the model of texture border processing exhibits sufficient *selectivity* to replicate a broad range of psychophysical results on human texture perception. In this section, the *stability* of the model equations is analysed in order to demonstrate that the model reaches a stable and well-defined equilibrated activation pattern in response to any input stimulus, independent of the (randomly chosen) initial values of cell activations.

As discussed above, the model consists of a hierarchy of bi-directionally linked areas which form several interconnected activity loops each enabling a recursive flow of information between two successive areas. The stability analysis of such a hierarchy of several recursive

and interconnected “sub-networks” is very demanding, and to the best of my knowledge no general solution or mathematical theorem exists which could be applied to that type of network architecture. As the comprehensive mathematical theory necessary to prove the global stability of the network is not yet available, I will characterise the dynamic behaviour of the model architecture in a step-by-step fashion. I will start with the demonstration that the model dynamics is bounded. Subsequently, the investigation focuses on the behaviour of a single layer of interconnected model cells. For this simplified architecture, the applicability of the Cohen-Grossberg theorem (Cohen and Grossberg, 1983) as well as of Hirsch’s theorem on global stability (Hirsch, 1989) is evaluated. The results serve as proof that the type of model cell dynamics employed here allow for the construction of globally stable recursive networks, at least when considering sufficiently straightforward network architectures. Finally, the general case of a hierarchy of several bi-directionally linked areas is considered and Hirsch’s theorem is utilized to derive stability conditions for a layer of model cells embedded in such a hierarchy. It is shown that the type of analysis used to demonstrate the stability of the one-layer network is still applicable in case of more complicated architectures. Consequently, the model dynamics doesn’t possess any property which would preclude a hierarchy of several model layers from being stable when selecting an appropriate set of model parameters. However, in order to enable mathematical treatability, the latter analysis crucially relies upon rather coarse and conservative approximations. The parameter range in which the stability of a hierarchy of model layers can be guaranteed based on this type of analysis is extremely restricted and the set of model parameters (listed in Appendix A.1) utilized in the simulation studies as presented in the next sections is clearly outside this range. However, the model never showed an unstable behaviour in any of the simulations. This may serve as hint that it is indeed stable even when a rigorous proof is not possible due to the current lack of an appropriate mathematical theory. It should be mentioned that the results of the stability analysis of the hierarchy of model layers are still helpful when empirically determining the model parameters as they characterize the impact of specific model parameters on the stability of the overall model. In particular, they indicate whether an

increase or decrease of a specific parameter helps to stabilize the model or rather tends to destabilise it.

In the following, I focus on a network in which the impact of top-down modulation is characterized by the equilibrated responses given by equation (2.2), p. 32:

$$x_{i\theta} = \frac{\beta_1 I_{i\theta}[1 + C h_{i\theta}]}{\alpha_1 + \delta_1 I_{i\theta}[1 + C h_{i\theta}]}$$

Using the equilibrated equation (2.2) instead of equation (2.1) reduces the number of first-order differential equations used to describe the model cell activity to one per model area, namely equation (2.3), p. 32:

$$\frac{\partial}{\partial t} y_{i\theta} = -\alpha_2 y_{i\theta} + \beta_2 \left\{ x * \Psi^{+} * \Lambda^{+} \right\}_{i\theta} - \left(\delta_2 y_{i\theta} + \zeta_2 \varepsilon(y_{i\theta}) \right) \cdot \left\{ x * \Psi^{-} * \Lambda^{-} \right\}_{i\theta}$$

In turn, this allows to test global stability of our network dynamics using the theorems of Cohen and Grossberg (1983) and Hirsch (1989), respectively. Using the equilibrated responses of equation (2.2) is equivalent to assuming that each cell combines bottom-up and top-down activity with negligible time delay compared to the horizontal interactions between neighboring cells within a model area. In contrast, the full first-order differential equation (2.1) acts as temporal low-pass filter. A physiological study by Girard et al. (2001) may help to justify this simplification. In that study, the conductance velocities of connections between V1 and V2 were investigated to see how fast cells in one area can drive recipient cells. The authors found that the distribution of delays (median velocity of 3.5 m/s) were comparable for feedforward and feedback connections. Unlike this pattern, delays of lateral horizontal responses distribute much broader yielding a median velocity of approximately 0.6 m/s. These findings led the authors conclude that the main processing sweeps (feedforward and feedback) act fast and on mainly the same temporal scale whereas lateral interactions act significantly slower.

A network using equation (2.2) instead of (2.1) will reach its final activation pattern in response to a constant input more quickly, but will behave similarly in all other aspects. However, it is important to keep in mind that this only holds as long as the time delay

introduced by the first-order differential equation isn't too long, as time delays in recurrent networks can cause instability. As discrete grids of model cells are used, the convolutions of activity x with the Gaussian kernels $\Lambda^{+/-}$, $\psi^{+/-}$ in equation (2.3) are substituted by discrete sums in the following (indices i and θ are combined to a single index for that purpose) and the constant parameters are recombined (see Appendix A.2). For stability analysis, equations (2.2) and (2.3) can then be rewritten as:

$$x_i = \frac{I_i[1+Ch_i]}{D+I_i[1+Ch_i]} \tag{2.5a}$$

$$\frac{\partial}{\partial t} y_i = -A\ y_i\ + \sum_{j=1}^{N} e_{ij} x_j\ - B y_i \sum_{j=1}^{N} d_{ij} x_j \tag{2.5b}$$

with:

$$A = \alpha_2; \quad B = \frac{\beta_1}{\delta_1}\delta_2; \quad D = \frac{\alpha_1}{\delta_1}; \quad e_{ij} = \frac{\beta_1}{\delta_1}(\beta_2\ c_{ij} - \zeta_2 \varepsilon(y_i)\ d_{ij})$$

$$c_{ij} \geq 0; \quad c_{ij} = c_{ji}; \quad \sum_{j=1}^{N} c_{ij} = 1; \quad c_{ii} = \max(c_{ij}) \quad \forall j \in [1, N]$$

$$d_{ij} \geq 0; \quad d_{ij} = d_{ji}; \quad \sum_{j=1}^{N} d_{ij} = 1; \quad d_{ii} = \max(d_{ij}) \quad \forall j \in [1, N]$$

The properties of c_{ij} and d_{ij} stem from the fact that they represent the weighting factors of the excitatory and inhibitory isotropic Gaussian kernels $\Lambda^{+/-}$, $\psi^{+/-}$. It is assumed here that a cell doesn't inhibit itself, i.e. e_{ii} >0.

2.3.1 Boundedness

As a prerequisite to the stability analysis of equations (2.5a,b) we demonstrate that the model output is bound within a certain range. We assume that (i) y_i is initialized at $t=0$ with some random value ≥ 0 representing the low spontaneous firing of the cell assembly, and that (ii) bottom-up and top-down inputs are never negative ($I_i \geq 0$, $h_i \geq 0$). Then the *lower bound* of the cell activity (eq. 2.5b) is given by 0: Once $y_i=0$ is reached, the step function $\varepsilon(y_i)$ embedded in the weights e_{ij} switches off any subtractive inhibition. Furthermore, the divisive inhibition term $\sum d_{ij} x$ is suppressed by multiplication with $y_i=0$. Consequently, y_i can never

fall below 0. Additionally, it can be shown that an *upper bound* y_{max} exists for which $\partial/\partial t\, y_i < 0 \Big|_{y_i > y_{\max}}$ (Appendix A.3). This upper bound is given by

$$y_{\max} \leq \frac{e_{ii} \cdot N^+}{A + B \cdot d_{ii} \cdot N^+} \tag{2.6}$$

with N^+ being the number of positive weights $e_{ij} > 0$. When only a single layer of model cell is used (i.e. $h_i = y_i$; see Fig. 2.4A) and input I_i is restricted to the range $[0, I_{max}]$, then it can be shown that

$$y_{\max} \leq \frac{e_{ii} \cdot N^+ \cdot \dfrac{I_{\max}[1 + C \cdot y_{\max}]}{D + I_{\max}[1 + C \cdot y_{\max}]}}{A + B \cdot d_{ii} \cdot N^+ \cdot \dfrac{I_{\max}[1 + C \cdot y_{\max}]}{D + I_{\max}[1 + C \cdot y_{\max}]}} \tag{2.7}$$

Equation (2.7) may be used to iteratively determine a more accurate upper bound y_{max}, starting with the initial approximation given by equation (2.6).

2.3.2 Test for the Applicability of the Cohen-Grossberg Theorem

The Cohen-Grossberg theorem (1983) allows to prove the global stability of a wide range of different neural network types which obey the following type of activation dynamics for which a global Liapunov function is known (Grossberg, 1988):

$$\frac{\partial}{\partial t} y_i = a_i(y_i) \left[b_i(y_i) - \sum_{j=1}^{N} d_{ij} f_j(y_j) \right] \tag{2.8}$$

A network of this form is globally stable if its functions a_i, b_i and f_i as well as its coefficients d_{ij} obey the technical conditions listed in Cohen and Grossberg (1983). These conditions include symmetry and non-negativity of the coefficients d_{ij}, positivity of a_i, and non-negativity of f_j.

The Cohen-Grossberg theorem can only be applied to networks consisting of a single layer of model cells. Consequently, I restrict the following analysis to the most basic form of model equations (2.5a,b) with $h_i = y_i$. Critically, in equation (2.8), the contributions of all neighboring cells to the model cell activation level are only inhibitory. In contrast, in our model dynamics, a mixed form of excitatory and inhibitory interactions is utilized. As a consequence, the

Cohen-Grossberg theorem is only applicable to equations (2.5a,b) when the excitatory kernel contributing to the coefficients e_{ij} in equation (2.5b) is approximated by a Dirac impulse function. Rearranging the terms in equations (2.5a,b) and substituting the excitatory center $\sum_{j=1}^{N} c_{ij} x_j$ by a term restricted to self-excitation (i.e., $c_{ii}\, x_i$) leads to the form given by equation (2.8):

$$
\begin{aligned}
\frac{\partial}{\partial t} y_i &= -A\, y_i \quad + \sum_{j=1}^{N} e_{ij} x_j \qquad - B y_i \sum_{j=1}^{N} d_{ij} x_j \\
&= -A\, y_i \quad + \sum_{j=1}^{N} \left(\tfrac{\beta_1}{\delta_1} \beta_2 c_{ij} - \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i) d_{ij}\right) x_j \quad - B\, y_i \sum_{j=1}^{N} d_{ij}\, x_j \\
&= -A\, y_i \quad + \tfrac{\beta_1}{\delta_1} \beta_2 \sum_{j=1}^{N} c_{ij}\, x_j \quad - (B\, y_i + \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i)) \sum_{j=1}^{N} d_{ij}\, x_j \\
&= (B\, y_i + \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i)) \Big[- \frac{A y_i}{B\, y_i + \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i)} + \frac{\tfrac{\beta_1}{\delta_1} \beta_2}{B\, y_i + \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i)} \sum_{j=1}^{N} c_{ij}\, x_j \quad - \sum_{j=1}^{N} d_{ij}\, x_j \Big] \\
&\underset{\substack{\uparrow \\ \boldsymbol{c_{ii}=1} \\ \boldsymbol{c_{ij}=0}\ \forall \boldsymbol{i \neq j}}}{=} a_i(y_i) \Big[- \frac{A y_i}{a_i(y_i)} \quad + \frac{\beta_1 \beta_2}{\delta_1 a_i(y_i)} c_{ii}\, x_i - \sum_{j=1}^{N} d_{ij}\, x_j \Big] \quad = a_i(y_i) \Big[b_i(y_i) \; - \sum_{j=1}^{N} d_{ij} f_j(y_j) \Big]
\end{aligned}
\tag{2.9}
$$

with:

$$
\begin{aligned}
a_i(y_i) &= B\, y_i + \frac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i) \\
b_i(y_i) &= - \frac{A y_i}{a_i(y_i)} + \frac{\beta_1 \beta_2}{\delta_1} c_{ii} \frac{f_i(y_i)}{a_i(y_i)} \\
f_i(y_i) &= x_i = \frac{I_i [1 + C y_i]}{D + I_i [1 + C y_i]}
\end{aligned}
$$

As mentioned above, a rigorous proof of the global stability of the network given by equation (2.9) would include testing if functions a_i, b_i and f_i obey the technical conditions listed in Cohen and Grossberg (1983). However, the proof would only hold for the simplified network which utilizes an excitatory center restricted to self-excitation. Based on the results of numerical simulations it was initially suggested by Cohen and Grossberg that the violation of this condition and the use of more general forms of center-surround interaction would probably *not* cause instability (Cohen and Grossberg, 1983b). Cohen, however, disproved this conjecture later on (Cohen, 1988) and demonstrated that a recurrent network obeying all

conditions but the restriction to self-excitation can exhibit oscillatory behaviour. This led me to use Hirsch's theorem to prove global stability of the general form of the network described in equations (2.5a,b). However, the above results demonstrate a major difference between the shunting competitive networks of Grossberg (1988) (see chapter 1.4) and our model dynamics: Compared to Grossberg's networks we use a more generalized form of center-surround interaction in which cells in close neighbourhood can excite each other. Clearly, the latter is a more plausible description of the cell interaction in the cerebral cortex. The second main difference is our usage of a first-order Naka-Rushton nonlinearity to describe the interaction between bottom-up input and top-down activity (see equation 2.5a).

2.3.3 Test for Global Stability Using Hirsch's Theorem

Hirsch (1989) proved a powerful theorem on the global stability of networks for which no Liapunov functions are known. He considers networks which can be written in the form

$$\dot{\mathbf{y}} = \mathbf{F}(\mathbf{y}, \mathbf{I}) \tag{2.10}$$

with vector $\mathbf{y}$ denoting the network activation pattern and vector $\mathbf{I}$ denoting the constant input. The idea of the proof is to demonstrate that two trajectories starting from the initial network activation patterns $\mathbf{y}_1(0)$ and $\mathbf{y}_2(0)$ converge to the same final solution when the initial patterns are sufficiently close to each other. Hirsch demonstrated that this is the case if the Jacobian matrix $\mathbf{A}$ of $\mathbf{F}(\mathbf{y},\mathbf{I})$ has the property

$$\langle \mathbf{A}\mathbf{y}, \mathbf{y} \rangle \leq -\mu \|\mathbf{y}\|^2 \quad \forall \mathbf{y} \in \mathbf{R}^n, \mu > 0 \tag{2.11}$$

for all input vectors used ($\langle\ \rangle$ and $\|\ \|^2$ denote the inner dot product and the square of the Euclidean norm, respectively). If two initial patterns which are sufficiently close to each other converge to the same solution then this is the case for all initial activation patterns, since there is a finite sequence between $\mathbf{y}_1(0)$ and $\mathbf{y}_2(0)$ in which successive pairs of initial activation patterns are close. Consequently, a network is globally asymptotic stable if equation (2.11) holds. Equation (2.11) is transformed by Hirsch using Gerschgorin's circle theorem to finally yield the condition

$$\frac{\partial \mathrm{F}_i}{\partial y_i} + \tfrac{1}{2} \sum_{j,j \neq i} \left| \frac{\partial \mathrm{F}_i}{\partial y_j} + \frac{\partial \mathrm{F}_j}{\partial y_i} \right| \leq -\mu \tag{2.12}$$

for the global stability of networks having the form depicted in equation (2.10). Hirsch uses Taylor expansion of $\mathbf{F}(\mathbf{y},\mathbf{I})$ to break it down into the first-order derivatives and a residual capturing the higher-order terms. Equation (2.12) ensures that the first-order terms exhibit a sufficiently steep decrease with increasing y_i so that the sum of the first-order terms and the residuals (for which the summation of the $\left|\partial \mathrm{F}_i / \partial y_j + \partial \mathrm{F}_j / \partial y_i\right|$ represents an upper boundary) is negative. This results in a convergence of the two trajectories which start from neighboring points $\mathbf{y}_1(0)$ and $\mathbf{y}_2(0)$ towards a common solution.

As $\left|\partial \mathrm{F}_i / \partial y_j + \partial \mathrm{F}_j / \partial y_i\right| \leq \left|\partial \mathrm{F}_i / \partial y_j\right| + \left|\partial \mathrm{F}_j / \partial y_i\right|$ it holds that:

$$\frac{\partial \mathrm{F}_i}{\partial y_i} + \tfrac{1}{2} \sum_{j,j \neq i} \left| \frac{\partial \mathrm{F}_i}{\partial y_j} + \frac{\partial \mathrm{F}_j}{\partial y_i} \right| \leq \frac{\partial \mathrm{F}_i}{\partial y_i} + \tfrac{1}{2} \sum_{j,j \neq i} \left(\left| \frac{\partial \mathrm{F}_i}{\partial y_j} \right| + \left| \frac{\partial \mathrm{F}_j}{\partial y_i} \right| \right) \tag{2.13a}$$

Consequently, the global stability of a network can also be proven by demonstrating

$$\frac{\partial \mathrm{F}_i}{\partial y_i} + \tfrac{1}{2} \sum_{j,j \neq i} \left(\left| \frac{\partial \mathrm{F}_i}{\partial y_j} \right| + \left| \frac{\partial \mathrm{F}_j}{\partial y_i} \right| \right) \leq -\mu \tag{2.13b}$$

which is sometimes more convenient. In case of a single layer of model cells with $h_i = y_i$ (Fig. 2.4A), substituting equation (2.5a) into equation (2.5b) and taking the partial derivatives of equation (2.5b) with respect to y_i as well as y_j yields

$$\frac{\partial F_i}{\partial y_i} = -A + (e_{ii} - B\, d_{ii}\, y_i) \frac{D C I_i}{(D + I_i[1 + C y_i])^2} - B \sum_{j=1}^{N} d_{ij} \frac{I_j[1 + C y_j]}{D + I_j[1 + C y_j]} \tag{2.14a}$$

and

$$\left| \frac{\partial F_i}{\partial y_j} \right| = \left| (e_{ij} - B\, d_{ij}\, y_i) \frac{D C I_j}{(D + I_j[1 + C y_j])^2} \right|, \tag{2.14b}$$

respectively. In the following, an upper bound of $\partial \mathrm{F}_i / \partial y_i$ is established by substituting y_i and the y_j using their minimal values $y_{min} = 0$:

$$\frac{\partial F_i}{\partial y_i} \leq -A + e_{ii} \frac{D C I_i}{(D + I_i)^2} - B \sum_{j=1}^{N} d_{ij} \frac{I_j}{D + I_j} \tag{2.15a}$$

Using the minimal and maximal values y_{min} and y_{max} (eq. 2.7) also allows to derive upper bounds of the $\left|\partial \mathrm{F}_i / \partial y_j\right|$:

$$\left|\frac{\partial F_i}{\partial y_j}\right| \leq \begin{cases} \left(|e_{ij}| + B\, d_{ij}\, y_{\max}\right) \dfrac{DC\, I_j}{(D + I_j)^2} & \forall\, e_{ij} \leq 0 \\ \max\left\{e_{ij},\, B\, d_{ij}\, y_{\max} - e_{ij}\right\} \cdot \dfrac{DC\, I_j}{(D + I_j)^2} & \forall\, e_{ij} > 0 \end{cases} \tag{2.15b}$$

Equations for upper bounds of the $\left|\partial \mathrm{F}_j / \partial y_i\right|$ can be easily obtained by exchanging indices *i* and *j* in equation (2.15b). In the following, I will use the right-hand sides of equations (2.15a) and (2.15b) to approximate $\partial \mathrm{F}_i / \partial y_i$ as well as the $\left|\partial \mathrm{F}_j / \partial y_i\right|$ and $\left|\partial \mathrm{F}_i / \partial y_j\right|$. By rearranging the terms in equations (2.15a,b), the sum of the derivatives in equation (2.13b) can thereby be approximated by a sum of functions which only depend on $I_1, \ldots, I_i, \ldots, I_n$, respectively:

$$\frac{\partial \mathrm{F}_i}{\partial y_i} + \frac{1}{2} \sum_{j, j \neq i} \left(\left|\frac{\partial \mathrm{F}_i}{\partial y_j}\right| + \left|\frac{\partial \mathrm{F}_j}{\partial y_i}\right| \right) \leq f(I_i) + \sum_{j, j \neq i} g_j(I_j) \tag{2.16}$$

Global stability of the network is then demonstrated by showing that

$$f(I_i) + \sum_{j, j \neq i} g_j(I_j) \leq -\mu \tag{2.17}$$

with:

$$\begin{aligned} f(I_i) &= -A + e_{ii} \frac{DC\, I_i}{(D + I_i)^2} - B\, d_{ii} \frac{I_i}{D + I_i} + \tfrac{1}{2} \sum_{j, j \neq i} \left|\frac{\partial F_j}{\partial y_i}\right| \\ &= -A + e_{ii} \frac{DC\, I_i}{(D + I_i)^2} - B\, d_{ii} \frac{I_i}{D + I_i} + K \frac{DCI_i}{(D + I_i)^2} \\ &= -A + \frac{L I_i - B d_{ii} I_i^2}{(D + I_i)^2} \end{aligned} \tag{2.18a}$$

$$\begin{aligned} g_j(I_j) &= -B\, d_{ij} \frac{I_j}{D + I_j} + \tfrac{1}{2} \left|\frac{\partial F_i}{\partial y_j}\right| \\ &= -B d_{ij} \frac{I_j}{D + I_j} + k_j \frac{DC I_j}{(D + I_j)^2} \\ &= \frac{l_j I_j - B d_{ij} I_j^2}{(D + I_j)^2} \end{aligned} \tag{2.18b}$$

$$K = \frac{1}{2}\sum_{j,j\neq i}\begin{cases}\left(|e_{ij}| + B\,d_{ij}\,y_{\max}\right) & \forall\, e_{ij} \le 0\\ \max\left\{e_{ij},\ B\,d_{ij}\,y_{\max} - e_{ij}\right\} & \forall\, e_{ij} > 0\end{cases} \qquad k_j = \frac{1}{2}\begin{cases}\left(|e_{ij}| + B\,d_{ij}\,y_{\max}\right) & \forall\, e_{ij} \le 0\\ \max\left\{e_{ij},\ B\,d_{ij}\,y_{\max} - e_{ij}\right\} & \forall\, e_{ij} > 0\end{cases}$$

$$L = e_{ii}\,DCI_i + K\,DC - BDd_{ii} \qquad l_j = k_j\,DC - BDd_{ij}$$

In order to ensure stability for all possible input vectors $\mathbf{I}$ with $I_j \in [0, I_{\max}]$ we have to find the input vector $\mathbf{I_{crit}}$ which maximizes the left-hand side of equation (2.17). This can be achieved by using the derivatives $df(I_i)/dI_i$ and $dg_j(I_j)/dI_j$ to calculate the inputs Inp_{j_max} which maximize the functions f and g_j, respectively:

$$Inp_{i_\max} = \frac{DL}{2BDd_{ii} + L} \qquad Inp_{j_\max} = \frac{Dl_j}{2BDd_{ij} + l_j}\ \forall\ j \in 1..N; j \neq i \tag{2.19}$$

As function g_j has exactly one maximum, it will monotonically decrease within the input range $[0, I_{max}]$ if $Inp_{j_max} < 0$. On the other hand, g_j will monotonically increase for input values $[0, I_{max}]$ if $Inp_{j_max} > 0$. Function f follows the same qualitative behaviour. Consequently, the input value I_{j_crit} which maximizes either g_j or f within the input range $[0, I_{max}]$ is given by:

$$I_{j_crit} = \begin{cases}0 & \forall\, Inp_{j_\max} < 0\\ Inp_{j_\max} & \forall\, Inp_{j_\max} \in [0, I_{\max}]\\ I_{\max} & \forall\, Inp_{j_\max} > I_{\max}\end{cases} \quad \forall j \in 1..N \tag{2.20}$$

The I_{j_crit} can then be used to calculate the maximum of the left-hand side of equation (2.17). If this maximum is negative, then the network is stable for all input vectors $\mathbf{I}$ with $I_j \in [0, I_{\max}]$.

Taken together, the following three steps are necessary to ensure global stability of a single-layer network with an input within the range $[0, I_{max}]$:

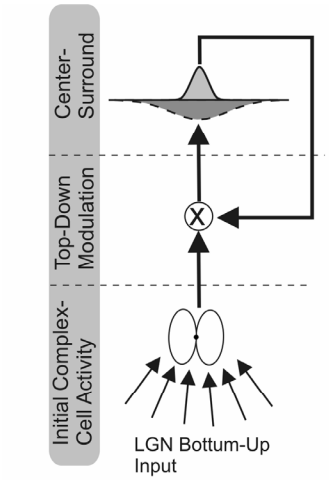
1. Calculate the maximal cell activation y_{max} using the equations (2.6) and (2.7).
2. Use y_{max} and equations (2.19) and (2.20) to determine the input vector which maximizes the left-hand side of equation (2.17).
3. Apply this input vector to the left-hand side of equation (2.17) and check if the result is negative.

The results are illustrated using a simple recursive network consisting of a feed-forward stage of complex cells to detect lines and edges in the input stimulus, followed by a layer of

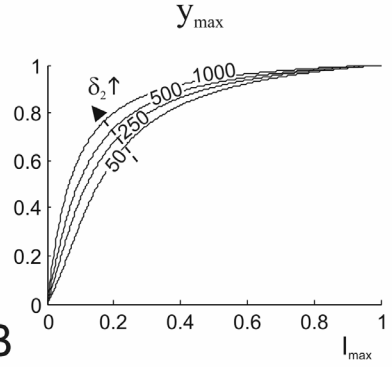
Figure 2.4 (right page): Analysis of the behaviour of a simple one-layer recurrent network. (A) Outline of the model architecture. Model LGN-, simple- and complex-cells as described in section 2.2 are used to pre-process the input image. The complex-cell activity is used as input to the recurrent processing stage. (B) The maximal complex-cell activity I_{max} is systematically varied and the maximal cell activation strength after center-surround normalization y_{max} is plotted against I_{max}. Several levels of shunting inhibition are investigated, controlled by parameter δ_2. A strong normalization of the cell activity (i.e. a high δ_2) results in a steep increase of y_{max} for low levels of I_{max}. (C) Plot of Hirsch's stability criterion against I_{max} for the levels of δ_2 used in (B). The network is stable for the range of I_{max} and δ_2 utilized, but would possibly become instable when increasing the impact of shunting inhibition further. (D) Example stimulus demonstrating the impact of different levels of shunting inhibition on the overall cell activation pattern of the network. The complex-cells signal the outlines of the bars, thereby being sensitive to the varying luminance differences between the bars and the background. Normalization of cell activity by shunting inhibition amplifies weak complex-cell activities and attenuates the differences between model cell activities signaling the borders of differently bright bars (second row).

---

recurrently connected model neurons (Fig. 2.4A; model parameters given in Table A.2). In Fig. 2.4B and 2.4C, I_{max} refers to the maximal complex cell activity used as bottom-up input to the layer of recurrently connected cells. This maximal bottom-up input is systematically increased from 0 to 1, and the resulting effect on y_{max} and the maximum of equation (2.12) is plotted. In order to study the impact of shunting inhibition on model behaviour, parameter δ_2 is systematically increased from 50 to 1000 (see eq. 2.3 and 2.4; δ_2 corresponds to parameter B in eq. 2.5b). As can be seen in Fig. 2.4B, y_{max} depends approximately linearly on the bottom-up input for low levels of I_{max}. However, for higher input levels, y_{max} is normalized to approach an upper boundary (1 in this model) due to shunting inhibition. Increasing the effect of shunting inhibition (i.e. increasing δ_2) results in a steeper increase of y_{max} for low levels of I_{max}, and y_{max} approaches the saturation level more quickly. In this case, the network amplifies weak bottom-up inputs more strongly. This is demonstrated in Fig. 2.4D. The stimulus consists of 5 bars which differ in brightness. The complex cells signal the outline of the bars, but retain the luminance differences between the bars. In the second row, the equilibrated model activation patterns after 9 iterations are shown. For a low δ_2 of 50, the

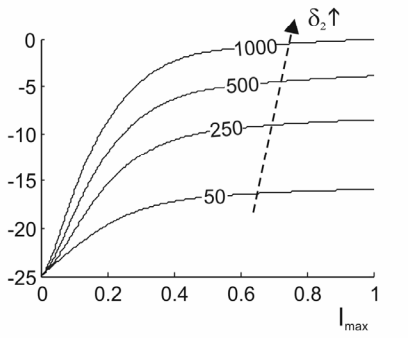


A

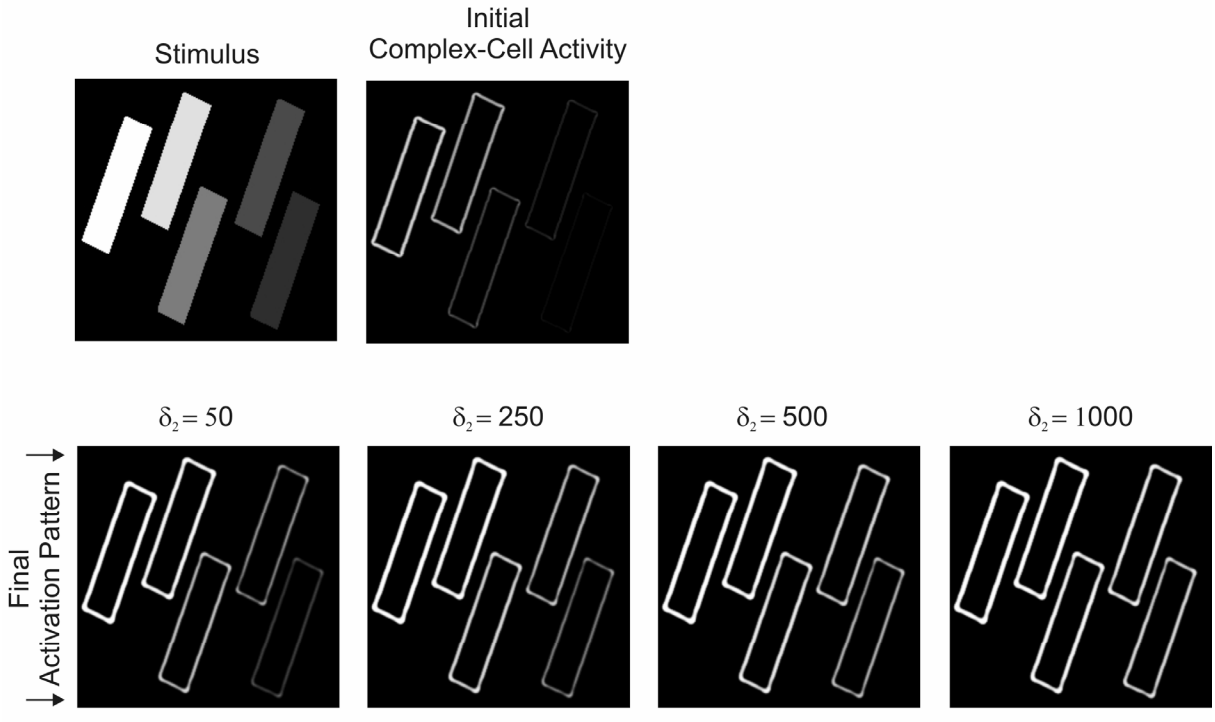


B

$$\max\left\{\frac{\partial F_i}{\partial y_i}+\frac{1}{2}\sum_{i,j\neq i}\left|\frac{\partial F_i}{\partial y_j}+\frac{\partial F_j}{\partial y_i}\right|\right\}$$



C



D

model behaves more or less like a linear feed-forward network and preserves the luminance differences. Increasing δ_2, however, results in a normalization of the model cell responses irrespective of the original input strength of the complex cells. In particular, for a δ_2 of 1000, even the outline of the darkest bar is clearly visible in the model activation pattern. In Fig. 2.4C it is demonstrated that the network is stable in the input range $I_{max} \in [0 \ldots 1]$ for all levels of δ_2. However, further increase of δ_2 would result in a putatively instable network, i.e. very strong shunting inhibition can destabilize the network.

2.3.4 Stability Analysis of a Hierarchy of Model Areas

In a recurrent model consisting of several subsequent model stages the output of a particular model area modulates the activity of the preceding area via top-down feedback, thereby exerting influence on its own input $\mathbf{I}$. Likewise, the output of a particular area influences the activity of the next area, i.e. it modulates its own feedback activity h_j. Clearly, such a situation is much more complex than a single layer of recurrently connected neurons, and beyond the scope of Hirsch's theorem. However, in case of an area embedded in a hierarchy of model areas, taking the partial derivatives of equation (2.5b) with respect to y_i and y_j allows us to determine upper bounds for the terms of Hirsch's theorem (eq. 2.12; Appendix A.4):

$$\frac{\partial F_i}{\partial y_i} \le -A + \frac{1}{D}\sum_{j=1}^{N}(|e_{ij}| + Bd_{ij}\frac{e_{ii}N^+}{A + Bd_{ii}N^+})\left([1 + Ch_{\max}]\max(|\tfrac{\partial}{\partial y_i}I_j|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_i}h_j|)\right) \tag{2.21}$$

$$\left|\frac{\partial F_i}{\partial y_j}\right| \le \frac{1}{D}\sum_{k=1}^{N}\left(|e_{ik}| + Bd_{ik}\frac{e_{ii}N^+}{A + Bd_{ii}N^+}\right)\left([1 + Ch_{\max}]\max(|\tfrac{\partial}{\partial y_j}I_k|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_j}h_k|)\right) \tag{2.22}$$

For the first area of the hierarchy, the I_k represent the external bottom-up input filtered by the cells' receptive fields ($\max(|\tfrac{\partial}{\partial y_j}I_k|) = 0$ in this case). For higher areas, the maximal output y_{max} of the preceding area together with the cells' receptive field properties determine the maximal input I_{max} to the next area. Likewise, the maximal output y_{max} of the next area is the maximal feedback activation h_{max} that the preceding area can receive.

The results rely on very coarse approximations. Furthermore, the terms $\max(|\frac{\partial}{\partial y_j} I_k|)$ and $\max(|\frac{\partial}{\partial y_j} h_k|)$ ($\forall j,k \in [1...\mathrm{N}]$) depend on the receptive field properties and model dynamics of the preceding and higher model areas and are unknown in first instance. Consequently, the above equations are only of limited use when trying to prove the global stability of the overall network. However, they allow us to determine the impact of changes of specific model parameters on the stability of the network and can thus be used to guide the empirical development of the model:

- Increase of the passive decay terms A and D decreases $\partial \mathrm{F}_i / \partial y_i$ and the $\left|\partial \mathrm{F}_j / \partial y_i\right|$, thereby stabilizing the network.
- Increase of factor C controlling the strength of feedback from higher model areas tends to destabilize the network.
- Increasing constant B which determines the strength of shunting inhibition may destabilize the network. Increasing N^+ (the number of positive weights) causes the same effect. However, the impact of B and N^+ is restricted as for large values an upper saturation level is reached (eq. 2.21 and 2.22): $B d_{ik}\, e_{ii} N^+ / (A + B d_{ii} N^+) \rightarrow e_{ii}\, d_{ik} / d_{ii}$ for $B,\, N^+ \rightarrow \infty$

Using the above results as guidelines, the constants of the equations of the model cell dynamics were empirically determined prior to the actual simulations in such a way that the whole network could reach a stable activation pattern quickly after onset of the input pattern (the constants are listed in Table A.1). This model configuration was subsequently kept unchanged through all numerical experiments and the network was well-behaved for the vast amount of different input images tested. This may serve as hint that the chosen parameter set is in effect stable, although a rigorous mathematical proof is not possible. In order to speed up processing, the differential equations 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 4-5 iterative cycles. Each simulation was continued

until iteration 9 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. In order to circumvent numerical boundary artifacts and to guarantee well-defined convolution results at image boundaries, white pixels were padded to each side of the input image prior to stimulus presentation. The number of rows and columns padded was three times the standard deviation of the widest Gaussian used. This is equivalent to presenting the stimulus on a uniform white background.

2.4 General Model Behavior

In the following, the general model behavior is exemplarily demonstrated using the final equilibrated model activation patterns in response to two selected texture stimuli, as shown in Fig. 2.5. For purposes of visualization, the activation patterns were summed up over all orientations at each spatial position and the resulting two-dimensional activity distributions were illustrated as gray-scale images, with the maximal activity of each model area coded as white.

The stimulus of Fig. 2.5A consists of a texture array composed of oriented line items in which a central vertical bar pops out, caused by an orientation contrast (OC) of 90° between the line elements of the bar and the texture background (image size 270x270 pixels). Additionally, a constant orientation shift of 20° between neighboring lines (BN, background noise) leads to apparent “flow patterns” and introduces orientation noise. Model V1 complex cells signal the outlines and orientations of the line elements of the stimulus arrangement. This V1 activity pattern serves as bottom-up input to model V2 bipole cells, which group it to form continuous smooth curves. The resulting V2 activation pattern corresponds well with the subjective impression of apparent “flow patterns” visible in the stimuli. Finally, the borders of the bar and the entire texture field are detected by model V4 cells, which respond to orientation discontinuities in the input delivered by model area V2. In the *full recurrent model* (Fig. 2.5A bottom row), the responses of model V2 cells at the borders are enhanced via modulatory feedback interaction from V4. This enhancement in turn results in stabilized and increased V4 cell activities signaling the outline of the pop-out bar. In contrast, model area V4 of the *feed-forward model* (Fig. 2.5A top row) contains significantly more spurious activations caused by the BN and the model cell activations signaling the borders of the pop-out bar are blurred and intermingled with the activations caused by orientation noise. In the recurrent model, the enhancement of V2 cell activity via feedback from V4 leads to an indirect and weaker modulation of model V1 cells in the range of 10%-20% of their mean

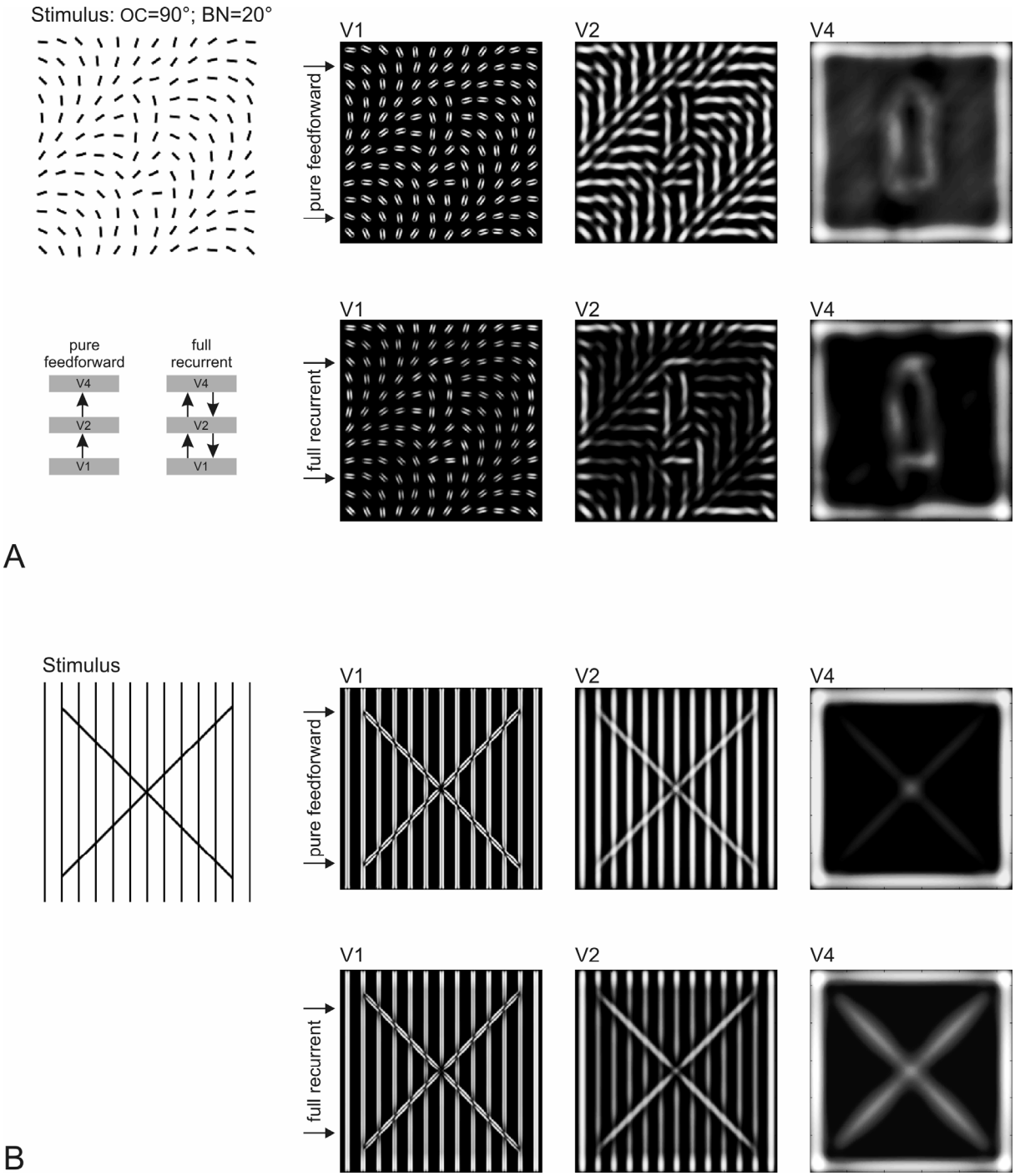


Figure 2.5: Equilibrated cell activation patterns in response to two selected stimuli, summed over all orientations for purposes of visualization. (A) The borders of the texture region and the pop-out bar are detected by model V4 cells. In the full recurrent model (bottom row), feedback from V4 results in enhanced model V2 cell responses at topographical positions of high V4 activity. Compared to the pure feed-forward model (top row), this enhancement helps to stabilize and increase V4 cell activities signaling the outline of the pop-out bar. (B) The borders of the field of vertical lines as well as the two oblique lines resembling an X are detected by model area V4. In the recurrent model (bottom row), feedback interaction helps to segregate the model V2 activities signaling the X from the activities due to the field of vertical lines. This in turn results in increased V4 responses to the X.

firing rate (this weak influence is hardly visible in the gray scale pictures; please refer to section 2.9 for quantitative results).

The stimulus of Fig. 2.5B consists of a field of regularly spaced vertical lines, superimposed by two oblique line elements representing an X (image size 270x270 pixels). The model V4 cells signal the borders of the field of vertical lines, but are insensitive to its homogeneous inner region. Additionally, they react to the orientation discontinuity caused by the X pattern. The functional impact of feedback from V4 is revealed by comparison of the activation patterns of the recurrent and feed-forward model variants: Model V2 cell activities signaling the X pattern are enhanced, helping to segregate them from the V2 activities due to regular vertical lines. The enhanced V2 activities, in turn, result in markedly increased V4 activities signaling the X pattern.

The model responses to the two selected stimuli examplarily demonstrate the main impact of feedback connections that enable a recurrent flow of activity: They help to focus the processing of the overall model on salient discontinuities, thereby enhancing and stabilizing the corresponding V4 activation patterns. A systematic investigation of the impact of feedback on model performance is subject of the next section.

2.5 Systematic Variation of Background Orientation Noise: Effect of Feedback

The goal of the following stimulation study was to characterize in detail the effect of orientation noise on the model's performance in texture boundary processing. Artificial texture stimuli were used in order to be able to systematically control and vary the orientation contrast at the border of an embedded pop-out texture region as well as the overall background orientation noise in the stimulus. In particular, the overall impact of a recurrent flow of activity was investigated by comparing the results of the full model with a pure feed-forward model. The main emphasis of this chapter is to link the model activation patterns with

the objective behavioral results on human texture processing obtained in psychophysical studies.

2.5.1 Stimuli & Analysis of Model Activation Patterns

The stimuli were adopted from psychophysical studies (Nothdurft, 1991, 1992) in order to be able to directly compare the activation patterns of our model with the results of these studies. Gray scale images of sizes 270x270 pixels containing texture arrays of 12x12 black line elements on a white background were utilized (line length 12 pixels; see Fig. 2.6A). In the central region a bar pops out, caused by a certain OC between the neighboring lines at the region border. Line positions were slightly varied to attenuate alignment effects. An additional constant orientation shift between neighboring line elements leads to apparent “flow patterns” and thus introduces orientation noise. The difficulty to detect the central pop-out bar can be controlled by means of the orientation contrast at the border and the background orientation noise (BN), allowing a systematic parametric examination of the activation pattern of the model network.

In the current set of simulations, BN was systematically varied between 0° and 30° in steps of 5° (example stimuli are depicted in Fig. 2.6A). For each value of BN, input-output curves of the model behavior in dependence on the border orientation contrast were measured. The minimal OC tested was determined by the BN (i.e. in case of BN = 15°, the minimal OC was 15°), and the maximal OC was 90°. A stepping of OC of 10° was used, except for BN values of 5°, 15° and 25°, for which the initial stepping was 5°.

Model V2 bipole cells group neighboring line elements to form continuous smooth activation patterns (see Fig. 2.5A), and also tend to group together elements of the pop-out bar and the background pattern, depending on the relative orientation of these elements to each other (for details see section 2.7). As this effect may influence the model V4 response to the pop-out bar, three different texture patterns with varying orientations of the line elements of the pop-out bar were tested for each combination of BN and OC: One texture pattern had elements which were maximally aligned with the long border of the pop-out bar (see Fig.

2.10A), the other pattern consisted of elements which were maximally non-aligned with the border and the third pattern had elements with an in-between arrangement. For further analysis of the model V4 activation pattern, the mean activation level in response to the three stimuli was calculated at each topographical position.

For each combination of BN and OC, the mean cell activation strength was determined for the V4 region corresponding to the inner pop-out bar (light gray region of model area V4 depicted in the legend of Fig. 2.6) as well as for the V4 region corresponding to the surround (darker gray region). These two values were plotted depending on OC for each value of BN. Additionally, the ratio between these two values was calculated and in turn plotted as a function of OC for each value of BN. In order to suppress high ratios resulting from very low mean surround activity levels, a constant α was added (as tonic activation) to the mean activation of the surround region.

Two model variants were considered, namely a full recurrent and a pure feed-forward model. The comparison of their activation patterns allowed to explore the functional role of the feedback connections in texture boundary processing.

2.5.2. Results

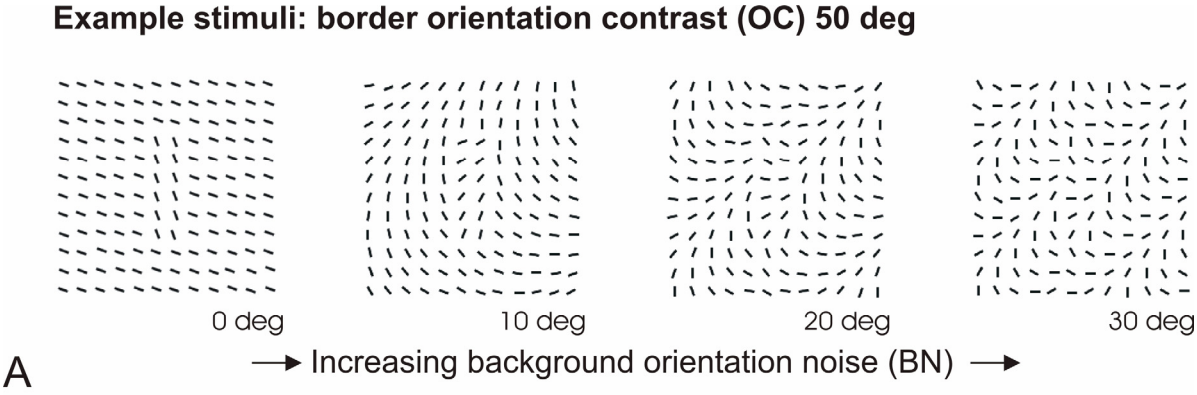
Recurrent Model

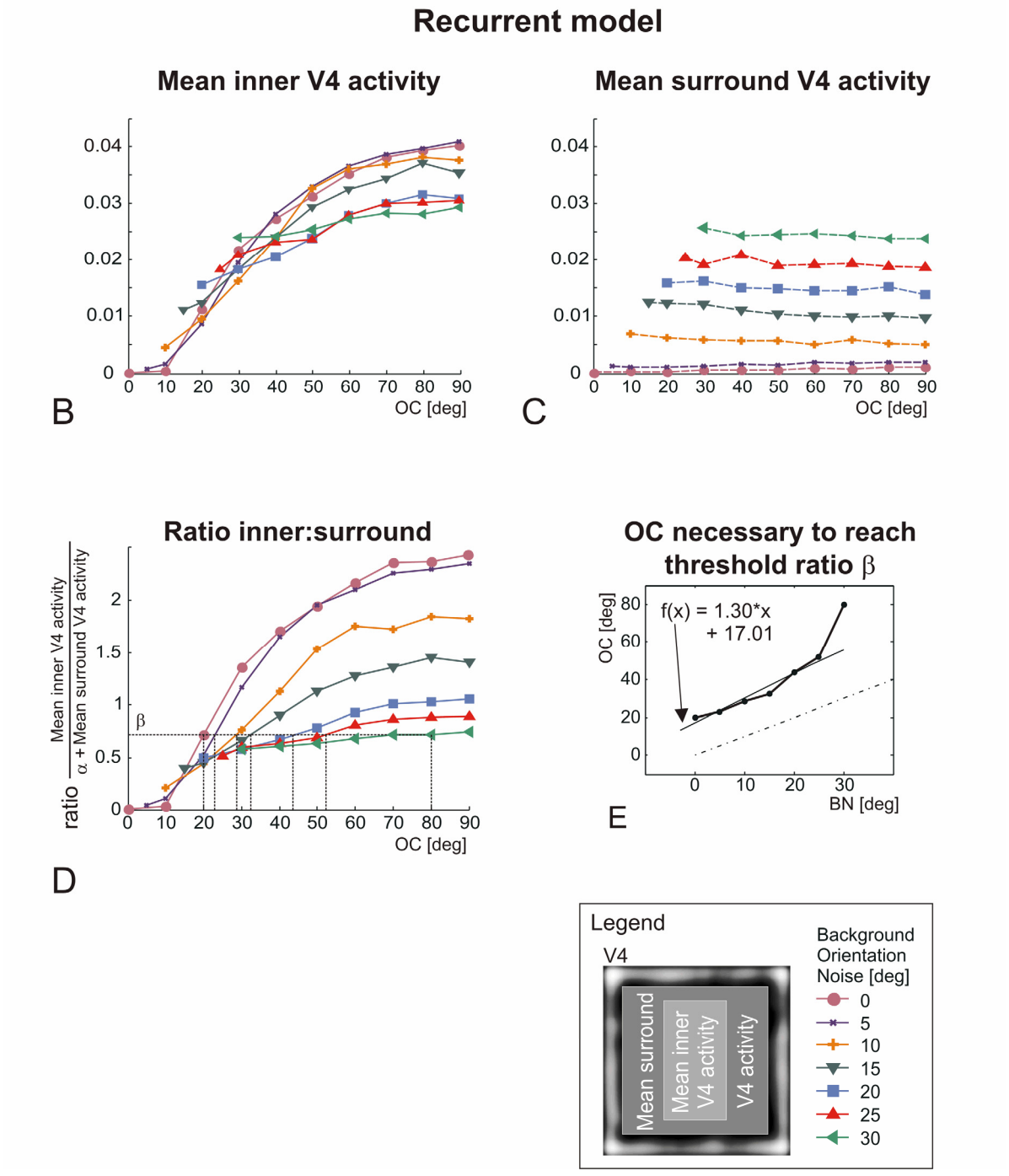
The input-output curves of the recurrent model in dependence on OC are depicted in Fig. 2.6B-E. For a BN of 0°, the mean inner V4 activity exhibits a sigmoidal dependency on OC (Fig. 2.6B: purple line with circles) with its steepest increase for OC-values from 10° to 30°. Higher values of BN result in increasingly flatter input-output curves of the inner V4 activity, indicating that the orientation noise tends to disturb the formation of model V4 activity that signals the border of the pop-out bar. In contrast to the sigmoidal curve of BN = 0°, the curve of, e.g. the highest BN value of 30° increases approximately linearly from 0.024 to 0.028 (Fig. 2.6B: green line with triangles).

Fig 2.6 (right page): Effects of a systematic variation of BN on model V4 activation patterns. (A) Example stimuli consisting of a pop-out bar defined by an OC of 50° embedded in various levels of orientation noise. The bar is visible for BN values up to 20°, but can hardly be detected for BN = 30°. (B) Mean model cell activity in the inner V4 region (marked as light gray rectangle in the legend), plotted in dependence on OC for various levels of BN. The mean inner V4 activity exhibits a sigmoidal dependency on OC for low levels of BN. However, the curves are markedly flatter for higher BN values. (C) Mean surround V4 activity (in the darker gray region in the legend), plotted in dependence on OC for various levels of BN. The level of surround activity is almost independent from OC, but increases monotonically with increasing BN. (D) Ratio between mean inner and surround V4 activity. The curves are sigmoidal for low BN values, but get flatter with increasing BN. The OC necessary to reach the reference ratio β increases with increasing BN. (E) Threshold OC necessary for reference ratio β in dependence on BN. The curve is almost linear for BN values up to 25°. However, a very high OC is necessary for BN = 30°, indicating that the model's ability to detect the bar gradually breaks down for orientation noise exceeding this level.

In Fig. 2.6C the dependencies of the mean V4 surround activity on OC are depicted. For a specific value of BN, the surround activity remains at a more or less constant level independent of OC. This level exhibits a monotonic increase with increasing BN. In particular, for a BN of 30°, the mean inner V4 activity is only moderately enhanced above the surround activity level even for high values of OC (green lines with triangles in Fig. 2.6B&C). Consequently, strong background orientation noise results in an unspecific activation of the model V4 cells, which in turn prevents the formation of V4 activity to signal the border of the pop-out bar.

The ratios between the mean inner and surround V4 activities are plotted in Fig. 2.6D. One can assume here that the texture border is detected by the model when the ratio between inner and surround V4 activity exceeds a given threshold value β. Additionally, a constant (or tonic activation) α is added to the surround activity level. This prevents unspecific noise-related V4 activations in the inner region from causing high ratios in the case of low surround activity levels. For vanishing background orientation noise, the mean inner V4 activity significantly exceeds zero for OC values greater than 20°. In consequence, the ratio for BN=0° and OC= 20° was taken as reference threshold value β (indicated as horizontal


Example stimuli: border orientation contrast (OC) 50 deg
0 deg
10 deg
20 deg
30 deg
A
Increasing background orientation noise (BN)


Recurrent model
Mean inner V4 activity
Mean surround V4 activity
0.04
0.03
0.02
0.01
0
0 10 20 30 40 50 60 70 80 90
OC [deg]
B
C
Ratio inner:surround
ratio
Mean inner V4 activity
α + Mean surround V4 activity
β
2
1.5
1
0.5
0
OC [deg]
D
OC necessary to reach threshold ratio β
f(x) = 1.30*x + 17.01
OC [deg]
80
60
40
20
0
0 10 20 30
BN [deg]
E
Legend
V4
Mean surround
Mean inner V4 activity
V4 activity
Background Orientation Noise [deg]
0
5
10
15
20
25
30

dashed line), and α=0.0154 was determined empirically so that the range of all curves was approximately centered around this value. Consequently, for all values of BN, a ratio greater than β indicates that the mean center activation strength significantly exceeds the mean surround activation, meaning that the model “detected” the pop-out bar.

The OC values for which the input-output curves reach the reference value β are monotonically shifted to the right with increasing BN. These threshold OC values are plotted against BN in Fig 2.6E. For BN up to 25°, the values increase approximately linearly. However, in case of BN = 30°, very high OC value are necessary to reach the reference threshold β (see green line with triangles in Fig. 2.6D).

Taken together, in our model, border orientation contrast in the input image results in V4 activity signaling the outline of the pop-out bar. The strength of this V4 activity exhibits a sigmoidal dependency on the OC in the input image. Background orientation noise, however, evokes unspecific V4 cell activations which in turn disturb the formation of V4 activity at the border of the pop-out bar. As a consequence, higher orientation contrasts are necessary to recover the models ability to detect the pop-out bar. For orientation noise exceeding approx. 30°, this ability gets gradually lost.

Feed-forward Model

In a model without feedback connections, V4 cells loose their ability to modulate cell activity in the lower model areas V1 and V2. In the following, the impact of this limitation on the model performance in texture border processing was systematically tested by repeating the above depicted simulations using a pure feed-forward model. The dependency of the mean inner V4 activity of the feed-forward model on the border orientation contrast is depicted in Fig. 2.7A. As in the recurrent model, the input-output curves exhibit sigmoidal slopes for low values of BN. However, the initial rising flanks of these curves are less steep compared to the input-output curves of the recurrent model. Moreover, the flattening of the input-output curves with increasing BN is notedly more pronounced than observed in the recurrent model.

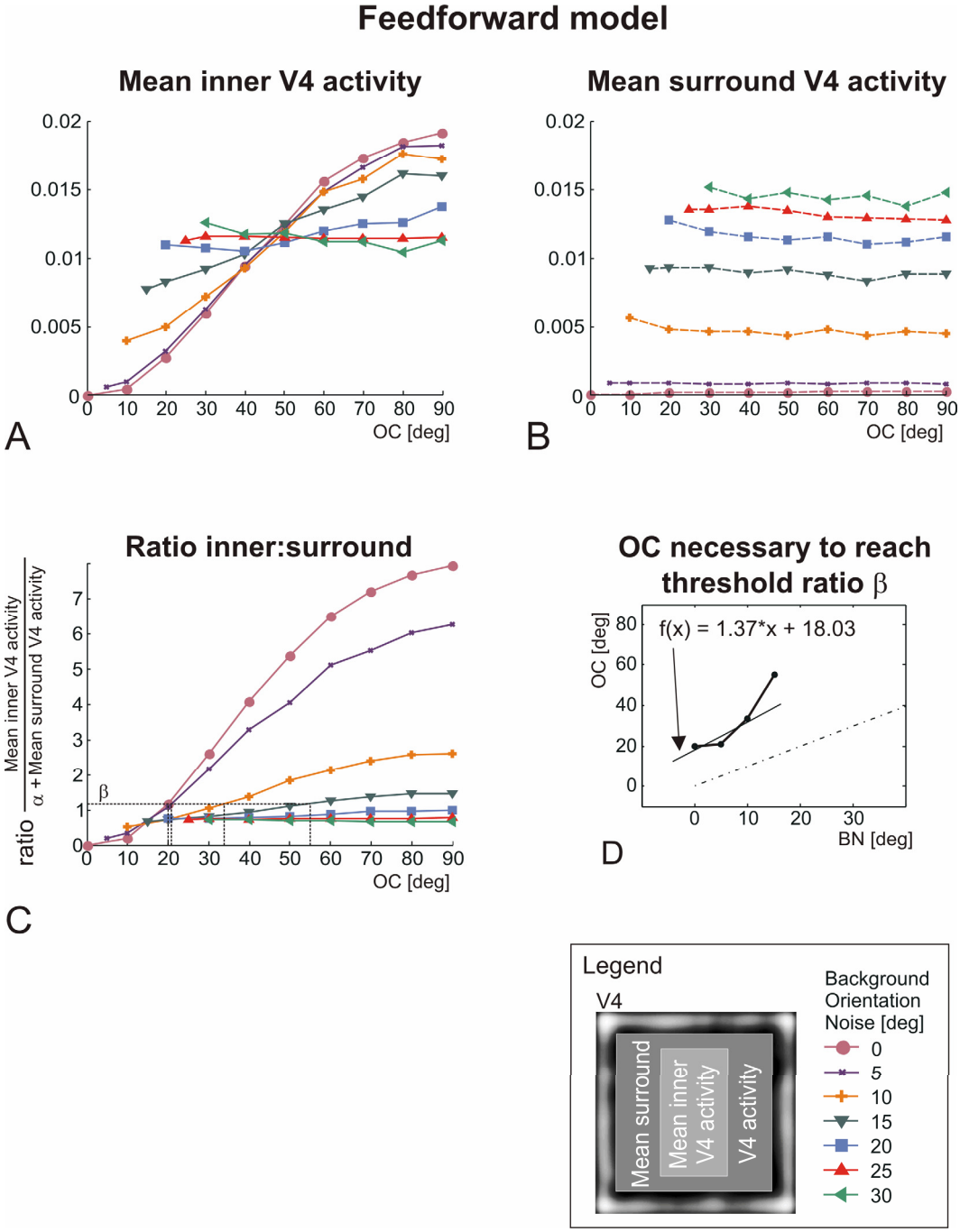


Figure 2.7: Effects of background orientation noise on V4 activation patterns of a pure feed-forward model. (A) Dependency of mean inner V4 activity (light gray region) on OC for varying levels of BN. Compared to the full recurrent model (Fig. 2.6), the curves get flat at lower values of BN. (B) Mean surround V4 activity (dark gray region) plotted in dependence on OC. It is almost independent of OC, but increases monotonically with increasing BN. For BN = 25° and BN = 30°, the surround activity is stronger than the inner V4 activity even for high values of OC. (C) Ratio between mean inner and surround V4 activity. Compared to the full recurrent model, the curves get flat at lower values of BN. Consequently, for BN values exceeding 20°, the reference ratio β is not reached even for the maximal OC of 90°. (D) Threshold OC necessary to reach reference ratio β in dependence on BN. The model's ability to detect the texture bar breaks down already for BN values exceeding 15°.

For BN values of 25° and 30°, the mean inner V4 activities do not increase at all with increasing OC (Fig 2.7A: lines with triangles pointing upwards and to the left).

In Fig. 2.7B the mean V4 surround activity of the feed-forward model is depicted. Comparably to the full recurrent model, the surround activity is more or less independent of OC for a specific value of BN and increases monotonically with increasing BN. However, for BN = 25° and 30°, the mean surround activity is always stronger than the mean inner activity for all values of OC (Fig 2.7A/B: lines with triangles pointing upwards and to the left). Clearly, the feed-forward model was unable to detect the orientation contrast at the border of the pop-out bar in these cases. Consequently, the curves depicting the ratios between inner and surround activity (Fig. 2.7C) remain approximately flat for BN values above 15°. Again, the ratio for BN=0° and OC= 20° was taken as reference threshold value β (α=0.0022). When introducing background orientation noise, this reference value was reached by the feed-forward model only for BN values up to 15° (see Fig. 2.7D).

Taken together, the behavior of the feed-forward model was comparable to that of the recurrent model for low values of BN. However, the model's ability to suppress unspecific activations evoked by orientation noise crucially depends on the modulation of cell activity in lower model areas via feedback from V4. Consequently, the performance of the feed-forward model in detecting the border of the pop-out bar breaks down for BN values above 15°. This in turn demonstrates that modulatory feedback enables the model visual system to achieve robustness of target detection over a wide range of stimulus conditions.

2.5.3. Relation to Psychophysics

The model replicates some of the key findings for human texture processing obtained in the psychophysical experiments of Nothdurft (1991, 1992). In his experiments subjects had to indicate the presence or absence of a pop-out bar while OC and BN were systematically varied. Correct response rates ranged between 50% (chance level) and 100%. His key observations were:

- For a BN of 0°, the curves indicating the number of correct responses (in %) exhibit a sigmoidal dependence on OC and approach 100% correct responses at OC = 30°. Likewise, in our model, the ratio between inner and surround V4 activity shows a sigmoidal dependence on OC with its steepest increase for OC values up to 30°.
- In the psychophysical experiments, the correct responses rates in dependence on OC are flattened with increasing BN. This is replicated by our model which has flatter ratios between the inner and surround V4 activity for higher values of BN.
- Nothdurft plotted the minimal OC necessary for a detection rate of 75% against BN. This curve increases almost linearly with increasing BN for values up to 20° - 25°. In our model, the minimal OC necessary to reach the reference ratio β exhibits an approximately linear dependence on BN for BN values up to 25°.
- Above a BN of 30°, subjects usually did not reach the 75% correct response level. Likewise, for BN = 30°, the inner V4 activity level of our model is only marginally increased above the surround activity level even for high values of OC. This indicates that the models ability to detect the orientation contrast at the border of the pop-out bar breaks down for BN values exceeding 30°.

Taken together, the V4 activation patterns of the full recurrent model exhibit exactly the same dependency on OC and BN as the detection performance of human observers, tracing back the psychophysical results to a putative neural architecture. Furthermore, based on our simulation results, we propose that suppression of orientation noise crucially depends on the feedback connections between visual areas V1, V2 and V4.

It should be noted that the initial activity patterns of the recurrent model are comparable to those of the feed-forward network. Consequently, the fast feed-forward processing suffices to segment simple and unambiguous image structures. The segregation of more complex scenes, however, crucially depends on recurrent processing mediated by the feedback connections and in turn takes more time to develop.

2.6 Functional Roles of Specific Model Connections

In the previous section it was demonstrated that feedback in general enhances the abilities of the model to successfully segregate noisy texture images. Now, the functional roles of specific model connections are investigated further. Five different model configurations are considered (see Fig. 2.8C): The full recurrent and the pure feed-forward models, variants with lesioned connections from V4 to V2 and from V2 to V1, respectively, and finally a model with additional direct bi-directional connections between V1 and V4. The last configuration is included in the study to determine the functionality of weak direct anatomical connections between human V1 and V4, which have been reported regularly (Felleman and van Essen, 1991, Nakamura et al., 1993, Rockland et al., 1994, Zilles and Clarke, 1997). The strength of these connections was chosen to be 25% of the total input strength to model areas V1 and V4, respectively.

Stimuli are texture arrays similar to those used in section 2.5 (see Fig. 2.8A). The BN is fixed at 20°, and the border OC is systematically varied between 0° and 90° in steps of 10°. Mean cell activities are determined in the inner and surround regions of model area V4. The ratio between these two values is finally plotted against OC for each model variant (Fig. 2.8B).

The comparison of the curves depicting the behavior of the full recurrent and pure feedforward variants reveals the clear impact of the recurrent flow of activity on model V4 activation patterns as already observed in the previous section (Fig. 2.8B; solid red and dotted green lines): Cycles of feedforward and feedback interaction enhance model V4 responses at the border of the texture bar and suppresses spurious activities evoked by orientation noise, resulting in higher activation ratios between the inner and surrounding regions.

Lesions of specific feedback connections help to determine their relative contributions to noise suppression via feedback modulation, as observed above. For example, the performance of a model without feedback from V4 to V2 is comparable to that of the pure

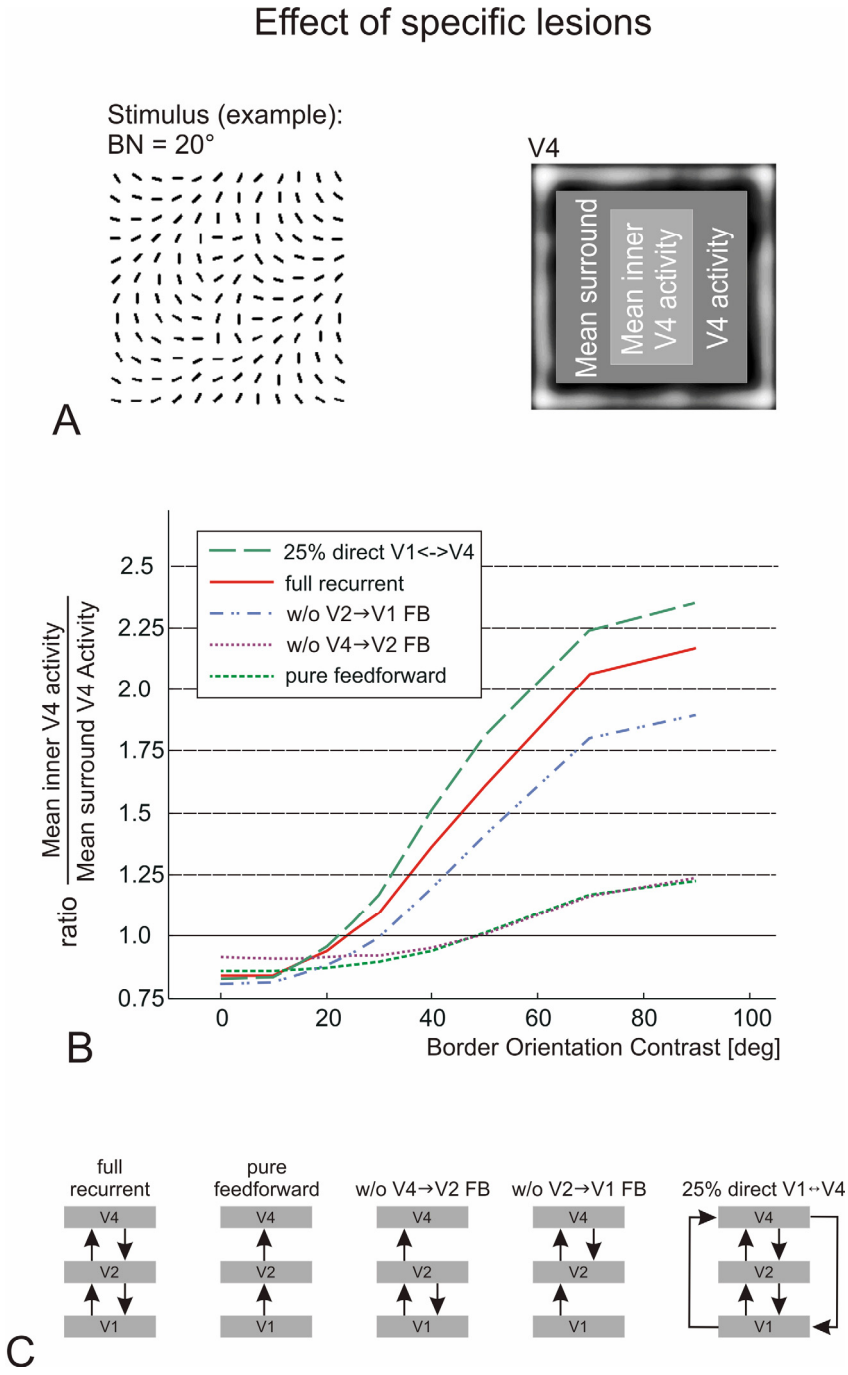


Figure 2.8: Systematic investigation of model V4 activation patterns in dependence on border OC. (A) In all simulations, BN was fixed at 20° while systematically increasing OC. The mean inner and surround activity levels in model V4 were determined for each value of OC. Their ratio was subsequently plotted against OC. (B) Dependencies of the inner-to-surround ratio on OC for the network configurations depicted in (C). High ratios crucially depend on the existence of feedback connections from model V4 to V2.

feedforward model (Fig. 2.8B; chain dotted line). In this model variant, V4 completely looses its ability to modulate cell activities in lower model areas and, in turn, is unable to enhance salient texture patterns at putative region boundaries. However, the capabilities of the network are markedly diminished even when the V4-V2 connections are left intact, but feedback from V2 to V1 is selectively lesioned (Fig. 2.8B; blue dotted and dashed line). In the full model, V2 cells with top-down modulated activity levels modulate V1 cell activity at corresponding spatial positions and in turn receive modulated bottom-up input. This effect boosts the impact of feedback activity from model area V4 on model area V2, resulting in a higher signal-to-noise ratio of the full recurrent model compared to the variant with lesioned V2-V1 connections.

Finally, direct bi-directional connections between model areas V4 and V1 in addition to the pair-wise V1-V2 and V2-V4 connectivities, respectively, result in a further increase of the inner-to-surround V4 activity ratio. This indicates that the direct information flow between model areas V1 and V4 helps to further decrease spurious V4 activations caused by orientation noise.

2.7 Functional Role of Model Area V2

In contrast to model areas V1 and V4, the functional role of model area V2 in texture boundary processing is less obvious. Model V1 cells map the input image into a distributed representation in several (in our case eight) orientation fields. Model V4 cells detect discontinuities in these orientation fields and establish region boundaries along different orientations. Therefore, both areas are mandatory for the segregation of oriented textures. However, the impact of the grouping mechanisms of model area V2 on the model abilities in texture processing are more difficult to characterize. In the following, this impact is first exemplarily demonstrated using a selected input image. Subsequently, the usage of psychophysical stimuli allows us to systematically characterize the functional role of model area V2 and to directly link it to human performance assessed in behavioral studies.

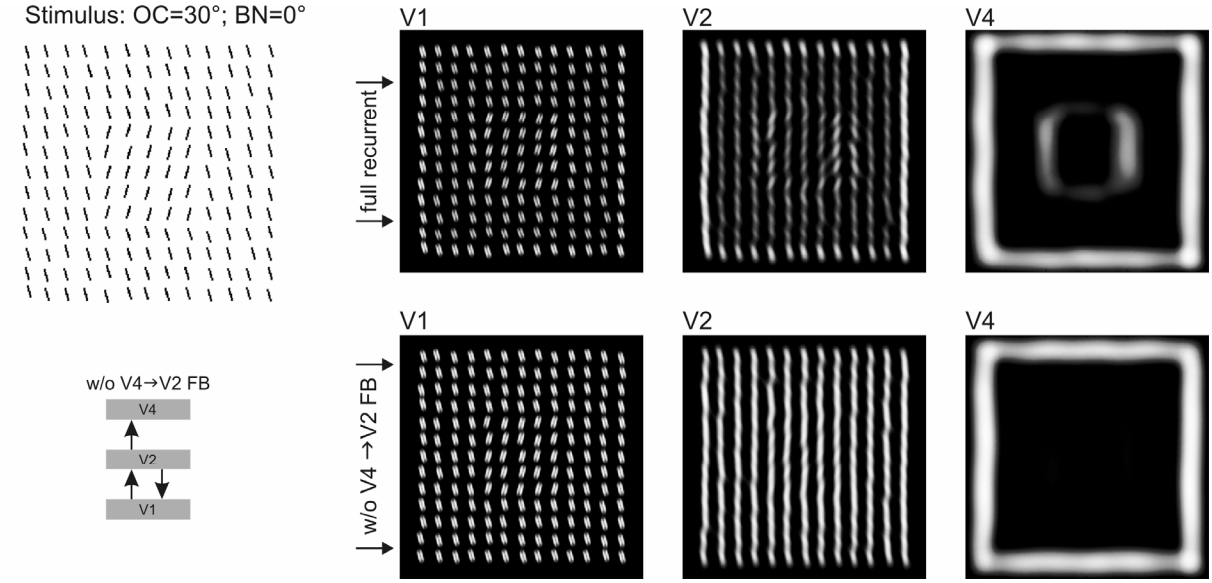


Figure 2.9: Equilibrated responses of the full recurrent network and a network without feedback from model area V4 to a central pop-out square (stimulus size 300x300 pixel). The responses in model area V4 to the horizontal borders of the square are attenuated due to alignment effects in model area V2. Without feedback from model V4, alignment leads to a complete suppression of the square region.

2.7.1 Example: Lesioning Feedback from Model Area V4 to V2

In the model, V2 bipole cells selectively enhance those V1 complex cell responses which form smooth flow lines from several adjacent line elements (see Fig. 2.5A). The influence of this alignment effect on the capabilities of the complete model in texture boundary processing can be best tested by comparing a model version with artificially lesioned V4-V2 feedback connections against the full recurrent model. After lesioning, model area V4 looses its ability to modulate the cell activity in lower areas, which in turn enables unhindered alignment of line elements by V2 bipole cells. The resulting activity patterns are exemplarily demonstrated in Fig. 2.9 using a stimulus which consists of lines which are tilted +15° in the background area and –15° in the pop out area with respect to the vertical. Model V2 cells tend to align the line elements of the background and the pop out region to vertical continuous activation patterns and thus reduce the initial orientation contrast. Model V4 cells attempt to enhance the initial contrast via feedback. After lesioning the model V4-V2 feedback connections, V4 looses its ability to preserve the contrast and the boundary

activations in V4 are no longer maintained (bottom row of Fig. 2.9). This finding is in accordance with the results of section 2.6 which demonstrate a decreased capability of the model to process stimuli containing orientation noise when the feedback from model area V4 to V2 was selectively lesioned. Likewise, direct connections between model areas V1 and V4 increase the capabilities of the overall model in texture boundary processing.

2.7.2 Relation to Psychophysics

The effects of element alignment on the human performance in texture segregation was investigated in detail in psychophysical studies (Nothdurft, 1991, 1992). In the following, we demonstrate that the results of these studies can be traced back to grouping mechanisms as mediated by area V2 in our model. For that purpose, we directly adapt the stimuli used in the study of Nothdurft (1992), which are texture arrays as already used in the previous sections. In order to study alignment effects, the background orientation noise was fixed at 10°, and the border OC was systematically varied between 10° and 60° in steps of 10°. Two sets of simulations were performed, namely one using texture patterns containing line elements which were maximally aligned with the long border of the pop-out bar, and one with texture arrays consisting of elements which were maximally non-aligned with the border. Alignment was controlled by the orientation of the line element in the gray dotted circle depicted in Fig. 2.10A, which was either vertical (aligned) or horizontal (non-aligned). For each value of OC, the mean cell activation strengths in the model V4 region corresponding to the inner pop-out bar as well as the region corresponding to the surround were determined. The ratios between these two values were plotted as a function of OC for both sets of simulations.

In general, the ratio between the inner and surround V4 activity exhibits a monotonically increasing dependency on OC (Fig. 2.10B). We can assume here that the pop-out bar is detected, when this ratio exceeds a given threshold which is significantly higher than 1 in order to gain noise robustness. For both the aligned and non-aligned condition, this is the

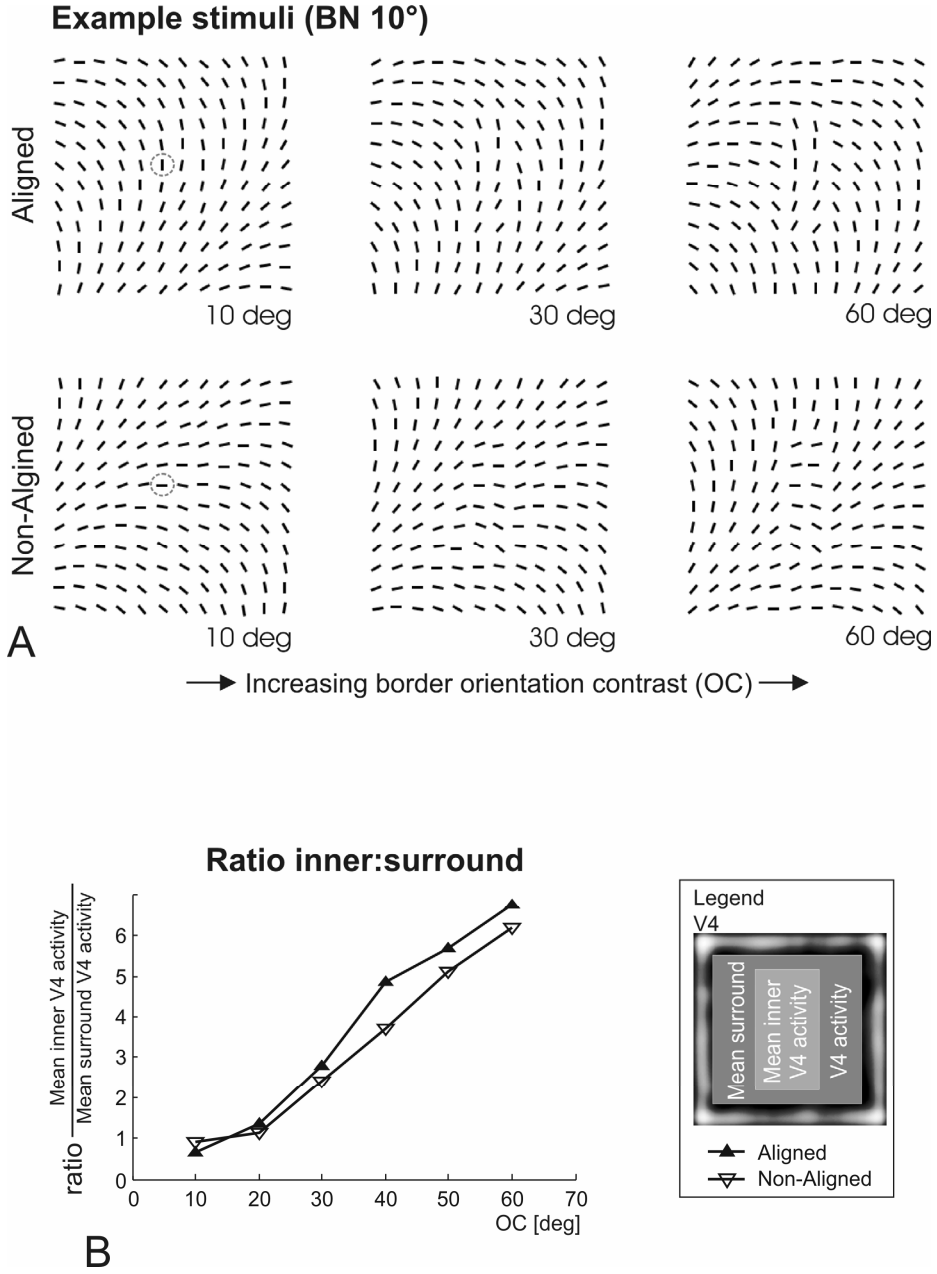


Figure 2.10: Effects of element alignment on model V4 activation strength. (A) Example stimuli of varying OC embedded in a BN of 10°. Alignment is controlled by the line element marked by the dashed circle, which is either parallel (aligned) or perpendicular (non-aligned) to the vertical border of the bar. (B) Ratios between mean inner and surround V4 activity in dependence on OC. For OC-values exceeding 20°, the model achieves higher inner-to-surround ratios for aligned stimuli compared to non-aligned stimuli.

case for OC-values greater than 30°. However, higher OC-values are necessary for non-aligned stimuli compared to aligned stimuli in order to reach the same inner-to-surround ratio.

For example, approximately the same value is achieved for aligned stimuli with an OC of 40° and for non-aligned stimuli with an OC of 50°.

We have demonstrated above that long-range groupings mediated by model V2 bipole cells constitute the neural origin of the observed alignment effect. Model V2 cells tend to align the elements of the background and the pop-out bar to continuous activation patterns and thus reduce the initial OC at the border of the bar. As a consequence, the cell activity in the inner V4 region is reduced. In the non-aligned condition, the line elements are oriented approximately perpendicular to the long borders of the bar, enabling the V2 bipole cells to act across these long borders and resulting in an attenuation of inner V4 activity. In contrast, in the aligned condition V2 bipole cells group line elements only across the short borders of the pop-out bar, which has a markedly weaker effect on model V4 activity.

Nothdurft (1992) investigated the effect of element alignment on the detectability of a pop-out bar. Subjects had to identify pop-out bars embedded either in aligned or in non-aligned stimuli. While keeping the BN fixed at 10°, detectability

- monotonically increased with increasing OC.
- was always higher in the aligned compared to the non-aligned condition for OC-values exceeding 20°.
- differed maximally between aligned and non-aligned stimuli for intermediate OC-values of 30° and 40° (weak visibility of the pop-out bar).

Nothdurft concluded that “the ability of the visual system to detect line continuations may *add* to border extraction from local contrast” (Nothdurft, 1992, p. 361). In our model, alignment effects lead to spatial long-range groupings by V2 bipole cells which in turn reduce the V4 activity at region borders in the case of a common alignment of the lines across these borders. This results in a ratio between inner and surround V4 activity exhibiting the same dependency on OC and on element alignment as the detectability measure in the psychophysical experiment. However, in our model, the ratio (i.e. the detectability) is decreased in the non-aligned condition rather than increased in the aligned condition. This indicates that the main effect of the ability of the visual system to detect line continuations is

more likely to disturb (rather than to support) the border extraction from local contrast. To conclude, the experimental findings of alignment effects that influence the detectability of a central pop-out bar can be traced back to the activity of model V2 bipole cells as the most probable neural origin.

2.8 Variations of Texture Density

In natural images the distances between the single items or sub-structures composing a texture pattern are not constant, but quite variable. In other words, the spatial frequency components usually vary between different texture patterns. On the other hand, neurons in human area V4 have receptive field sizes which increase monotonically with the distance from the fovea, but are rather constant for a given eccentricity (Smith et al., 2001). Consequently, the question arises if the ability to detect texture borders stays more or less the same across a wide range of different texture element distances (or *texture densities*), or if it is rather variable and, e.g. a systematic function of the texture density. In this chapter, the effect of variation of texture density (as studied by Nothdurft, 2000c) on model activation patterns is determined by comparing the simulation results in response to homogenous texture fields of varying density, with a central pop-out target line being either present or absent. Comparison of the simulation results with results from psychophysical studies again allows to trace back data on human performance in texture processing to the putative underlying neural mechanisms.

The stimuli were gray scale images of sizes 300x300 pixels containing texture arrays composed of oblique black lines (45°) on a white background. The line length was fixed at 8 pixels and the elements' center-to-center distances were systematically varied using spacings of {14, 17, 20, 23, 28, 35, 47, 71} pixels (see Fig. 2.11A). Line positions were slightly varied to avoid alignment effects. Two sets of simulations were performed, one with and one without a central target element defined by an orientation contrast (OC) of 90°. The

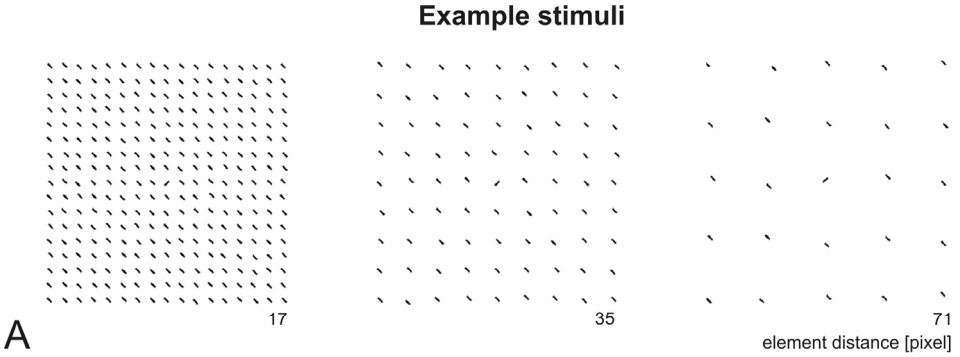
Figure 2.11 (right page): Effects of varying texture density on model V4 activation patterns. (A) Example stimuli of varying texture density containing a central pop-out line element. (B) Equilibrated V4 activation patterns in response to the example stimuli. In case of very high texture densities (element distance 17), the central pop-out line evokes low model V4 activity in the region defined by the dashed line. Medium texture densities (element distance 35) result in high V4 activity in this region and low texture densities (element distance 71) lead to slightly lower, but still clear-cut activities. However, for low texture densities, the line elements composing the texture background also evoke V4 activity blobs. (C) Equilibrated V4 activation patterns in response to stimuli equivalent to those depicted in (A), but with the central pop-out target line being replaced by an iso-oriented background line. The mean inner V4 activity in the region defined by the dashed line increases monotonically with decreasing texture density. (D) Mean inner V4 activity in response to a target line (left) or to an iso-oriented background line (right), plotted in dependence on texture density. (E) The difference between the mean inner V4 activities of the target and non-target trials reveals the amount of V4 activity which stems specifically from the orientation contrast of the pop-out line.

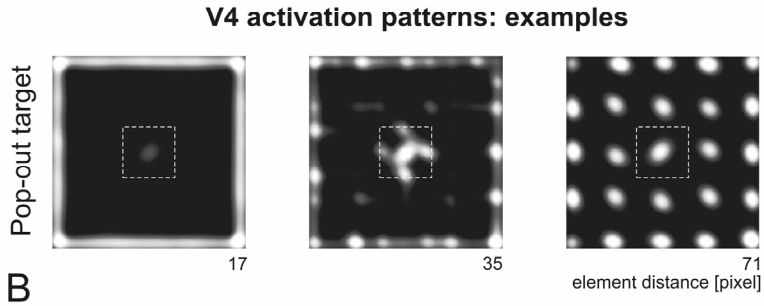
mean inner V4 activity (Fig. 2.11B&C: region defined by the dashed white lines) was determined and plotted against element distance for both sets of simulations.

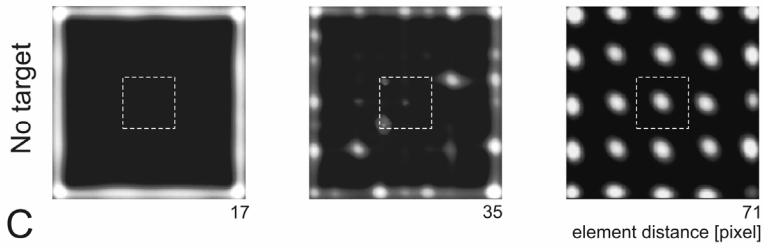
2.8.1 Simulation Results

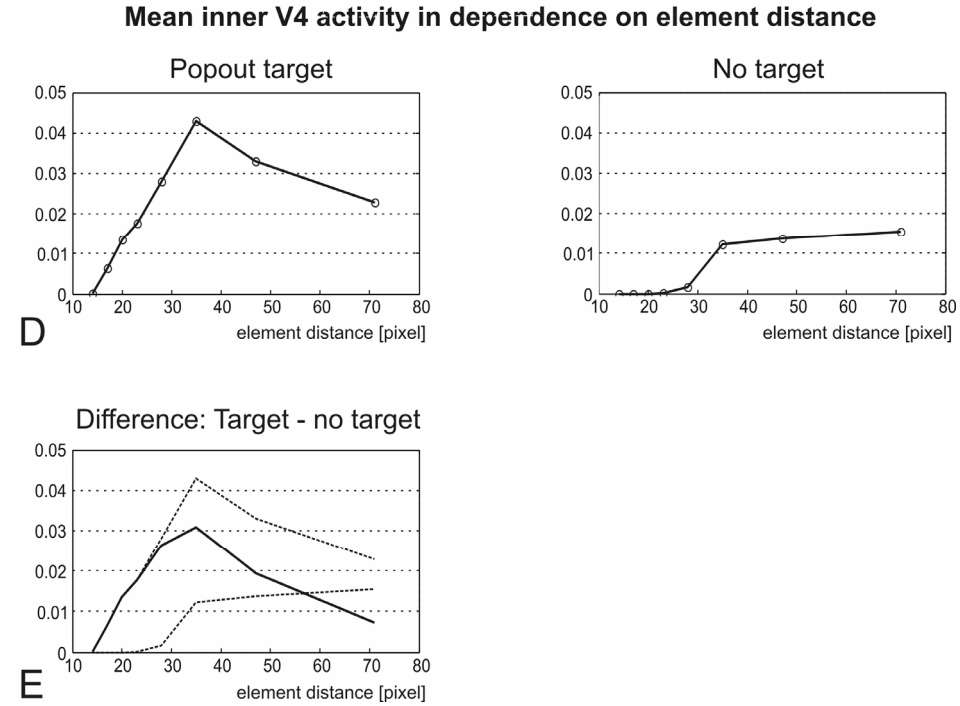
Simulations with Target Element

For very high texture densities, the intra-areal center-surround competition between cells in the lower model areas results in attenuated V1 and V2 cell responses to all line elements, except to those at the outer texture borders. This effect also diminishes the V1 and V2 responses to the target line element. In consequence, for high texture densities, model area V4 receives only weak bottom-up activity signaling the presence of a target, in turn resulting in a low mean inner V4 activity (see element distance 17 in Fig 2.11B). The influence of center-surround competition in model areas V1 and V2 diminishes with increasing element distance, leading to a clear detection of the target line by model V4 cells. Furthermore, model V4 cells also respond to the background lines surrounding the target. This effect is caused by a partial loss of lateral inhibition by an iso-oriented line at the position of the target line. The combination of both effects results in a high mean inner V4 activity for medium texture


Example stimuli
17
35
71
element distance [pixel]
A


V4 activation patterns: examples
Pop-out target
17
35
71
element distance [pixel]
B


No target
17
35
71
element distance [pixel]
C


Mean inner V4 activity in dependence on element distance
Popout target
No target
0.05
0.04
0.03
0.02
0.01
0
10 20 30 40 50 60 70 80
element distance [pixel]
D
Difference: Target - no target
E

densities (element distance 35 in Fig. 2.11B). For low texture densities, only the V4 activity caused by the target line remains within the inner V4 region defined by the dotted lines, resulting in a decay of mean inner V4 activity (element distance 71 in Fig. 2.11B). Taken together, the mean inner V4 activity exhibits an approximate inverted U-shaped dependency of response on element distance in presence of a target line (Fig. 2.11D left). In our simulations, the size and shape of the inner V4 region was determined somewhat arbitrarily. Altering the region will rescale the curve and shift the point of the maximum. However, the general shape of the curve will remain inversely U-shaped with its maximum at medium texture densities.

Missing Target Element

In case of the target line being absent, the mean inner V4 activity monotonically increases with increasing element distance (Fig 2.11C&D right). This effect is caused by the limited spatial extend of the model V4 receptive fields, particularly of the lateral inhibitory subfields that pool the iso-oriented surround. Moreover, the impact of V4 intra-areal center-surround competition decreases with increasing element distance.

The specific effect of a pop-out target on V4 cell activation is determined by taking the difference between the mean inner V4 activities with and without a target line and plotting the resulting values against element distance (Fig. 2.11E). The resulting curve has a clear peak for medium texture densities.

2.8.2 Relation to Psychophysics

The effect of variations of texture density on the saliency of a pop-out target has been investigated in a psychophysical study (Nothdurft, 2000c). In that experiment, subjects had to rate the saliency of a target defined by an OC of 90° while systematically varying texture density. The key results were:

- The saliency of the pop-out target was low for high densities, it increased to a maximal value for medium densities and decreased again for low density values. In our model, the

mean V4 activity in the topographical region surrounding the target element exhibits the same dependency on texture density (Fig 2.11D left).

- When no pop-out target was present, the iso-oriented texture elements themselves were rated psychophysically as increasingly salient at low texture densities. Likewise, in our model, the inhibitory effect of the iso-oriented surround on the mean inner V4 activity vanishes for low texture densities (Fig. 2.11D right).
- In order to isolate the saliency caused by the orientation contrast, the difference between the saliency ratings of the target and non-target trials was determined and plotted against texture density. The resulting curve increases from low to medium element distances and decreases afterwards to approach zero at high distance values. In our simulations, the curve depicting the difference between the mean inner model V4 activities in response to the target and non-target stimuli exhibited the same general dependency on texture density (Fig. 2.11E).

Using the results of our simulations, the psychophysical data on the effects of texture density (Nothdurft, 2000c) may be interpreted as follows: Let us assume that the V4 activation pattern is “read out” by a decision stage using spatial receptive fields with some kind of center-surround organization. Then the mean inner V4 activity is a gradual measure, or signature, of the “saliency” of the topographical region corresponding to the pop-out target: The higher the mean activity in this region, the higher the probability that the decision stage rates the underlying input image as containing a “salient” element.

2.9 Modulation of V1 Responses by Feedback Activity

Electrophysiological studies indicate that cells in V1 exhibit increased firing-rates when texture borders fall on their receptive fields (Gallant et al., 1995, Lamme et al., 1998). As the receptive fields of V1 cells are small and cannot integrate information over wide ranges in the visual field, feedback from higher visual areas has been proposed to be a putative source of the observed modulatory effects. In the following, the effect of indirect feedback from model

area V4 (via V2) on model V1 cell activity is demonstrated. Furthermore, the impact of weak direct bi-directional connections between V1 and V4 on V1 activity is investigated (Felleman and van Essen, 1991, Nakamura et al., 1993, Rockland et al., 1994, Zilles and Clarke, 1997). A stimulus containing a central pop-out square defined by a border OC of 90° was utilized (Fig. 2.12). Since gray scale pictures can only provide a coarse impression of the absolute height of cell response levels, quantitative values of V1 cell activities were displayed as follows: The mean V1 activation levels in each column of spatial locations within the region indicated by the dashed lines was calculated. This leads to activation patterns along the abscissa as shown in the bottom pictures of Fig. 2.12.

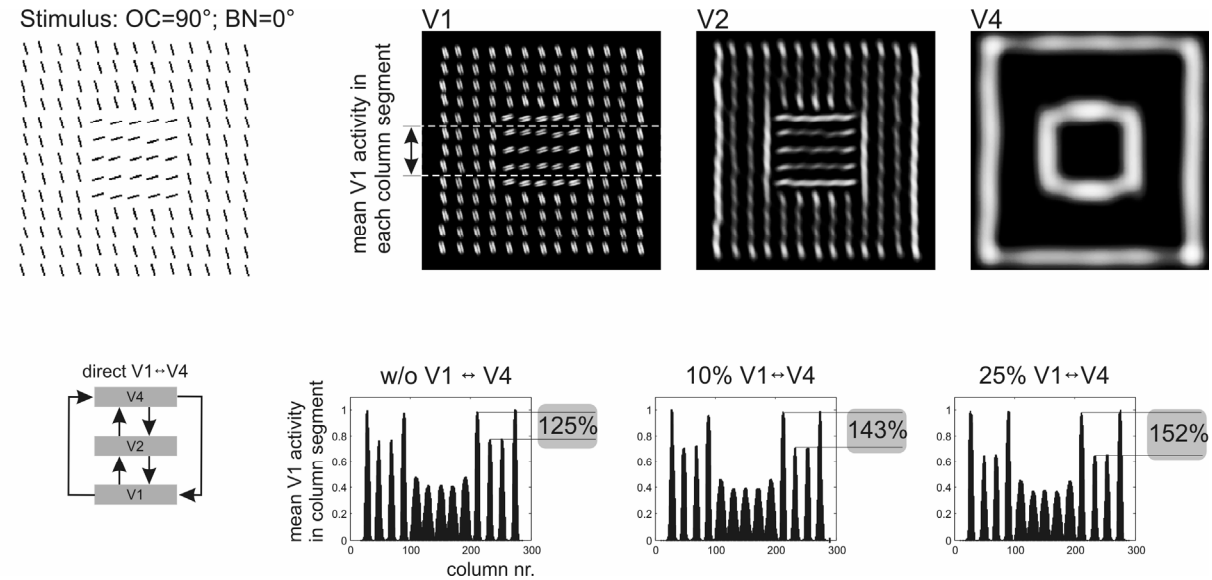


Figure 2.12: Effects of indirect feedback from model area V4 as well as of additional direct connections between V4 and V1 on V1 cell activation strength (stimulus size 300x300 pixel). Model V1 complex cell responses of each column segment in the region denoted by the dotted white lines were summed up, leading to the response profiles depicted in the three lower illustrations. The activity of the background line elements at the border of the square was compared with the cell activity evoked by the next column of lines. The V1 activity at the texture border is increased via feedback from V4. Additional direct connections between V1 and V4 significantly increase the observed border enhancement effect.

The relative strength of the bi-directional connections between model areas V1 and V4 was varied between 0%, 10% and 25% (numbers in percent of total modulatory input strength to V1 and V4, respectively) and the difference between the activities of the background line

elements at the border of the square as well as the adjacent line column was determined (Fig. 2.12). When considering the model without direct V1-V4 connections, the indirect interactions via model V2 enhance the model V1 complex cell responses by 25% at the texture square border. The direct bi-directional connections between model areas V1 and V4 further increase the responses of the complex cells at the border (additional 18% increase at 10% direct V1-V4 connection strength; 27% increase at 25% direct connection strength as compared to the model without direct V1-V4 connections). In other words, 10% direct V1-V4 interaction nearly doubles the border enhancement effect in V1 compared to the model without V1-V4 connections.

2.10 Segmentation of Real-world Images: An Example

The main goal of this modeling approach was to trace back human behavioral results on texture processing to their putatively underlying visual areas and neural mechanisms. Consequently, emphasis was placed on incorporating biologically plausible receptive field properties and cell dynamics into the model as well as using a connectivity scheme between the model areas which resembles the known anatomical connections in the human visual cortex. The usage of stimuli adapted from psychophysical studies enabled us to systematically investigate and characterize the model behavior and to directly link the model activation patterns with the psychophysical results. Achieving a high performance in segmenting real-world textures was not of high priority when developing and testing the model. Furthermore, the psychophysical stimuli used in the previous sections show rather idealized structures compared to real-world textures, making it difficult to estimate the computational capacities of the model in segmenting more complex images.

However, a series of simulations using real-world pictures containing regions of oriented textures was successfully conducted, demonstrating that the model has at least limited capabilities in segmenting also more complex pictures. Consider, for example, the responses

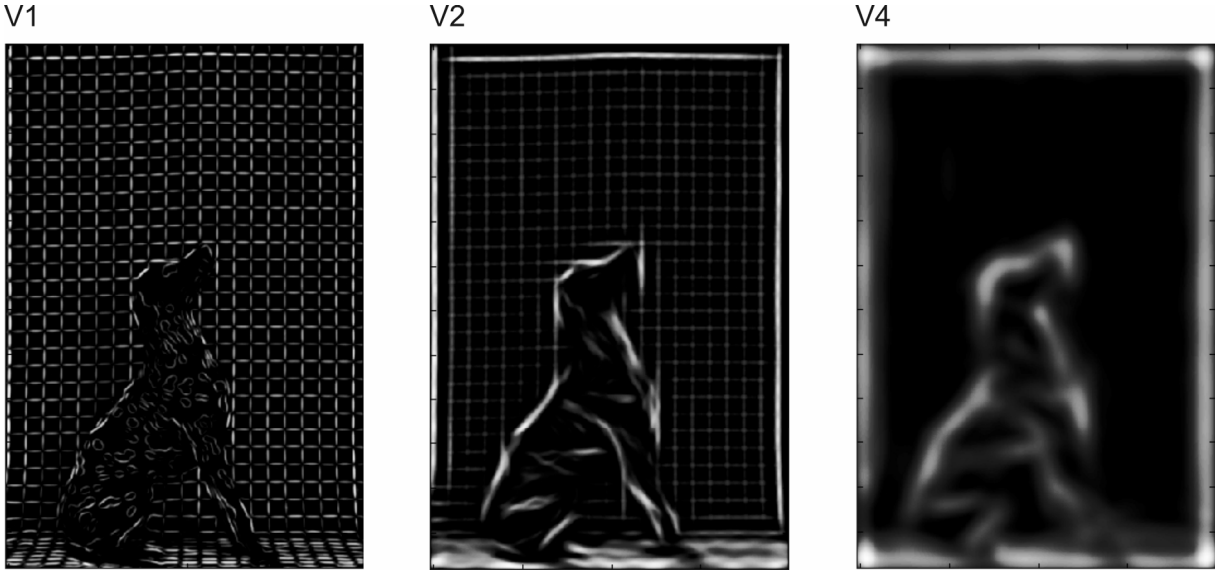


Figure 2.13: Model responses to the image of Fig. 2.1. The cells of model area V1 do not succeed in signaling a smooth continuous outline of the dog. However, model V4 cells detect the change in texture and recover major parts of the border of the dog as well as the border of the entire image. Model V2 activity at figural borders is enhanced by model V4 activity via modulatory feedback interaction.

of the full recurrent model to the Dalmatian (Fig. 2.1) in Fig. 2.13. The dog cannot be segregated from ground by detecting a continuous luminance transition to delineate its outline, but has to be identified by means of the changing texture. Consequently, model V1 cells sensitive to luminance transitions that lie at the border between the dog and the background show cluttered and fragmented activities. Model areas V2 and V4 succeed in recovering the shape of the dog in large parts by filling in the gaps of activity patterns from V1 and detect the change between the regular texture in the background and the random black dots of the dog's skin. Model V4 cells detect the border of the dog as well as the border of the entire image and in turn enhance the responses of model V2 cells at these positions via modulatory feedback interaction.

2.11 Main Discussion

Since the early works of Beck (1966) and Julesz (1965), respectively, texture processing has been intensively studied by many authors using psychophysical, physiological, and functional imaging methods that have led to a vast amount of data. In the following discussion, these experimental results will be related to the key features and simulation results of the neural model of texture boundary processing, with the goal to integrate them into a coherent computational framework. In particular, the close link between the psychophysical results of Nothdurft (1991, 1992, 2000c) and the neural model will be demonstrated, allowing to trace back many of his key results to the putatively underlying neural mechanisms. Furthermore, the physiological data on texture processing will discussed in the context of the model framework. An emphasis will be placed on the functional roles of areas V1 and V4 in texture processing, which are controversially discussed in the literature. The relation between this neural model and previous modeling approaches to human texture processing will be subject of the next section. The discussion concludes with the limitations of the model and putative future developments.

2.11.1 Key Model Features

In the neural model outlined and tested in the preceding sections, the positions of texture boundaries as well as pop-out target lines are detected via recurrent processing within a hierarchy of bi-directionally linked topographical areas. Model V4 cells use anisotropic receptive fields consisting of an excitatory center surrounded by two lateral inhibitory subfields to compare the orientation-selective input delivered by area V2. A line element evokes high V4 activity when it is *not* surrounded by approximately iso-oriented lines, independent whether it is a single target line embedded in a field of distractors or a line element belonging to the (continuous) border of a texture region. Consequently, model V4 cells gradually signal *local discontinuities* in element orientation and/or element density. The pooling of bottom-up input is followed by a process of intra-areal center-surround competition

to sharpen the initial activation patterns and to normalize cell activities in a local neighborhood. The spatial distance at which neighboring line elements exert suppressive influence on each other is restricted by the V4 receptive field sizes and the range of center-surround competition, which both act in spatial regions of approximately equal size.

The resulting model V4 activity is finally fed back to modulate the initial model V2 activation pattern, enabling a recurrent flow of activity. Both the intra-areal center-surround competition and the process of feedback modulation help to enhance V4 activations due to salient pattern arrangements while at the same time suppressing ambiguous cell activities. This results in a context-dependent retuning of a cell's sensitivity to orientation discontinuities. Center-surround competition in model areas V1 and V2 acts in a spatial and orientational neighborhood. As a consequence, a stimulus that contains only minor background orientation noise will activate only specific orientation channels. In turn, this results in a weaker impact of center-surround competition on model V1 and V2 cell activation strength and in an increased sensitivity of model V4 cells to low orientation contrasts. In contrast, noisy stimuli activate a broad range of orientation channels, which compete with each other to attenuate the V2 bottom-up input to model area V4. In this case, higher orientation contrasts are necessary to evoke model V4 activity. Excitatory feedback from model area V4 results in an orientationally and spatially specific enhancement of model V2 activity, in turn giving those V2 cells an advantage in the process of center-surround competition which signal putative "salient" element configurations. As demonstrated in the result sections, this effect focuses the processing of the overall model on important image regions, increases the V4 cells' sensitivity to salient orientation discontinuities and suppresses noise-related cell activities.

2.11.2 Relation to Psychophysics

The computational experiments presented in this paper allow to link several key results of the psychophysical studies of Nothdurft (1991, 1992, 2000c) on human texture processing with activation patterns in area V4 of our model. It was demonstrated that reading out model V4

activation patterns using simple shunting (i.e. divisive) center-surround mechanisms is sufficient to reproduce Nothdurfts findings on (i) effects of background orientation noise, (ii) alignment effects and (iii) effects of varying texture density. This enables us to trace back his results to the putative underlying neural mechanisms and to the relevant target areas.

First, it is demonstrated that the increase of orientation contrast necessary to detect a pop-out bar embedded in a background of increasing orientation noise can be accounted for by taking the ratio between inner model V4 activity (border of the bar) and surrounding V4 activity (orientation noise). The resulting quantitative salience measure resembles the behavioral curves observed in psychophysical experiments (Nothdurft, 1991, 1992). The ratio has to reach a certain threshold in order to get a clear-cut model V4 activity signaling the border of a pop-out bar and in turn enabling its detection (section 2.5). In particular, the OC necessary to reach the threshold ratio increases almost linearly for BN values up to 25°. For higher values of BN, the model's ability to detect the border of the pop-out bar breaks down even for a maximal OC of 90°, thereby resembling the performance of human observers. As already outlined in section 2.11.1, this behavior can be explained by (i) the increasing number of V1 and V2 orientation channels activated at increasing levels of orientation noise and tending to camouflage salient orientation discontinuities, and (ii) feedback from model area V4 in order to concentrate processing on "salient" image regions, thereby counteracting the effects of orientation noise for BN values up to 25°.

Second, it is shown that the effects of aligned line items, which compose a texture field, on the detectability of an embedded pop-out bar (Nothdurft, 1992, p. 360) is most likely caused by anisotropic mechanisms for contour integration and completion, as observed in contour neurons of V2 (v. d. Heydt et al., 1984). In our model, the common grouping of both line elements that belong to the pop-out bar and those of the background to continuous "flow patterns" by model V2 bipole cells tends to reduce the orientation contrast "seen" by model V4 cells (section 2.7). In particular, model V4 activity signaling the outline of a pop-out bar is notedly reduced when V2 bipole cells are able to group line elements across the long border,

but not the short border of the bar (section 2.7.2). This dependency of the “detectability” of the bar on the direction of the grouping directly replicates the results of Nothdurft (1992).

Third, it is demonstrated that the effects of texture density on the saliency of a pop-out target most likely depends on V4 receptive field sizes, the range of V4 intra-areal center-surround competition and a V4 receptive field organization of orientation selective cells having an excitatory center and lateral inhibitory surrounds (chapter 2.8). The increased saliency ratings of both pop-out and non-pop-out lines at low texture densities can be traced back to the limited receptive field sizes and range of center-surround competition. The peak saliency ratings of pop-out lines at medium texture densities depend on the receptive field organization. The simulation results on the effects of texture density also allow to interpret the findings of a recent psychophysical study (Meinecke and Donk, 2002) on visual search efficiency in texture fields of varying homogeneity. Stimulus homogeneity was controlled by replacing an increasing number of the line elements of a homogeneous texture with blank gaps. In that study, it was demonstrated that the detection of a perpendicular target line embedded in an inhomogeneous field of iso-oriented distractors follows a U-shaped dependency on the number of line elements: Detection performance is high for stimuli containing only a few lines, decreases with increasing element number and starts to increase again for high element numbers. Based on the simulation results, we suggest that pre-attentive segmentation processes are inefficient for low element numbers (i.e. low texture densities), resulting in a separate V4 activity blob for each line element. In this case, detection performance depends on the capability of subsequent attentive processes and scales inversely with the number of line elements. For high element numbers, however, pre-attentive texture border processing achieves segmentation of the stimulus and isolated V4 activity at the position of the target line enables efficient search and fast target detection (see also Thielscher and Neumann, 2005). This interpretation is supported by ERP results (Schubö et al., 2001, Schubö, 2002) which complement the psychophysical study by Meinecke and Donk (2002): Differences in posterior N2 amplitude (a brain potential specific to pre-attentive processing) between trials with and without a target line were only observed

for high element numbers. This indicates that pre-attentive processing is only efficient when texture density is sufficiently high.

Taken together, the simulation results presented here clearly support Nothdurft's view that human texture processing is based on the detection of salient feature discontinuties by demonstrating that a biologically plausible implementation of this principle is capable to replicate several of his key findings. His results can be linked to the putatively underlying neural mechanisms, target areas and key processing principles. This includes, e.g. V4 receptive field sizes and organization which determine the range of texture density effects. Furthermore, center-surround interactions and activity modulation via feedback connections is demonstrated to be crucial for the suppression of orientation noise.

2.11.3 Relation to Physiological and Neuroimaging Data:

The Functional Roles of Areas V1 and V4

In a study characterizing the responses of cells in the primary visual cortex to texture stimuli, Lamme and coworkers (1999) reported that multi-unit activity (i.e. the average activity of several neighboring cells) is enhanced at positions belonging to the surface of a central pop-out square. This indicates that cells as early as in area V1 are involved in texture processing. However, this modulation was significantly delayed relative to the onset of the cell responses and was abolished by ablation of extrastriate visual areas. In consequence, although the results demonstrate that V1 is involved in texture processing, they also show that a hierarchy of several visual areas is necessary for texture segmentation and figure-ground separation. In the following, two major stages in this hierarchy, namely areas V4 and V1, are addressed and the available physiological data on their roles in texture processing is related to the neural model of texture boundary processing.

The Role of Area V4 in Texture Segmentation

Several ablation studies demonstrate that monkeys with lesioned area V4 were severely impaired in the perception of texture-defined contours as well as illusory contours (de Weerd

et al., 1996, Merigan, 1996, Merigan, 2000). Interestingly, in the study of Merigan (2000), the V4 lesioned monkeys were still able to detect the presence of high orientation contrasts of 90° in the stimuli. However, they were no longer able to identify the orientation of a pop-out bar composed of several line elements, i.e. they could not recover the shape of the bar outlined by the texture border. The involvement of higher cortical areas in texture segregation is also demonstrated by a functional brain-imaging study of Kastner and coworkers (Kastner et al., 2000). In contrast to lower visual areas, activities in areas V4 and TEO were significantly increased during presentation of texture stimuli containing several region borders compared to homogeneous textures. It is important to note that functional brain-imaging studies that investigate the perception of illusory or motion-defined contours showed the same tendency of increased activation strength in higher visual areas as compared to lower areas (Reppas et al., 1997, Mendola et al., 1999).

The above cited studies clearly demonstrate that higher ventral visual areas (and in particular V4) are engaged in texture processing. However, these studies do not allow to identify the functional roles of the involved areas. For example, the mechanism of texture segmentation can be further subdivided with respect to two functional aspects, namely the detection of texture borders followed by the representation of texture surfaces. It is difficult to conclude from the above cited studies to which proportion a specific higher visual area contributes to one of these two subtasks. Furthermore, it remains unclear to what extent lower areas such as V1 are involved. Based on the results of our simulation study, I propose that within the hierarchy of ventral visual areas, V4 integrates the pre-processed information from areas V1 and V2 to build a coherent neural representation of surface boundaries. Based on this border representation, V4 and probably even higher areas, such as TEO, can build up a robust representation of surfaces and the background.

One of the key assumptions of the neural model, namely that cells in V4 have a receptive field layout which makes them reactive to texture borders is supported by two electrophysiological studies. In the study of Pasupathy and Connor (2001) it was demonstrated that a significant amount of V4 cells selectively responds to *boundaries* of

complex shapes defined by luminance transitions at specific positions in the stimuli. Pollen et al. (2002) investigated the interaction of bar and grating stimuli in single cell V4 responses. The key observations were that (i) individual neurons are predominantly selective to single input orientations and that (ii) RF obey a spatial center-surround organization. This indicates that V4 cells show stereotyped RF properties through selective integration of input activities from cells belonging to a single specific orientation channel. These findings support the layout of the model V4 cells which (i) integrate input activations from specific orientation fields generated at previous model stages, and (ii) use a spatial center-surround organization of the RF to generate boundary activities at texture discontinuities.

A recent electrophysiological study demonstrates that saccades during a free viewing visual search task were directed toward the topographic locations of high V4 activity (Mazer and Gallant, 2003). In half of the neurons exhibiting an increased activity level in case of being at the target location of the next saccade, this increase was determined solely by (bottom-up) stimulus features. The activity of the other half was additionally modulated by tasks demands, demonstrating a feature-based attentional modulation. The authors concluded that "the spatial distribution of activity in V4 encodes the retinotopic locations of salient features throughout the visual field" (Mazer and Gallant, 2003, p. 1248). Restricted to the orientation domain, the V4 activation patterns of the model presented here exactly replicate these findings: Model V4 activity is a gradual signal of the saliency of the underlying local orientation discontinuities in the input image, either evoked by a single target line or due to region borders. Although it was beyond the scope of the current modeling investigations, the general model architecture and dynamics clearly allow to integrate task-dependent mechanisms, as observed in the electrophysiological study. A feature-specific modulation of V4 cell activity could be achieved by biasing model V4 center-surround competition to, e.g. a specific orientation channel using modulatory feedback interactions from higher model areas which represent visual areas involved in guiding attention (Itti et al., 1998). Taken together, the modeling approach is able to link the psychophysical results of Nothdurft (1991, 1992, 2000c) on target salience and detectability, as discussed above, with the electrophysiological

results of Mazer and Gallant (2003) by proposing a putative neural implementation of salience computations with the result being represented by model V4 activation patterns.

The Role of Area V1 in Texture Segmentation

Several studies using single- or multi-unit recordings in the primary visual cortex of monkeys give evidence that V1 is involved in the processing of texture boundaries. For example, Gallant et al. (1995) reported that activity of a subsample of orientation selective neurons is significantly enhanced at the border of a bar defined by line items with a border orientation contrast of 90°. An enhancement of V1 cell activity at texture borders together with a suppression of cell activity by homogeneous texture surrounds was reconfirmed by the studies of Nothdurft et al. (1999, 2000). Likewise, Lamme and coworkers showed that multi-unit activity is strongly enhanced at the border of a square that was defined by an orientation contrast of 90° (Lamme et al., 1998, Lamme et al., 1999). Lamme and coworkers reported that the enhancement effect was still observable when substantial parts of the extrastriate cortex were ablated (Lamme et al., 1998). This lead these authors to the conclusion that cells in V1 can detect texture boundaries even without top-down input, probably based upon processes mediated by intra-areal horizontal long-range connections.

Hupé et al. reported that the response modulation in V1 is not affected by the inactivation of V2 by GABA injections, which indicates that modulation of V1 mainly depends on lateral interactions within V1 (Hupe et al., 2001). However, the stimulus used there was a single pop-out line being either parallel or perpendicular to the surrounding. Suppression instead of enhancement effects were observed as the main surround effect. It might be that the chosen stimulus is not the best configuration to study feedback via long-range connections. Other studies using similar stimuli, or examining the effect of an isolated presentation of the surround on spontaneous cell activities confirm the initial inhibitory component. However, they also report an additional late facilitatory component (Li et al., 2000, Li et al., 2001). In that paper, it was proposed by the authors that facilitation is mediated by feedback

connections from extrastriate areas, whereas horizontal connections within V1 contribute to the initial suppression.

Taken together, the above citied studies indicate that V1 appears to be involved in texture border detection and that it seems to able to signal texture borders even when feedback from higher visual areas is absent (Lamme et al., 1998). This is surprising given the rather small receptive field sizes in V1. Horizontal long-range projections within V1 have been proposed to be the likely candidates underlying the observed border enhancement effects. These projections span cortical distances of up to 8 mm and preferentially link cells having the same orientation preference in an anisotropic pattern (Gilbert and Wiesel, 1989, Hirsch and Gilbert, 1991, Kapadia et al., 1995, Kapadia et al., 2000). Computational models trying to elucidate the functional role of these connections have been suggested that are able to detect texture borders in the case of high border orientation contrasts and homogeneous background texture regions (Li, 1998, Hansen and Neumann, 1999). Hansen and Neumann showed that this circuitry of V1 long-range interaction is capable of contour enhancement and reliable corner detection, thus building a preliminary stage for perceptual grouping and figure-ground segmentation. Li's model of striate cortex also incorporates long-range connections to achieve contour enhancement. In addition, the model is able to detect the borders of simplified texture arrangements with high orientation contrasts. In most of the electrophysiological studies cited above, stimuli were employed that either consisted of a single perpendicular line element placed in an iso-oriented surround, or of texture regions with perpendicular line elements with respect to the adjacent regions. In these cases of localized high border orientation contrasts, the contour enhancement within V1 together with the partial loss of lateral inhibition by iso-oriented lines at the texture border can account for the observed border enhancement effect.

However, there are some major factors that limit the putative role of V1 cells in texture border detection. First, their spatial field of activity integration is rather small even when considering horizontal long-range connections. Cavanaugh et al. (2002a, 2002b) calculated cortical magnification factors representing the receptive field sizes of V1 cells in terms of cortical

distances in mm. They demonstrated that the majority of receptive field radii fell within a range of approximately 3 to 4 mm on the cortical map. Horizontal long-range connections in V1 span up to distances of 8 mm (Gilbert and Wiesel, 1989, Hirsch and Gilbert, 1991). In other words, the diameter of activity integration via horizontal long-range connections is (very roughly estimated) maximal 2 to 3 times the diameter of a cell's receptive field. However, this is still about 3 to 4 times smaller than the receptive field of a V4 cell (Smith et al., 2001). The second limiting factor stems from the spatial anisotropy of the V1 long-range connections. It has been demonstrated that the response of a V1 complex cell to an optimally oriented line element inside its receptive field is markedly influenced by the relative orientation and spatial position of surrounding lines (Kapadia et al., 1995, Kapadia et al., 2000). In particular, facilitation can be obtained by appropriately aligning the surrounding lines along the axis of the cell's orientation preference. In consequence, it is likely to assume that the enhancement of V1 complex cell responses at texture borders via horizontal long-range interactions is notably affected by the orientation of the line elements constituting the texture regions. However, a robust identification of texture borders crucially depends on the reliable identification of meaningful orientation discontinuities, and should be widely independent of this effect. Third, an enhanced cell firing rate in V1 is not sufficient for an unambiguous representation of region borders. Higher firing rates can be obtained by a variety of image configurations (e.g. line segments constituting a non-closed smooth contour; Kapadia et al., 1995) not necessarily representing a region border. However, an unambiguous border representation is necessary in that in can serve as starting point for figure-ground segmentation.

Taken together, I propose that horizontal long-range projections within V1 implement a powerful and multi-purpose preprocessing mechanism, e.g. to enhance the signal-to-noise ratio, which may suffice for texture border detection in those cases where high border orientation contrasts occur isolated and localized, and where background regions show no orientation shift. The reliable formation of texture boundaries in the case of background noise requires integration utilizing a wide spatial and orientation neighborhood and is thus more

likely the function of a higher cortical area. In this case, an observed border enhancement effect in V1 should be mainly due to feedback from higher areas (see section 2.9). This view is also supported by the findings of Nothdurft who investigated texture density effects not only for the orientation domain, but also other features such as luminance and motion (Nothdurft, 2000c). He observed that the density effects differed substantially with feature modality. In particular, the difference between the saliency curves of orientation- and motion-defined pop-out bars led him to suggest that “orientation and motion are processed in different subsequent areas (for example, in V4 and MT)”, but not jointly within V1.

This study exclusively focuses on the detection of texture borders by cells in higher visual areas and does not consider horizontal connections within V1. This restriction allowed to investigate the contribution of higher areas along the ventral form pathway in texture processing and to facilitate the localization of the neural origin of observed modulatory effects. Furthermore, more elaborate texture patterns could be utilized this way. It is important to note that the model of texture boundary processing is *not* incompatible to models using V1 long-range connections. Instead, the model of Hansen and Neumann (1999) shares the same kind of cell dynamics and processing principles as presented in this paper and could easily be integrated in the hierarchy of model areas, just as the V2 long-range integration adopted from Neumann and Sepp (1999). In this case, the long-range computations in model area V1 would create some kind of low-level “saliency map” (as proposed in Li, 2002), indicating the likely important image locations by an enhanced cell firing rate. This map would serve as pre-processed input for the higher model areas (in particular V4), allowing a faster detection of the region borders in case of simple input stimuli. However, the current modeling approach serves as a clear-cut examination of the likely role of higher visual areas in texture boundary processing, which is a necessary prerequisite to study the interaction of horizontal V1 long-range processing and feedback modulation from higher model areas in further investigations.

2.11.4 Other Models

The proposed modeling approach is related to two different classes of models, namely (i) models of recurrent V1-V2 interaction and (ii) models of human texture processing. Models of recurrent V1-V2 interaction simulate processes of robust contour formation observed in early vision (Grossberg and Mingolla, 1985, Neumann and Sepp, 1999, Mansson, 2000). They can reproduce the induction of illusory contours, as observed in V2 contour neurons (v. d. Heydt et al., 1984), as well as the spatial grouping of texture items and the high saliency of aligned bars. The computational capabilities of these models are preserved and incorporated in our approach, resulting in a model being in accordance with a more general architecture of form processing in the ventral visual stream. In particular, our model demonstrates how grouping of aligned contour items by V1-V2 interaction influences the capabilities of the overall model in texture boundary detection.

Many computational models which try to elucidate the neural mechanisms underlying texture segregation use the output energy of a set of filters with different spatial frequency channels (scales) and orientations for further processing such as response normalization, output combination from different channels, etc. (Bergen, 1991, Graham et al., 1993, Lee, 1995, Kehrer, 1997, Ontrup et al., 2001). Some of these models incorporate rather sophisticated higher processing stages to demonstrate similar response properties to certain types of textures corresponding with human perception. However, it still remains a matter of debate whether a biological equivalent of a bank of differently scaled filters exists in the visual system that is applied to generate a representation of spatial frequency channels from the input luminance signal. Neurons that exhibit different spatial frequency selectivities are arranged in regular cortical maps within V1 (Issa et al., 2000). However, as all of these neurons have small receptive fields, these maps are *not* equivalent to a filter bank employing spatial filters to the input image that have sizes ranging over several octaves. Instead, a combination of the output of several V1 neurons by neurons in higher visual areas would be necessary to compute the spatial low-frequency components in the input image. The above depicted modeling approaches fail to propose biologically plausible neural mechanisms for

such a computation of a spatial frequency representation distributed over several cortical areas. They directly apply multi-scale filtering to the input luminance distribution. In contrast, in the model proposed in this thesis, cells utilize filters of increasing sizes along the processing pathway in different higher-order model areas. However, neurons in higher model areas do not represent the low-frequency components in the input image but their functionality is to compute line groupings and texture borders.

Many models trying to elucidate the neural mechanisms which underlie human texture processing propose a two-stage strategy (e.g., Malik and Perona, 1990, Bergen, 1991, Landy and Bergen, 1991, Graham et al., 1993, Kehrer and Meinecke, 2003): First, the input image is analysed using a set of filters having different orientations and (in some cases) different spatial frequencies. Subsequently, a second processing stage detects activity changes in the output of the first filter bank in order to establish borders between different texture regions. For example, in the model of Malik and Perona (1990), initial V1 responses undergo a non-linear center-surround interaction computed over a neighborhood in the spatial and orientation domain, which is similar to the stage of V1 shunting competition of the model outlined here. Furthermore, in that model, these responses are then spatially blurred and gradients are measured in each orientation field separately. Finally, at each spatial location the maximum response is selected over all orientation channels. This cascade of processing stages is a simplified version of the V1-V4 feedforward path in the model presented here. Taking the gradient is a computational simplification of the functionality implemented by the model V4 cell kernels. The modeling approaches utilizing a two-stage strategy have parallels with a general scheme of texture segregation of Beck (1982). Based on the results of various psychophysical investigations, Beck conjectured the existence of several processing levels involved in texture segregation, namely detection, linking, difference encoding, and (threshold-based) decision for segmentation. The model suggested in this thesis incorporates the overall functionality of this proposed hierarchy of processing units. For example, feature detection and linking into boundary fragments is achieved by

model cells of areas V1 and V2, respectively. The functionality of difference detectors and decision units is realized by the function of model area V4.

Unlike the approach suggested here, the models cited above are not capable to explain either alignment effects or the possible functional role of feedback from higher visual areas. Unlike the selection of maximal responses in the model of Malik and Perona or Beck's proposal of a hard-coded decision stage, segmentation in the model outlined here emerges from V4 center-surround competition generating a map of graded responses and in turn enhancing compatible activity distributions at previous processing stages. The feedback connections enable a recurrent flow of activation, thereby creating more computational flexibility and circumventing simple all-or-none decisions necessary in feed-forward models to generate localized results. The activity pattern in a specific model area is modulated via feedback activity from higher areas in a context-sensitive way. This mechanism enables the overall model to focus processing on the important parts of a visual scene, and thereby gaining robustness against noise and pattern perturbations. The key contributions of feedback mechanisms are that they mediate the graded enhancement of boundaries defined by texture orientation contrast, together with the suppression of spurious activities that were generated by orientation noise. As demonstrated in section 2.5, only the model with intact feedback is able to account for the performance of human observers in the detection of texture bars.

In summary, I propose that a variety of visual areas including V1, V2 and V4 is involved in the robust determination of texture boundaries for segmentation of surface regions. It is the assumption of the full hierarchy of areas that permits to construct a biologically plausible model of texture processing. The approach outlined here preserves the capabilities of recurrent V1-V2 models in contour formation, and results in a model being in accordance with a more general architecture of the ventral visual pathway of form processing. In contrast to other models of texture processing, this approach can trace back fundamental psychophysical and electrophysiological results to the likely underlying neural architecture,

including the neural mechanisms of texture border detection, alignment effects and texture density effects.

2.11.5 Model Limitations and Putative Future Developments

The major goal of the model presented in this thesis was to achieve robust detection of borders between oriented textures. The focus on this issue allowed to employ a model of area V2 which is simplified in comparison with more elaborated models of recurrent V1-V2 processing for contour integration (Grossberg and Mingolla, 1985, Neumann and Sepp, 1999; see Neumann and Mingolla 2001 for a review of models). These approaches include more sophisticated mechanisms of bipole cell integration, which have better capabilities in completing curved contours. However, the activation patterns of the model area V2 utilized here show satisfactory alignment effects. In addition, the restriction to simpler receptive fields permits to reduce the computational complexity while keeping the same computational principles.

When testing the model with real-world images, it was revealed that the inability to segment some of theses images did mostly stem from an insufficient representation of the basic texture elements by model V1 cells. The reason is that these cells all react optimally to the same, fixed spatial frequency component. This kind of cells was sufficient for the simulations presented here, with the main emphasis to demonstrate the link between the model and psychophysical and physiological data. However, in order to improve the performance of the overall model and to allow for a more flexible and robust processing of luminance discontinuities in real-world pictures, the model V1 cells could be substituted by a set of cells having different spatial frequency selectivities. Furthermore, in order to further stabilize the V1 activation patterns and to study the interaction of texture border processing in higher visual areas and grouping mechanisms mediated by horizontal V1 long-range projections (as already outlined in section 2.11.3), the model of Hansen and Neumann (1999) could be integrated in this modeling approach.

2.12 Conclusion

The detailed neural mechanisms underlying texture segregation in the human visual system are still largely unknown and a matter of ongoing research. The goal of the first part of this thesis was to decipher these mechanisms by developing a biologically plausible computational model based on anatomical and physiological data. A hierarchical recurrent model consisting of bi-directionally connected areas V1, V2 and V4 was constructed. This model integrates electrophysiological data revealing enhanced firing-rates of V1 cells at texture borders and physiological and functional imaging data demonstrating the importance of area V4 for texture processing into a common conceptual framework. The systematic investigation of the model behavior using stimuli adapted from psychophysical studies allowed to tightly link the model activation patterns with human behavioral results on texture processing, in turn revealing the key processing mechanisms and key areas which underlie certain aspects of human texture perception. Most importantly, the functional role of the dense feedback connections between the model areas was demonstrated to be the suppression of texture noise, allowing the overall model to focus processing on important parts of the image. In particular, intact feedback connections from V4 were shown to be crucial in order to raise the model performance to the level of human observers. Furthermore, texture density effects were traced back to a putative receptive field organization in area V4 in combination with the range of V4 intra-areal center-surround competition. Finally, effects of texture element alignment on human performance in texture segregation were shown to likely result from anisotropic mechanisms of contour groupings in area V2.

Although the focus was on segmentation of oriented textures, I think that the general ideas of border detection and enhancement via model V4 cells and feedback modulation are not restricted to the orientation domain, but can be applied to other visual features as well. For example, Bergen and Adelson (1988) showed that the responses of cells with simple linear center-surround receptive fields can effectively code the size of texture elements

independent of their exact shape. The output of these cells can in turn provide the input to the model of recurrent texture processing, particularly V4 cells, in order to establish borders between texture elements of variable size. Generally speaking, the model V4 processing stage measures gradients in the activity maps of the lower model areas. Here, these activity maps represent the distribution of oriented texture items in the input picture. However, taking the idea of gradient measurements one step further, the input could also be luminance or color information as coded, e.g. by the blob-regions of area V1. In consequence, the combination of model V4 cells and feedback interaction could act as general-purpose scheme for the robust detection of salient discontinuities in the fields of pre-processed input representing the distribution of different features in the input image.

3 Globally Consistent Depth Sorting of Overlapping 2D Surfaces in a Model of Local Recurrent Interactions

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 which 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; 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 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.

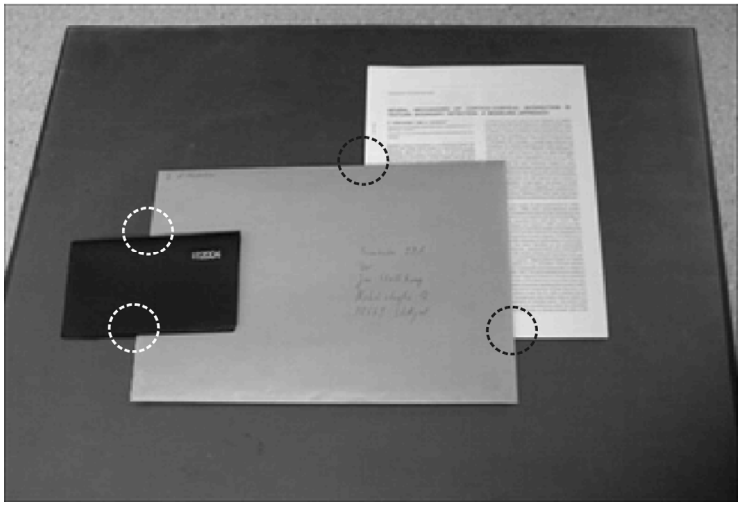


Figure 3.1: An example of overlapping real-world objects (diary, envelope, paper). T-junctions occur at those positions where the contours of two objects overlap (indicated as dotted circles). The contour corresponding to the cross-bar of the T belongs to the object being locally in the foreground and the stem of the T belongs to a surface contour that continues behind the occluding object.

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, Gonzalez and Krause, 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. 3.1) are hints allowing 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 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 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, I 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. It is demonstrated that within this framework a globally consistent depth sorting of overlapping surfaces can be obtained on the basis of local interactions. The current modeling 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. Accordingly, the chapter starts with a description of the V1-V2 model of contour processing (section 3.1), which was extended by a simple feed-forward scheme for the detection of corners and T-junctions. The model is in great parts identical to model stages V1 and V2 of the previously presented model of texture boundary processing, so the presentation is kept very brief. Then the model of recurrent depth processing is presented (section 3.2), starting with the presentation of the general scheme of depth sorting (subsection 3.2.1) and continuing with the neural model used to implement that scheme (subsections 3.2.2 – 3.2.6). Results of the simulations performed with the model are depicted in section 3.3. The presentation ends with a discussion of the model and the results in section 3.4.

3.1 Recurrent V1-V2 Contour Processing

A model of hierarchically organized visual areas LGN, V1 and V2 is used for the initial processing of surface contours (Fig 3.2A). As the model stages are in great parts identical to the corresponding stages of the previously presented model of texture boundary processing, the reader is referred to sections 2.2 and 2.3 for a description of the receptive field properties and the model cell dynamics. Here, only the differences between the two models are highlighted and the feed-forward stages for the detection of corners and T-junctions are introduced. Most obviously, the model employed in this chapter lacks area V4 (Fig 3.2A). The missing feedback from model area V4 resulted in shifted tuning curves of the cells in the lower model areas V1 and V2. In order to counteract this effect, the constants determining the cell dynamics and receptive field properties were adapted accordingly. The values of the modified constants are listed in Appendix B.1.

Furthermore, the V2 bipole cells of the V1-V2 model not only pool bottom-up input from

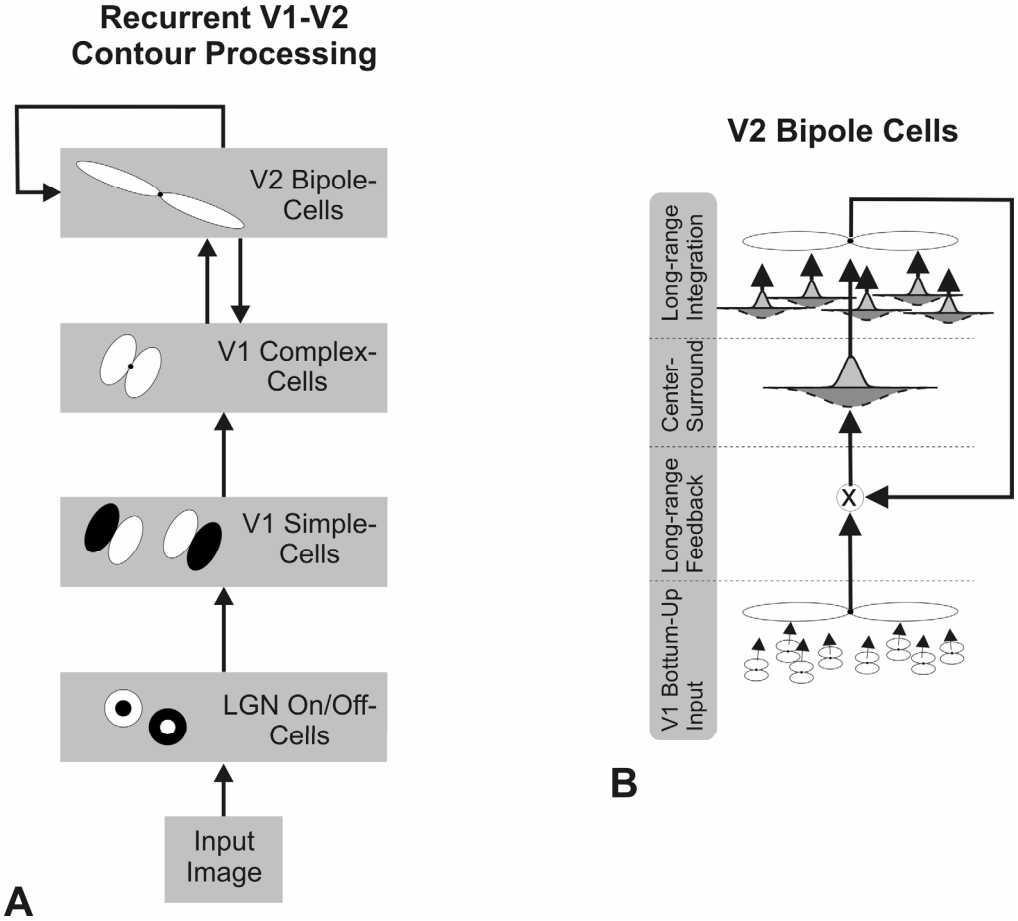


Figure 3.2: Recurrent V1-V2 contour processing. (A) Outline of the hierarchy of model areas and model cells. The receptive field kernels are sketched on the left (white: excitatory subfields; black: inhibitory subfields). The field sizes are not drawn to scale. (B) Model V2 bipole cells integrate the input of appropriately aligned V1 and V2 cells. Here, the two subfields of the cell are depicted twice in order to help visualizing the computational stages used to determine the final V2 cell activity: Once for the pooling of V1 bottom-up activity (bottom row) and again for the pooling of intra-areal V2 activity (top row). The initial bipole cell activity is determined by the (driving) V1 bottom-up input pooled by the two prolated subfields. The cell activity is then modulated by excitatory V2 intra-areal input which is also pooled by the two subfields. This activity, determined by driving V1 input and modulated by long-range V2 activity, finally undergoes a stage of center-surround competition.

model area V1 but use their elongated subfields to additionally sample the activity of neighboring V2 cells (Fig. 3.2B): First, the bottom-up activity delivered by V1 complex cells is pooled by the two subfields and combined by a soft-AND-gate mechanism to only generate significant responses when both fields are excited simultaneously (Grossberg and Mingolla, 1985, Neumann and Sepp, 1999). Subsequently, the initial V2 bipole cell response to the driving V1 bottom-up input is modulated by the activity of neighboring V2 cells. More

specifically, the V2 activity pooled by the two elongated subfields of a bipole cell modulates the initial cell response to the V1 input in a multiplicative fashion, thereby helping to normalize the V2 cell activation strength along contours. While the bottom-up mechanism enables the cells to complete fragmented contours and to respond to illusory contours, the V2 horizontal long-range interaction helps to stabilize the initial contour representation in particular for noisy contours, e.g. at surface boundaries with varying luminance levels in the background. The V2 activity determined by driving V1 bottom-up input and modulatory V2 long-range interaction finally undergoes a stage of center-surround competition to determine the output activity of model area V2.

Apart from the details outlined above, the model used for preprocessing the initial luminance distribution in the input image is identical to model areas LGN to V2 of the model of texture processing in chapter 2. A more comprehensive description of the functional properties of the V1-V2 model of contour processing with respect to, e.g. grouping and illusory boundary formation can be found in (Neumann and Sepp, 1999).

Detection of Corners and T-Junctions

The position as well as orientation of corners and T-junctions is determined in a feed-forward process using the steady-state activation patterns of model V1 complex cells and model V2 long-range activity.

- *Model end-stop cells*: End-stop cells in this model signal the position and orientation of contour endings. They pool the model V1 complex cell activity using excitatory and inhibitory subfields (Fig. 3.3A). 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.2.
- *Detection of corners and T-junctions*: As indicated in Fig. 3.3B, the likely position of contour corners is signaled by the product of the end-stop activities at orientation θ with

the activities at the perpendicular orientation $\theta + \pi/2$. Similarly, the product of the end-stop activities at orientation θ and the V2 long-range activity at orientation $\theta + \pi/2$ signals candidate positions of T-junctions. 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. The model equations are depicted in Appendix B.2.

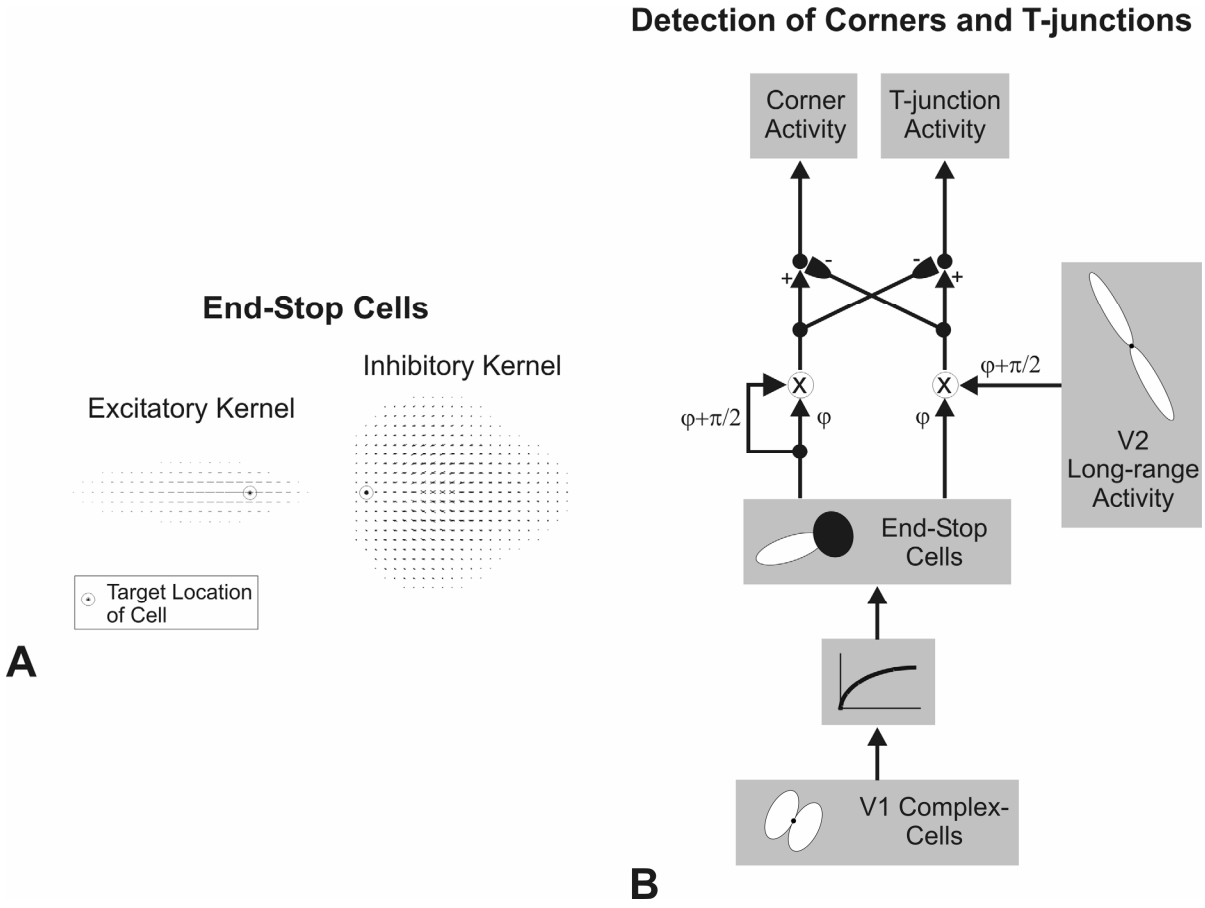


Figure 3.3: Detection of corners and T-junctions. (A) Receptive field kernels of the end-stop cells used to filter V1 complex cell activity. (B) Feed-forward scheme to detect the position of corners and T-junctions. V1 complex cell activity is normalized and filtered by end-stop cells. Subsequently, 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 approx. 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.

3.2 Recurrent Depth Processing

The following chapter outlines the neural model for the depth sorting of 2D surface contours. It starts with a theoretical scheme describing how overlapping contours can be arranged in depth in a globally consistent way using local junction information. Subsequently, the neural mechanisms necessary to integrate this scheme into an overall model architecture are discussed. Mathematical details of the implementation of these mechanisms are described in the remaining sections of this chapter.

3.2.1 A Recursive Scheme for Depth Sorting of Surface Contours Using Local Relative Depth Cues

In the following, a computational scheme is described which allows to sort surface contours according to their positions in depth. It uses local relative cues (such as T- or X-junctions) to recursively determine a globally consistent representation of the depth of the contours. When trying to use the T-junctions which a contour forms with other, intersecting contours in order to determine its exact position in depth, one recognizes that the result mostly remains ambiguous. Consider, for example, the stimulus in the top row of Figure 3.4 consisting of several overlapping surfaces. On the right side, the corresponding surface contours and the positions of the T-junctions are depicted. The white surface patch in the center of the image has two T-junctions which indicate that it is behind other contours and one showing that it is in front of another contour. 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. Generally speaking, the number of local T-junctions which indicate that a contour is in the front of or behind other contours merely depends on the layout of the visual scene, but mostly does not contain information about its exact depth position. There are two exceptions to that rule, namely the contours belonging to surfaces which 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. So the recursive scheme starts with determining

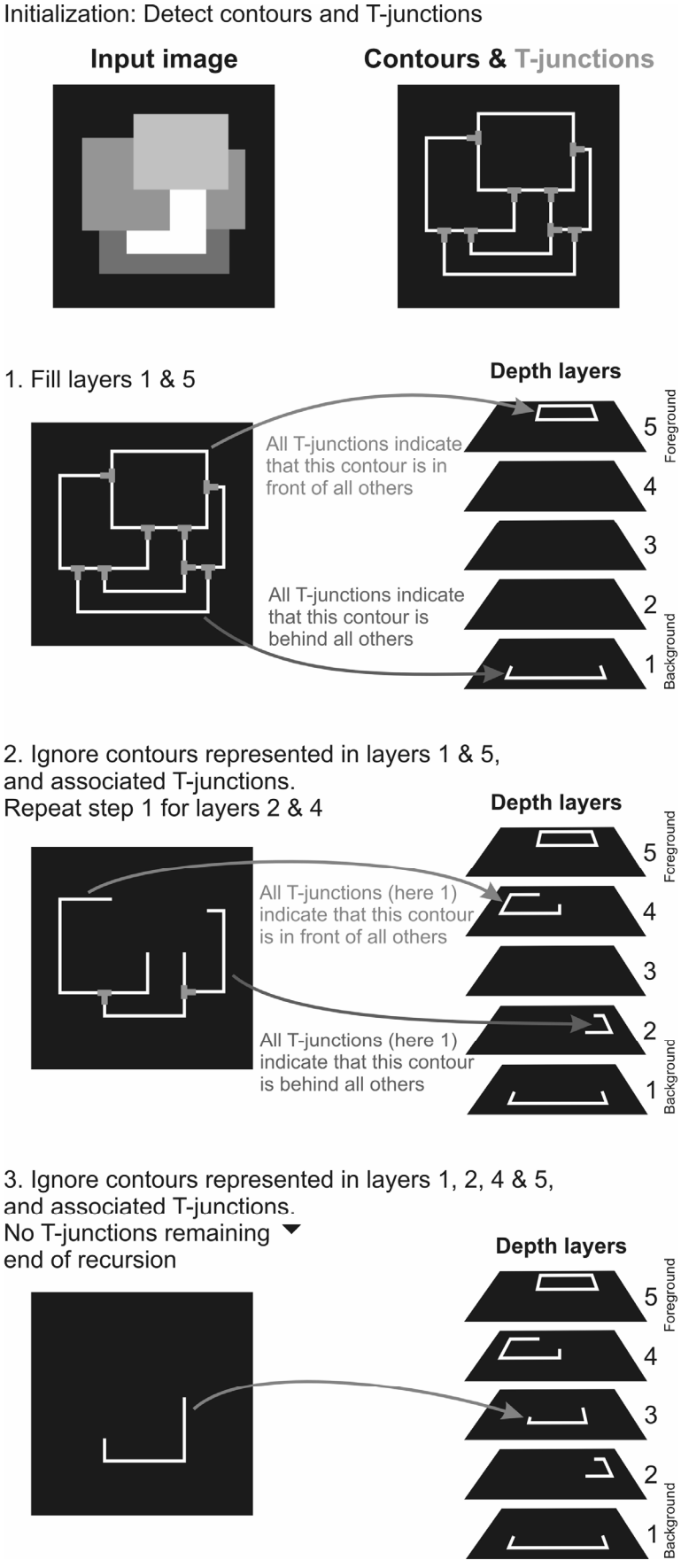


Figure 3.4: Recursive scheme to determine a globally consistent interpretation of the depth of surface contours using relative depth cues.

these contours (second row of Fig. 3.4) and transferring them to depth layers representing the foreground (layer 5) and background (layer 1), respectively. Subsequently, they and their T-junctions are deleted from the set of contours which still have ambiguous depth positions (third row of Fig. 3.4). When considering the remaining surfaces, there will again be contours which are either in the front of or behind all other remaining contours, which is now unambiguously indicated by their T-junctions. These contours are stored in depth layers 2 and 4 and deleted from the set of remaining contours (forth row of Fig. 3.4). Given that we initially provided a sufficient number of layers, this scheme can be recursively continued until all contours are stored in their appropriate depth layer.

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: A contour is in the background of all other contours and consequently arranged in layer $\#1$ if local depth cues are absent that signal any other contours continuing behind it. Similarly, if no depth cues are available that signal another contour continuing in front of a contour, that one has to be in the foreground and is arranged in layer $\#N$ (contours owned by an isolated surface not overlapping with others will be simultaneously assigned to layers $\#1$ and N in this way).
- 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 considered and the first step is repeated for the remaining contours and depth cues: Remaining contours which 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 will 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 allows to successively resolve 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.

3.2.2 Outline of the Overall Model Architecture and Key Processing Mechanisms

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 necessary key structures and mechanisms to integrate this computational scheme into a neural architecture. The neural model builds upon the activation pattern of the model of recurrent V1-V2 interaction as the underlying representation of the contours in the image. The T-junction detectors signal topographical positions at which activation patterns representing different surface contours intersect and allow to determine local figure-ground relationships.

A basic building block of the neural model is its mechanism to encode the depth of a contour. Based on physiological findings of cell pools in the visual cortex which exhibit different disparity profiles (Poggio et al., 1988), the model uses a stack of several topographical 2D maps and signals the depth of a contour by the activity of neurons at corresponding spatial positions within the corresponding depth map. In the above depicted scheme of depth sorting, contours are recursively assigned to depth layers, starting with the contours belonging to surfaces in the fore-and background. This recursive scheme requires that a set

of those contours is maintained that have depth positions which are still unresolved and ambiguous. Once a contour is assigned to a layer, it and the corresponding T-junctions are excluded from the set. In the neural model, both the representations of this set as well as of the contours that have already been assigned to a specific position in depth have to be commonly signaled by the activation patterns of the neurons within the stack of depth layers. This is achieved by the integration of two inhibitory mechanisms into the model, one acting between depth layers and one acting within each layer, as described below. In combination, they enable the activation patterns which signal contours with ambiguous depth positions to be continuously passed on from the outer to the inner depth layers, thereby representing the neural equivalent to the discrete set used in the abstract sorting scheme.

The inhibitory mechanisms acting within a layer are based on the dipole cell dynamics introduced by Grossberg (1991). A dipole cell consists of two antagonistic channels (ON and OFF), which 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 the neighboring cells are pooled by the cell's receptive fields and constitute the two inputs to the dipole (Fig. 3.5A). 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 resets, the input to one of the channels becomes very strong, in turn resetting the cell. By this, a self-contained wave of activity is triggered in the two-dimensional map of model cells. In the model of depth sorting, this effect enables to inhibit (or reset) the ON-channels of cells representing contours not belonging to the actual depth layer via locally restricted interactions between neighboring dipoles in the topographical map. The local activity of T-junction detectors initiates these waves of OFF-channel activity which then propagate autonomously within the depth layer. V2 bipole cells control the input gain of the dipole channels to restrict the propagation of the waves along surface contours having a different position in depth.

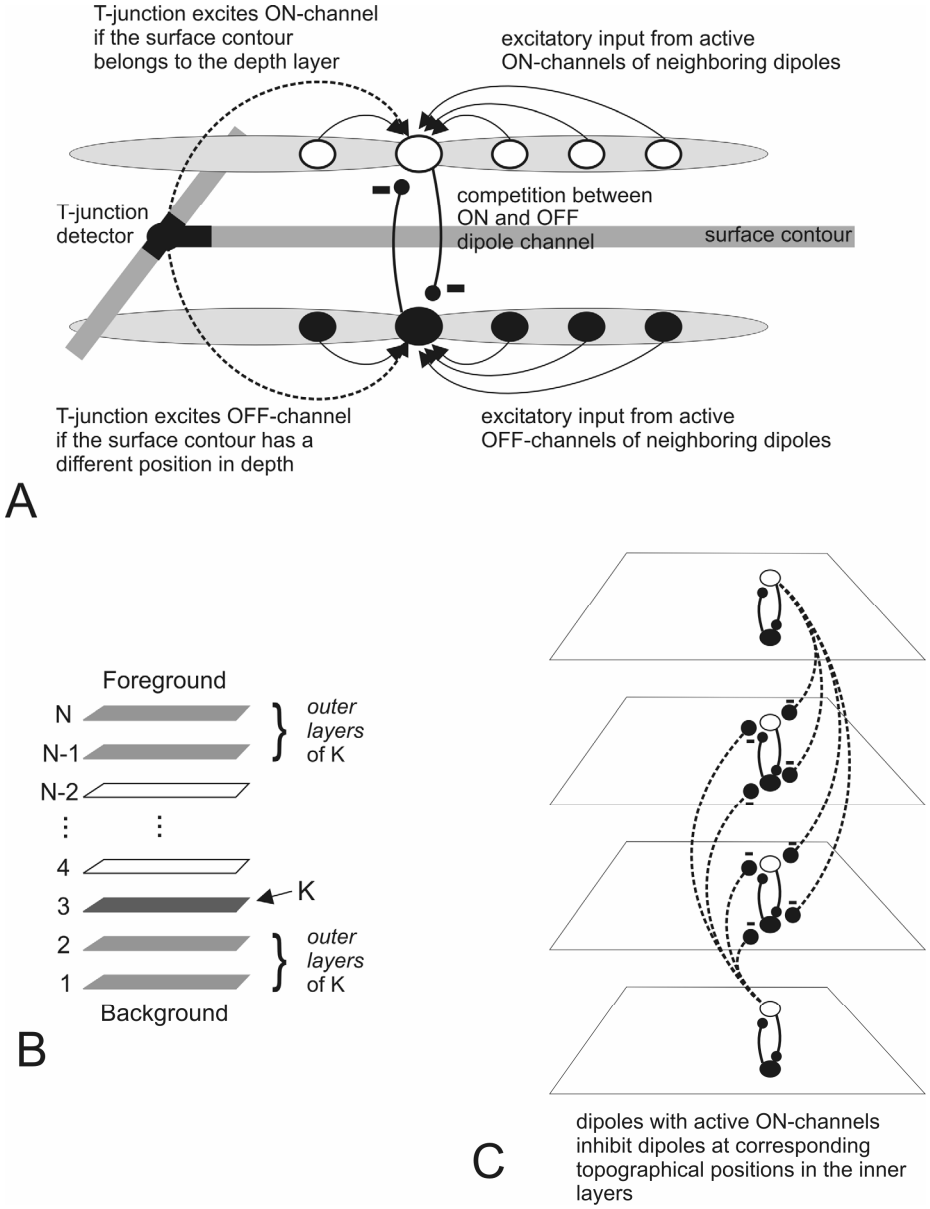


Figure 3.5: Key model mechanisms. (A) Flow of activity within a depth layer: A dipole cell pools the ON- and OFF-channel activities in the neighborhood using elongated receptive fields. The ON- and OFF-channels of the dipole continuously compete against each other to signal the channel receiving the 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 self-contained 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) Example depicting the outer layers of an intermediate depth layer #K. For layer #3, the outer layers are given by #1, 2, N-1 and N. (C) Interaction between depth layers: Active ON-channels of dipoles in the outer layers inhibit the dipole activity in the inner layers. This mechanism causes activity patterns representing contours having intermediate depth positions to be transferred from the outer to the inner depth layers.

The distribution of the dipole activity patterns across depth layers is controlled by inhibitory mechanisms between the layers (Fig. 3.5C). The release of inhibition enables the model to continuously transfer the activity patterns which represent contours with an ambiguous position from the outer to the inner depth layers. More precisely, neurons with active ON-channels in a certain layer of the stack inhibit all neurons at the corresponding topographical positions of those layers which represent more medial depth positions. Initially, the activation patterns of the two outermost depth layers represent *all* contours, thereby indicating that the depth positions of all contours are unknown in the beginning. Consequently, the activity of the neurons in the inner layers of the stack are initially suppressed by inhibitory interaction. Waves of OFF-channel activity triggered by T-junction detectors reset those neurons in the outer layers which signal contours being not in the fore- or background, respectively. These cells loose their inhibitory impact on the cells in the inner layers, which, in turn, signal those contours having an intermediate depth position. More specifically, since the depth layers neighboring the outermost layers again inhibit the more medial layers, the activation patterns of 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 the inhibitory mechanisms acting within and between layers enables to continuously pass on those activities from the outer to the inner depth layers which represent contours having more intermediate, but still ambiguous positions in depth.

In the following sections, we will have a closer view on the key components of the model of depth sorting, starting with the dipole dynamics and the interaction of neighboring dipoles to allow a recurrent flow of activity within a model layer. Subsequently, the mathematical equations determining the inhibitory interactions between the model layers are discussed. The last section outlines the mechanisms by which the T-junction detectors exert influence on the neighboring dipoles.

3.2.3 Dipole Dynamics: A Close-up

A model dipole 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. 3.6A). 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 and 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.

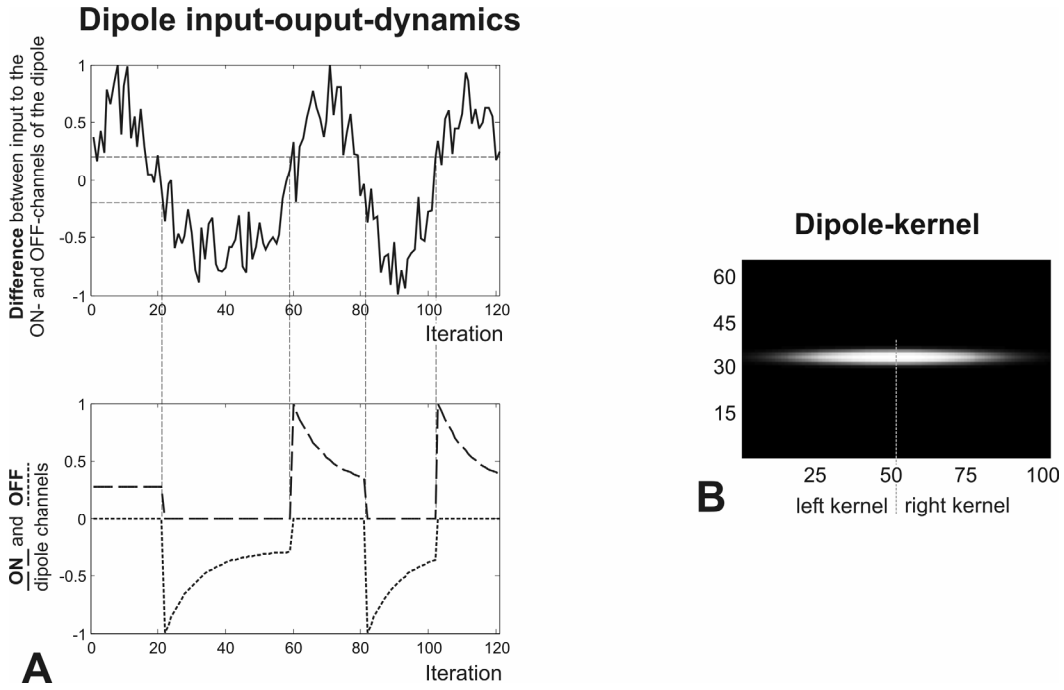


Figure 3.6: (A) Example time course of the dipole input-output dynamics. The difference between the two input channels is continuously calculated and the dipole resets when the difference either exceeds or falls below a threshold value (indicated by the two dashed lines in the upper plot). The formerly 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 doesn't change its sign again, the output of the active channel slowly decays to asymptotically reach a baseline level. (B) Spatial layout of the dipole kernels used to integrate activity of neighboring dipoles. The kernels are cut into a left and a right half.

The functionality of the dipoles as described above is captured by the following equations. First, the thresholded and have-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 $d_{t-1}^{ON/OFF}$ and the result is stored in a_t^{ON} and a_t^{OFF}:

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

$[x]^+ := \max\{x,0\}$ stands for have-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 B.3 in Appendix B.3. In equation (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 $a_t^{ON/OFF}$ signal those points in time at which an antagonistic rebound should be triggered. They are subsequently used to update the internal activity states $b_t^{ON/OFF}$ of both channels:

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

As long as one input remains stronger than the other one (both a_t stay at zero in this case), 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). In this case, the b_t-activation corresponding to the weaker input stays at zero (see eq. 3.3). When 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}. The internal activity states b_t compete against each other via mutual subtractive inhibition, resulting in the suppression of the weaker dipole channel:

$$d_t^{ON} = b_t^{ON} \times \left[C_{gain} \left[b_t^{ON} - b_t^{OFF} \right]^+ \right]^{\leq 1}$$
$$d_t^{OFF} = b_t^{OFF} \times \left[C_{gain} \left[b_t^{OFF} - b_t^{ON} \right]^+ \right]^{\leq 1} \quad (3.3)$$

$\left[x\right]^{\leq 1} := \min\{x,1\}$ denotes a bounded linear transfer function. The activities d_t represent the outputs of the dipole channels and are subsequently pooled by the receptive fields of the neighboring dipoles to constitute their new input activities g_t, thereby creating a recurrent flow of activity in each depth layer.

3.2.4 Interactions between Neighboring Dipoles

Dipoles use anisotropic receptive fields (or kernels) to pool the activity delivered by the ON- and OFF-channels of their neighboring dipoles (Fig. 3.6B). As dipoles are not orientation selective *per se*, they use the activity of the model V2 bipole cells to adaptively shape their receptive fields: Take, for example, a horizontal contour which maximally activates the horizontally oriented V2 bipole cells at corresponding spatial positions. Consequently, the dipole cells at these positions use horizontally elongated receptive fields to pool dipole activity in their neighborhood which corresponds to the contour.

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 in the range [0, π). Subsequently, the activity of the kernel which 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 y^{V2_Norm}:

$$e_{ti\theta}^{ON} = d_{ti}^{ON} \times y_{i\theta}^{V2_Norm}$$
$$e_{ti\theta}^{OFF} = d_{ti}^{OFF} \times y_{i\theta}^{V2_Norm} \quad (3.4)$$

Subscripts t, i and θ represent time, the spatial location and orientation, respectively. In the computational implementation of the model, activities y^{V2_Norm} and $e_t^{ON/OFF}$ 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 y^{V2_Norm} is derived from the steady-state V2 bipole activity y^{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 B.3; eq. B.11). The maps of modulated dipole activities $e_{ti\theta}^{ON/OFF}$ selectively signal the dipole activities at contour positions. As activity $e_{ti\theta}^{ON/OFF}$ results from the multiplication of the dipole activities with the orientation-selective activity y^{V2_Norm}, the orientation of the underlying contours is also encoded in the activity pattern of $e_{ti\theta}^{ON/OFF}$. Subsequently, the dipoles integrate the activities $e_{ti\theta}^{ON/OFF}$ in their neighborhood using elongated receptive field kernels K (Fig 3.6B):

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

Ψ_f denotes a Gaussian weighting function in the orientation domain, K is the kernel in the spatial domain, and $*$ is the convolution operator. K is modeled as modified anisotropic Gaussian kernel which is normalized to yield steeper flanks at the boundaries of the kernel (Appendix B.3; eq. B.12). 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 $e_{ti\theta}^{ON/OFF}$ from that 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 $f_{ti\theta}^{ON/OFF}$ are finally weighted with the normalized V2 bipole activity y^{V2_Norm} to adaptively shape the dipole kernels: The multiplication with y^{V2_Norm} selects those activities f which were pooled by anisotropic kernels K which correspond best to the orientation of the underlying contour. The weighted activities are summed over all orientations, resulting in the new input activities $g_{(t+1)}^{ON/OFF}$ (see eq. 3.1) of the dipole at time $t+1$:

$$\begin{aligned} g_{(t+1)\ i}^{ON} &= \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{ON} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) \\ g_{(t+1)\ i}^{OFF} &= \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{OFF} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) \end{aligned} \tag{3.6}$$

The elongated layout of the kernels K in equation (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. At surface corners, however, the contours elements 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 which 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. 3.7). In order to (i) allow the waves of activity to flow along contour corners and to (ii) prevent crosstalk between aligned contours of neighboring surface patches, the end-stop cell activities 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. 3.7; the mathematical details are given in Appendix B.3).

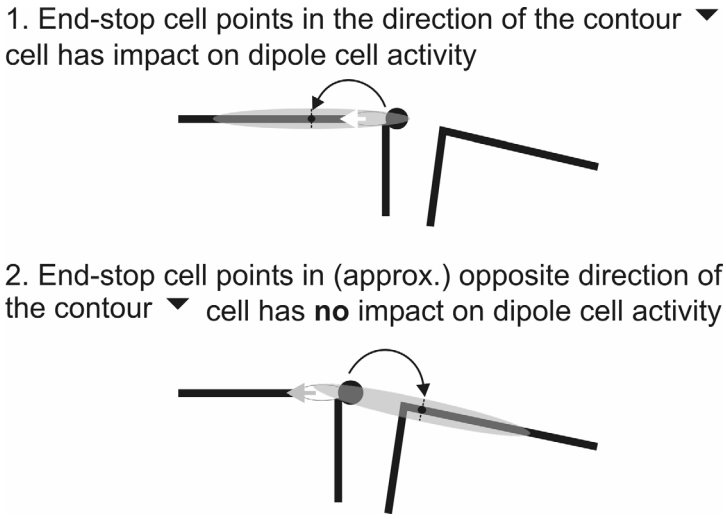


Figure 3.7: A dipole cell uses the left and right halves of its spatial kernel to selectively integrate output activity at contour corners only if the corresponding end-stop cells point approximately in the direction of the dipole cell.

The six equations outlined above describe the mechanisms of dipole dynamics and dipole interactions necessary for a recurrent flow of activity within the depth layers, in turn enabling the self-contained waves of OFF-channel activity to autonomously propagate along contours. In the following sections, the last two key components of the model are highlighted which 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.

3.2.5 Interaction between Model Layers

As already outlined above, inhibitory mechanisms between the depth layers continuously transfer those activity patterns from the outer to the inner depth layers which represent contours having more intermediate positions in depth (see section 3.2.2; Fig. 3.5B/C). These mechanisms suppress the output activities of dipoles in the more medial depth layers which are at the topographical positions of those dipoles in the outer layers which have activated ON-channels. For example, in a model with N depth layers as depicted in Figure 3.5B, the activity in layer $K=3$ is suppressed by ON-channel activity in layers 1, 2 as well as N-1 and N. In other words, 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 given by:

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

The inhibition of dipole output activity in layer $\#K$ is incorporated into the model using a modified version of equation (3.4):

$$\begin{aligned} e_{ti\theta}^{ON} &= \left[d_{ti}^{ON} - C_{ol} \times \sum_{j \in outerlayers} d_{ti}^{ON_layer(j)} \right]^{+} \times y_{i\theta}^{V2_Norm} \\ e_{ti\theta}^{OFF} &= \left[d_{ti}^{OFF} - C_{ol} \times \sum_{j \in outerlayers} d_{ti}^{ON_layer(j)} \right]^{+} \times y_{i\theta}^{V2_Norm} \end{aligned} \quad (3.8)$$

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. By this mechanism, waves of

OFF-channel activity in the outer layers result in a release of inhibition in the more medial layers which transfers the activation patterns corresponding to contours at intermediate depth positions continuously from the outer to the inner model layers.

3.2.6 The Impact of T-junctions on Dipole 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 which signal contours not belonging to the actual depth layer. More specifically, they locally increase the OFF-channel input g^{OFF} to those dipoles in their neighborhood which correspond to contours *not* belonging to the actual depth layer and thereby induce the waves of activity running along that contours. Additionally, they locally increase the input to the ON-channels of the remaining dipoles in the neighborhood to prevent the waves of activity to unintentionally spread along the intersecting contours.

The T-junction activity undergoes a stage of normalization and Gaussian filtering in the spatial and orientation domain (for details please refer to Appendix B.4), resulting in the two activity distributions q_{ti}^{foregr} and q_{ti}^{backgr}: Activity q_{ti}^{foregr} signals those topographical positions at which a contour element is locally in the foreground of another contour (i.e. the top of the T), whereby activity q_{ti}^{backgr} represents the positions of contours being locally in the background (i.e. the stem of the T). In the background layer $\#1$, activity q_{ti}^{foregr} modulates the input g^{OFF} to the dipoles to trigger waves of OFF-channel activity resetting all dipoles corresponding to contours being in the foreground of other contours. Accordingly, the modified equation (3.6) reads:

$$\begin{aligned} g_{(t+1)\ i}^{ON} &= \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{ON} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) + q_{ti}^{backgr} \\ g_{(t+1)\ i}^{OFF} &= \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{OFF} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) + q_{ti}^{foregr} \end{aligned} \tag{3.9}$$

Correspondingly, activity q_{ti}^{backgr} exerts influence on the OFF-channels of the dipoles in the foreground layer $\#N$ to trigger waves resetting contours being locally in the background:

$$g_{(t+1)\ i}^{ON} = \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{ON} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) + q_{ti}^{foregr}$$
$$g_{(t+1)\ i}^{OFF} = \sum_{k=1}^{n_{orient}} \left(f_{ti(k-1)\times\pi/n_{orient}}^{OFF} \times y_{i(k-1)\times\pi/n_{orient}}^{V2_Norm} \right) + q_{ti}^{backgr} \quad (3.10)$$

In the intermediate layers, activities q_{ti}^{foregr} and q_{ti}^{backgr} are suppressed by active ON-channels of dipoles in the outer layers at the corresponding spatial positions (analogous to the inhibition between the dipoles across depth layers, as described in section 3.2.5; please refer to Appendix B.4 for details). By this inhibition, the T-junction information corresponding to contours having intermediate depth positions is continuously transferred from the outer to the inner depth layers. This prevents, e.g. T-junction information stemming from the intersections of the contour in the foreground with other contours to exert influence on the activation patterns in the intermediate depth layers. For layers $\#1$ to $\lfloor N/2 \rfloor$, equation (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. Similarly, for layers $\#\lceil N/2 \rceil + 1$ to N, equation (3.10) is used to specify how activity q_{ti}^{backgr} resets the dipoles corresponding to contours being locally in the background.

The computational stages and mechanisms of this modelling approach 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). 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 which 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 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 and the previous sections, we developed the equations specifying the three key components of the model which (i) allow for a propagation of activity within model layers via recurrent interactions between cells in a locally restricted neighborhood, (ii) distribute the model activity between the depth layers via local inhibitory interactions between cells in different layers at corresponding topographical positions and (iii) enable the T-junction detectors to trigger waves of activity resetting those dipole activations corresponding to contours having different depth positions. 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.

3.3 Simulation Results

In the following, simulation results are used to demonstrate the computational capabilities of the combined models of contour processing (section 3.1) and recurrent depth processing (section 3.2) 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.

3.3.1 Pre-Processing

In the following, the key properties of the model of recurrent V1-V2 contour processing followed by the feed-forward stages for T-junction and corner detection (section 3.1) are outlined using the stimulus depicted in Fig. 3.8. The final equilibrated V1 and V2 model activation patterns after 7 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 4 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, which 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. Taken together, the overall model 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 of recurrent V1-V2 contour processing constitutes the input to the feed-forward stages for T-junction and corner detection. V1 complex cell activity is linearly filtered by model end-stop cells to determine contour endings (Fig. 3.8: 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. 3.8: 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. 3.8: last row left). After preprocessing, 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.

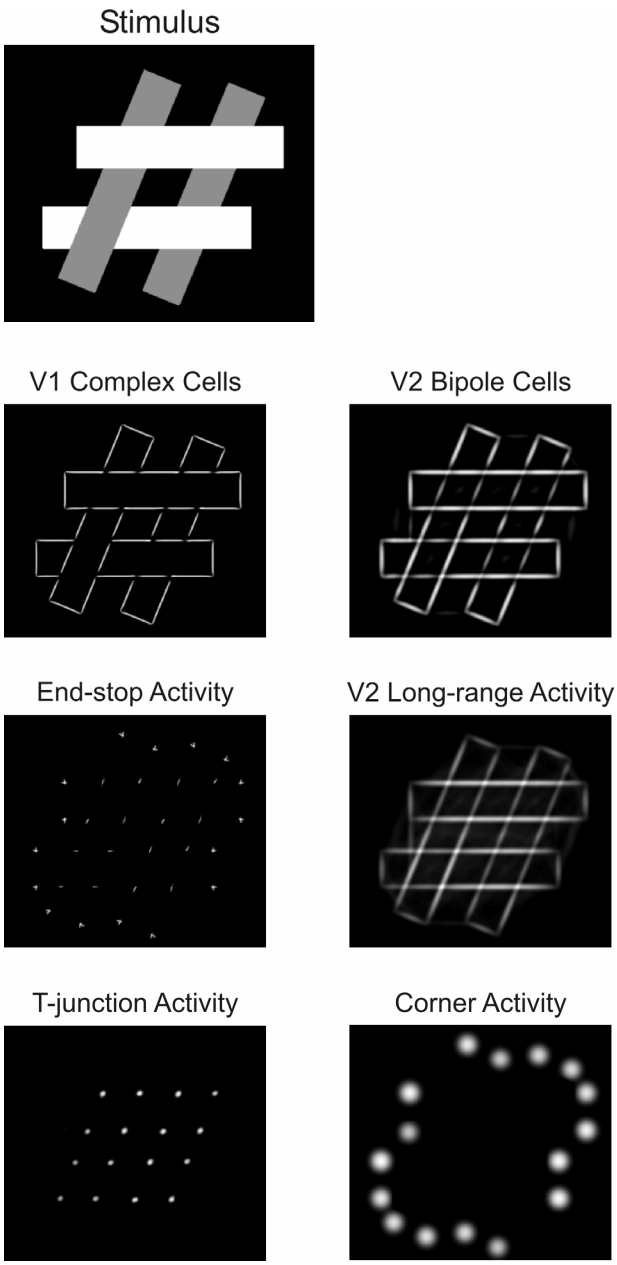


Figure 3.8: Example for the processing of contours by the model of recurrent V1-V2 interaction. The physical outlines of the four gray or white rectangles are signaled by V1 complex cells. V2 bipole cells subsequently insert illusory contours to bridge gaps created by overlapping surfaces. The model 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 section 3.1).

3.3.2 Recurrent Depth Processing

The functionality of the neural mechanisms of the model of recurrent depth processing (section 3.2) 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 model 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 Model Layer

T-junction activity signals local foreground-background relations at crossings of surface contours. In the model of recurrent depth processing, this local information is distributed along contours, as depicted exemplarily for the background layer in Fig. 3.9. 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. 3.8) and subsequently normalized to yield a contour mask being approximately 1 at contour positions and 0 otherwise. Here and in the following sections, 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. 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. 3.9 top row: Iteration 1). The high OFF-channel activities of the reset dipoles are pooled by neighboring dipole cells, which in turn reset themselves (Fig. 3.9 second row: Iteration 4).

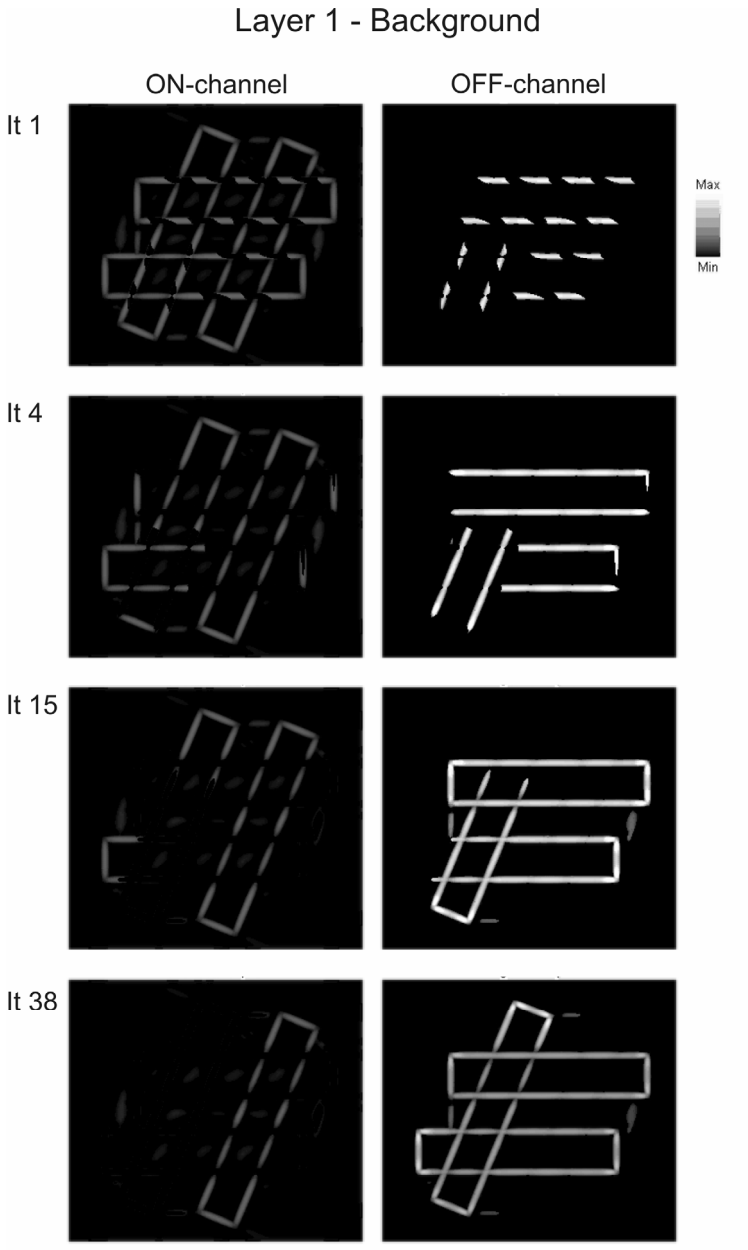


Figure 3.9: Recurrent model of depth processing: Development of the dipole activity pattern in the background model layer in response to the stimulus of Fig. 3.8. Only the dipoles positions corresponding to surface contours are shown by applying a mask based on thresholded model V2 bipole activity. Iteration 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&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.

This results in waves of OFF-channel activity running 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. 3.9 bottom row: Iteration 38). Successful completion of fragmented contour elements by the model 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 allows to increase 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 section 3.2.4 & Fig. 3.7).

The flow of dipole activity in the foreground layer (Fig. 3.10 right column) is contrary to that in the background layer: Initially, T-junction activity locally resets the dipoles corresponding to contours which are in the background of another contour (Fig. 3.10 right column: Iteration 1). This information is distributed along the contours by waves of OFF-channel dipole activity (Fig. 3.10 right column: Iteration 15&19). Finally, ON-channel activities remain only at positions corresponding to contours being in the foreground of all others (Fig. 3.10 right column: Iteration 38).

Interaction between depth layers

Inhibitory connections from the outer to the inner model layers enable interaction between the layers to develop a consistent depth representation (Fig. 3.10). Dipole cells in outer layers with active ON-channels inhibit the dipole outputs in the inner layers at corresponding spatial positions (chapter 3.2.5). Consequently, the ON- and OFF-channel activities of the

dipoles in depth layers 2 and 3 are initially suppressed (Fig. 3.10 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.

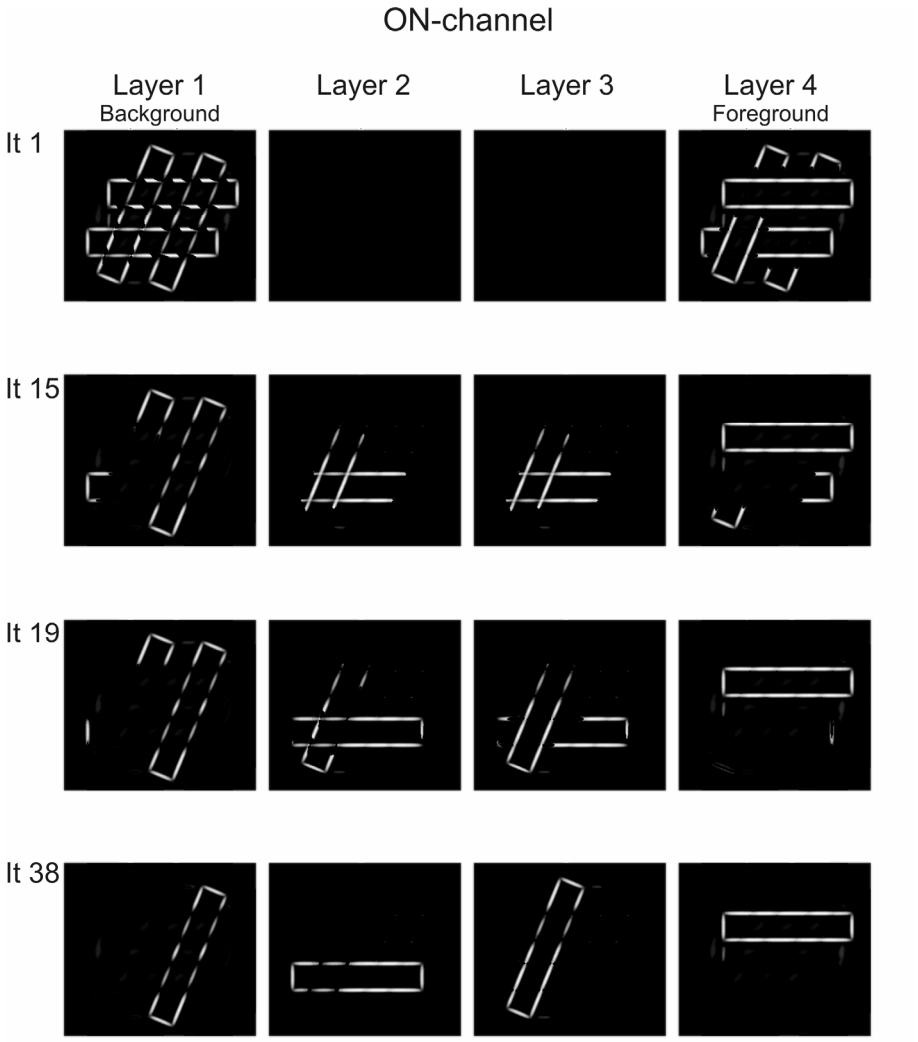


Figure 3.10: Development of the activation pattern of the model of recurrent depth processing in response to the stimulus of Fig. 3.8. Iteration 1: Only the dipoles in layers 1 (background) and 4 (foreground) are active and suppress the dipole activity in the intermediate layers. Iteration 15: T-junction 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.

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. 3.10 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 model 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. 3.10 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. 3.10 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 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 depths takes more time to develop.

3.3.3 Further Simulation Results

In the previous sections, 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 6 different relative depths is considered (Fig. 3.11). 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 section 3.2.1).

The stimulus is preprocessed by the model of recurrent V1-V2 interaction, as depicted in the upper right area of Fig. 3.11. The contours of the surface patches are signaled by model V2 bipole cells. The subsequent feed-forward stages filter 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 of recurrent depth processing. In Fig. 3.11, 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 chapter 3.3.2). 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

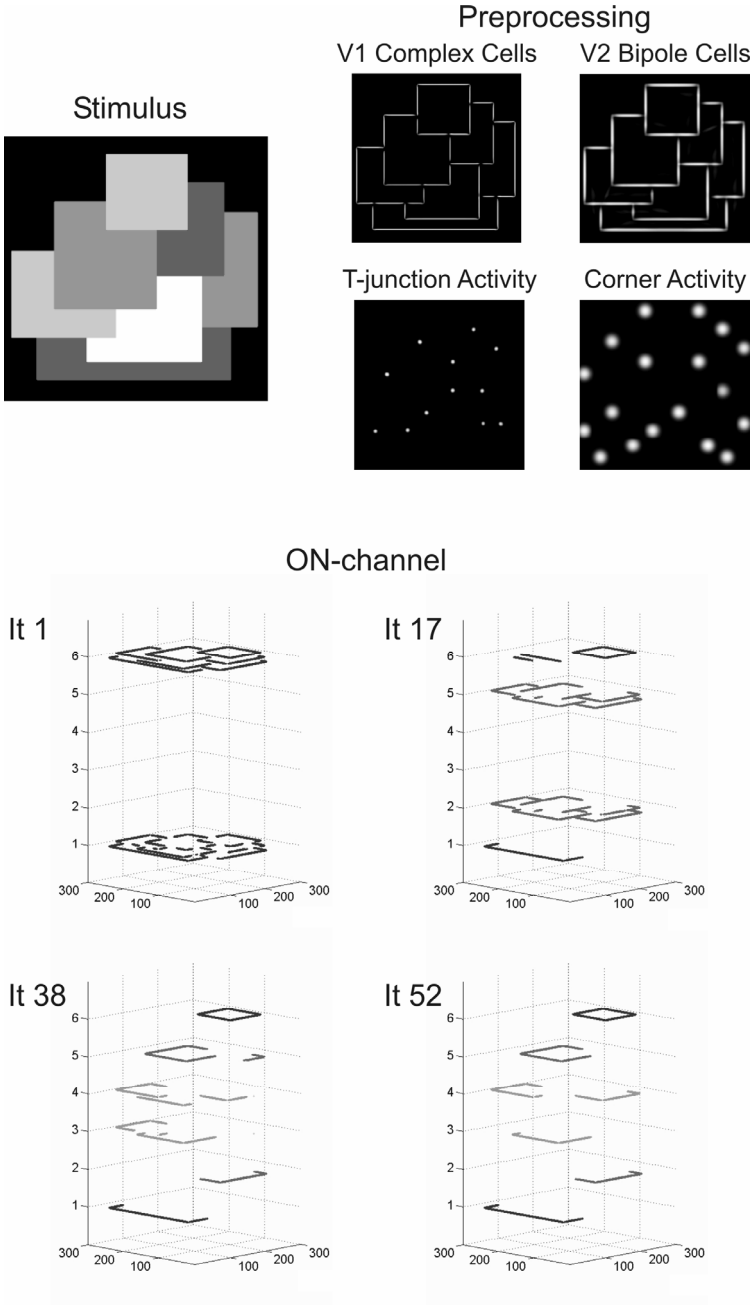


Figure 3.11: Model responses to overlapping surface patches arranged in 6 depth levels. Preprocessing: The model of recurrent V1-V2 interaction signals the surface contours, followed by the feed-forward stages detecting the positions of T-junctions and corners in the input image. Model of recurrent depth processing: Dipoles with active ON channels are shown at several points in time. 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 nor 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 layers 2 and 5 is released. Iteration 38: At positions without ON-channel activity in layers 1, 2, 5 & 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.

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

In Figure 3.12, the model reactions to the real-world picture of Figure 3.1 are shown. V2 bipole cells signal the outlines of the three objects and of the underlay. The activation patterns of the model of V1-V2 interaction constitutes the input to the feed-forward processing stages 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 of recurrent depth processing which 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 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.

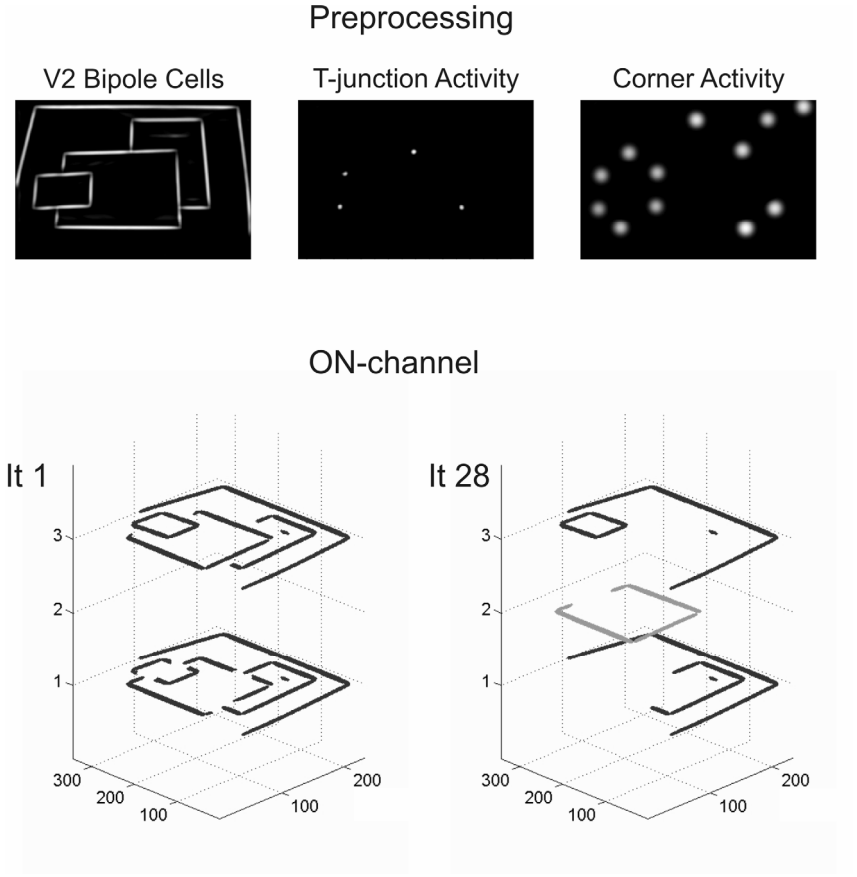


Figure 3.12: Model responses to the real-world stimulus of Fig. 3.1. The preprocessing network signals the positions of contours, corners and T-junctions in the input image. The model of recurrent depth processing successfully achieves the correct depth sorting of the three central overlapping objects. 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.

The final example demonstrates that interaction of the model of recurrent V1-V2 contour processing and the model of recurrent depth processing can achieve successful figure-ground segmentation of stimuli based on illusory contours, such as a Kanizsa square (Fig. 3.13): 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 colinearly aligned contour fragments. Consequently, V2 activation patterns additionally signal the

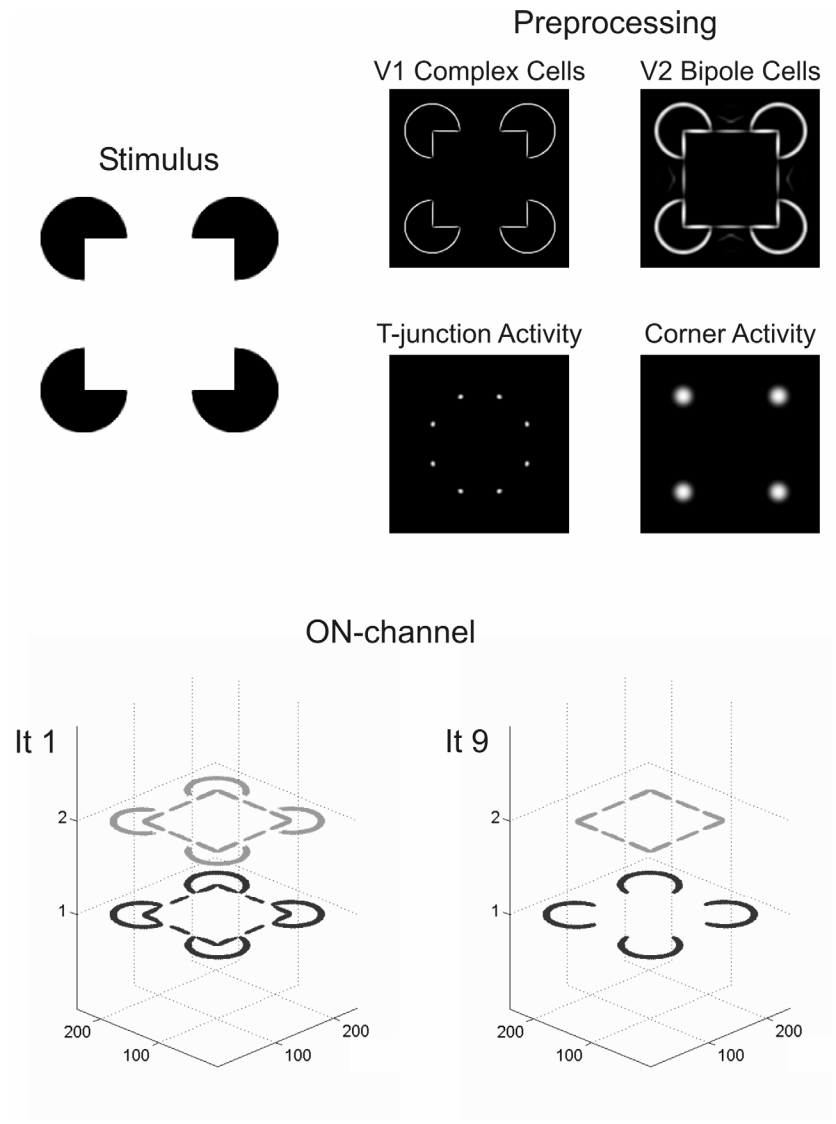


Figure 3.13: Model responses to a Kanizsa square. Model V2 bipole cells respond to the illusory contours constituting the square. The subsequent feed-forward stages detect the “illusory” T-junctions created by the overlap of the outline of the Kanizsa square (as signaled by V2 bipole cells) and the remaining 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.

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 feed-forward stages following the model of recurrent V1-V2 interaction. Finally, the depth relations between the contours are resolved by the model of recurrent

depth processing using the “illusory” T-junction information as cue. To conclude, in our model architecture, illusory contours are able to transform physical corners to “illusory” T-junctions, thereby enabling a depth segmentation of the resulting contour representation.

3.4 Discussion

3.4.1 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 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 section 3.2.4). 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 early visual areas, either by horizontal activity propagtion 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. Furthermore, 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.

3.4.2 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 model 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 abrupt changed input stimulation an antagonistic rebound of dipole cell activation is elicited which, in turn, enables the overall system to quickly reset itself. 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 sections 3.2.1&3.3.2). 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 to unintentionally obtain a biasing advantage due to cross-talk between the depth layers and results in an overall robust model behavior.

3.4.3 General Model Framework

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 and Duke, 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 than disparity information (Howard and Duke, 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 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 (section 3.3.3).

3.4.4 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 basing 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 which 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 section 3.4.1, 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 cells is sensitive to the 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 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 and 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 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 model layers which, in turn, guarantee robust model behavior.

3.4.5 Model Properties, Limitations and Future Extensions

The aim of this study was to demonstrate that 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 details of the model which were outside the central scope of the study. For example, the most obvious limitation is the exclusive usage of T-junctions as depth cues. A straight-forward extension in future versions may therefore be the integration of X-junction information. Adelson and Anandan (1990) as well as Anderson (1997) proposed rules for the classification of X-junctions that allow to determine which junctions are consistent with a transparency interpretation. This classification scheme could be integrated into our model by additional mechanisms basing on V1 simple cell activity 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 the model hierarchy proposed here, binocular matching would be used to create a common representation of surface boundaries independent of 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. I 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 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.

Taken together, in the model proposed here, 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 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.

4 Summary

One of the main goals of early visual processes is the reliable segregation of figure and background. The human visual system utilizes a multitude of visual cues such as luminance, color, movement, depth and texture information to solve that task. Early research started with the question how luminance information (the most basic feature) is processed by the visual system. For example, Hubel and Wiesel (1959) discovered that a majority of cells in the primary visual cortex is sensitive to abrupt luminance changes having a certain orientation. It was proposed that this response property helps to focus early visual processes on surface *borders* (instead of on the surfaces themselves), allowing to *discount the illuminant*. A variety of questions arise from that view. First of all, if the surface information itself is discarded in a first step, how is it reconstructed (or "*filled in*") later on? A border physically belongs to the surface which is in the fore- and not in the background. However, the information of *border ownership* is not contained in the initial representation of luminance discontinuities, but has to be recovered in a second step to allow for an unambiguous reconstruction of the surfaces. Consequently, before being able to *"fill in"* the surface information, one has to solve the problem of *border ownership*.

The visual system uses a variety of different features for segmentation. Consequently, a further question is how the processing of the different cues is integrated in a coherent framework. A directly related question is whether the processing of features other than luminance is also based on an initial representation of discontinuities representing putative borders.

In this thesis, I tackled some of the problems outlined above and presented solutions for them, with the focus of my work being on the processing of luminance and texture cues. In a first step, I developed a biologically plausible model of human texture segmentation which integrates the processing of luminance and texture information into a *common* computational framework. This model allows to link a variety of key psychophysical and physiological

results on human texture processing with the model cell activation patterns. In turn, the human behavioural results can be traced back to their putative neural origins and processing principles. The model is based upon the two key computational properties that (i) texture segmentation is based on boundary detection (Nothdurft, 1985, Nothdurft, 1991, Gallant et al., 1995, Lamme et al., 1998, Lamme et al., 1999, Nothdurft et al., 1999, Nothdurft et al., 2000), and (ii) texture boundaries are detected mainly on the basis of a large scenic context that is analyzed by higher cortical areas within the ventral visual pathway (de Weerd et al., 1996, Merigan, 1996, Kastner et al., 2000, 2000). The model implements texture boundary processing as a distributed task within the hierarchy of visual areas V1, V2 and V4. The model mechanisms were motivated by models of V1-V2 interaction for contour completion and illusory contour formation (Neumann and Sepp, 1999, Grossberg and Raizada, 2000, Ross et al., 2000). The computational capabilities of these approaches are incorporated and preserved by our model, thereby integrating the processing of luminance and texture information for surface boundary finding into a common framework. In particular, it shares the key processing principles of the model proposed in (Neumann and Sepp, 1999), namely (i) the bi-directional information flow between cortical areas, enabling the modulation of cell responses via feedback from higher visual areas, and (ii) the intra-areal normalization of cell responses by means of center-surround competition in a spatial and orientational neighborhood. The combination of these two principles realizes a *context-selective gain enhancement* or *soft-gating mechanism*: Initial cell activities, which match the activation pattern in the next model area, are enhanced via excitatory feedback connections and in turn inhibit cells in their neighborhood via center-surround competition. This results in a facilitation of bottom-up activity matching the “expectation” of the cells in the higher model area (Grossberg, 1980, Mumford, 1994). As cells in higher areas integrate information over wide parts of the input image, the overall process of recurrent activity interaction enables a context-selective enhancement of salient texture arrangements, while at the same time spurious and perceptually irrelevant activities are suppressed.

As recurrent networks tend to exhibit chaotic and instable trajectories, the analysis of the model dynamics is a necessary part of a comprehensive characterization of the model behaviour. In chapter 2.3 of this thesis, the boundedness of the model dynamics is proven as a first step. Subsequently, the global stability properties of the model equations are demonstrated using Hirsch's theorem (Hirsch, 1989). In particular, the specific impact of the single model parameters on the stability of the network is discussed. In the subsequent chapters 2.4 to 2.10, the model behaviour in response to texture stimuli is investigated in detail using a comprehensive series of simulations and the link between the model activation patterns and physiological and psychophysical data is outlined. The main results are:

- Cells in monkey V1 signal the presence of a texture border within their receptive fields by an enhanced firing rate (e.g. Gallant et al., 1995). Based on the simulation results, feedback from area V4 acting indirectly via V2 is identified as a putative source of the observed enhancement effects.
- A electrophysiological study demonstrated that "the spatial distribution of activity in V4 encodes the retinotopic locations of salient features throughout the visual field" (Mazer and Gallant, 2003, p. 1248). It is shown that the model can account for these findings and model area V4 acts as saliency map for the orientation domain.
- The psychophysical results of Nothdurft (1991) which characterize how orientation noise effects the detectability of a pop-out bar have been shown to correlate with the activation patterns in model area V4. In particular, intact feedback connections have been demonstrated to be crucial to raise the model performance to that of human observers. This suggests that the main functional role of feedback in human texture processing may be noise suppression.
- It has been also shown psychophysically that the relative orientation of the line elements with respect to the texture border exerts influence on the detectability of a pop-out region (Nothdurft, 1992). This effect is most likely caused by anisotropic mechanisms for contour integration and completion, as observed in contour neurons of V2 (v. d. Heydt et al., 1984). In our model, the common grouping of both line elements that belong to the pop-

out bar and those of the background to continuous “flow patterns” by model V2 bipole cells effects the orientation contrast “seen” by model V4 cells.

- Modulation of texture density results in varying saliency ratings of an embedded pop-out line with a peak for medium density values (Nothdurft, 2000c). The origin and extent of this effect can be traced back to a combination of V4 receptive field sizes, the range of V4 intra-areal center-surround competition and a V4 receptive field organization of orientation selective cells having an excitatory center and lateral inhibitory surrounds.

Two key predictions of the model have been confirmed in fMRI-studies. First, in case of no orientation noise, the detectability of a pop-out bar increases monotonically with increasing border orientation contrast and saturates at about 30°. The model prediction that this psychophysical curve directly correlates with the activity of neurons in visual area V4 has been confirmed in the first fMRI-experiment (Grön et al., 2003). The second fMRI study confirmed that the dependence of the saliency of a pop-out line on texture density is reflected by a corresponding activation of neurons in V4 (Kölle et al., 2004). This suggests that V4 is the neural origin of the observed behavioural effects. Taken together, the combination of modelling- and fMRI-studies strengthens the view that human visual area V4 plays a crucial role in texture border processing. In this thesis, it was demonstrated through simulations that the model of texture border processing can account for a wide variety of psychophysical and physiological data. In particular, it was successfully used to generate new predictions.

Determining the position and orientation of region borders in the retinal image is only the first step which is necessary to reconstruct the objects in the visual scene. Once surface borders have been determined, all boundaries belonging to the same surface have to be grouped together and segregated (or segmented) from the boundaries of other objects, i.e. the problem of *boundary ownership* has to be solved. It is well established that the human visual system uses depth information as a major cue to segregate those items from each other that belong to distinct objects (Nakayama et al., 1989, Kovacs et al., 1995, Baumann et al., 1997b). In the second part of this thesis, I presented 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 and to solve the boundary ownership problem. A major problem is that many depth cues such as disparity or occlusion information are locally restricted, i.e. their information is only available at sparse locations in the visual scene. Furthermore, they often signal the relative depth relation between two objects, but not the absolute position of an object in depth. In consequence, in order to obtain a globally consistent interpretation of the depth relations between the objects in a visual scene, a mechanism is required that allows to globally propagate the local and relative information of the depth cues. Electrophysiological studies demonstrate that neurons in visual area V2 possess the ability to use occlusion cues for figure-ground segmentation (Baumann et al., 1997b, Zhou et al., 2000). This indicates that mechanisms for depth processing might already be integrated at rather early stages of visual processing. However, neurons in early visual areas process the visual information in high resolution and have therefore rather small receptive fields. Furthermore, they interact only in a restricted spatial neighbourhood (Gilbert and Wiesel, 1989). Therefore, the critical question arises how a framework of neural interaction would have to look like in order to allow these neurons to exchange their local information and to achieve a globally consistent interpretation of the depth relations in the visual scene.

The second part of this thesis discusses a computational model in which depth information from local relative cues (in our case T-junctions) is propagated along surface contours using only local recurrent interactions between neighbouring neurons. It is demonstrated that within this framework local interactions are sufficient to obtain a globally consistent depth sorting of overlapping surfaces, in turn solving the border ownership problem. The model is an extension of a previously developed model of recurrent V1-V2 interaction for contour processing and illusory contour formation (Neumann and Sepp, 1999, Grossberg and Raizada, 2000, Ross et al., 2000) and uses the contour representation created by this model as a basis for its recursive scheme of depth sorting. It consists of a stack of several topographical 2D maps and signals the depth of a contour by the activity of neurons at corresponding spatial positions within the corresponding depth map. The model dynamics is

determined by three key mechanisms: First, dipole cells (Grossberg, 1991) consisting of two mutually competing channels use anisotropic receptive fields to integrate the activities of those neighbouring dipoles which correspond to contour positions. This allows for a propagation of activity within a depth layer along contours via recurrent interactions between dipoles in a locally restricted neighbourhood. Second, local inhibitory interactions between cells in different layers distribute the model activity within the 3D stack of depth layers. Third, T-junction detectors trigger the propagation of activity within a layer. This results in a reset of those dipole activations which correspond to contours having a depth position not represented by the actual depth layer.

In simulations, it is demonstrated that the recursive flow of activity within the model results in a stable final activation pattern and that this pattern signals a globally consistent depth sorting of the surface contours in the input image, even in case of several overlapping surfaces. It is also shown that the combination of the model of V1-V2 contour processing (Neumann and Sepp, 1999) with the mechanisms of recurrent depth processing can successfully achieve the figure-ground segmentation of stimuli based on illusory contours, such as a Kanizsa square. The intersections of the illusory contours which form the Kanizsa square and the curved contours of the Pacman-inducers act as illusory T-junctions which trigger the recurrent depth sorting. Taken together, the simulations show that the three key mechanisms integrated into the model are sufficient to allow for a globally consistent depth sorting. The number of overlapping contours which can be separated from each other by the model is determined by the number of depth layers and is, in theory, unlimited. Therefore, it would be interesting to determine the number of overlapping surfaces which humans can automatically (or pre-attentively) distinguish from each other. To my knowledge, up to now, no psychophysical study addressing this question exists. In the electrophysiological studies which examine the response properties of V2 cells to depth cues (Baumann et al., 1997a, Heider et al., 2000, Zhou et al., 2000, Heider et al., 2002) either one isolated patch presented on a uniform background or two overlapping surfaces patches were used as stimuli. However, it might be useful to study cell responses for three for more overlapping surfaces in

future investigations. In the model of depth processing, dipole cells can signal more than two depth positions, depending on the number of model layers. In consequence, it would be interesting to test whether cortical V2 cells exist which selectively signal intermediate depth positions. Furthermore, in the model, the local inhibitory interactions between the cells in different layers result in a recursive scheme of depth sorting, in which 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 how the latencies of border ownership-related response differences depend on the depth position. Finally, the model achieves the depth sorting using only locally restricted interactions. This could be tested experimentally by examining the depth sensitivity of V2 neurons when the feedback connections from higher model areas are deactivated or lesioned.

Taken together, in this thesis, I give solutions to two problems of early vision while restricting the modelling to the processing of luminance and texture information. First, I tackle the question how the neural mechanisms of the human visual system could look like which combine the processing of several distinct visual features with each other. Starting with the previously developed model of recurrent V1-V2 interaction for border processing (Neumann and Sepp, 1999), I propose a biologically plausible model which integrates the processing of luminance- and texture borders into a common computational framework. A key component of the model is the V4 processing stage which is sensitive to gradients in the activity maps of the lower model areas. These activity maps represent the distribution of luminance information or of oriented texture items in the input picture. However, I claim that the general idea of border detection and enhancement via model V4 cells and feedback modulation is not restricted to these two features, but can be extended to other visual cues such as color information or texture elements differing in size (Bergen and Adelson, 1988). In consequence, the proposed model architecture could act as general-purpose scheme for the robust detection of salient discontinuities in the activity maps representing the distribution of different features in the input image.

Second, I demonstrate how depth information can be used to solve the border ownership problem when the initial representation of objects is based on surface boundaries. The representation of luminance borders created by the model of V1-V2 interaction is utilized by a successive stage of recurrent depth processing. This stage propagates local T-junction information stemming from overlapping surfaces along the boundaries and creates a globally consistent depth sorting of the surfaces in the input image. Only local recurrent interactions between neighboring cells are used within the model to achieve this depth sorting, thereby demonstrating that depth processing may be already implemented in rather early visual areas.

Appendix

A Model of Texture Border Processing: Supplements

A.1 Receptive Field Equations

In the following, the equations describing the receptive field properties are outlined which determine the initial activation I of a cell. The letter y denotes the final activation of the cells in the *prior* area 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.

	parameters for x				parameters for y							
									σ of			
	α_1	β_1	δ_1	C	α_2	β_2	δ_2	ζ_2	ψ^+	Λ^+	ψ^-	Λ^-
V1	12.0	0.73	3.7	25	1.0	2.8	5.0	3.5	0.2	1.0	2.0	3.0
V2	12.0	0.85	4.2	20	1.0	2.9	50.0	3.1	0.2	2.0	2.0	6.0
V4	-	-	-	-	1.0	10.6	1000	9.9	0.2	8.0	2.0	24.0

Table A.1: Parameters of the three-stage cell dynamics and standard deviations σ of the Gaussians pooling the center and surround activity in the spatial and orientational domains (see eq. 2.1 to 2.4). The values were kept constant through all experiments.

Preprocessing: LGN ON/OFF and V1 Simple Cells

LGN cell activity $x^{on/off}$ is determined by convolution of the input image with a difference-of-Gaussians operator followed by a half-wave rectification:

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

with: $\sigma_{Center} = 0.8$; $\sigma_{Surround} = 3\ \sigma_{Center}$

The symbols $\Lambda_{Center/Surround}$ denote isotropic 2D Gaussian kernels (standard deviations σ_{Center} and $\sigma_{Surround}$ define their sizes in pixels) and $[x]^{+}$ and $*$ denote have-wave rectification and the convolution operator, respectively. Model V1 simple cells exist for two polarities (dark-light, dl; light-dark, ld) and eight orientations. They have elongated ON and OFF subfields to pool the input delivered by appropriately aligned LGN ON or OFF cells:

$$\begin{aligned} p_{\theta}^{on/off_left} &= x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,-\tau_y,\theta} \\ p_{\theta}^{on/off_right} &= x^{on/off} * \Lambda_{\sigma_x,\sigma_y,0,\tau_y,\theta} \end{aligned} \tag{A.2}$$

with: $\sigma_y = 0.8$; $\sigma_x = 3.5\ \sigma_y$; $\tau_y = 1.6\ \sigma_y$

$n_{orient} = 8$; $\theta = 0,\ \pi/n_{orient},\ \ ...,\ (n_{orient}\text{-}1)\ \pi/n_{orient}$

The subfields are modeled by anisotropic 2D Gaussian weighting functions $\Lambda_{\sigma x,\, \sigma y,\, 0,\, +/-\tau y,\, \theta}$. Their size in y- and x-direction is defined by the standard-deviations σ_x and σ_y. They are shifted perpendicular to their main axis by $+/-\tau_y$ and then rotated by θ. The activity of the subfields is subsequently fed into a soft-AND-circuit which combines additive and multiplicative interactions between the subfield activities to determine the response s of the simple cell and results in a more localized simple cell activation at luminance borders compared to the pure summation of subfield responses (please refer to Neumann et al. 1999 for details). For example, the activation $s_{i\theta}^{ld}$ of a V1 simple cell sensitive for light-dark polarity is given by:

$$s_{i\theta}^{ld} = \frac{\mathrm{A_s}(p_{i\theta}^{on_left} + p_{i\theta}^{off_right}) \ + \ 2\mathrm{B_s}\ p_{i\theta}^{on_left}\ p_{i\theta}^{off_right}}{\mathrm{A_s D_s} + \mathrm{E_s}(p_{i\theta}^{on_left} + p_{i\theta}^{off_right})} \tag{A.3}$$

with: $\mathrm{A_s} = 1.0$; $\mathrm{B_s} = 10000.0$; $\mathrm{D_s} = 0.05$; $\mathrm{E_s} = 100.0$

Spatial locations in the topographical maps are expressed by the index i. The first term in the numerator denotes the additive excitatory interaction between the ON and OFF subfields, while the second term defines the amount of activation resulting from their multiplicative interaction. The denominator normalizes the response by the total sum of subfield responses.

V1 Complex Cells

The initial V1 complex cell response $I_{i\theta}^{V1}$ 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:

$$I_{i\theta}^{V1} = \mathrm{A_c}([s_{i\theta}^{ld} - s_{i\theta}^{dl}]^+ + [s_{i\theta}^{dl} - s_{i\theta}^{ld}]^+) \tag{A.4}$$

with: $\mathrm{A_c} = 0.1$

V2 Bipole Cells

Model V2 bipole cells use two prolated subfields $K^{left/right}$ aligned along the axis of the cell's orientation preference to pool the input delivered by appropriately aligned V1 complex cells. V1 complex cell activity y^{V1} (represented by a 3D matrix: 2D – space, 1D – orientation) is blurred in the orientation domain (using a convolution with a 1D isotropic Gaussian ψ_f) and convolved with the 2D spatial weighting functions $K^{left/right}$ to determine the subfield activities $f^{left/right}$:

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

with: $\Psi_f : \ \sigma_{f_orient} = 0.3$

The weighting functions $K^{left/right}$ determining the spatial layout of the subfields are modeled as anisotropic Gaussians, which are cut off in the central part of the cell by means of a sigmoid function. The partial overlap of the subfields in the center of the cell defines the classical receptive field:

$$K_{i\theta}^{left} = \Lambda_{\sigma_{k_x},\sigma_{k_y},\tau_{k_x},0,\theta}(\vec{x}_i) \cdot \frac{1}{1+\exp(-\mathrm{A_k}\vec{x}_i \bullet \binom{\cos\theta}{\sin\theta} - \mathrm{B_k})}$$

$$K_{i\theta}^{right} = \Lambda_{\sigma_{k_x},\sigma_{k_y},-\tau_{k_x},0,\theta}(\vec{x}_i) \cdot \frac{1}{1+\exp(+\mathrm{A_k}\vec{x}_i \bullet \binom{\cos\theta}{\sin\theta} + \mathrm{B_k})} \qquad \text{(A.6)}$$

$\vec{x}_i$: cartesian coordinates of point i

with: $\sigma_{k_x} = 22.0$; $\sigma_{k_y} = 1.0$; $\tau_{k_x} = 2.0$; $\mathrm{A_k} = 0.8$; $\mathrm{B_k} = 9.0$

The subfield activations are combined using a soft-AND-gate, in turn only generating significant initial V2 cell responses $I_{i\theta}^{V2}$ when both subfields are excited simultaneously:

$$I_{i\theta}^{V2} = \frac{\mathrm{A_t}(f_{i\theta}^{left} + f_{i\theta}^{right}) \; + \; 2\mathrm{B_t} \; f_{i\theta}^{left} \; f_{i\theta}^{right}}{\mathrm{A_t D_t} + \mathrm{E_t}(f_{i\theta}^{left} + f_{i\theta}^{right})} \qquad \text{(A.7)}$$

with: $\mathrm{A_t} = 2.3$; $\mathrm{B_t} = 2600.0$; $\mathrm{D_t} = 0.15$; $\mathrm{E_t} = 100.0$

In accordance to the three-stage model cell dynamics, the initial V2 cell activities subsequently undergo feedback modulation by V4 activity and intra-areal center-surround competition to yield the V2 output activity y^{V2}.

Model V4 cells

Model V4 cells consist of one central excitatory subfield q^{center} and two lateral inhibitory elongated subfields $q^{left/right}$ all sensitive to the same input orientation. The lateral inhibitory fields are shifted perpendicularly to their main axis. The differences between the V2 bipole cell activities y^{V2} pooled by the center field and the left and right inhibitory subfields are determined. The sum of the two half-wave rectified differences is the initial activation of the V4 cell. For each of the eight V1 complex cell orientations (denoted by the angle θ), V4 cells in eight orientations exist (denoted by the angle φ).

The initial activation $I_{i\theta\varphi}^{V4}$ of a model V4 cell is given by:

$$I_{i\theta\varphi}^{V4} = [q_{i\theta\varphi}^{center} - C \; q_{i\theta\varphi}^{left}]^+ \; + \; [q_{i\theta\varphi}^{center} - C \; q_{i\theta\varphi}^{right}]^+ \qquad \text{(A.8)}$$

with:

$$q_{\varphi}^{center} = y^{V2} * \Psi_q * \Lambda_{\sigma_{q_x},\sigma_{q_y},0,0,\varphi}$$
$$q_{\varphi}^{left} = y^{V2} * \Psi_q * \Lambda_{\sigma_{q_x},\sigma_{q_y},0,-\tau_{q_y},\varphi} \quad \text{(A.9)}$$
$$q_{\varphi}^{right} = y^{V2} * \Psi_q * \Lambda_{\sigma_{q_x},\sigma_{q_y},0,\tau_{q_y},\varphi}$$

$\sigma_{q_x} = 22.0; \quad \sigma_{q_y} = 8.0; \quad \tau_{q_y} = 16.0; \quad C = 1.25$

$\Psi_q: \ \sigma_{q_orient_\theta} = 0.4$

$\varphi = 0, \ \pi/n_{orient}, \ ..., \ (n_{orient}\text{-}1) \ \pi/n_{orient}$

As V4 constitutes the highest model area, the initial activation $I_{i\theta\varphi}$ is not modulated via feedback, but is directly fed into the center-surround competition denoted by equation (2.4). In this competition, all V4 cells sensitive to the same V1 orientation interact in a spatial and orientational (denoted by φ) neighborhood. No interaction between different V1 orientation channels exists in model area V4. The final responses y^{V4} of all V4 cells sensitive to the same V1 orientation are summed up after center-surround competition and fed back to V2:

$$y_{i\theta}^{V4} = \sum_{k=1}^{n_{orient}} y_{i\theta\ (k-1)*\pi/n_{orient}}^{V4} \quad \text{(A.10)}$$

A.2 Model Cell Dynamics: Reformulation of Equations for Stability Analysis

For stability analysis, the indices i and θ in equations (2.2) and (2.3) are combined to a single index i, and the re-indexed equation (2.2)

$$x_i = \frac{\beta_1 I_i[1 + C h_i]}{\alpha_1 + \delta_1 I_i[1 + C h_i]} \tag{A.11}$$

is used to reformulate equation (2.3):

$$\begin{aligned}
\frac{\partial}{\partial t} y_i &= -\alpha_2 y_i + \beta_2 \sum_{j=1}^{N} c_{ij} x_j - \left(\delta_2 y_i + \zeta_2 \varepsilon(y_i)\right) \sum_{j=1}^{N} d_{ij} x_j \\
&= -\alpha_2 y_i + \sum_{j=1}^{N} (\beta_2 c_{ij} - \zeta_2 \varepsilon(y_i) d_{ij}) \frac{\beta_1 I_i[1 + C h_i]}{\alpha_1 + \delta_1 I_i[1 + C h_i]} - \delta_2 y_i \sum_{j=1}^{N} d_{ij} \frac{\beta_1 I_i[1 + C h_i]}{\alpha_1 + \delta_1 I_i[1 + C h_i]} \\
&= -\alpha_2 y_i + \sum_{j=1}^{N} (\tfrac{\beta_1}{\delta_1} \beta_2 c_{ij} - \tfrac{\beta_1}{\delta_1} \zeta_2 \varepsilon(y_i) d_{ij}) \frac{I_i[1 + C h_i]}{\frac{\alpha_1}{\delta_1} + I_i[1 + C h_i]} - \tfrac{\beta_1}{\delta_1} \delta_2 y_i \sum_{j=1}^{N} d_{ij} \frac{I_i[1 + C h_i]}{\frac{\alpha_1}{\delta_1} + I_i[1 + C h_i]} \\
&= -A y_i + \sum_{j=1}^{N} e_{ij} x_j^* - B y_i \sum_{j=1}^{N} d_{ij} x_j^*
\end{aligned} \tag{A.12}$$

with:

$$x_i^* = \frac{I_i[1 + C h_i]}{D + I_i[1 + C h_i]}$$

$$A = \alpha_2; \quad B = \frac{\beta_1}{\delta_1} \delta_2; \quad D = \frac{\alpha_1}{\delta_1}; \quad e_{ij} = \frac{\beta_1}{\delta_1} (\beta_2 c_{ij} - \zeta_2 \varepsilon(y_i) d_{ij})$$

The weighting factors c_{ij} represent the convolution of the excitatory isotropic Gaussian kernels Λ^+ (two-dimensional) and ψ^+ (1D), resulting in a three-dimensional isotropic kernel. Likewise, d_{ij} represent the convolution of the inhibitory isotropic kernel Λ^- and ψ^-. It therefore follows that

$$\begin{aligned}
&c_{ij} \geq 0; \;\; c_{ij} = c_{ji}; \;\; \sum_{j=1}^{N} c_{ij} = 1; \;\; c_{ii} = \max(c_{ij}) \;\; \forall j \in [1, N] \\
&d_{ij} \geq 0; \;\; d_{ij} = d_{ji}; \;\; \sum_{j=1}^{N} d_{ij} = 1; \;\; d_{ii} = \max(d_{ij}) \;\; \forall j \in [1, N]
\end{aligned} \tag{A.13}$$

It is assumed here that a cell doesn't inhibit itself, i.e. $e_{ii} > 0$.

A.3 Boundedness

In the following, it is shown that an upper bound y_{max} for the cell activation level y_i exists by demonstrating that $\partial/\partial t\, y_i < 0 \Big|_{y_i > y_{\max}}$. In order to establish an upper bound, equation (2.5b) can be approximated as

$$\begin{aligned} \frac{\partial}{\partial t} y_i &= -A\ y_i + \sum e_{ij} x_j\ - B y_i \sum d_{ij} x_j\ \ = -A\ y_i\ + \sum (e_{ij} - B y_i d_{ij}) x_j \\ &\le -A\ y_i\ + \sum_{j=1}^{N^+} (e_{ij}^+ - B y_i d_{ij}^+) x_j \end{aligned} \tag{A.14}$$

with e_{ij}^+ denoting the positive weights $e_{ij} > 0$, d_{ij}^+ standing for the d_{ij} at the spatial positions of e_{ij}^+, and N^+ being the number of positive weights e_{ij}^+. The weighting factors e_{ij}^+ are created by subtraction of a wide inhibitory Gaussian kernel (given by factors d_{ij} weighted by a constant) from a narrow excitatory kernel (given by weights c_{ij}). Consequently, the new kernel described by factors e_{ij}^+ will be narrower than the original excitatory and inhibitory kernels. In other words, the decrease of the weighting factors e_{ij}^+ from the center to the surround will be steeper for the new kernel. This relation between the widths of the new and the original *inhibitory* kernel can be expressed as

$$\frac{e_{ii}}{e_{ij}^+} > \frac{d_{ii}}{d_{ij}^+} \tag{A.15}$$

It follows with equation (A.15) that

$$(e_{ij}^+ x_j - B y_i d_{ij}^+) \le (\frac{e_{ii}}{d_{ii}} d_{ij}^+ - B y_i d_{ij}^+)\ = \underbrace{\frac{d_{ij}^+}{d_{ii}}}_{\le 1} (e_{ii} - B y_i d_{ii})\ \le (e_{ii} - B y_i d_{ii}) \tag{A.16}$$

Substituting equation (A.16) in (A.14) yields

$$\begin{aligned} \frac{\partial}{\partial t} y_i\ &\le -A\ y_i\ + \sum_{j=1}^{N^+} (e_{ij}^+ - B y_i d_{ij}^+) x_j\ \le -A\ y_i\ + \sum_{j=1}^{N^+} (e_{ii} - B y_i d_{ii}) x_j \\ &= -A\ y_i + (e_{ii} - B y_i d_{ii}) \sum_{j=1}^{N^+} x_j\ \le -A\ y_i + (e_{ii} - B y_i d_{ii}) \cdot N^+ \cdot x_{\max} \end{aligned} \tag{A.17}$$

As the upper bound y_{max} requires

$$\partial / \partial t \, y_i \overset{!}{=} 0 \Big|_{y_i = y_{\max}} \tag{A.18}$$

setting equation (A.17) to 0, and solving for y_i finally yields

$$y_{\max} \leq \frac{e_{ii} \cdot N^+ \cdot x_{\max}}{A + B \cdot d_{ii} \cdot N^+ \cdot x_{\max}} \tag{A.19}$$

Using $x_{\max} = 1$ for $I_i \to \infty$ and $h_i \to \infty$ (see eq. 2.5a), y_{max} becomes

$$y_{\max} \leq \frac{e_{ii} \cdot N^+}{A + B \cdot d_{ii} \cdot N^+} \tag{A.20}$$

When a single layer of model cells is used, then $h_i = y_i$ and equation (A.19) can be written (using eq. 2.5a) as

$$y_{\max} \leq \frac{e_{ii} \cdot N^+ \cdot \dfrac{I_{\max}[1 + C \cdot y_{\max}]}{D + I_{\max}[1 + C \cdot y_{\max}]}}{A + B \cdot d_{ii} \cdot N^+ \cdot \dfrac{I_{\max}[1 + C \cdot y_{\max}]}{D + I_{\max}[1 + C \cdot y_{\max}]}} \tag{A.21}$$

Equation (A.21) can be used to iteratively determine a maximal cell activation level y_{max}, provided that the input I_i is restricted to the range $[0, I_{max}]$.

A.4 Stability Analysis of a Hierarchy of Model Areas: Supplements

In the following, upper bounds for $\partial \mathrm{F}_i / \partial y_i$ and the $\left|\partial \mathrm{F}_i / \partial y_j\right|$ are derived for a model layer embedded in a hierarchy of recurrently connected layers. Substituting eq. (2.5a) into (2.5b), and taking the partial derivative of eq. (2.5b) with respect to y_i results in

$$\begin{aligned}
\frac{\partial F_i}{\partial y_i} &= -A + \sum_{j=1}^{N}(e_{ij} - Bd_{ij}y_i)\frac{\partial}{\partial y_i}\frac{I_j[1+Ch_j]}{D+I_j[1+Ch_j]} - B\sum_{j=1}^{N} d_{ij}\frac{I_j[1+Ch_j]}{D+I_j[1+Ch_j]} \\
&\le -A + \sum_{j=1}^{N}(e_{ij} - Bd_{ij}y_i)\frac{\partial}{\partial y_i}\frac{I_j[1+Ch_j]}{D+I_j[1+Ch_j]}\Bigg|_{I_j \to 0} \\
&= -A + \sum_{j=1}^{N}(e_{ij} - Bd_{ij}y_i)\frac{D[1+Ch_j]\frac{\partial}{\partial y_i}I_j + DCI_j\frac{\partial}{\partial y_i}h_j}{(D+I_j[1+Ch_j])^2} \qquad \text{(A.22)} \\
&\le -A + \frac{1}{D}\sum_{j=1}^{N}(|e_{ij}| + Bd_{ij}y_{\max})\left([1+Ch_{\max}]\max(|\tfrac{\partial}{\partial y_i}I_j|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_i}h_j|)\right) \\
&\le -A + \frac{1}{D}\sum_{j=1}^{N}(|e_{ij}| + Bd_{ij}\frac{e_{ii}N^+}{A+Bd_{ii}N^+})\left([1+Ch_{\max}]\max(|\tfrac{\partial}{\partial y_i}I_j|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_i}h_j|)\right)
\end{aligned}$$

Likewise, taking the partial derivatives of eq. (2.5b) with respect to y_j results in

$$\begin{aligned}
\left|\frac{\partial F_i}{\partial y_j}\right| &= \left|\sum_{k=1}^{N}(e_{ik} - Bd_{ik}y_i)\frac{\partial}{\partial y_j}\frac{I_k[1+Ch_k]}{D+I_k[1+Ch_k]}\right| \\
&= \left|\sum_{k=1}^{N}(e_{ik} - Bd_{ik}y_i)\frac{D[1+Ch_k]\frac{\partial}{\partial y_j}I_k + DCI_k\frac{\partial}{\partial y_j}h_k}{(D+I_k[1+Ch_k])^2}\right| \\
&\le \frac{1}{D}\sum_{k=1}^{N}\left|e_{ik} - Bd_{ik}y_i\right|\left([1+Ch_k]\,|\tfrac{\partial}{\partial y_j}I_k| + CI_k\,|\tfrac{\partial}{\partial y_j}h_k|\right) \qquad \text{(A.23)} \\
&\le \frac{1}{D}\sum_{k=1}^{N}\left(|e_{ik}| + Bd_{ik}y_{\max}\right)\left([1+Ch_{\max}]\max(|\tfrac{\partial}{\partial y_j}I_k|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_j}h_k|)\right) \\
&\le \frac{1}{D}\sum_{k=1}^{N}\left(|e_{ik}| + Bd_{ik}\frac{e_{ii}N^+}{A+Bd_{ii}N^+}\right)\left([1+Ch_{\max}]\max(|\tfrac{\partial}{\partial y_j}I_k|) + CI_{\max}\max(|\tfrac{\partial}{\partial y_j}h_k|)\right)
\end{aligned}$$

parameters for x				parameters for y							
								σ of			
α_1	β_1	δ_1	C	α_2	β_2	δ_2	ζ_2	ψ^+	Λ^+	ψ^-	Λ^-
20.0	13.0	38.0	9	25.0	6.0	50 250 500 1000	5.6	0.5	1.6	0.8	5.0

Table A.2: Parameters of the one-layer network used for the stability analysis in section 2.3.3.

B Model Determining the Depth of Surface Contours: Supplements

B.1 Recurrent V1-V2 Contour Processing

As the model stages used for preprocessing the initial luminance distribution in the input image are in great parts identical to the corresponding stages of the model of texture boundary processing presented in chapter 2, the reader is referred to sections 2.2, 2.3 and Appendix A for a description of the receptive field properties and the model cell dynamics. In the following tables, only the updated parameters and constants of the V1-V2 model of contour processing are listed:

	parameters for x				parameters for y							
	α_1	β_1	δ_1	C	α_2	β_2	δ_2	ζ_2	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

Table B.1: Updated parameters of the three-stage cell dynamics and standard deviations σ of the Gaussians pooling the center and surround activity in the spatial and orientational domains as used in chapter 3 (see eq. 2.1 to 2.4).

Model cell	Equation	Parameters / Constants
LGN ON/OFF cells	A.1	$\sigma_{Center} = 0.8$; $\sigma_{Surround} = 3\ \sigma_{Center}$
V1 simple cells	A.2/A.3	$\sigma_y = 0.8$; $\sigma_x = 3.0\ \sigma_y$; $\tau_y = 0.8\ \sigma_y$ $A_s = 1.0$; $B_s = 10000.0$; $D_s = 0.05$; $E_s = 100.0$ $n_{orient} = 8$; $\theta = 0,\ \pi/n_{orient},\ \ldots,\ (n_{orient}\text{-}1)\ \pi/n_{orient}$
V1 complex cells	A.4	$A_c = 0.1$
V2 bipole cells	A.5/A.7	$A_t = 1.0$; $B_t = 50000.0$; $D_t = 0.15$; $E_t = 100.0$ Ψ_f : $\sigma_{f_orient} = 0.25$ As in the model of recurrent texture processing, the initial activation $I_{i\theta}^{V2}$ of a V2 bipole cell is calculated using eq. A.7 (with the parameters given above). Additionally, the two subfields of a bipole cell also pool the final V2 activity y^{V2} after center-surround competition. This is done by applying eq. A.7 a second time, but using y^{V2} instead of y^{V1} as input: $$LR_{i\theta}^{V2} = \frac{A_t(f_{i\theta}^{left} + f_{i\theta}^{right}) \ + \ 2B_t\ f_{i\theta}^{left}\ f_{i\theta}^{right}}{A_t D_t + E_t(f_{i\theta}^{left} + f_{i\theta}^{right})}$$ with $f^{left/right} = y^{V2} * \Psi_f * K^{left/right}$
V2 bipole cells (cont'd)	A.6	As in the model of recurrent texture processing, the spatial subfields K are modeled as modified anisotropic Gaussians which are cut off in the central part of the cell using a sigmoid function: $$K_{i\theta}^{left_unnorm} = \Lambda_{\sigma_{k_x},\sigma_{k_y},\tau_{k_x},0,\theta}(\vec{x}_i) \cdot \frac{1}{1+\exp(-A_k \vec{x}_i \bullet (\frac{\cos\theta}{\sin\theta}) - B_k)}$$ $$K_{i\theta}^{right_unnorm} = \Lambda_{\sigma_{k_x},\sigma_{k_y},-\tau_{k_x},0,\theta}(\vec{x}_i) \cdot \frac{1}{1+\exp(+A_k \vec{x}_i \bullet (\frac{\cos\theta}{\sin\theta}) + B_k)}$$ $\vec{x}_i$: cartesian coordinates of point i Additionally, the K are subsequently normalized to yield steeper flanks at the boundaries of the kernels and the area of each K is scaled to 1: $$K_{i\theta}^{left} = \frac{K_{i\theta}^{left_unnorm}}{\alpha_K + K_{i\theta}^{left_unnorm}} \times \left(\sum_j \frac{K_{j\theta}^{left_unnorm}}{\alpha_K + K_{j\theta}^{left_unnorm}} \right)^{-1}$$ $$K_{i\theta}^{right} = \ldots$$ $\sigma_{k_x} = 18.0$; $\sigma_{k_y} = 1.25$; $\tau_{k_x} = 16.0$; $A_k = 0.5$; $B_k = 2.0$; $\alpha_K = 0.0004$

Table B.2: Updated parameters and modifications of model stages LGN to V2 of the model of recurrent V1-V2 contour processing.

B.2 Feed-forward Scheme for the Detection of Corners and T-Junctions

End-stop Cells: Model Equations

A cell's end-stop behavior is modeled using the following equations. First, the V1 complex cell activity is normalized in the orientation domain at each spatial position:

$$y_{i\theta}^{V1_Norm} = \frac{y_{i\theta}^{V1}}{\alpha + \sum_{k=1}^{n_{orient}} y_{i\ (k-1)\times\pi / n_{orient}}^{V1}} \tag{B.1}$$

$\alpha = 0.01$

The normalized activity is filtered by the excitatory and inhibitory subfields (Fig 3.3A). The activity pooled by the excitatory subfield is denoted by:

$$\begin{aligned} 0 \le \theta < \pi: \quad Act_{i\theta}^{ex} &= \left\{ y^{V1_Norm} * \Psi_{es} * \Lambda^{ex} \right\}_{i\theta} \\ \pi \le \theta < 2\pi: \quad Act_{i\theta}^{ex} &= \left\{ y^{V1_Norm} * \Psi_{es} * \Lambda^{ex} \right\}_{i(\theta-\pi)} \end{aligned} \tag{B.2}$$

$\Psi_{es} : \sigma_{es_orient} = 0.35$

$\Lambda^{ex}_{\sigma_x,\sigma_y,\tau_x,0,\theta} : \sigma_x = 8.0;\ \ \sigma_y = 1.5;\ \ \tau_x = -3.0$

$\theta = 0,\ \ \pi / n_{orient},\ \ ...,\ \ (2n_{orient} - 1)\pi / n_{orient}$

The two ranges $[0,\ \pi)$ and $[\pi,\ 2\pi)$ stem 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 inhibitory subfield consists of three anisotropic and shifted Gaussians, each being sensitive to slightly different input orientations:

$$\begin{aligned} 0 \le \theta < \pi: \quad Act_{i\theta}^{inh} &= C_{cen} \left\{ y^{V1_Norm} * \Psi_{es} * \Lambda^{inh_cen} \right\}_{i\theta} + C_{lat} \left\{ y^{V1_Norm} * \Psi_{es} * \Lambda^{inh_lat1} \right\}_{i(\theta+1)} \\ &\quad + C_{lat} \left\{ y^{V1_Norm} * \Psi_{es} * \Lambda^{inh_lat2} \right\}_{i(\theta-1)} \end{aligned} \tag{B.3}$$

$\pi \le \theta < 2\pi:$ ...

$C_{cen} = 2.0;\ \ C_{lat} = 1.2$

$\Lambda^{inh_cen}_{\sigma_x,\sigma_y,\tau_x,0,\theta} : \sigma_x = 6.0;\ \ \sigma_y = 3.0;\ \ \tau_x = 10.0$

$\Lambda^{inh_lat1} / \Lambda^{inh_lat2} : \sigma_x = 4.0;\ \ \sigma_y = 4.0;\ \ \tau_x = 6.0;\ \ \tau_y = +/- 4.0;$

The have-wave rectified difference between the excitatory and inhibitory subfield activities is weighted with y^{V1_Norm} to restrict the end-stop activity to spatial positions corresponding to contours. Subsequently, the full orientation range of $[0, 2\pi)$ is collapsed to the range $[0, \pi)$ by pooling the activities corresponding to orientations θ and $(\theta+\pi)$:

$$\begin{aligned} es_{i\theta}^{left} &= \left[Act_{i\theta}^{ex} - Act_{i\theta}^{inh}\right]^{+} \times y_{i\theta}^{V1_Norm} \\ es_{i\theta}^{right} &= \left[Act_{i(\theta+\pi)}^{ex} - Act_{i(\theta+\pi)}^{inh}\right]^{+} \times y_{i\theta}^{V1_Norm} \\ es_{i\theta} &= es_{i\theta}^{left} + es_{i\theta}^{right} \end{aligned} \tag{B.4}$$

$\theta = 0,\ \pi / n_{orient},\ ...,\ (n_{orient} - 1)\pi / n_{orient}$

Detection of Corners and T-junctions: Model Equations

The product of the end-stop activities at orientations θ and $\theta + \pi/2$, subsequently summed over all orientations, signals the likely positions of contour corners in the input image:

$$Act_i^{corner} = \sum_{k=1}^{n_{orient}/2} es_{i\ (k-1)\times\pi / n_{orient}}^{sm} \times es_{i\ (k-1)\times\pi / n_{orient}+\pi/2}^{sm} \tag{B.5}$$

with $es_{i\theta}^{sm} = \left\{es * \Psi_{corner} * \Lambda^{corner}\right\}_{i\theta}$

$\Psi_{corner} : \sigma_{corner_orient} = 0.1;\ \ \Lambda^{corner} : \sigma_{corner} = 1.5$

Similarly, the product of the end-stop activities at orientation θ and the V2 long-range activity at orientation $\theta + \pi/2$ signals candidate positions of T-junctions:

$$Act_i^{T} = \sum_{k=1}^{n_{orient}} es_{i_{(k-1)\times\pi / n_{orient}}}^{sm} \times LR_{i\ (k-1)\times\pi / n_{orient}+\varphi}^{V2_sm} \tag{B.6}$$

with $LR_{i\theta}^{V2_sm} = \left\{LR^{V2} * \Psi_T * \Lambda^{T}\right\}_{i\theta}$

$\varphi = \pi/2 \quad \forall k \le n_{orient}/2;\quad \varphi = -\pi/2 \quad \forall k > n_{orient}/2$

$\Psi_T : \sigma_{T_orient} = 0.3;\ \ \Lambda^{T} : \sigma_T = 0.1$

The initial activities Act^{corner} and Act^{T} undergo a stage of mutual subtractive inhibition:

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

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

$\Lambda^{sm} : \sigma = 6.0$
$C_T = 0.1; \;\; C_{Cor} = 8.0$
$\alpha_{Cor} = 0.0035; \;\; \alpha_T = 0.03;$

Orientations are assigned to T-junctions by weighting $T^{unorient}$ with end-stop activities es^{sm}:

$$T_{i\theta} = \frac{T_i^{unorient} \times es_{i\theta}^{sm}}{\alpha_{T2} + T_i^{unorient} \times \sum_{k=1}^{n_{orient}} es_{i\,(k-1)\times\pi / n_{orient}}^{sm}} \tag{B.9}$$

$\alpha_{T2} = 0.09$

The end-stop activities at contour corners are isolated by weighting es^{left} and es^{right} (eq. B.4) with corner activity Cor:

$$\begin{aligned} es_{i\theta}^{Corner_left} &= es_{i\theta}^{left} \times Cor_i \\ es_{i\theta}^{Corner_right} &= es_{i\theta}^{right} \times Cor_i \end{aligned} \tag{B.10}$$

B.3 Interactions between Neighboring Dipoles: Supplements

Deriving the Normalized V2 Bipole Activity y^{V2_Norm} from y^{V2}

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

$$y_{i\theta}^{V2_Norm} = \frac{C_{Eq} \left[y_{i\theta}^{V2} - \beta_{Eq} \sum_{k=1}^{n_{orient}} y_{i(k-1)\times\pi / n_{orient}}^{V2} \right]^+}{\alpha_{Eq} + \left[y_{i\theta}^{V2} - \beta_{Eq} \sum_{k=1}^{n_{orient}} y_{i(k-1)\times\pi / n_{orient}}^{V2} \right]^+} \tag{B.11}$$

$n_{orient} = 8$ (# of orientations); $\theta = 0, \pi / n_{orient}, ..., (n_{orient} - 1)\pi / n_{orient}$
$C_{Eq} = 1.6; \;\; \alpha_{Eq} = 0.015; \;\; \beta_{Eq} = 0.1$

Spatial Layout of Kernels K

Kernel K 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:

$$K_{i\theta} = \frac{\Lambda_{\sigma_x,\sigma_y,0,0,\theta}(i)}{\alpha_K + \Lambda_{\sigma_x,\sigma_y,0,0,\theta}(i)} \times \left(\sum_j \frac{\Lambda_{\sigma_x,\sigma_y,0,0,\theta}(j)}{\alpha_K + \Lambda_{\sigma_x,\sigma_y,0,0,\theta}(j)} \right)^{-1} \tag{B.12}$$

$\Lambda_{\sigma_x,\sigma_y,0,0,\theta}$: $\sigma_{_x} = 13.0;\ \sigma_{_y} = 0.7$
$\alpha_K = 0.004$

The kernel is cut into the two halves K^{left} and K^{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. 3.6B & 3.7; details are depicted below):

$$K_{i\theta}^{left} = \begin{cases} K_{i\theta} & \forall\ \vec{x}_i \cdot \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} < 0 \\ 0 & \forall\ \vec{x}_i \cdot \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \geq 0 \end{cases} \qquad K_{i\theta}^{right} = \begin{cases} K_{i\theta} & \forall\ \vec{x}_i \cdot \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \geq 0 \\ 0 & \forall\ \vec{x}_i \cdot \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} < 0 \end{cases} \tag{B.13}$$

$\vec{x}_i$: cartesian coordinates of point i

Integration of Dipole Activity at Contour Corners

As outlined in section 3.2.4, the recurrent interactions between dipoles result in waves of activity running 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 enhancing the maximal value C_{max} of the internal activity b_t (eq. 3.2) at spatial positions corresponding to corners:

$$C_{\max}(i) = C_{max_const} + C_{corner_amp} \times Cor_i \tag{B.14}$$

$C_{max_const} = 1.0;\ C_{corner_amp} = 2.4$

Index i depicts the spatial location. Cor_i denotes the activity of the corner detectors (see eq. B.7). Furthermore, the dipole activities $e_t^{ON/OFF}$ are enhanced by end-stop activities at contour corners and the spatial kernels K uses two the sub-kernels $K^{left/right}$ to selectively

pool those dipole activities which correspond to end-stop cells pointing in the direction of the contour (see Fig. 3.7). The modified version of equation (3.5) then reads:

$$f_{ti\theta}^{ON/OFF} = \left\{ \left[(1 + C_{es} es^{Corner_left}) \times e_t^{ON/OFF} \right] * \Psi_f * K^{left} \right\}_{i\theta} + \left\{ \left[(1 + C_{es} es^{Corner_right}) \times e_t^{ON/OFF} \right] * \Psi_f * K^{right} \right\}_{i\theta} \quad \text{(B.15)}$$

$C_{es} = 10.0$

es^{Corner_left} and es^{Corner_right} denote the activities of end-stop cells at contour corners (eq. B.10). In consequence, the terms $(1 + C_{es} es^{Corner_left/Corner_right})$ selectively enhance the impact of dipole output activities $e_t^{ON/OFF}$ at contour corners. K^{left} selectively pools dipole activity e_t modulated by es^{Corner_left}, while K^{right} selectively pools dipole activity e_t modulated by es^{Corner_right}. This restricts the impact of es^{Corner_left} and es^{Corner_right} to the contour to which the actual corner belongs to and prevents a wave of activity to jump over to neighboring contours.

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}	100000
3.5	Std-Dev of Gaussian Ψ_f: σ_{f_orient} =	0.25
3.8	C_{ol}	7.2

Table B.3: Constants of the equations delineating the model of recurrent depth sorting in chapter 3.2.

B.4 Impact of the Model T-junction Detectors on Model Dipole Activity

In section 3.2.6 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, I will outline (i) how activations $q_{ti}^{foregr/backgr}$ 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 section 3.2.6, activations $q_{ti}^{foregr/backgr}$ 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 running 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:

$$T_{ti\theta}^{\mathrm{mod}} = T_{i\theta} \times \frac{1}{\left(1 + \left[t - \beta_q\right]^+\right)^2} \tag{B.16}$$

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 $\#\mathrm{K}$, 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:

$$u_{ti} = \left[\frac{1 - \beta_u \times \sum_{j \in outerlayers} d_{ti}^{ON_layer(j)}}{\alpha_u + \sum_{j \in outerlayers} d_{ti}^{ON_layer(j)}}\right]^+ \tag{B.17}$$

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

$$v_{ti} = [u_{ti} - \beta_v \, u_{(t-1)i}]^+ \tag{B.18}$$

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

$$\begin{aligned} w_{ti} &= ([\beta_w w_{(t-1)i} + C_v v_{ti}]^{\le 1})^2 \\ T_{ti\theta}^{\mathrm{mod}} &= T_{i\theta} \times [C_w w_{ti}]^{\le 1} \end{aligned} \tag{B.19}$$

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 y^{V2_Norm} and summed over all orientations:

$$\begin{aligned} p_{ti}^{foregr} &= \sum_{k=1}^{n_{orient}} \left(y^{V2_Norm}_{i(k-1)\times\pi/n_{orient}} \times \left\{ T_t^{\mathrm{mod}} * \Psi_p * \Lambda \right\}_{i((k-1)\times\pi/n_{orient}+\pi/2)} \right) \\ p_{ti}^{backgr} &= \sum_{k=1}^{n_{orient}} \left(y^{V2_Norm}_{i(k-1)\times\pi/n_{orient}} \times \left\{ T_t^{\mathrm{mod}} * \Psi_p * \Lambda \right\}_{i(k-1)\times\pi/n_{orient}} \right) \end{aligned} \tag{B.20}$$

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

$$q_{ti}^{foregr/backgr} = \frac{C_T \; p_{ti}^{foregr/backgr}}{\alpha_q + p_{ti}^{foregr/backgr}} \tag{B.21}$$

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

Bibliography

Adelson, E. H. and Anandan, P., 1990. Ordinal Characteristics of Transparency. AAAI Workshop on Qualitative Vision. 77-81.

Anderson, B. L., 1997. A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. Perception. 26, 419-453.

Banks, M. S., Gepshtein, S. and Landy, M. S., 2004. Why is Spatial Stereoresolution So Low? The Journal of Neuroscience. 24, 2077-2089.

Baumann, R., van der Zwan, R. and Peterhans, E., 1997. Figure-ground segregation at contours: A neural mechanism in the visual cortex of the alert monkey. European Journal of Neuroscience. 9, 1290-1303.

Bayerl, P. and Neumann, H., 2004. Disambiguating Visual Motion Through Contextual Feedback Modulation. Neural Computation. 16, 2041–2066.

Baylis, G. C. and Driver, J., 2001. Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. Nat Neurosci. 4, 937-942.

Beck, J., 1966. Effect of orientation and of shape similarity on perceptual grouping. Perception & Psychophysics. 1, 300-302.

Beck, J., 1982. Textural Segmentation. In: Beck, J. (Ed.), Organization and Representation in Perception. Lawrence Erlbaum Associates, Hillsdale, pp. 285-317.

Beck, J., 1983. Texture Segmentation, Second-Order Statistics and Textural Elements. Biological Cybernetics. 48, 125-130.

Bergen, J. R. and Adelson, E. H., 1988. Early vision and texture perception. Nature. 333, 363-364.

Bergen, J. R., 1991. Theories of Visual Texture Perception. In: Regan, D. (Ed.), Spatial Vision, vol.10. Macmillan Press, pp. 115-134.

Boselie, F., 1994. Local and global factors in visual occlusion. Perception. 23, 517-528.

Cavanaugh, J. R., Bair, W. and Movshon, J. A., 2002a. Nature and Interaction of Signals From the Receptive Field Center and Surround in Macaque V1 Neurons. Journal of Neurophysiology. 88, 2530-2546.

Cavanaugh, J. R., Bair, W. and Movshon, J. A., 2002b. Selectivity and Spatial Distribution of Signals From the Receptive Field Surround in Macaque V1 Neurons. Journal of Neurophysiology. 88, 2547-2556.

Cohen, M. A. and Grossberg, S., 1983. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems on Man and Cybernetics. SMC-13, 815-826.

Cohen, M. A. and Grossberg, S., 1983b. Some global properties of binocular resonances: Disparity matching, filling in, and figure ground synthesis. In: Caelli, T. and Dodwell, P. (Eds.), Figural synthesis. Hillsdale, NJ, pp. 117-151.

Cohen, M. A., 1988. Sustained Oscillations in a Symmetric Cooperative-Competitive Neural Network: Disproof of a Conjecture About Content Addressable Memory. Neurol Networks. 1, 217-221.

Conway, B. R., Hubel, D. H. and Livingstone, M. S., 2002. Color contrast in macaque V1. Cerebral Cortex. 12, 915-925.

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

de Weerd, P., Desimone, R. and Ungerleider, L. G., 1996. Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. Visual Neuroscience. 13, 529-538.

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, A. K., Fries, P. and Singer, W., 2001. Dynamic Predictions: Oscillations and Synchrony in Top-Down Processing. Nat Rev Neurosci. 2, 704-716.

Felleman, D. J. and van Essen, D. C., 1991. Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cerebral Cortex. 1, 1-47.

Finkel, L. H. and Sajda, P., 1992. Object Discrimination Based on Depth-from-Occulsion. Neural Computation. 4, 901-921.

Frien, A. and 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, 1466-1478.

Fukushima, K., 2001. Recognition of partly occluded patterns: a neural network model. Biol Cybern. 84, 251-259.

Gallant, J. L., van Essen, D. C. and Nothdurft, H. C., 1995. Two-dimensional and three-dimensional texture processing in visual cortex of the Macaque monkey. In: Papathomas, T. V. (Ed.), Early Vision and Beyond. The MIT Press, Cambridge.

Gilbert, C. D. and Wiesel, T. N., 1989. Columnar specificity of intrinisic horizontal and corticocortical connections in cat visual cortex. Journal of Neuroscience. 9, 2432-2442.

Girard, P., Hupe, J.M. and Bullier, J., 2001. Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. J Neurophysiol. 85, 1328-1331.

Graham, N., Sutter, A. and Venkatesan, C., 1993. Spatial-frequency- and Orientation-Selectivity of Simple and Complex Channels in Region Segregation. Vision Research. 33, 1893-1911.

Grön, G., Neumann, H., Spitzer, M., Wunderlich, A. and Thielscher, A., 2003. Texture segmentation in human perception: a combined modeling and fMRI study. Proceedings of the Proceedings of Human Brain Mapping, pp.

Grossberg, S. and Grunewald, A., 2002. Temporal dynamics of binocular disparity processing with corticogeniculate interactions. Neural Networks. 15, 181-200.

Grossberg, S. and Mingolla, E., 1985. Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. Perception & Psychophysics. 38, 141-171.

Grossberg, S. and Mingolla, E., 1985. Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. Perception & Psychophysics. 38, 141-171.

Grossberg, S. and Raizada, R. D. S., 2000. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Vision Research. 40, 1413-1432.

Grossberg, S. and Williamson, J. R., 2001. A neural model of how horizontal and interlaminar conncetions of visual cortex develop into adult circuits that carry out perceptual grouping and learning. Cerebral Cortex. 1, 37-58.

Grossberg, S., 1973. Contour enhancement, short-term memory, and constancies in reverberating neural networks. Studies in Applied Mathematics. 52, 213-257.

Grossberg, S., 1980. How does a brain build a cognitive code? Psychological Review. 87, 1-51.

Grossberg, S., 1988. Nonlinear neural networks: Principles, mechanisms, and architectures. Neural Networks. 1, 17-61.

Grossberg, S., 1991. Why do parallel cortical systems exist for the 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 Networks. 6, 463-483.

Hansen, T. and Neumann, H., 1999. A model of V1 visual contrast processing utilizing long-range connections and recurrent interactions. Proceedings of the ICANN 99, Edinburgh, UK, pp.

Heider, B., Meskenaite, V. and Peterhans, E., 2000. Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours. European Journal of Neuroscience. 12, 4117-4130.

Heider, B., Spillmann, L. and Peterhans, E., 2002. Stereoscopic Illusory Contours - Cortical Neuron Responses and Human Perception. J Cogn Neurosci. 14, 1018-1029.

Heitger, F., v. d. Heydt, R., Peterhans, E., Rosenthaler, L. and Kübler, O., 1998. Simulation of neural contour mechanisms: Representing anomalous contours. Image and Vision Computing. 6, 407-421.

Hirsch, J. A. and Gilbert, C. D., 1991. Synaptic physiology of horizontal connections in the cat's visual cortex. Journal of Neuroscience. 11, 1800-1809.

Hirsch, M. W., 1989. Convergent Activation Dynamics in Continuous Time Networks. Neural Networks. 2, 331-349.

Howard, I. P. and Duke, P. A., 2003. Monocular transparency generates quantitative depth. Vision Research. 43, 2615-2621.

Howard, I. P., 2003. Neurons that respond to more than one depth cue. Trends Neurosci. 26, 515-517.

Hubel, D. H. and Wiesel, T. N., 1959. Receptive fields of single neurones in the cat's striate cortex. Journal of Physiology. 148, 574-591.

Hubel, D. H. and Wiesel, T. N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology. 160, 106-154.

Hubel, D. H. and Wiesel, T. N., 1965. Receptive fields, and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol. 28, 229-289.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P. and Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature. 394, 784-787.

Hupe, J., James, A., Girard, P. and Bullier, J., 2001. Response Modulations by Static Texture Surround in Area V1 of the Macaque Monkey Do Not Depend on Feedback Connections from V2. Journal of Neurophysiology. 85, 146-163.

Issa, N. P., Trepel, C. and Stryker, M. P., 2000. Spatial frequency maps in cat visual cortex. Journal of Neuroscience. 20, 8504-8514.

Itti, L., Koch, C. and Niebur, E., 1998. A model of saliency-based visual attention for rapid analysis. IEEE PAMI. 20, 1254-1259.

Julesz, B., 1965. Texture and visual perception. Scientific American. 212, 38-48.

Julesz, B., 1981. Textons, the elements of texture perception, and their interactions. Nature. 290, 91-97.

Julesz, B., 1986. Texton Gradients: The Texton Theory Revisted. Biological Cybernetics. 54, 245-251.

Kandel, E. R., Schwartz, J. H. and Jessell, T. M., 2000. Principles of neural science. McGraw-Hill, New York.

Kapadia, M. K., Ito, M., Gilbert, C. D. and Westheimer, G., 1995. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron. 15, 843-856.

Kapadia, M. K., Westheimer, G. and Gilbert, C. D., 2000. Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception. Journal of Neurophysiology. 84, 2048-2062.

Kastner, S., de Weerd, P. and Ungerleider, L. G., 2000. Texture Segregation in the Human Visual Cortex: A Functional MRI Study. Journal of Neurophysiology. 83, 2453-2457.

Kastner, S., de Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R. and Ungerleider, L. G., 2001. Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. Journal of Neurophysiology. 86, 1398-1411.

Kehrer, L. and Meinecke, C., 2003. A space-variant filter model of texture segregation: parameter adjustment guided by psychophysical data. Biol Cybern. 88, 183-200.

Kehrer, L., 1997. The central performance drop in texture segmentation: a simulation based on a spatial filter model. Biological Cybernetics. 77, 297-305.

Kellman, P. J. and Shipley, T. F., 1991. A theory of visual interpolation in object perception. Cognit Psychol. 23, 141-221.

Kellman, P. J., Yin, C. and Shipley, T. F., 1998. A common mechanism for illusory and occluded object completion. J Exp Psychol Hum Percept Perform. 24, 859-869.

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

Kobatake, E. and Tanaka, K., 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. Journal of Neurophysiology. 71, 856-867.

Kölle, M., Thielscher, A., Herrnberger, B., Meier, C., Neumann, H., Spitzer, M. and Grön, G., 2004. The visual area V4 and pop out search: an approach combining fMRI and neural network modeling. Proceedings of the Proceedings of Human Brain Mapping, Budapest Hungary, pp.

Kovacs, G., Vogels, R. and Orban, G. A., 1995. Selectivity of Macaque Inferior Temporal Neurons for Partially Occluded Shapes. J Neurosci. 15, 1984-1997.

Kumaran, K., Geiger, D. and Gurvits, L., 1996. Illusory surface perception and visual organization. Network: Computation in Neural Systems. 7, 33-60.

Lamme, V., Rodriguez-Rodriguez, V. and Spekreijse, H., 1999. Seperate Processing Dynamics for Texture Elements, Boundaries and Surfaces in Primary Visual Cortex of the Macaque Monkey. Cerebral Cortex. 9, 406-413.

Lamme, V., Super, H. and Spekreijse, H., 1998. Feedforward, horizontal, and feedback processing in the visual cortex. Current Opinion in Neurobiology. 8, 529-535.

Landy, M. S. and Bergen, J. R., 1991. Texture segregation and orientation gradient. Vision Research. 31, 679-691.

Lee, T. S., 1995. A bayesian framework for understanding texture segmentation in the primary visual cortex. Vision Research. 35, 2643-2657.

Lennie, P., Krauskopf, J. and Sclar, G., 1990. Chromatic mechanisms in striate cortex of macaque. Journal of Neuroscience. 10, 649-669.

Li, W., Thier, P. and Wehrhahn, C., 2000. Contextual Influence in Orientation Discrimination of Humans and Responses of Neurons in V1 of Alert Monkeys. Journal of Neurophysiology. 83, 941-954.

Li, W., Thier, P. and Wehrhahn, C., 2001. Neural responses from beyond the classic receptive field in V1 of alert monkeys. Experimental Brain Research. 139, 359-371.

Li, Z., 1998. A neural model of contour integration in the primary visual cortex. Neural Computation. 10, 903-940.

Li, Z., 2002. A saliency map in primary visual cortex. Trends in Cognitive Sciences. 6, 9-16.

Liter, J. C. and Bülthoff, H. H., 1996. An Introduction to Object Recognition), pp. 1-12. MPI für biologische Kybernetik, Tübingen.

Liu, X. and Wang, D. L., 1999. Perceptual Organization based on Temporal Dynamics. Paper presented at the IJCNN'99, Washington DC, USA.

Malik, J. and Perona, P., 1990. Preattentive texture discrimination with early vision mechanisms. Journal of the Optical Society of America A. 7, 923-932.

Mansson, J., 2000. Occluding Contours: A Computational Model of Suppressive Mechanisms in Human Contour Perception. Proceedings of the Lund University Cognitive Studies, Lund, 81

Marr, D. and Poggio, T., 1979. A computational theory of human stereo vision. Proceedings of the Royal Society of London B. 204, 301-328.

Mazer, J. A. and Gallant, J. L., 2003. Goal-Related Activity in V4 during Free Viewing Visual Search: Evidence for a Ventral Stream Visual Salience Map. Neuron. 40, 1241-1250.

Meinecke, C. and Donk, M., 2002. Detection performance in pop-out tasks: Nonmonotonic changes with display size and eccentricitiy. Perception. 31, 591-602.

Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K. and Tootell, R. B. H., 1999. The Representation of Illusory and Real Contours in Human Cortical Visual Areas Revealed by Functional Magnetic Resonance Imaging. The Journal of Neuroscience. 19, 8560-8572.

Merigan, W. H., 1996. Basic visual capacities in shape discrimination after lesions of extrastriate area V4 in macaques. Visual Neuroscience. 13, 51-60.

Merigan, W. H., 2000. Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. Visual Neuroscience. 17, 949-958.

Mignard, M. and Malpeli, J. G., 1991. Paths of information flow through visual cortex. Science. 251, 1249-1251.

Mumford, D. B., 1994. Neuronal Architectures for Pattern-theoretic Problems. In: Koch, C. and Davis, J. (Eds.), Large-Scale Neuronal Theories of the Brain. MIT Press, pp. 125-152.

Naka, K. and Rushton, W., 1966. S-potentials from colour units in the retina of fish (cyprinidae). Journal of Physiology. 185, 536-555.

Nakamura, H., Gattass, R., Desimone, R. and Ungerleider, L. G., 1993. The Modular Organization of Projections from Areas V1 and V2 to Areas V4 and TEO in Macaques. The Journal of Neuroscience. 13, 3681-3691.

Nakayama, K., Shimojo, S. and Ramachandran, V. S., 1990. Transparency: relation to depth, subjective contours, luminance, and neon color spreading. Perception. 19, 497-513.

Nakayama, K., Shimojo, S. and Silverman, G. H., 1989. Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. Perception. 18, 55-68.

Neumann, H. and Mingolla, E., 2001. Computational neural models of spatial integration in perceptual grouping. In: Shipley, T. F. and Kellman, P. J. (Eds.), From fragments to objects - Segmentation and grouping in vision. Elsevier, Amsterdam, pp. 353-400.

Neumann, H. and Sepp, W., 1999. Recurrent V1-V2 interaction in early visual boundary processing. Biological Cybernetics. 81, 425-444.

Nishina, S., Okada, M. and Kawato, M., 2003. Spatio-temporal dynamics of depth propagation on uniform region. Vision Research. 42, 2493-2503.

Nothdurft, H. C., 1985. Sensitivity for structure gradient in texture discrimination tasks. Vision Research. 25, 1957-1968.

Nothdurft, H. C., 1991. Texture segmentation and pop-out from orientation contrast. Vision Research. 31, 1073-1078.

Nothdurft, H. C., 1992. Feature analysis and the role of similarity in preattentive vision. Perception & Psychophysics. 52, 355-375.

Nothdurft, H. C., 2000c. Salience from feature contrast: variations with texture density. Vision Research. 40, 3181-3200.

Nothdurft, H. C., Gallant, J. L. and van Essen, D. C., 1999. Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. Visual Neuroscience. 16, 15-34.

Nothdurft, H. C., Gallant, J. L. and van Essen, D. C., 2000. Response profiles to texture border patterns in V1. Visual Neuroscience. 17, 421-436.

Ohzawa, I., DeAngelis, G. C. and Freeman, R. D., 1997. The Neural Coding of Stereoscopic Depth. NeuroReport. 8, iii-xiii.

Ontrup, J., Wersing, H. and Ritter, H., 2001. A computational feature binding model reproducing human texture perception. submitted.

Pasupathy, A. and Connor, C. E., 2001. Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation. The Journal of Neuroscience. 86, 2505-2519.

Pessoa, L., Mingolla, E. and Neumann, H., 1995. A contrast- and luminance-driven multiscale network model of brightness perception. Vision Research. 35, 2201-2223.

Peterhans, E. and Heitger, F., 2001. Simulation of Neuronal Responses Defininig Depth Order and Contrast Polarity at Illusory Contours in Monkey Area V2. J Comput Neurosci. 10, 195-211.

Peterhans, E. and v. d. Heydt, R., 1991. Subjective contours - Bridging the gap between psychophysics and physiology. Trends in Neurosciences. 14, 112-119.

Peterhans, E., 1997. Functional Organization of Area V2 in the Awake Monkey. In: Rockland, K. S. et al. (Eds.), Extrastriate Cortex in Primates, vol.12. Plenum Press, New York.

Pianta, M. J. and Gillam, B. J., 2003. Paired and unpaired features can be equally e.ective in human depth perception. Vision Research. 43, 1-6.

Poggio, G. F., Gonzalez, F. and Krause, F., 1988. Stereoscopic Mechanisms in Monkey Visual Cortex: Binocular Correlation and Disparity Selectivity. J Neurosci. 8, 4531-4550.

Pollen, D. A., Przybyszewski, A. W., Rubin, M. A. and Foote, W., 2002. Spatial Receptive Field Organization of Macaque V4 Neurons. Cerebral Cortex. 12, 601-616.

Przybyszewski, A. W., Gaska, J. P., Foote, W. and Pollen, D. A., 2000. Striate cortex increases contrast gain of macaque LGN neurons. Vis Neurosci. 17, 485-494.

Regan, D., Erkelens, C. J. and Collewijn, H., 1986. Visual field defects for vergence eye movements and for stereomotion perception. Invest Ophthalmol Vis Sci. 27, 806-819.

Reppas, J. B., Niyogi, S., Dale, A. M., Sereno, M. I. and Tootell, R. B., 1997. Representation of motion boundaries in retinotopic human visual cortical areas. Nature. 388, 175-179.

Riesenhuber, M. and Poggio, T., 1999. Hierarchical Models of Object Recognition in Cortex. Nature Neuroscience. 2, 1019-1025.

Rockland, K. S., Saleem, K. S. and Tanaka, K., 1994. Divergent feedback connections from areas V4 and TEO in the macaque. Visual Neuroscience. 11, 579-600.

Ross, W. D., Grossberg, S. and Mingolla, E., 2000. Visual cortical mechanisms of perceptual grouping: Interacting layers, networks, columns, and maps. Neural Networks. 13, 571-588.

Rubin, N., 2001. Figure and ground in the brain. Nat Neurosci. 4, 857-858.

Rubin, N., 2001b. The role of junctions in surface completion and contour matching. Perception. 30, 339-366.

Sajda, P. and Finkel, L. H., 1995. Intermediate-Level Visual Representations and the Construction of Surface Perception. J Cogn Neurosci. 7, 267-291.

Salin, P. A. and Bullier, J., 1995. Corticocortical connections in the visual system: structure and function. Physiol Rev. 75, 107-154.

Sandell, J. H. and Schiller, P. H., 1982. Effect of cooling area 18 on striate cortex cells in the squirrel monkey. J Neurophysiol. 48, 38-48.

Schubö, A., 2002. Is preattentive processing in visual search similar to preattentive processing in texture segmentation? In: Bülthoff, H. H. et al. (Eds.). Proceedings of the TWK, Tübingen. Knirsch Verlag, pp. 191.

Schubö, A., Meinecke, C. and Schröger, E., 2001. Automaticity and attention: investigating automatic processing in texture segmentation with event-related brain potentials. Cogn Brain Res. 11, 341-361.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R. and Tootell, R. B. H., 1995. Borders of multiple visual areas in humans revealed by functional MRI. Science. 268, 889-893.

Shadlen, M. N. and Movshon, J. A., 1999. Synchrony Unbound: A Critical Evaluation of the Temporal Binding Hypothesis. Neuron. 24, 67-77.

Shimojo, S. and Nakayama, K., 1990. Real world occlusion constraints and binocular rivalry. Vision Research. 30, 69-80.

Shipley, T. F. and Kellman, P. J., 1990. The role of discontinuities in the perception of subjective figures. Percept Psychophys. 48, 259-270.

Shipley, T. F. and Kellman, P. J., 1992. Strength of visual interpolation depends on the ratio of physically specified to total edge length. Percept Psychophys. 52, 97-106.

Singh, M. and Huang, X., 2003. Computing Layered Surface Representations: An Algorithm for Detecting and Separating Transparent Overlays. Paper presented at the IEEE CVPR'03, Wisconsin USA.

Smith, A. T., Singh, K. D., Williams, A. L. and Greenlee, M. W., 2001. Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. Cerebral Cortex. 11, 1182-1190.

Sporns, O., Tononi, G. and Edelman, G. M., 1991. Modeling Perceptual Grouping and Figure-Ground Segregation by Means of Active Reentrant Connections. PNAS. 88, 129-133.

Spratling, M. W. and Johnson, M. H., 2001. Dendritic Inhibition Enhances Neural Coding Properties. Cerebral Cortex. 11, 1144-1149.

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

Thielscher, A. and Neumann, H., 2005. Neural mechanisms of human texture processing: Texture boundary detection and visual search. Spatial Vision. 18, 227-257.

Thomas, O. M., Cumming, B. G., & Parker, A. J., 2002. A specialization for relative disparity in V2. Nat Neurosci. 5, 472-478.

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B., Sereno, M. I. and Dale, A. M., 1997. Functional Analysis of V3A and Related Areas in Human Visual Cortex. The Journal of Neuroscience. 17, 7060-7078.

Tootell, R. B. H., Silverman, M. S., Switkes, E. and deValois, R. L., 1982. Deoxyglycose analysis of retinotopic organization in primate striate cortex. Science. 218, 902-904.

Ungerleider, L. G. and Haxby, J. V., 1994. 'What' and 'where' in the human brain. Current Opinion in Neurobiology. 4, 157-165.

v. d. Heydt, R., Heitger, F. and Peterhans, E., 1993. Perception of occluding contours: Neural mechanisms and a computational model. Biomedical Research. 14, 1-6.

v. d. Heydt, R., Peterhans, E. and Baumgartner, G., 1984. Illusory contours and cortical neuron responses. Science. 224, 1260-1262.

Williams, L. R. and Hanson, A. R., 1996. Perceptual Completion of Occluded Surfaces. Computer Vision and Image Understanding. 64, 1-20.

Williamson, J. R., 1996. Neural Network for Dynamic Binding with Graph Representation: From, Linking, and Depth-from-Occlusion. Neural Computation. 8, 1203-1225.

Witkin, A. P. and Tenenbaum, J. S., 1983. On the role of structure in vision. In: Beck, J. et al. (Eds.), Human and Machine Vision. Academic Press, New York.

Zeki, S. M., 1993. A Vision of the Brain. Blackwell Science, Oxfort.

Zhou, H., Friedman, H. S. and v. d. Heydt, R., 2000. Coding of border ownership in monkey visual cortex. J Neurosci. 20, 6594-6611.

Zilles, K. and Clarke, S., 1997. Architecture, Connectivity, and Transmitter Receptors of Human Extrastriate Visual Cortex - Comparison with Nonhuman Primates. In: Rockland, K. S. et al. (Eds.), Extrastriate Cortex in Primates, vol.12. Plenum Press, New York.