# A Neural Model of Early Vision: Contrast, Contours, Corners and Surfaces

**Contributions toward an Integrative Architecture of Form and Brightness Perception**

**Thorsten Hansen**


**University of Ulm**
**Faculty of Computer Science**
**Dept. of Neural Information Processing**

# A Neural Model of Early Vision: Contrast, Contours, Corners and Surfaces

Contributions toward an Integrative Architecture of Form and Brightness Perception

# A Neural Model of Early Vision: Contrast, Contours, Corners and Surfaces

## Contributions toward an Integrative Architecture of Form and Brightness Perception

**Thorsten Hansen**

aus Leer

## Dissertation

zur Erlangung des Doktorgrades Dr. rer. nat.

2003


**University of Ulm**
**Faculty of Computer Science**
**Dept. of Neural Information Processing**

| | |
|---|---|
| Dekan: | Prof. Dr. Günther Palm |
| Erster Gutachter: | Prof. Dr. Heiko Neumann |
| Zweiter Gutachter: | Prof. Dr. Günther Palm |
| | |
| Tag der Promotion: | 22. September 2002 |

**Abstract**

The thesis is concerned with the functional modeling of information processing in early and mid-level vision. The mechanisms can be subdivided into two systems, a system for the processing of discontinuities (such as contrast, contours and corners), and a complementary system for the representation of homogeneous surface properties such as brightness.

For the robust processing of oriented *contrast signals*, a mechanism of *dominating opponent inhibition* (DOI) is proposed and integrated into an existing nonlinear simple cell model. We demonstrate that the model with DOI can account for physiological data on luminance gradient reversal. For the processing of both natural and artificial images we show that the new mechanism results in a significant suppression of responses to noisy regions, largely independent of the noise level. This adaptive processing is further examined by a stochastic analysis and numerical evaluations. We also show that contrast-invariant orientation tuning can be achieved in a purely feedforward model based on inhibition. DOI results in a sharpening of the tuning width of model simple cells which are in accordance with empirical findings. The results lead to the proposal of a new functional role of the dominant inhibition as observed empirically, namely to sharpen the orientation tuning and to allow for robust contrast processing under suboptimal, noisy viewing conditions.

For the processing of *contours*, a model of *recurrent colinear long-range interaction in V1* is proposed. The key properties of the model are excitatory long-range interactions between cells with colinear receptive fields, inhibitory unoriented short-range interactions and modulating feedback. In the model, initial noisy feedforward responses which are part of a more global contour are enhanced, while other responses are suppressed. We show for a number of artificial and natural images that the recurrent long-range processing results in a selective enhancement of coherent activity at contour locations. The competencies of the model are further quantitatively evaluated using two measures of contour saliency and orientation significance. The model also qualitatively reproduces empirical data on surround suppression and facilitation. We further suggest and examine a model variant using early feedback of grouped responses, showing an even stronger enhancement of contour saliency compared to the standard model. These results may suggest a functional role for the layout of different feedback projections in V1.

Regions of intrinsically 2D structures such as *corners and junctions* are important for both biological and artificial vision systems. We propose a novel scheme for the robust and reliable representation and detection of junction points, where junctions are characterized by high responses at multiple orientations within an model hypercolumn. The recurrent long-range interaction results in a robust extraction of orientation information. A measurement of circular variance is used to detect and localize junctions. We show for a number of artificial and natural images that localization accuracy and positive correctness of the junction detection is improved compared to a purely feedforward computation of contour orientations. We also use ROC analysis to compare the new scheme with two other junction detector schemes based on Gaussian curvature and the structure tensor, showing that the new approach performs superior to the standard schemes.

*Brightness surfaces* are reconstructed by a diffusive spreading or filling-in of sparse contrast measurements, which is locally controlled by contour signals. A mechanism of *confidence-based filling-in* is proposed, where a confidence measure ensures a robust selection of sparse contrast signals. We show that the model with confidence-based filling-in can generate brightness surfaces which are invariant against size and shape transformations and can also generate smooth brightness surfaces even from noisy contrast data, in contrast to standard filling-in. The model can also account for psychophysical data on human brightness perception. We further suggest a new approach for the reconstruction of reference levels, where sparse contrast signals are modulated to carry an additional luminance component. We show for a number of test stimuli that the newly proposed scheme can successfully predict human brightness perception.

Overall, we show that basic tasks in early vision can be robustly and efficiently implemented by biologically motivated mechanisms. This leads to a deeper understanding of the functional role of the particular mechanisms and provides the basis for practical applications in technical vision systems.

**Zusammenfassung**

Die Dissertation behandelt die Modellierung der Informationsverarbeitung bei der frühen visuellen Wahrnehmung. The verschiedenen Mechanismen lassen sich zwei Systemen zuordnen: ein System zur Verarbeitung von Diskontinuitäten (Kontraste, Konturen, Ecken) und ein komplementäres System zur Repräsentation homogener Bereiche (Oberflächen).

Im Bereich der *Kontrastverarbeitung* wird motiviert durch empirische Daten ein Mechanismus der dominanten opponenten Inhibition (DOI) vorschlagen und in ein bestehendes nichtlineares simple cell Modell integriert. Der neue Mechanismus erlaubt die Simulation experimentell gemessener Antworten auf Hell-dunkel-Balken. Bei der Verarbeitung von natürlichen und synthetischen Bildern führt der Mechanismus zu einer signifikanten Unterdrückung von Rauschen, weitgehend unabhängig von der Höhe des Rauschlevels. Diese adaptive Verarbeitung wird in einer stochastischen Analyse und in umfangreichen numerischen Evaluationen ausführlich untersucht. Weiter wird gezeigt das kontrast-invariantes Orientierungstuning durch eine reine feedforward Verarbeitung realisiert werden kann. DOI führt dabei zu einer Verschärfung der Tuningkurven in Übereinstimmung mit empirischen Daten. Aufgrund der Ergebnisse wird als eine neue funktionelle Rolle der empirisch beobachteten dominanten Inhibition die Verschärfung des Orientierungstunings und die robuste Verarbeitung von orientierten Kontrastsignalen vorgeschlagen.

Die *Konturverarbeitung* wird realisiert durch ein rekurrentes Modell langreichweitiger, kolinearer Verbindungen im primären visuellen Kortex. Wesentliche Mechanismen sind exzitatorische langreichweitige Interaktionen zwischen Zellen mit kolinearen rezeptiven Feldern, inhibitorische orientierungsunspezifische kurzreichweitige Interaktionen und modulatorisches Feedback. In dem Modell werden initiale, verrauschte feedforward Messungen durch modulatorisches Feedback verstärkt, wenn sie Teil einer zusammenhängenden Kontur sind, andernfalls abgeschwächt. Diese Verstärkung kohärenter Konturaktivität wird für eine Reihe von natürlichen und synthetischen Bildern gezeigt. Zur quantitativen Evaluierung der Modelleigenschaften wird ein Maß für Salienz und Orientierungsvarianz verwendet. Das Modell erlaubt ausserdem die Reproduktion empirischer Daten zu Kontexteffekten. Weiter wird eine Modellvariante vorgeschlagen und untersucht, bei der eine frühe Rückkopplung der gruppierten Kontursignale verwendet wird, was zu einer noch größeren Verstärkung der Salienz von Konturen führt.

Regionen mit intrinsisch zweidimensionaler Struktur wie *Ecken und Kreuzungspunkte* spielen eine wichtige Rolle für biologische und artifizielle Sehsysteme. Wir schlagen ein neues Schema für die robuste Repräsentation und Detektion von Ecken und Kreuzungspunkten vor, basierend auf starken Antworten für mehrere Orientierungen an einem Ort, d. h. innerhalb einer kortikalen Hyperkolumne. Diese Orientierungsinformation kann robust durch rekurrente langreichweitige Interaktionen generiert werden. Ausgehend von dieser Repräsentation wird ein Maß für die Orientierungsvarianz zur Detektion und Lokalisation der Kreuzungspunkte verwendet. Wir zeigen für synthetische und natürliche Bilder, dass die Lokalisation und die positive Korrektheit durch die langreichweitige Interaktion verbessert wird, verglichen mit einer reinen feedforward Verarbeitung. Eine ROC Analyse zeigt eine bessere Detektionsleistung des neuen Verfahrens im Vergleich mit zwei anderen Verfahren, basierend auf Gaußscher Krümmung und dem Strukturtensor.

*Helligkeitsoberflächen* werden durch eine laterale Diffusion von spärlichen Kontrastsignalen rekonstruiert, deren Ausbreitung lokal durch Kontursignale geblockt wird. Durch einen Mechanismus des konfidenzbasierten Filling-in werden Positionen, an denen keine Kontrastsignale vorliegen, nicht als Eingabe für den Diffusionsprozess genutzt. Auf diese Weise kann auch bei spärlichen Kontrasten eine Helligkeitsrepräsentation unabhängig von der Form und Größe der auszufüllenden Fläche erreicht werden, und eine glatte Helligkeitsoberfläche kann selbst bei verrauschen Kontrastsignalen generiert werden. Das Modell simuliert ausserdem die menschliche Wahrnehmung für eine Reihe von Helligkeitsillusionen. Weiterhin wird eine neuer Ansatz zur Generierung von Referenzniveaus vorgeschlagen, basierend auf spärlichen, luminanzmodulierten Kontrastsignalen. Wir zeigen für verschiedene Teststimuli, dass dieses Schema zutreffende Voraussagen in Übereinstimmung mit der menschlichen Helligkeitswahrnehmung macht.

Zusammenfassend wird in der Arbeit gezeigt, dass sich wesentliche Aufgaben der frühen visuellen Verarbeitung unter Verwendung biologisch motivierter Verarbeitungsmechanismen realisieren lassen. Dies führt zu einem Verständnis der funktionellen Rolle der verwendeten Prinzipien und bildet gleichzeitig die Grundlage für die Umsetzung und Anwendung in technischen Systemen.

# Acknowledgments

First I would like to thank the supervisor of my thesis, Prof. Dr. Heiko Neumann. Heiko provided a rich source of inspiration, ideas and knowledge and has considerably shaped the topics and methods of the thesis. His steady support and motivation, the frequent pointing to and supplying of relevant literature, along with numerous discussions were extremely helpful and encouraging for my work.

Second I would like to thank Prof. Dr. Günther Palm. Besides kindly undertaking the second expert's report of the thesis, Günther gave several valuable hints to improve the presentation of the work. As head of department, he generously provided financial support for visiting several conferences, and for fast computing equipment and hardware upgrades.

I would like to acknowledge the contributions of the members of our local vision group. I thank Dr. Gregory Baratoff who provided helpful ideas during many discussions, especially concerning the evaluation of contrast processing methods. I also thank Karl O. Riedel for numerous important discussions on various aspects of filling-in. Special thanks go to Christian Toepfer for inspiring conversations on computer vision and beyond. Finally I would like to acknowledge the valuable comments of Dr. Ingo Ahrns, Matthias S. Keil, Wolfgang Sepp and Axel Thielscher.

I thank my colleagues at the Department of Neural Information Processing for a nice and friendly atmosphere, and especially Stefan Sablatnög for constant and fast support during technical problems.

Further I would like to thank Stefanie Buchmayer for proofreading parts of the document with respect to lingual and grammatical correctness.

My final thanks are devoted with love to my parents, to my sons Paul and Jonas and especially to my wife Susanne for all her love and encouragement.

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| 1D | one-dimensional |
| 2AFC | two-alternative forced choice |
| 2D, 3D | two-, three-dimensional |
| AOS | additive operator splitting |
| BCS | boundary contour system |
| CC | cooperative-competitive |
| COC | Craik-O'Brien-Cornsweet |
| DLD | dark-light-dark |
| DOI | dominating opponent inhibition |
| DoG | difference of Gaussians |
| EPSP | excitatory postsynaptic potential |
| FCS | feature contour system |
| FFT | fast Fourier transform |
| GABA | gamma-aminobutyric acid |
| GCL | ganglion cell layer |
| HRP | horseradish peroxidase |
| HWHH | half width at half height |
| INL | inner nuclear layer |
| IPL | inner plexiform layer |
| IPSP | inhibitory postsynaptic potential |
| LDL | light-dark-light |
| LGN | lateral geniculate nucleus |
| LM | secondary visual area lateromedial (rat) |
| MT | middle temporal area of the cortex |
| ONL | outer nuclear layer |
| OPL | outer plexiform layer |
| PCG | preconditioned conjugate gradients |
| PSP | postsynaptic potential |
| RF | receptive field |
| RGC | retinal ganglion cell |
| ROC | receiver operator characteristics |
| SOR | successive overrelaxation |
| V1 | primary visual cortex |
| V1, V2,..., V5 | areas of the visual cortex |

# Notation

## Constants, Sets and General Identifiers

| | |
|---|---|
| e | Euler's number |
| i | imaginary unit, $\mathrm{i} = \sqrt{-1}$ |
| $\pi$ | $\pi = 3.14159265358979323846264338327950288419 7\ldots$ |
| | |
| $\mathbb{N}$ | natural numbers 1, 2, 3, ... |
| $\mathbb{N}_0$ | natural numbers including zero |
| $\mathbb{R}$ | real numbers |
| $\mathbb{R}^+$ | positive-valued real numbers |
| $\mathbb{C}$ | complex numbers |
| | |
| $x, y$ | space |
| $t$ | time |
| $\theta$ | orientation |

## Differentiation Operators

| | |
|---|---|
| $\partial$ | partial derivation |
| $\partial_t$ | partial derivation with respect to $t$, $\partial_t := \frac{\partial}{\partial t}$ |
| $\nabla$ | Nabla operator, generalized differentiation operator |
| $\Delta$ | Laplacian, $\Delta = \nabla^2$ |
| div | divergence operator |
| grad | gradient operator |
| $\delta$ | functional derivation |

## Statistics and Stochastics

| | |
|---|---|
| $\overline{x}$ | mean of $x$ |
| std | standard deviation |
| $\mathbf{x}$ | random variable |
| $E\{\mathbf{x}\}$ | expected value or mean of a random variable $\mathbf{x}$ |
| $f_{\mathbf{x}}$ | density function for a random variable $\mathbf{x}$ |

## Linear Algebra

| | |
|---|---|
| $(a_{ij})$ | entries of matrix $A$ |
| $A^{-1}$ | inverse of matrix $A$ |
| $A^{\mathrm{T}}$ | transpose of matrix $A$ |
| $\det A$ | determinant of matrix $A$ |
| $I$ | unit matrix |
| $\lambda_1, \lambda_2$ | eigenvalues |
| $\mathbf{v}_1, \mathbf{v}_2$ | eigenvectors |

## Signal Processing and General Functions

| | |
|---|---|
| $\lvert\cdot\rvert$ | absolute value |
| $[\cdot]^+$ | half-wave rectification |
| $*$ | convolution |
| $\star$ | correlation |
| $g$ | 1D Gaussian |
| $G$ | 2D Gaussian |
| $G_\sigma$ | 2D Gaussian with standard deviation $\sigma$ |
| $B$ | bipole filter |
| DoG | difference-of-Gaussians |
| DooG | difference-of-offset-Gaussians |
| LoG | Laplacian-of-Gaussians |
| $H$ | Heaviside function |
| circvar$(\cdot)$ | circular variance |
| osgnf$(\cdot)$ | orientation significance |

## Signal Detection Theory

| | |
|---|---|
| $t_{\mathrm{p}}, f_{\mathrm{p}}$ | true positive and false positive rate |
| $t_{\mathrm{n}}, f_{\mathrm{n}}$ | true negative and false negative rate |
| $P_{\mathrm{N}}, P_{\mathrm{SN}}$ | noise distribution and signal-plus-noise distribution |
| $\mu_{\mathrm{N}}, \mu_{\mathrm{SN}}$ | mean of noise distribution and signal-plus-noise distribution |
| $d'$ | distance between signal and signal-plus-noise distribution |
| $c$ | decision criterion |
| erf$(\cdot)$ | error function |
| erfinv$(\cdot)$ | inverted error function |
| $z(\cdot)$ | $z$-transformation, $z(p) := \mathrm{erfinv}(p)$ |

## Model Variables

| | |
|---|---|
| $I$ | input image |
| $I_{\mathrm{c}}, I_{\mathrm{s}}$ | center and surround filtered input image |
| $X$ | contrast-sensitive signals (nonzero DC level) |
| $X_{\mathrm{on}}, X_{\mathrm{off}}$ | contrast-sensitive signals for the on and off domain |
| $K$ | contrast signals (zero DC level) |
| $K_{\mathrm{on}}, K_{\mathrm{off}}$ | contrast signals for the on and off domain |
| $L$ | luminance signal |
| $R_{\mathrm{on}}, R_{\mathrm{off}}$ | on and off simple cell subfield |
| $S$ | simple cell |
| $S_{\mathrm{ld}}, S_{\mathrm{dl}}$ | light-dark and dark-light simple cell |
| $C$ | complex cell |
| $C_{\mathrm{pool}}$ | pooled complex cell responses |
| $C_{\mathrm{background}}$ | response of pooled complex cells at a background location |
| $C_{\mathrm{edge}}$ | response of pooled complex cells at an edge location |
| $V$ | combination of feedforward and feedback responses |
| $W$ | long-range responses |
| $J$ | junction responses |
| $J_{\mathrm{LR}}, J_{\mathrm{GC}}, J_{\mathrm{ST}}$ | junction responses based on long-range interaction, Gaussian curvature, and the structure tensor |
| $B$ | boundary activity |
| $P$ | permeability |
| $Z$ | confidence |
| $Z_{\mathrm{on}}, Z_{\mathrm{off}}$ | confidence values for the on and off domain |
| $U$ | filled-in brightness |
| $U_{\mathrm{on}}, U_{\mathrm{off}}$ | filled-in brightness for the on and off domain |
| $O$ | final brightness prediction |

Chapter 1

# Introduction

> And God said: "Let there be light," and there was light.
>
> Genesis 1:3

Vision is one of the most important senses of human beings. By vision, we can, for example, enjoy a beautiful sunset or a masterful art work, recognize a familiar face of a person standing at the other side of the room, localize a cup of tea to reach for, judge small deviations from perpendicular orientation of a picture at the wall, and faithfully determine subtle differences of object color and brightness under highly changing illumination conditions. All these various things can be done immediately and effortlessly, without consciousness strain of any nerve, and often in parallel. The direct nature of visual sensations disguises the complexity of the underlying processes. A large fraction of the cortex, probably more than one third of the primate brain, is concerned with visual processing.

A closer look at the very first stages of vision may elucidate the problems and the complexity of the task. Light that impinges on the retina stimulates an array of photoreceptors, coarsely similar to the pixels senses by a video camera. Based on an ever changing distribution of measured light intensities the brain has to extract invariant properties of the external world, such as objects at different distances and angles, often obscured and partly occluded. Even the first steps in this process, such as the extraction of contrast, the formation of edges and lines, the spotting of corner and junction points, and the representation of homogeneous surface qualities are far from trivial. Early approaches in computer vision have revealed the complexity of visual processing and the effort necessary to solve even the seemingly most basic tasks. Consider the situation as sketched in Fig. 1.1. In this tomographic image of a human head one can easily see edges, corners, and regions of homogeneous gray surfaces. The underlying distribution of pixel intensities, however, strongly deviates from an ideal signal: edges and corners are not straight or smooth but subject to noise, and regions which are perceived as areas of homogeneous color show considerable variations of image intensities.

Visual processing is not accessible to introspection. Empirical sciences such as psychophysics and physiology have strongly advanced our knowledge of the underlying processes. Computational and modeling approaches are helpful by integrating a multitude of findings from different disciplines into a common framework.

The present work deals with the computational modeling of early and midlevel visual processing. In particular, we address the robust detection of oriented contrasts, the grouping of contours, the extraction of corners and junctions and the representation of homogeneous surface qualities. Inspired by earlier work of Grossberg and coworkers we suggest how the various tasks can be solved within a unified architecture of interacting subsystems. The potential impact of such computational models is twofold. First, computational models integrate different empirical hypothesis and findings into a precise algorithmic description, accessible to analysis and thorough evaluation. Moreover, computational models allow to validate or even suggest the underlying functional role of an empirically discovered mechanism or wiring scheme. Second, the functional mechanisms

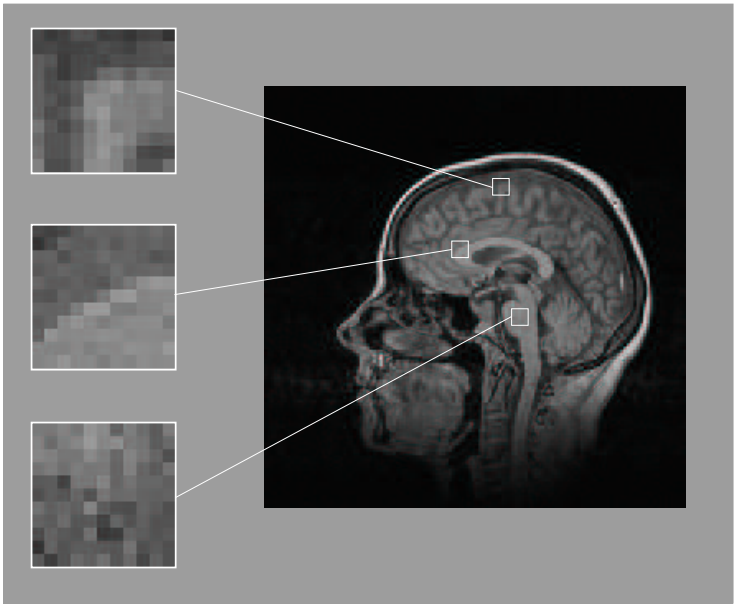

**Fig. 1.1:** A tomographic image of a human head serves to illustrate the intrinsic problems of early visual processing. In the tomographic image *(right)*, one can easily recognize edges, corners, and surfaces. However, the close-ups of the particular regions *(left inset images)* reveal that the underlying distribution of pixel intensities deviates considerably from an ideal signal: they are noisy, not smooth, and subject to high fluctuations.

found in mammalian vision, once algorithmically detailed and tested, can help to improve the performance of technical systems.

In the following, we shall first motivate the notion of vision as a constructive and creative process. We shall then give a brief survey of two important theoretical approaches to vision, the ecological approach by Gibson and the computational approach by Marr. We shall then provide an overview of the model components and the overall architecture of the present work. Finally we shall summarize the main contributions of the thesis and outline the overall organization of the thesis.

## 1.1 The Creative and Constructive Nature of Vision

At first glance one may conjecture that no particularly complex or advanced processing is involved in vision and perception in general. The only task seems to be a simple sensing of everything that is already present in the outside world. In this view perception is simply the assembling of sensations impressed on the tabula rasa of the mind. The inadequacy of such ideas becomes plain when one remembers that, in a strict sense, there is actually nothing like brightness, color, lines and corners, or even objects, in the world outside our brains, but only distributions of physical energy and matter. The tremendous task of vision now is to organize and transform the transient stimulations by electromagnetic radiation as received by the two retinae into a stable percept of a coherent, three-dimensional world.

Historically, the importance of the active and creative nature of vision has first been fully appreciated by the Gestalt psychologists. According to their central tenet that the "whole is different from the sum of its parts", the Gestalt psychologists proposed that vision actively organizes the sensual input into a coherent whole, or Gestalt. The Gestalt is not a property of the object but reflects the a priori assumptions of the brain on what is to be seen in the world. The Gestalt

psychologists formulated a number of rules according to which the sensual data are organized, such as grouping by proximity, similarity, closure, good continuation or common fate (Fig. 1.2).

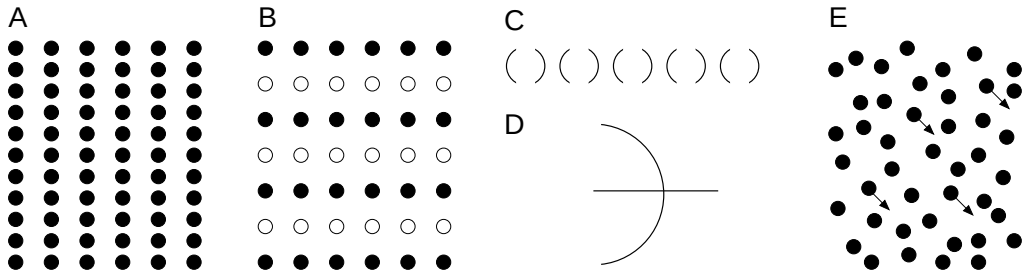


**Fig. 1.2:** Illustration of the Gestalt principles of perceptual organization. Grouping by *(A)* proximity, *(B)* similarity, *(C)* closure, *(D)* good continuation, and *(E)* common fate. (Partly adapted from Wertheimer, 1923 and Rock and Palmer, 1990.)

More recently, the importance of a priori assumptions in vision have been motivated by the view of vision as inverse optics (Poggio et al., 1985). Optics is the process which projects 3D objects onto a 2D image; inverse optics then denotes the inverse process of recovering a 3D representation from a 2D image. Such inverse problems are inherently ill-posed and cannot be solved based on the incoming data alone: there exists no unique solution, nor is the solution guaranteed to be stable. Thus, additional assumptions and constraints on the proper nature of the solution have to be imposed to solve the problem.

Visual illusions are particularly instructive to illustrate the creative nature of vision and help to reveal the heuristics used in visual processing. Visual illusions are neither amusing nor annoying failures of the visual system nor, in the words of Külpe (1893), "subjective perversions of the contents of objective perceptions". Instead, visual illusions reflect "information processing mechanisms that are normally adaptive" (Gregory, 1968) and provide important cues to unravel the underlying assumptions, constraints and processes involved in vision (Eagleman, 2001). This contemporary view of illusions can be dated back to von Helmholtz (1911), who stated that

> The study of what are called illusions of the senses is a very prominent part of the senses; for just those cases which are not in accordance with reality are particularly instructive for discovering the laws of those means and processes by which normal perception originates.

A number of visual illusions and their potential implications shall be detailed in the following. For example, illusory dark and bright regions in the Mach bands and the Herman grid have stimulated the proposal of lateral interactions between neurons. Brightness effects of simultaneous contrast stress the importance of the contextual surround in visual processing. Border contrast effects such as the Craik-O'Brien-Cornsweet (COC) effect point toward filling-in processes involved in human brightness perception. Illusory contour stimuli like the Kanizsa triangle or the Ehrenstein figure may reveal general grouping mechanisms involved in the completion of occluded objects.

To summarize the above considerations, vision is a creative process which constructs an inside representation of the outside world. This constructive process is based both on the incoming sensory data and on a priori assumptions. Failures of these assumptions for particular, often artificial stimuli as revealed by visual illusions help to discover the construction strategies employed by the visual system.

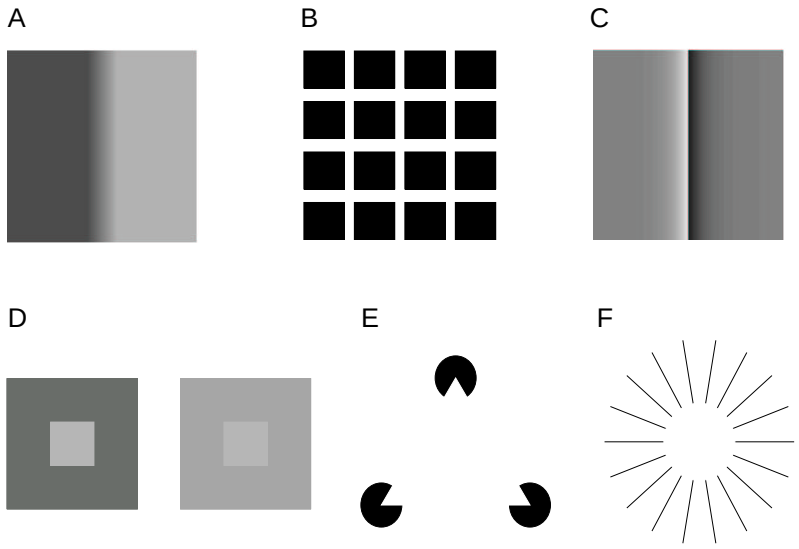


**Fig. 1.3:** Visual illusions of brightness perception and contour formation. *(A)* Mach bands. The small dark and bright bands confining the central transition from darker to brighter gray are illusory. *(B)* Hermann grid. Illusory dark patches are seen between the black squares. *(C)* Craik-O'Brien-Cornsweet (COC) effect. Both regions adjacent to the central high contrast flank have the same physical luminance value. *(D)* Simultaneous contrast. The same central square appears brighter on the dark background and darker on the brighter background. *(E)* Kanizsa triangle. An illusory triangle of increased brightness is seen. *(F)* Ehrenstein figure. An illusory circle of increased brightness is seen.

## 1.2 Theoretical Approaches to Vision

The empirical sciences of biology and psychology and their particular disciplines such as anatomy, physiology or psychophysics have gathered a wealth of data on visual perception. The focus of the empirical sciences is traditionally limited to a *description* of the phenomena, but does address only superficially, if at all, the *explanation* of the observed phenomena. The important question *why* a phenomenon occurs and what its purpose and overall functional role might be remains open. In the domain of vision, this explanatory gap has stimulated theoretical approaches to visual perception. Two important and prominent approaches, the ecological approach by Gibson and the computational approach by Marr and shall be detailed in the following.

### 1.2.1 Ecological Approach

The ecological approach to vision is strongly related to the work of Gibson (1979). Instead of treating more philosophical questions on the qualities of sensations or the distinction of sensation and perception, Gibson focused on the role of the senses as channels for the perception of the outside world. The important thing to understand, then, is how the invariant properties of the outside world can be extracted based on a continually changing stimulation. According to Gibson, such questions cannot be answered by studying the perception of highly artificial stationary stimuli within a laboratory environment. Instead, one has to consider an exploring observer, actively moving and looking around in a natural environment. The role of vision is not to sense everything which is principally available, or to reconstruct and represent a 3D world of objects. Rather, only the information relevant for a particular task at hand has to be considered. Gibson introduced the concept of *affordances* of the environment, "what it offers the animal, what it provides or furnishes, either for good or ill" (Gibson, 1979, p. 127). For example, surfaces are not perceived in terms of their qualities such as roughness, brightness, or slant, rather than in terms of various affordances such as "stand-on-able", "climb-able", or "sit-on-able".

Another idea central to the work of Gibson is the notion of *direct perception*. The behaviorally relevant invariants or affordances of the natural environment can be directly extracted or "picked-up" from the ambient array of light. Gibson postulates the existence of higher-order invariants present in the optic array, which directly supply the observer with the necessary information, without the need of intermediate representations or complex processing. Gibson suggests that

> the perceptual system simply extracts the invariants from the flowing array; it *resonates* to the invariant structure or is *atuned* to it. (Gibson, 1979, p. 249; emphases in the original)

One important contribution of Gibson's theory is the emphasis on the analysis of vision on a functional level, based on ecological constraints. The ecological approach forced to reconsider the stimulus properties in terms of the relevance for the observer, and stimulated new psychophysical research paradigms. In computer vision, the notion of an actively, exploring observer leads to the important research direction of active vision (e.g., Aloimonos et al., 1988). The idea of direct perception, however, remains controversy. Direct perception underestimates both the complexity of the information-processing tasks and the influence of experience and a priori assumption on perception. A critical discussion of the theory of direct perception can be found in Ullman (1980) and Fodor and Phylyshyn (1981). Recently it has been suggested by Norman (2002) that both the ecological approach of direct perception and the constructive-representational approaches can be related to distinct perceptual systems, the "ventral" and the "dorsal" system (Mishkin et al., 1983) which are engaged in two different visual tasks, namely visually guided identification and motor control, respectively.

### 1.2.2 Computational Approach

In this well-known and influential book "Vision", Marr details a novel, integrated approach to the understanding of vision (Marr, 1982). Based on the notion that vision is "first and foremost an information processing task" (Marr, 1982, p. 3), an overall framework of vision is formulated, which involves two complementary and dual components, namely the understanding of the *processes* involved and the *representations* these processes use and create. Besides introducing the information-processing view in the study of vision, the important and central concern of Marr's work is the notion that a complete understanding of vision involves three different levels of explanation. These levels include the computational theory, the algorithmic representation, and the implementation in neural or silicon hardware (Tab. 1.1).

**Table 1.1.** Levels of understanding an information processing system as suggested by Marr (1982, p. 25).

| **Computational theory** | **Representation and algorithm** | **Hardware implementation** |
|---|---|---|
| What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out? | How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation? | How can the representation and algorithm be realized physically? |

At the first level, the overall computational theory has to be formulated in terms of an abstract mapping of information, considering its appropriateness for the task at hand. At the second level, which due to Marr can be coarsely related to psychophysics, the precise algorithms and representations by which the task can be solved have to be understood. At the third and most basic level, which can be coarsely related to anatomy and physiology, the physical realization of the proposed algorithms and representations have to be addressed. The first and most abstract level of the computational theory is of critical importance, since the "nature of computations

that underlie perception depends more on the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented" (Marr, 1982, p. 27). While Marr's contribution is generally acknowledged, the importance of the computational level is still far from common place for empirical scientists, persisting in the notion that "the proper way of understanding the brain is to study the brain" (Zeki, 1993, p. 119).

Marr then identifies the overall task of vision as to "reliably derive properties of the world from images of it" (Marr, 1982, p. 23) and details a hierarchical architecture to solve the task. Central to this architecture are three different and gradually more abstract levels of representations, namely the primal sketch, the $2^1/_2$-D sketch and the 3D model representation. While this architecture can be criticized on a number of grounds, because it does not considers, e.g., the role of a priori knowledge and feedback, or the parallel processing in different streams, or an active, exploring observer moving around in the environment, the most important contributions of Marr survive: namely, to introduce the computational level into the study of vision, and to devise a new integrated approach, where the full understanding of a visual process has to be accomplished on three different levels.

### 1.2.3 Summary of Theoretical Approaches to Vision

The common and important idea of the two approaches reviewed above is the need to understand vision on a functional level. This idea underlies Gibson's notion of the extraction of invariant properties of the external world from sensory information, and has been made rigorous in Marr's emphasis on the level of the computational theory necessary for the understanding of vision. The importance of considering the functional level has also governed the more low and mid-level models developed in the present work. For example, we have suggested a mechanism of dominating opponent inhibition for contrast processing, which allows the visual system to robustly extract the relevant edge information from noisy stimulation as occurring under weak illumination conditions. Likewise, the brightness models based on filling-in of contrast information allows to discount the illuminant and to represent the invariant reflectance properties of objects.

## 1.3 Model Components and Sketch of the Overall Architecture

The present work deals with the modeling of early and midlevel visual information processing. In particular, we have examined and modeled the processing of contrast, contours, corners and surfaces. We show how such basic features can be efficiently computed and represented within an integrated architecture based on biological mechanisms.

The overall architecture is inspired by the BCS/FCS architecture developed by Grossberg and colleagues (e.g., Cohen and Grossberg, 1984; Grossberg and Todorović, 1988). At the first stage, contrast information is extracted from the raw image data and processed in a hierarchy of levels based on simple and complex cells. Oriented contrast information then feeds into a stage of contour processing, were localized contrast measurements are grouped to form coherent, stable contours. At the next stage, corners and junctions are detected based on a pruned and coherent representation of contours. Finally, a dense brightness surface is reconstructed from sparse unoriented contrast data; contours serve to signal local discontinuities in the brightness surface. The overall architecture is depicted in Fig. 1.4.

Throughout this work, the various mechanisms that are postulated and implemented are motivated by and based on empirical findings. We shall point toward the particular empirical motivations in the respective chapters. In the following, we shall examine the empirical basis of an overall

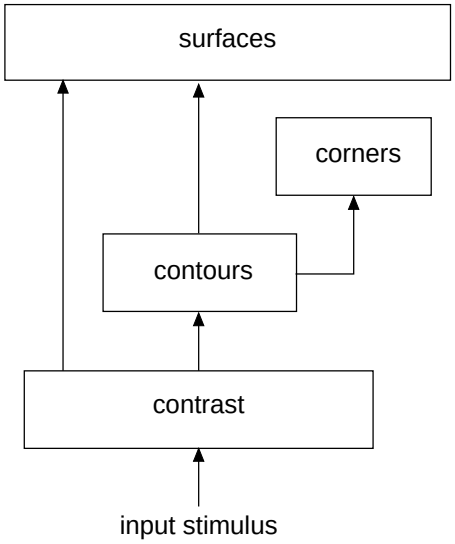


**Fig. 1.4:** Sketch of the overall model architecture.

principle of the suggested architecture, namely the processing within two segregated systems of form and brightness perception.

### Evidence for Separate Form and Brightness Systems

Distinct perceptual subsystems can be identified for the processing of visual information: one system that is concerned with the processing of discontinuities in the visual field, such as contrast and contours, and a complementary system that assigns surface properties to homogeneous regions.

Psychophysical evidence for the existence of two distinct systems for the processing of contour and surface information comes from the studies of so-called "phantom contours" (Rogers-Ramachandran and Ramachandran, 1998). In these experiments two images, each showing two fields of black and white disks on a gray background, but with opposite contrast polarity, flicker in counterphase at a high frequency of 15 Hz (Fig. 1.5). Under this stimulation subjects perceive a phantom border separating the two field, but cannot discriminate the temporal phase of the spots, i.e., the surface characteristics. Instead of alternating black and white disks, flickering spots are perceived. The surface characteristics can be seen only when the stimulus flickers at frequencies below 7 Hz. The results provide evidence for a fast, polarity-invariant system for the extraction of contours, and a slower, polarity-sensitive system for the assignment of surface color.

The psychophysical findings are paralleled by a physiological study on texture processing and figure-ground segregation (Lamme et al., 1999). Recordings in V1 show that the late components of cell responses ($> 80$ ms) correlate with boundary formation and are followed by filling-in or coloring of surface information between the edges.

The study of illusory contours has also provided evidence for the existence of two separate systems (Kanizsa, 1976, 1979). It has been shown that illusory contours can be produced by inducers with opposite contrast polarity (Prazdny, 1983; Shapley and Gordon, 1983). Consequently, as pointed out by Shapley and Gordon (1987), the illusory contour—like the shape or form of objects in general—does not depend on the sign of the contrast, while the apparent brightness does (Heinemann, 1955, 1972; Shapley and Enroth-Cugell, 1984). These results are confirmed by other studies which showed an independence of perceived brightness of the illusory figures and perceived sharpness of the illusory contours (Lesher, 1995; Petry et al., 1983). The results found for illusory contours can be most likely transferred to real contours, since both share many functional

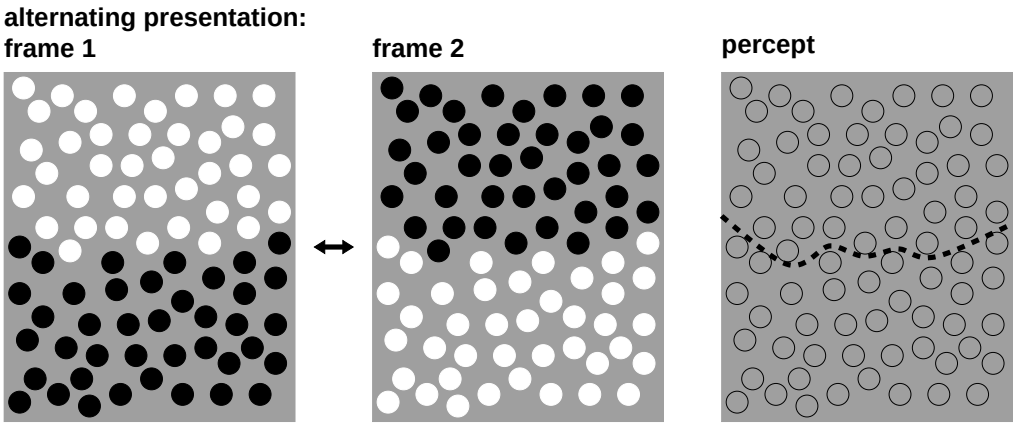


**Fig. 1.5:** Sketch of the experiment by Rogers-Ramachandran and Ramachandran (1998) generating the percept of phantom contours. When two images of black and white dot fields of opposite contrast polarity flicker at high frequency of about 15 Hz *(frame 1 and 2)*, a "phantom contour" is seen *(percept, dashed bold line)*, though the different surface colors of the two dot fields cannot be discerned *(percept, open circles)*.

properties, e.g., illusory contours can be used as targets or masks in visual masking experiments (Reynolds, 1981), can cause motion aftereffects (Smith and Over, 1979), or can produce geometric illusions (Farné, 1968; Gregory, 1972; Meyer and Garges, 1979; Pastore, 1971). For example, illusory contours can generate the railroad track or Ponzo illusion (Fig. 1.6).

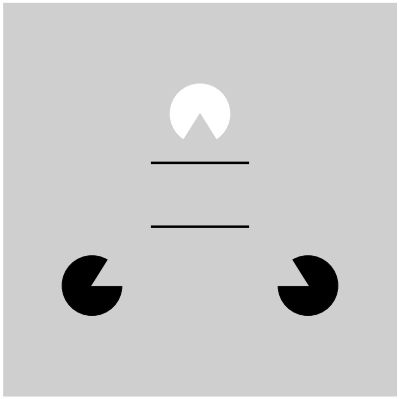

**Fig. 1.6:** Illusory contours give rise to the Ponzo illusion: the top horizontal line appears to be longer than the bottom line, though both lines have the same length. Modified after Kanizsa (1976) by using inducers with opposite contrast polarity.

To summarize, the findings reviewed above indicate that two distinct systems exist in human vision: a fast, polarity-insensitive system concerned with the processing of discontinuities such as contours, and a slower, polarity-sensitive surface system.

## 1.4 Contributions of the Thesis

In the following we shall detail the contributions of the present work. The contributions are organized according to the four main model components dealing with the processing of contrast, contours, corners and surfaces.

### Contrast

The processing of contrast begins with the extraction of raw, unoriented contrast signals by a center-surround operator similar to retinal ganglion cells. In accordance with biological findings, contrast signals are modeled in two domains of on and off contrasts, signaling light increments and decrements, respectively. For the further processing of contrast signals we have suggested a new mechanism of dominating opponent inhibition (DOI). DOI is based on a push-pull interaction of contrast signals from opposite domains and postulates a stronger weighting of the inhibitory (or "pull") signal from the opponent domain. Such a stronger weighting of the inhibitory input is in accordance with a number of physiological findings. The outcome of the DOI interaction then feeds into a previously suggested nonlinear simple cell circuit (Neumann and Pessoa, 1994).

In a first set of simulations we show that the model with DOI can account for empirical data of simple cell responses to luminance gradient reversal (Hammond and MacKay, 1983), which have not been successfully modeled before. With the same parameter settings we have then applied the model for the processing of noisy artificial and real world camera images. The results show that the sharpness of the response and the robustness to noise is increased with DOI. Moreover we show that the suppression of noise is largely invariant against changes of the noise level, leading to the interpretation of DOI as an adaptive threshold. This property of adaptive suppression is further examined by a stochastic analysis, showing that the mean response to homogeneous regions is antiproportional to the standard deviation of the noise process. Next, the properties of DOI are evaluated in a set of numerical simulations under extensive parameter variations. In these numerical studies we determine an optimal value for the amount of inhibition in the DOI interaction, and show that the model circuit remains sensitive to small contrast changes. Finally we address the intensely debated generation of contrast-invariant orientation tuning of cortical simple cells. We show that contrast-invariant orientation tuning can be generated within a purely feedforward model based on inhibition between complementary channels. In particular, we show that the proposed model exhibits contrast-invariant orientation tuning. The new mechanism of DOI causes a sharpening of the tuning curves, resulting in biologically realistic tuning widths.

Overall, we have introduced a biologically plausible mechanism of dominating opponent inhibition, which can account for empirical findings on simple cells, regarding contrast invariant orientation tuning and responses to luminance gradient reversal. The application of the model to the processing of images suggests a functional role of the mechanism, namely the adaptive suppression of noise, which allows for a more robust extraction of oriented contrast information under suboptimal viewing conditions.

### Contours

Initial contrast measurements are often fragmented and noisy. Based on empirical findings we have developed a model for contour grouping in primary visual cortex (V1), utilizing recurrent long-range interaction between cells with colinear aligned receptive fields. The core component of the model is the recurrent interaction between two bidirectionally linked layers. The excitatory, colinear long-range interaction implements the a priori assumptions, providing template shapes of frequently occurring contours. The sensory data of initial contrast measurements as carried by the feedforward path are matched against these templates. Coherent local measurements which fit into a more global context are selectively enhanced, while other noisy measurements are suppressed.

We show for a number of noisy artificial and natural images that the proposed circuit successfully groups local contrast measurements and enhances the coherent contours. In this process, amplitude differences along the contour are equalized such that gaps can be closed as long as some nonzero initial activity is present. Next, the competencies of the model are quantitatively evaluated using two measures of contour saliency and orientation significance. We show that both the contour saliency and the orientation significance are enhanced during recurrent long-range processing. The

model circuit is also evaluated regarding the processing of curved stimuli. Here we demonstrate that the model can enhance curved contours to a certain degree, depending on curvature. Further, we have probed the model with stimuli of fragmented contours and textures as used in an empirical study by Kapadia et al. (1995). The model responses qualitatively account for the empirical findings. In particular, effects of surround inhibition by randomly oriented bars and long-range excitation by colinear flankers on the response to a central bar element are successfully simulated. Finally, we have examined a model variant using early feedback. Compared to the standard model, contour saliency is higher while orientation significance is lower, suggesting two different functional roles of the different kinds of feedback loops.

Overall, we have shown that coherent contours can be extracted from noisy initial contrast measurement by biological mechanisms of recurrent, colinear long-range interactions.

### Corners

Intrinsically 2D signal variants such as corners and junctions are invariant against moderate changes of view point and viewing distance and provide important cues for a number of higher level visual tasks such as tracking or object recognition. The novel scheme for the detection of 2D signal variations developed in this work is based on the notion that corners and junctions are characterized by high activity in multiple orientations at a particular location. Such oriented activity is represented as a model hypercolumn and can be robustly extracted by recurrent, colinear long-range interactions for contour grouping, as introduced above. In the proposed scheme, corners and junctions are implicitly characterized by distributed activity within a hypercolumn. A measure of circular variance is used to read out the distributed information and explicitely localize corner and junction points.

In a first set of simulations the novel junction detection scheme is evaluated for a number of generic junction configurations such as L-, T-, X-, Y-, W- and $\Psi$-junctions. The localization performance of the new detection scheme based on recurrent long-range interactions is compared with results obtained for a representation as generated by a purely feedforward model of complex cells. Results show that the localization accuracy is improved by the recurrent long-range interaction. Next, we show for a number of artificial and natural images that positive correctness of detected junction points is higher for a representation based on recurrent long-range interaction than based on feedforward complex cell processing. In a second set of simulations we compare the new scheme with two widely used junction detector schemes in computer vision, based on Gaussian curvature and the structure tensor. We employ receiver operator characteristic (ROC) analysis for a threshold-free evaluation of the different junction detector schemes. The results obtained for both artificial and natural images show that the new approach performs superior to the standard schemes.

Overall, we have shown that junctions can be robustly and reliably represented by a suggested biological mechanism based on a distributed hypercolumnar representation and recurrent colinear long-range interactions. Further, we have shown how ROC analysis can be used for the evaluation of junction detectors.

### Surfaces

The processing of luminance discontinuities at the stages of contrast, contour and corner processing is complemented by a second stream for the processing and representation of homogeneous surface qualities such as brightness. Cells at the first stages along the visual pathway, such as retinal ganglion cells, primarily respond to luminance discontinuities, but not within homogeneous regions. In the present work we have focused on the question how a dense brightness surface can be generated based on sparse, local measurements. We show that this task can be accomplished by a mechanism of confidence-based filling-in. Unlike other filling-in approaches, a confidence measure

is employed which allows to distinguish valid contrast responses at luminance discontinuities from invalid noisy or zero-valued responses.

First the competencies of the new mechanism of confidence-based filling-in are evaluated in comparison to standard filling-in. We show in a number of simulations that only confidence-based filling-in can generate brightness predictions from sparse contrast data which are invariant against changes of both the shape and the size of the area to be filled-in. Next the brightness reconstruction for noisy artificial and natural images is assessed. Confidence-based filling-in can successfully discount the illuminant and is more robust against noise than standard filling-in, resulting in a smooth brightness surface even for noisy stimuli. We also examine the processing of stimuli that give rise to visual illusions. It is demonstrated that confidence-based filling-in can account for a number of phenomena, in particular simultaneous contrast and remote border contrast effects as occurring for Craik-O'Brien-Cornsweet (COC) stimuli. Finally we address the generation of reference levels by filling-in models. We propose that sparse contrast signals are locally gain-controlled by a luminance signal. We demonstrate that confidence-based filling-in of such luminance-modulated contrast signals can account for a number of challenging stimuli, in particular luminance staircase and luminance pyramid, COC effect and COC sequences, and simultaneous contrast.

Overall, we have proposed a full 2D model of human brightness perception based on a newly proposed mechanism of confidence-based filling-in. We have shown that the proposed mechanism can generate a smooth, dense brightness surface from sparse contrast data as signaled by retinal ganglion cells. Moreover, the proposed mechanism exhibits basic invariance properties and can account for a number of visual illusions.

## 1.5 Organization of the Thesis

In this section we give an outline of the overall organization of the thesis.

In Chap. 2 we give an introductory survey of basic neurobiological findings of early visual information processing. We shall describe the flow of information along the primary visual pathway, examine the transformations that occur and review the basic underlying mechanisms.

In Chap. 3 we address the extraction of oriented contrast information from the raw input stimulus. We give a detailed survey of empirical studies on contrast processing in mammals, focusing on the generation of orientation selectivity from unoriented LGN input by cortical simple cells. Next, we group the various models of simple cells into two main categories and review important models within each category. The extraction of oriented contrast information at luminance discontinuities, or edge detection, is also extensively studied in computer vision. We point out the basic schemes and review important approaches, such as the Canny edge detector. After this more general material, we present the proposed simple cell model and suggest a new mechanism of dominating opponent inhibition. We study the competencies of the new mechanism both numerically and analytically, and demonstrate the empirical relevance of the new scheme.

In Chap. 4 we address the grouping of initial contrast measurements to coherent contours. First, we present a survey of empirical findings regarding lateral long-range connections and recurrent processing in early vision. The survey covers a broad variety of disciplines, ranging from anatomy to physiology, psychophysics and statistics. Next, we provide an in depth review of the rich literature on computational approaches to contour grouping. We point toward the basic mechanisms and suggest an overall classification framework. A number of important schemes are discussed and characterized within the suggested framework. After this more general considerations we introduce a new approach for contour grouping based on colinear, recurrent long-range interactions. The competencies of the new scheme are examined and quantitatively evaluated. Further we show that the model can successfully account for an empirical study which examined the influence of

surrounding textures to a central bar element. Finally we introduce a model variant using early feedback and discuss the competencies in comparison with the standard model.

In Chap. 5 we deal with the extraction of intrinsically 2D signal variations such as corners and junctions. We propose a new scheme where junctions are implicitly characterized by strong responses for more than one orientation within an orientation hypercolumn. A measurement of circular variance is used to extract the corner and junction points from the distributed hypercolumnar representation. We compare detection results based on a purely feedforward representation to detection results as obtained from the recurrent-long range interaction for contour grouping, as introduced above. Detection and localization properties are evaluated for a variety of artificial and natural images. Finally, we evaluate the performance of the new scheme in comparison with two other widely used approaches to corner detection, based on Gaussian curvature and the structure tensor. We use ROC analysis for a threshold-free evaluation of the different junction detection schemes.

In Chap. 6 we examine the computation and representation of brightness surfaces. We show how a dense brightness representation can be generated from sparse contrast signals by a new mechanism of confidence-based filling-in. The competencies of the new scheme are evaluated, and important invariant properties are demonstrated. Further, we show that confidence-based filling-in can account for a number of brightness illusions. Finally, we address the generation of reference levels by filling-in models. We propose a new scheme using confidence-based filling-in of luminance-modulated contrast signals that successfully account for a variety of brightness phenomena.

In Chap. 7 we summarize the results of the present work and point toward future investigations.

Chapter 2

# Neurobiology of Early Vision

The visual system is the most complex of all sensory systems, and a huge part of the human brain is involved in vision. Despite of its complexity, the study of the visual system has attracted the effort of numerous scientists, and significant progress has been achieved in the past decades. The visual system shares properties with other sensory systems like the somatic sensory system. Therefore, the study of the visual system allows to identify common principles of sensory information processing in particular and cortical organization and functioning in general. The empirical findings in neurobiology, especially physiology, serve as motivation and guiding lines for computational models of visual processing.

In this chapter we provide an overview of the neural systems involved in the processing of visual information. The thesis is concerned with models of early vision and particularly investigates the processing of static gray level stimuli. Consequently, the review focuses on the early stages of visual information processing and does not cover higher order functions such as object recognition. Also, the processing of color and motion is only marginally covered. An extensive description of the visual system and neural science in general can be found in Kandel et al. (1991) or Purves et al. (1997). Further introductory descriptions can be found in, e.g., Coren et al. (1994) or Zeki (1993). A detailed review which focuses on the temporal aspects of neural coding at the early stages of visual processing is given by Victor (1999).

In the next sections we describe the flow of visual information. The review follows the primary direction of flow, starting from the focusing of light by the optical apparatus of the eye and the subsequent transduction in the retina, up to the segregated processing of different modalities such as color, motion or depth by the visual cortex. We shall describe both the anatomical and physiological properties of cell. The anatomy describes the types of neurons and wiring patterns in the visual system, which can be viewed as the cortical "hardware", while the physiology describes the response properties of neurons as part of the cortical "software".

## 2.1 Overall Anatomy of the Primary Visual Pathway

In this section an overview of the anatomy of the early visual system is provided. The overall anatomical structure of the primary visual cortex is sketched in Fig. 2.1. In the retina, electromagnetic radiation within a certain frequency band, the "visible light", is transduced into a neural code of spike patterns. The primary projection from the retina has its target in the lateral geniculate nucleus (LGN) which is part of the thalamus. From the LGN, projections go the primary visual cortex (V1). Two major output streams arise from V1. The feedforward stream projects to higher visual cortical areas, whereas the feedback stream projects back to the LGN and to other subcortical areas. Before reviewing the neural part of the visual system, we start with an outline of the structure of the eye.

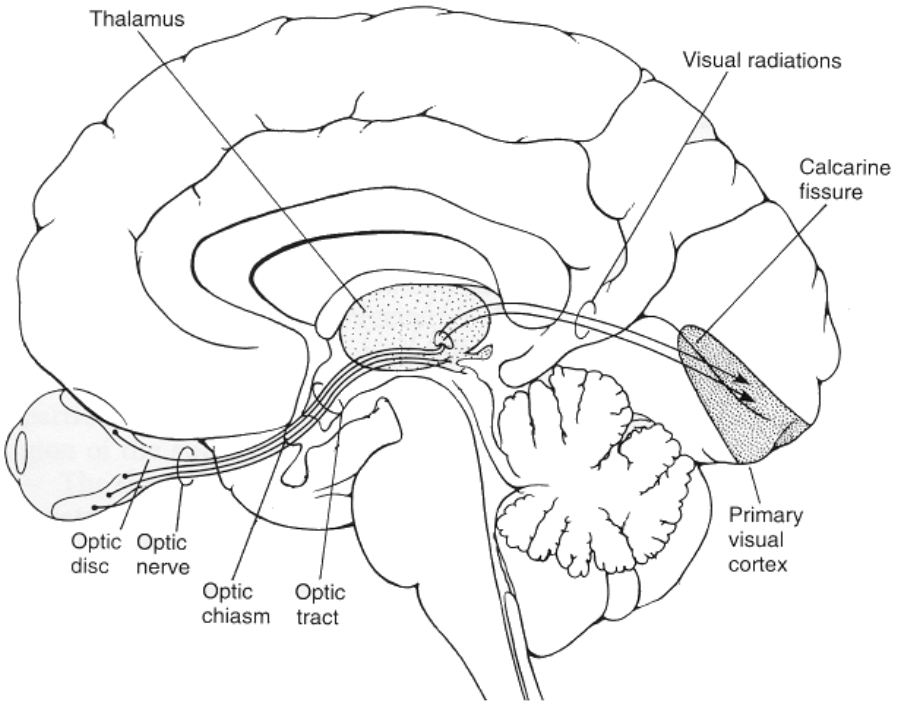


**Fig. 2.1:** The primary visual pathway. (From Kandel et al., 1991.)

## 2.2 The Structure of the Eye

The sense of sight is mediated through a fascinating organ, the eye. The eyes lie in protective bony sockets approximately half way down the head and have a spherical structure with 20–25 mm in diameter.

The supporting wall of the eyeball is formed by the *sclera*, which is seen as the "white" of the eyes. As the eye is an outer part of the central nervous system, the sclera is in continuity with the dura, the protective covering of the brain. The next intermediate layer adjacent to the sclera is the *choroid*, which is a vascular layer of blood vessels and large branched pigment cells. The third layer is the *pigment epithelium*. The pigment epithelium has two functions. First, cells in the pigment epithelium contain the black pigment melanin which absorbs light and thus decreases light scatter within the eye which would degrade the visual image. Second, the pigment epithelium assists the metabolic processes of the photoreceptors, in particular the photopigment regeneration (resynthesis). The retina is the innermost internal layer of the eye and contains the eye's receptor sheet, where the transduction of light into a neural code takes place.

Before entering the retina, light travels through the optic apparatus of the eye. Light enters the eye through the *cornea*, which is a clear, domelike window of about 13 mm in diameter. The cornea serves as a fixed lens, gathering and concentrating the incoming light rays. Since the cornea bulks forward, the visual field is extended slightly behind the eyes.

The part of the eye which usually captures our attention is a beautifully colored circular muscle, the *iris*. The pigmentation of the iris determines the "color" of our eyes. The iris has a circular opening, the *pupil*, that allows light to enter the eye. The pupil appears black because of the light absorbing pigment epithelium. The size of the pupil is controlled by circular muscle of the iris and determines the amount of light that enters the eye. Depending on the illumination, the diameter of the pupil varies between 2 mm in bright light and may dilate to more than 8 mm in the dark, resulting in a sixteenfold change in the area of the aperture. The function of the iris and the pupil

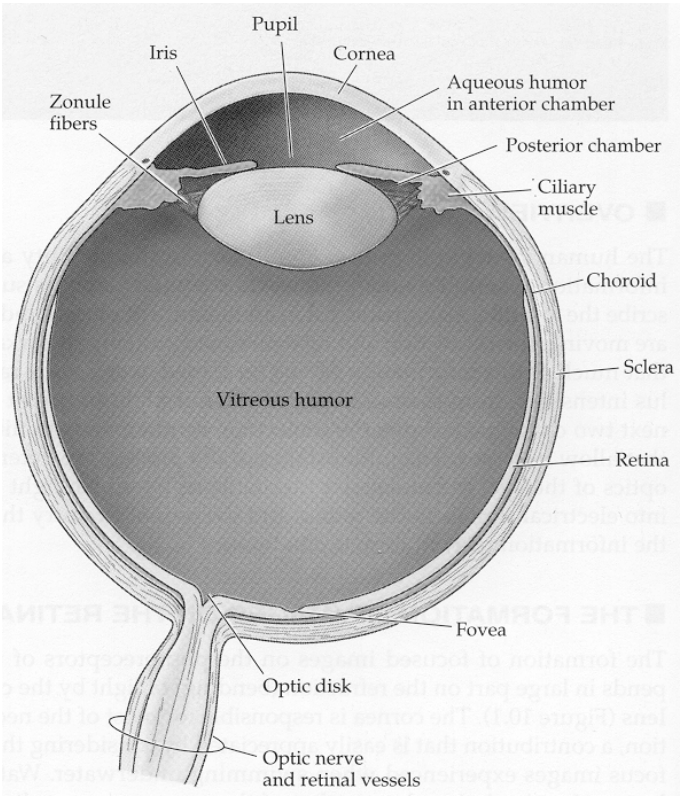


**Fig. 2.2:** Anatomy of the human eye. (From Purves et al., 1997.)

is similar to the diaphragm of a camera. A reduction of the size of the pupil limits the amount of light that reaches the retina. At the same time, a small pupil reduces optical aberrations and also increases the depth of field (or depth of focus), i.e., the range of distances at which objects are seen sharp and unblurred. Under dim illumination, the visual acuity is limited by the number of gathered photons rather than the optical aberrations. An adjustable pupil thus allows to take advantage of a better illumination condition by improving the optical response of the eye, while retaining the ability to gather an increased amount of light at dim illumination conditions.

The crystalline *lens* is located directly behind the pupillary aperture. Like the cornea, the lens is responsible for the refraction of the light that enters the retina. Unlike the cornea, however, the lens can change its shape and thus its refractive power, so that objects at different distances can be brought into focus at the retina. The focusing of an image by changing the shape of the crystalline lens is called *accommodation*. The shape of the lens is controlled by the *ciliary muscles*. When the ciliary muscles relax, the pressure of the fluid in eyeball and the tension of the zonal fibers connecting the lens to the inside wall of the eye cause a flattening of the lens. Under these conditions, distant objects are in focus. Contraction of the ciliary muscles results in a more spherical or convex shape of the lens, enabling focus at near.

The *anterior chamber* between cornea and iris and the *posterior chamber* between iris and lens are filled with a watery fluid called *aqueous humor*.

The large chamber of the eyeball is filled with a clear gel, the *vitreous humor*. The vitreous humor contains organic debris or floaters of clumped cells or strands of the vitreous gel. These floaters or *mouches volantes* (flying gnats) can be seen as floating shadows when fixating a bright surface, such as the blue sky or a white wall.

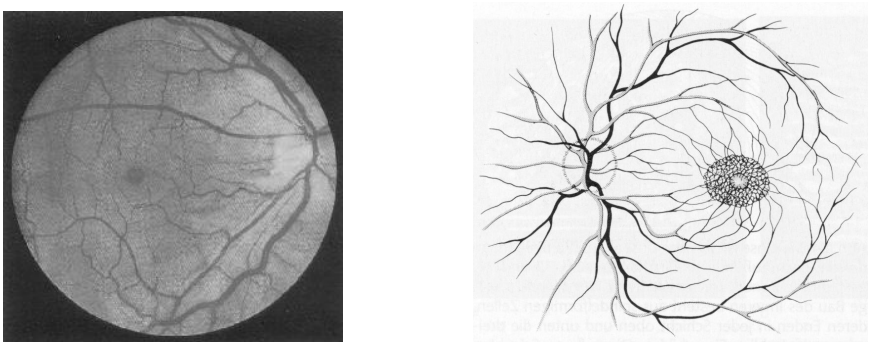

**Fig. 2.3:** The surface of the retina when viewed through an ophthalmoscope *(left)* (modified after Purves et al., 1997), and a sketch of the blood vessels in the retina *(right)* (from von Campenhausen, 1993). The blood vessels enter and exit at the optic disk. The macula lutea and its center, the fovea, lies temporal to the optic disk and does not contain any blood vessels.

Having passed through the optical apparatus of the eye, the cornea, the pupil and the lens, and having traversed the vitreous humor, light eventually enters a layer of neurons at the back of the eye, the *retina*. The retina contains the photoreptors, where transduction of light into a physiological signal is initiated. When viewed through an ophthalmoloscope, besides the netlike system of blood vessels that cover the inner cavity of the eye, two distinct regions can be seen with the plain eye, the *macula lutea* and the *optic disk*. The macula lutea is a yellow patch which lies near the optical axis. The central part of the macula is a small depression, the fovea. The fovea is the central point for image focus, defining the visual axis. At the fovea the photoreceptors are most densely packed, so that that fovea is the region of highest resolution. About 15° nasal to the fovea lies the optic disk, where the optic nerve leaves the retina. The optic disk contains no photoreceptors and defines the blind spot. Usually, one is unaware of the blind spot, because light coming from a single point of the visual field never falls on the blind spot of both eyes. We are usually unaware of the blind spot, since light of the binocular portion of the visual field never enters the blind spots of both eyes. The blind spot, which is remarkably large, can be perceived in a visual experiment shown in Fig. 2.4.

**Fig. 2.4:** Blind spot demonstration. The blind spot in the left eye can be demonstrated by shutting the right eye and fixating the upper cross from a distance of about 30 cm with the left eye. If the stimulus is moved back and forth slightly, the circle on the left suddenly disappears when falling on the blind spot. If the lower cross is fixated, the gap in the black line disappears when imaged on the blind spot. (Redrawn from Kandel et al., 1991.)

Besides the photoreceptors, the retina contains a complex networks of neurons, which are involved in the initial processing of the visual signal. The structure of the retina is detailed below.

**Table 2.1.** Optical and neural elements in the eye and their basic function in keywords.

| element | function |
|---|---|
| *optical elements* | |
| 1. cornea | gathering lens: gathering of light |
| 2. pupil and iris | optic diaphragm: control of the amount of light, reduction of optic aberrations, determination of depth of field |
| 3. crystalline lens | focusing lens: focusing of near and far objects |
| *neural element* | |
| 1. retina | transduction of light into a neural code, initial processing, segregation of different modalities |

## 2.3 Retina

In this section, major properties and functions of the primate retina are reviewed. More detailed descriptions are given by, e.g., Sperling (1990), Kaplan et al. (1990), Wässle and Boycott (1991), or Kolb et al. (2000).

The retina is the first processing stage in the visual pathway. In the retina, physical light intensities are transduced into a neural code. Besides this sensory function, the retina plays an important role in the initial visual processing, such as adaption or contrast detection, and in the segregation of visual modalities such as on- and off-responses into parallel pathways.

Despite its peripheral location, the retina is actually part of the central nervous system. Thus, the retina is a "window to the brain" (Wässle, 1996) in two perspectives: it allows for the perception of the visual world by the brain and, conversely, studying the retina gives insight into the functional mechanisms of the brain. Compared to other regions of the brain, however, the retina has a relatively simple structure. The primate retina contains five major cell types: photoreceptors, bipolar cells, horizontal cells, amacrine cells and ganglion cells. The neurons of the retina are

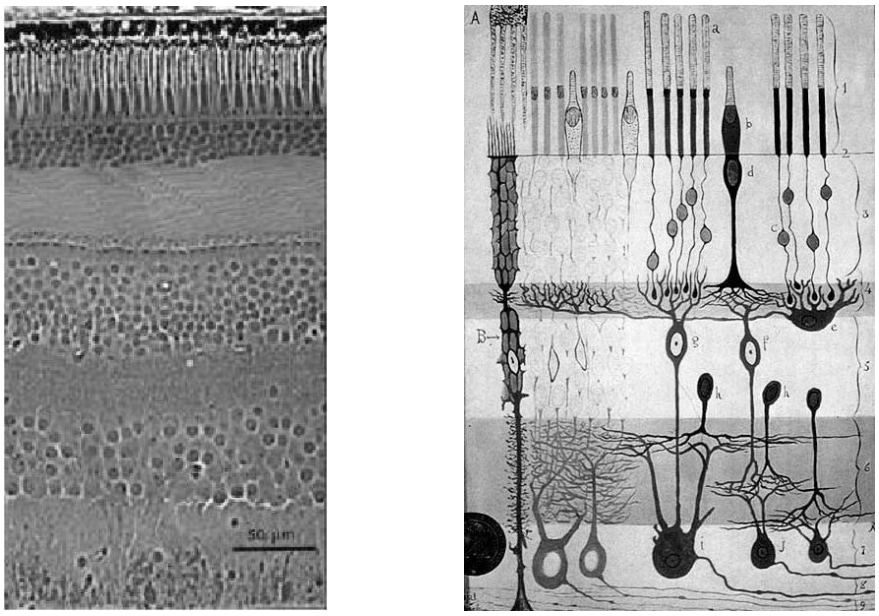


**Fig. 2.5:** Vertical cross-sections through the retina. *Left:* Light micrograph of a vertical section through the human retina (From Dowling, 1987.). *Right:* Drawing of a vertical cross-section of the retina based on Golgi-stained cells. (From Cajal, 1892.)

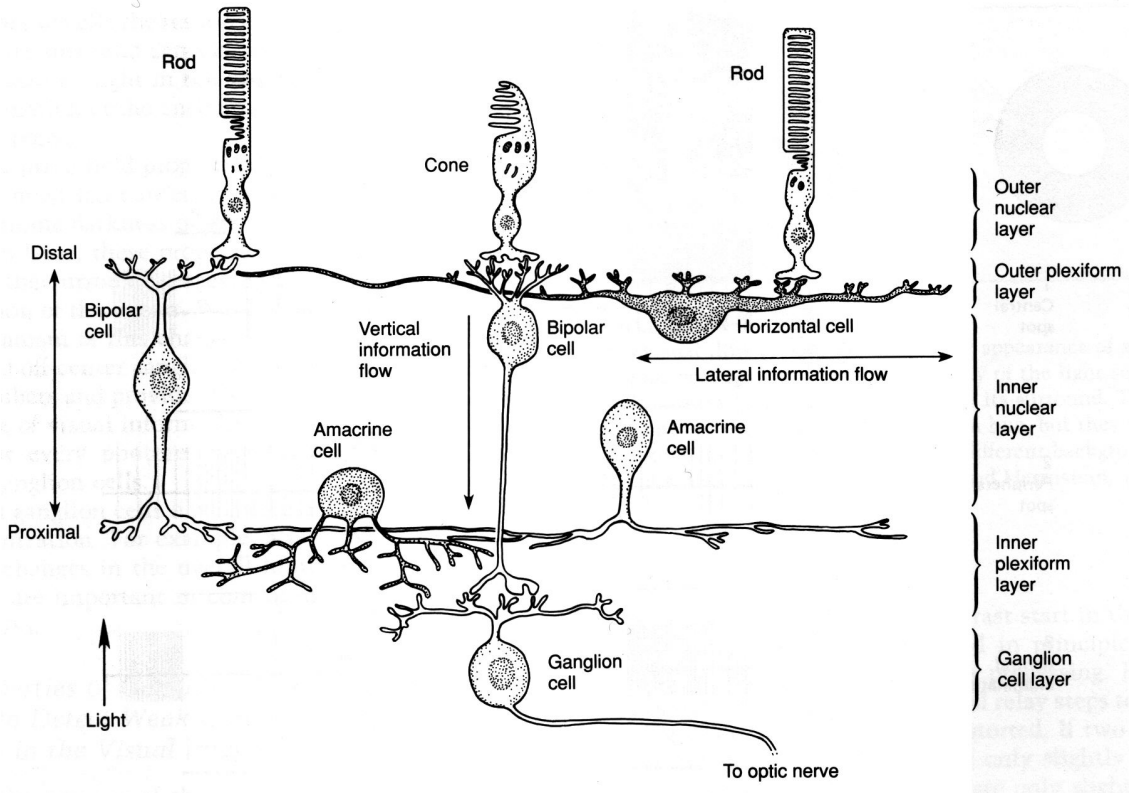


**Fig. 2.6:** The structure of the retina. (From Kandel et al., 1991.)

organized into five distinct layers: three layers of cell bodies, the outer nuclear layer (ONL) containing photoreceptors, the inner nuclear layer (INL) with horizontal, bipolar and amacrine cells, and the ganglion cell layer (GCL). Besides the nuclear layers, the retina has two layers of synaptic connections, the outer plexiform layer (OPL) and the inner plexiform layer (IPL). A schematic cross-section of the retina is shown in Fig. 2.6. The main flow of the visual signal goes from photoreceptors via bipolar cells to ganglion cells which transmit the signal through the optic nerve to other parts of the brain. The signal also spreads laterally in a network of horizontal and amacrine cells and is fed back to the distal layers by amacrine cells.

The orientation of the retina has the remarkable feature that incoming light has to pass all layers before striking the photoreceptors. The proximal layers are, however, relatively transparent, allowing for a minimal absorption or scattering of light. Additionally, in the fovea, the region of highest acuity, the proximal neurons are shifted aside, enabling the least distorted reception of light. In the following, we review basic properties of the various types of neurons found in the retina.

### 2.3.1 Photoreceptors

The human retina has two types of photoreceptors, rods and cones. Rods are specialized for night vision (scotopic), while cones are specialized for day vision (photopic). Cone-mediated vision is of greater spatial and temporal resolution. In the retina of humans and macaque, three different kinds of cones exist with peak sensitivity at different wavelengths, providing the basis color vision. The three kinds of cones contain different visual pigments with different but overlapping spectra,

having absorption maxima at short, middle and long wavelength (420 nm, 531 nm, 558 nm, respectively).

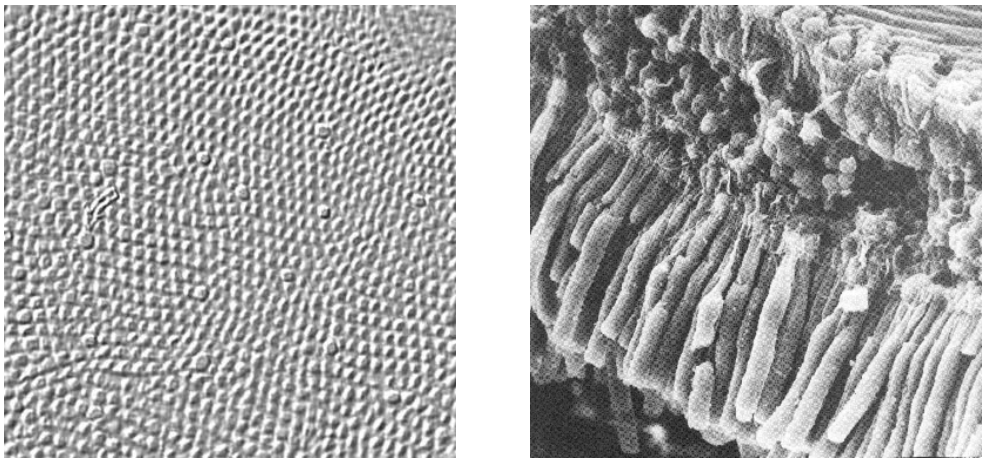

**Fig. 2.7:** *Left:* A tangential cross section through the human fovea shows the regular, hexagonal mosaic of cone receptors. (From Kolb et al., 2000.) *Right:* Vertical view of photoreceptors. (From Coren et al., 1994.)

The higher spatial resolution of cones is not caused by their larger number. In fact, there are about 20 times more rods than cones in the retina. Rather, the better spatial resolution of the cone system is caused by two other factors. First, the cones are concentrated in the fovea, where they are most densely packed, since no rods are present, and the visual image is least distorted. Second, unlike the rod system, where many rods synapse on the same bipolar cell, the cone system in the fovea is non-convergent: one photoreceptor contacts one bipolar cell which in turn feeds one ganglion cell.

To detect dim light, rods temporally average or integrate photons, so that the effect of photon absorption within a 100 ms interval summates. This averaging does not allow the rod system to detect fast light flicker beyond 12 Hz, while the cone system can resolve up to 55 Hz. The poor temporal resolution of the rod system as observed physiologically has direct consequences for the human perception. Signals from rods may arrive as much as 1/10 second later than those from cones (MacLeod, 1972), and rod detected motion stimuli appear to move at about 75% of the speed of a cone-detected reference stimulus (Gegenfurtner et al., 1999). In addition to the temporal averaging, the rod system exhibits a spatial averaging, where many rods provide input to a single ganglion cell. The different properties of the rod and cone system reflect the tradeoff between high resolution and good detection. The cone system operates under good illumination conditions, where light of different wavelengths can be detected with good spatial and temporal resolution. At dim light good sensitivity is most important to allow for the detection of weak signals. Thus, good temporal and spatial resolution or the detection of different wavelengths is sacrificed in the rod system in favour of excellent detection. A similar strategy seems to be pursued at higher visual stages such as the primary visual cortex (V1), where the RFs of cells have been shown to increase at lower contrast (Sceniak et al., 1999).

Rods and cones are not equally distributed across the retina. In the fovea, only cones are present, while their number rapidly decreases outside the fovea to a fairly even density in the peripheral retina. The number of rods, on the other hand, rapidly increases with the distance from the fovea, reaching a peak at about 20° from the foveal pit Fig. 2.8.

The differences of rods and cones are summarized in Tab. 2.2.

The transformation of light into a neural code is initiated in the outer segment of the photoreceptors, which is made of folded, stacked membranous disks which are packed with the photopigment

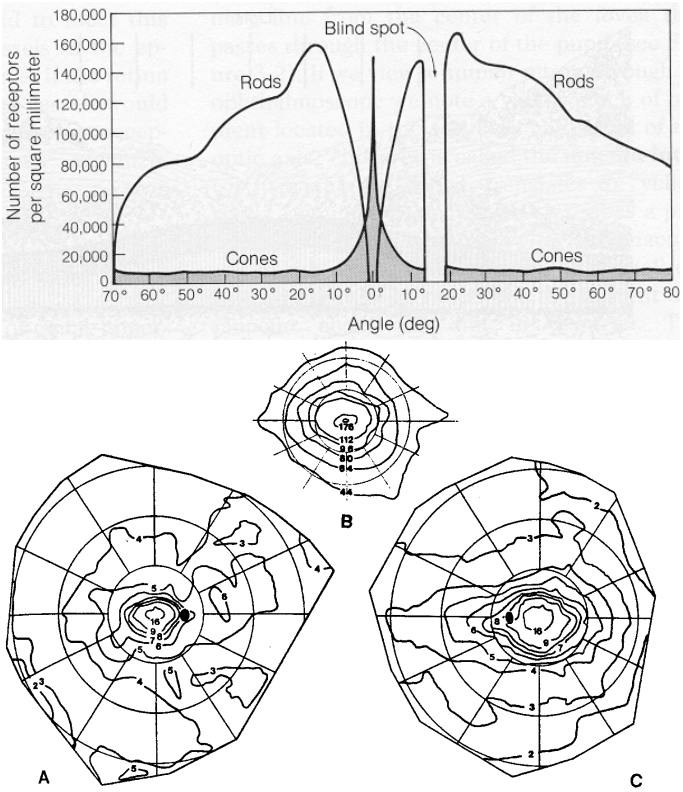


**Fig. 2.8:** Distribution of rods and cones in the human retina along the horizontal meridian (after Osterberg, 1935) and the two dimensional distribution of cones (from Curcio et al., 1987).

**Table 2.2.** Differences between rods and cones and between their associated neural systems (after Kandel et al., 1991).

| | Rods | Cones |
|---|---|---|
| sensitivity | high (night vision) | low (day vision) |
| amplification | high (single photon detection) | low |
| saturation | in daylight | only in intense light |
| temporal resolution | low | high |
| sensitivity | to scattered light | to direct axial light at most |
| convergence | high | low, not in the fovea |
| color vision | no | yes (three types of cones) |

retinal. The photopigment contains a light absorbing component, the retinal. In 1958 Georg Wald and coworkers found that light causes a change of the 11-*cis* isomer of the retinal to the all-*trans* isomer. This isomerization of the retinal is the primary event of visual process and the only light dependent process. The all-*trans* retinal triggers a cascade of biochemical events, which finally leads to the closing of $Na^+$ channels, resulting in a hyperpolarization of the photoreceptor. The biochemical cascade leading to the hyperpolarization of the photoreceptor has been most extensively studied for rods and is believed to be the same for cones.

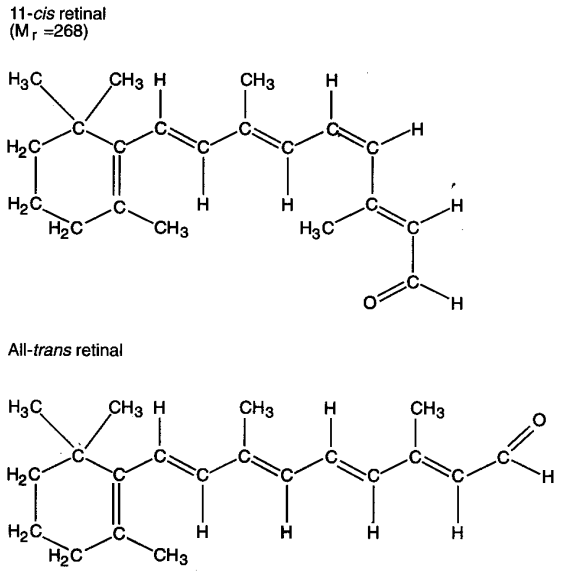


**Fig. 2.9:** Isomerization of 11-*cis* retinal to all-*trans* retinal. The isomerization is caused by absorption of a photon and is the initial event in visual transduction. (From Kandel et al., 1991.)

The photoreceptors, especially the rods, have a remarkable sensitivity. In a dark adapted rod, a single photon can prevent the flux of millions of $Na^+$ ions, resulting in a hyperpolarization of about 1 mV. Thus, the dimmest light ever imaginable, namely a single photon, can be detected by human rods. This property has been concluded by Hecht et al. (1942) based on psychophysical studies and has been confirmed by electrode recordings in monkey rods (Baylor et al., 1979; Schneeweis and Schnapf, 1995).

The human visual system can operate in dim light as well as bright sunlight, where illumination changes over orders of a magnitude. This process of light *adaption* involves many changes in the eye and the retina (such as contraction of the pupil), but the most important change occurs in the photoreceptors. A bright light maximally hyperpolarizes a cone to −70 mV, so that the cone cannot respond to further increases in light intensity. If the background illumination is maintained, the cone slowly depolarizes to a membrane potential between −70 and −40 mV, so that another subsequent increase of light can again be signaled by a hyperpolarization of the cone.

### 2.3.2 Horizontal Cells

The horizontal cells are located in the INL, and have dendrites and axons within the OPL. Horizontal cells make both pre- and post-synaptic contacts with cones. Horizontal cells integrate light from a larger area and have an inhibitory influence on postsynaptic cones. Horizontal cells are therefore thought to contribute to the antagonistic surround of bipolar cells. Horizontal cells make gap junctions gap junction between dendrites of other horizontal cells of the same type. The gap junctions allow a lateral flow of signals within a syncytial network and cause an increase of

receptive field size, which is well above the size as would be determined by the extend of the dendritic field of a horizontal cell. In primates, three type of horizontal cells are known with different morphological characteristics and chromatic preference of connections, i.e., they are selective for a specific subset of cone types (S, M, or L).

### 2.3.3 Bipolar Cells

The bipolar cells are the main route of the visual signal from the photoreceptors to the ganglion cells (*direct* or *vertical pathway*). In the layered architecture of the retina, the bipolar cells are the key interneuron, bridging the two plexiform layers. Bipolar cells have dendrites in the OPL where they receive input signals from rods and cones, and an axon in the IPL that provides output to ganglion cells or amacrine cells. Similar to photoreceptors, bipolar cells respond to light by graded changes in the membrane potential rather than by firing spike trains.

Two basic types of bipolar cells can be identified, rod bipolar cells and cone bipolar cells. These two kinds are distinguished by the type of photoreceptor which provides the input to the bipolar cells. Cone bipolar cells can be further subdivided based on a physiological difference into on-center and off-center bipolar cells. An on-center bipolar cell is depolarized by a light stimulus on the center but hyperpolarized by light on the surround. The opposite is true for off-center bipolar cells. The surround of a bipolar receptive field is determined by input from surrounding cones, which are mediated by horizontal cells (*lateral pathway*). Bipolar cells thus have an antagonistic center-surround organization which largely determines the antagonistic center-surround organization found at the level of ganglion cells. The on/off-dichotomy is caused by the different response of the two kinds of bipolar cells to the same neurotransmitter released by the photoreceptors.

### 2.3.4 Amacrine Cells

Amacrine cells form a dense network in the IPL, making synapses with the processes of bipolar, ganglion, and other amacrine cells. The term “amacrine” was introduced by Cajal, meaning cells lacking an axon. Amacrine cells serve to shape the responses of ganglion cells by integrating, modulating and interposing the visual signal. Amacrine cells can be divided into about 40 different subtypes with different morphology and cytochemistry. Amacrine cells come in all shapes, sizes and stratification patterns, and probably every neurotransmitter of the central nervous system is to be found in some amacrine cell.

The reason for such a large variety of amacrine cells may origin from a basic difference of the retina compared to the cortex. In the cortex, almost every area receives external modulatory input from other regions. The retina, however, receives no external modulatory input. Instead, its modulatory system has independently developed through the internal amacrine circuitry (Wässle and Boycott, 1991).

By modulation the response of retinal ganglion cells is adjusted to compensate for variations caused by the environment. For example, the spectral composition of light is shifted to the red part in the morning and in the evening, while shorter wavelengths are predominant at noon. The visual system largely compensates for this spectral shift, and the modulatory system mediated by amacrine cells might be involved in the initial part of this compensation. Another example are differences in the incident illumination of the upper and lower part of the retina. Under natural viewing conditions, the lower part of the retina receives light from the sky with orders of a magnitude brighter than the light received by the upper part of the retina which comes from the ground. A special, at first sight peculiar form of amacrine cells (somatostatin-immunoreactive) has cell bodies largely confined to the lower retina, and axonlike processes projecting to the upper retina. The projection system of this type of amacrine cells might be used to equilibrate these differences in incident illumination.

### 2.3.5 Rod and Cone Pathways to Ganglion Cells

As already noted, cone signals are transmitted rather directly via bipolar cells to ganglion cells. For rod signals, two different and more complicated pathways exist. In the *first* pathway, which is believed to be taken at *twilight*, rod signals can be transmitted to cones via gap junctions. These signals then can be relayed to ganglion cells through the cone pathway. In the *second* pathway, taken at very dim illumination such as *starlight*, rods connect a special type of amacrine cells (AII) via rod bipolar cells. AII amacrine cells directly send output to off-ganglion cells and indirectly onto on-ganglion cells via cone bipolar cells. When the second pathway under starlight is taken, ganglion cells are no longer inhibited by surround illumination, so that they function no longer as a contrast rather than a light detector. The convergence of rod and cone signals to ganglion cells via different pathways allows ganglion cells to operate over the full range of illumination.

### 2.3.6 Ganglion Cells

Ganglion cells are the output neurons of the retina. The whole visual world we perceive is encoded in the firing patterns of retinal ganglion cells. Ganglion cells get their input from the retinal interneurons, namely bipolar and amacrine cells. Axons of ganglion cells are bundled in the optic nerve and project to higher visual areas of the brain, primarily to the LGN.

#### Basic properties of ganglion cells receptive fields

Following the pioneering work of Hartline (1938), who shared the 1967 Nobel prize in physiology and medicine with R. Granit for the first electrical recording of responses of individual ganglion cells to light stimuli, Kuffler (1953) and Barlow (1953) discovered that ganglion cells respond in a characteristic manner to light within a specific area of the retina, its receptive field (RF). The organization of the receptive field of a ganglion cell has three properties. *Circularity:* RF are roughly circular. *Antagonistic center-surround organization:* The RF is divided into a circular center and a surrounding annulus. The subfields of the RF are antagonistic, i.e., a light spot directed to the surround has a complementary effect on the response of the cell compared to a light spot that is directed to the center. *Parallel on- and off-pathways:* two distinct kinds of ganglion cells exist, on-center ganglion cells that respond best to bright spot with a dark annulus, and *off-center* ganglion cells which respond best to the opposite kind of stimulus, i.e., a dark spot surrounded by a light annulus. The two pathways are *parallel*: the retina contains roughly equal numbers of on- and off-center ganglion cells, and every photoreceptor sends output to both pathways.

The stratification pattern of ganglion cell dendrites within the IPL correlates with the physiological distinction into on and off channels. The IPL can be subdivided into two functionally discrete sublaminae. Off-center ganglion cell dendrites branch in the more distal sublamina a closer to the INL, whereas on-center cells stratify in the more proximal sublamina b closer to the GCL.

The size of the RF center s not uniform but varies with the degree of eccentricity from the fovea. In the fovea, the RF *center* has a diameter of a few minutes of arc, whereas in the periphery, the size increases up to 5°.

#### Contrast Detection

The antagonistic center-surround organization causes ganglion cells to respond optimally to *contrast*. Ganglion cells respond best to different illumination of center and surround, but exhibit only a relatively weak response to diffuse illumination of the entire RF. The response to contrast rather than to absolute intensity is an important principle in the visual system. The absolute

intensity is to a large degree determined by the intensity of the light source. On the contrary, the contrast between adjacent regions is invariant against variations of the illumination and encodes important scene properties like, e.g., object borders. Thus, the response to contrast enables the visual system to discount the illuminant and to yield a roughly constant perception of object brightness over a wide range of illumination conditions. The important role of contrast causes a dependence of perceived brightness on the brightness of adjacent regions, as demonstrated in various visual illusions, e.g., simultaneous contrast.

The origin of contrast detection in the retina helps to reduce transmission errors of small luminance differences that would occur if absolute intensities would be directly signaled to the brain. The existence of two distinct pathways signaling light increments and decrements, respectively, enhances the dynamic range over which light changes can be reliably encoded.

### Contrast Sensitivity Function and Difference of Gaussians Model

Because of the circular center-surround organization of receptive fields, ganglion cells show different responses to sine gratings depending on the spatial frequency of the grating. If the frequency is too low, both center and surround are equally illuminated by the same dark or bright band of the grating, resulting in little or no response. Likewise, if the frequency is too high, the on- and off-regions are stimulated by an approximately equal number of small dark or light bands, again resulting in little response. Between these two extrema lies the optimal frequency, where the width of a stripe is confined to the central region of the RF. In this case, there exists high excitation of the center and only small inhibition of the surround which is mainly covered by stripes of the other polarity, resulting in a vigorous response of the cell. The contrast sensitivity function reflects this dependence of cell response on spatial frequency and has a characteristic inverted U-shape.

Difference of Gaussians (or DoG, Marr (1982)) models of ganglion cells have been successfully employed to fit contrast sensitivity functions (Enroth-Cugell and Robson, 1966; Rodieck, 1965). In the DoG model, the antagonistic center-surround organization of the RF of an on ganglion cell is modeled by subtracting a surround Gaussian with large standard deviation from a center Gaussian with small standard deviation (or vice versa for an off ganglion cell).

### P and M cells

Besides the dichotomy of on- and off-center cells, a further dichotomy of P and M cells exists. These cells can be classified by their projections to different regions of the LGN. The M cells project to the two magnocellular layers of the LGN (ventral), whereas P cells project to the four parvocellular layers of the LGN (dorsal). M cells are also called parasol, P$\alpha$ or A and account for 10% of ganglion cells. P cells are also called P$\beta$ or B and account for about 80% of ganglion cells, with the remaining 10% being of another type which projects to the superior colliculus and the pretectum. The anatomical classification corresponds to distinct physiological properties: M cells show a phasic, non-color-specific (broad band) response and have large RFs, while P cells have a tonic, color-specific response and small RFs. M and P cells have their morphological correlates in ganglion cells with large, sparse branching dendritic trees and small, dense branching dendritic trees, respectively. Cells with similar properties like the P and M cells in primates have been first described in the cat retina, where they are termed Y and X cells.

The properties of P and M cells are summarized in Tab. 2.3.

In many vertebrates such as birds, reptiles, fish and some mammals, directionally selective ganglion cells are found, which respond best to small stimuli moving in a preferred direction at a preferred speed. In primates and cats, however, such directionally selective cells are absent.

**Table 2.3.** Different properties of P and M cells in the monkey retina (cf. Coren et al., 1994; Kaplan et al., 1990).

| | P cells | M cells |
|---|---|---|
| *Anatomy* | | |
| percentage of cells | 80% (10% other) | 10% (10% other) |
| cell body | small ("parvo") | large ("magno") |
| dendritic arborization | small, dense | large, sparse |
| LGN projection layer | parvo | magno |
| *Physiology* | | |
| receptive field | small | large |
| spatial resolution | similar | |
| response to light steps | tonic (sustained) | phasic (transient) |
| luminance contrast gain | low | high |
| conduction rate | slow | fast |
| spatial summation | linear "X" | 75% linear "X"<br>25% nonlinear "Y" |
| spectral selectivity | yes | no |
| scotopic vision | no | yes |
| *Function* | | |
| proposed functionality | detailed form analysis<br>spatial analysis<br>color vision | motion detection<br>temporal analysis |

#### Ganglion cell coverage

Given the diversity of ganglion cell types, the question arises whether any part of the visual field is processed by all classes of ganglion cells within the corresponding part of the retina. For the on-/off dichotomy of cat ganglion cells, this has been shown to be the case (Wässle et al., 1981a,b). The dendrites of both on- and off- types of ganglion cells form a dense network that completely covers the retina, providing an efficient tiling with minimal overlap. Most likely, this is true for all classes of ganglion cells. Each part of the visual field is processed in parallel by subsets of ganglion cells with different functional properties.

### 2.3.7 Dual Systems in the Retina

In the retina, different modalities of the visual world, such as color and motion, have to be processed under different illumination conditions. In this section we review the dual systems in the retina that allow for this processing.

To cope with different illumination conditions, two different kinds of photoreceptors and associated systems have evolved, the rod system concerned for scotopic vision and the cone system for photopic vision. These two systems have different properties adequate for the different and contradicting demands of high sensitivity at scotopic vision and high acuity at photopic vision.

Under either kind of vision, two fundamental systems exist. A system of two different kinds of ganglion cells, one concerned with the processing of fine detail and the other concerned with the processing of coarse, moving stimuli. The two systems feed into segregated, parallel P and M pathways. Each pathway is subdivided by a second dual system of distinct on-center and off-center cells, which signal light increments and decrements, respectively. The segregation of information in parallel pathways allows the visual system to be highly sensitive under different kinds of illumination and stimulus conditions.

The dual systems that exist in the retina are summarized in Tab. 2.4.

**Table 2.4.** Dual systems in the retina with functional characteristics given in braces.

| | | |
|---|---|---|
| rods (scotopic vision) | — | cones (photopic vision) |
| magno cellular M (motion) | — | parvo cellular P (form and color) |
| on cells (brightness) | — | off cells (darkness) |

## 2.4 Lateral Geniculate Nucleus LGN

The lateral geniculate nucleus (LGN) is the major target of retinal ganglion cells. The LGN is a substructure of the thalamus. There are two LGN, one within each cortical hemisphere, representing the contralateral part of the visual field (see Sec. 2.5.1). The LGN has a complex, six-layered structure with roughly as many cells as there are ganglion cells in the retina, namely 1.5–1.8 million cells (Fig. 2.10). The six layers can be divided into four dorsal layers of small cells, where the retinal P pathway terminates, and two ventral layers of large cells, where the retinal M pathway terminates. Thus, the populations of P and M cells, which are spatially intermingled in the retina to allow any part of the visual field to be processed by each population, are segregated in the LGN and also in the primary cortex. Besides the P/M segregation, each layer receives input from one eye only, and the order of ipsilateral (I) and contralateral (C) layers from dorsal to ventral is C–I–C–I for the parvocellular layers and I–C for the magnocellular layers. The LGN shows a precise topographic organization: adjacent cells within each layer have adjacent RFs, and cells in different layers along a line orthogonal to the layer boundaries have roughly the same RF location.

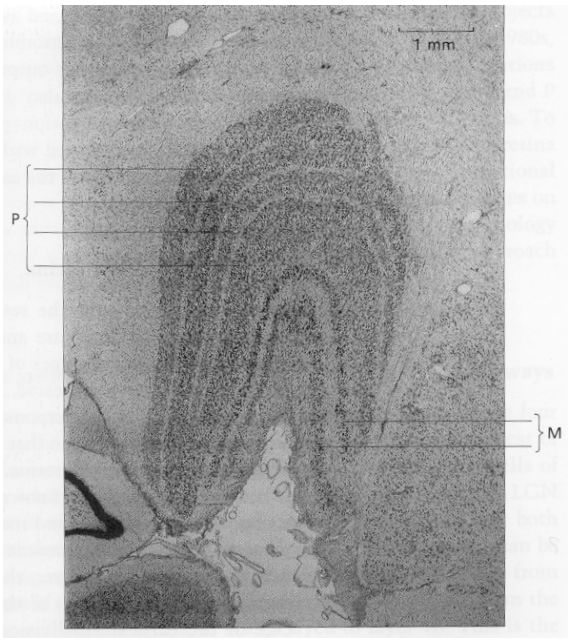


**Fig. 2.10:** Lateral geniculate nucleus. (From Zeki, 1993.)

The spatial and spectral properties of LGN cells are quite similar to retinal ganglion cells. However, the LGN is not merely a relay station between retina and cortex, since it receives most of its input from cortical feedback projections. It has been suggested that the LGN serves as a filter of incoming information or is involved in the attentional selection of salient information in the visual field to be processed by the brain (Crick, 1984).

## 2.5 Primary Visual Cortex V1

The LGN projects along the optic radiations to an area of the neocortex, the primary visual cortex or V1. The primary visual cortex, preferably in cats, is also called *area 17* following Brodmann's labeling scheme based on cytoarchitecture, or *striate cortex* because of a prominent white stripe made of myelinated axons. In macaque, the primary visual cortex covers a total area of 1300 $mm^2$ and contains about 200 000 neurons per $mm^2$ which results in a total of 260 millions cells. Compared to the number of 2 million LGN input neurons in macaque, this is an huge increase which suggests that population coding is used by the cortex for the representation of information.

#### Classes of Cell Populations

Like cells within other parts of the cortex, the V1 neurons can be divided into two basic classes, *pyramidal* and *nonpyramidal* cells. *Pyramidal cells* are large, have long spiny dendrites and are excitatory. They make up the majority of 80% of cortical cells. Pyramidal cells form projections to other regions of the brain as well as intracortical projections to different layers within an area. *Nonpyramidal cells* are small and comprise the local interneurons whose axons are confined to the same cortical area. Nonpyramidal cells with spiny dendrites are excitatory, those with smooth dendrites are inhibitory.

### 2.5.1 Retinal and Cortical Representations of the Visual Field

As we have already seen in the previous section, the visual field is represented in the cortex in a highly ordered manner. Each visual hemifield is represented in the contralateral hemisphere of the visual cortex. The representation is topographic, (or retinotopic) meaning that adjacent spots in the visual field (or retina) are represented by adjacent neurons within a cortical map (Hubel and Wiesel, 1977).

The distinctive connections of the retina to the cerebral hemispheres which occur in such a highly ordered fashion are detailed in the following.

Light that falls on the retina passes the cornea, the lens and the vitreous humor. The lens causes an optical inversion of the visual field on the retina, i.e., the projection reverses up-down and left-right. The upper half of the visual field is projected to the lower (or ventral) half of the retina, and the lower half of the visual field is projected to the upper (or dorsal) half of the retina. The right half of the visual field is projected onto the left half of each retina (or left hemiretina), i.e., onto the temporal hemiretina of the left eye and onto the nasal hemiretinal of the right eye. And vice versa for the left half of the visual field.

The output of the two retina is transmitted by two optic nerves, which split at the optic chiasm in a specific way: output from the nasal hemiretinae crosses over and projects to the opposite (contralateral) hemisphere of the brain, while output from the temporal hemiretinae does not. Thus, the right visual field is represented in the left cortical hemisphere and vice versa. Beyond the optic chiasm, the pathway is called optic tract and projects to the LGN. Due to the segregation at the optic chiasm, the LGN of each hemisphere processes only information from the contralateral visual field. The retina projects to the LGN in a detailed, point-to-point manner: fibers that start from neighboring retinal neurons contact neighboring LGN cells. The visual field is not represented isometrically in the LGN. The central part of the visual field has a proportionally larger representation than the peripheral parts of the visual field. Further, the retina projects to different layers of the LGN, resulting in a multifold representation of each point of the visual field in each layer of the LGN.

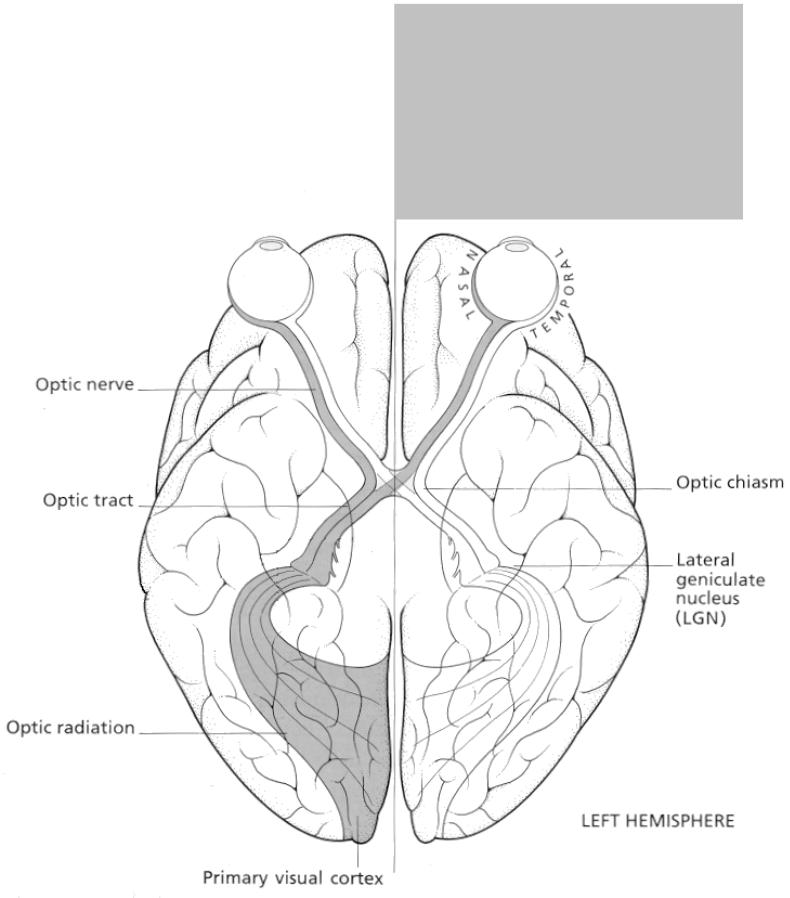


**Fig. 2.11:** Retinal and cortical representation of the visual field. (From Zeki, 1993.)

Output from the LGN travels in the optic radiation (also called bundle of Gratiolet) to terminate in the primary visual cortex V1, which is situated in the occipital lobe at the back of the brain. Each LGN projects only to the ipsilateral V1 in a topographic manner.

To summarize, the left and right part of the visual field are topographically represented in segregated maps in the contralateral side of the visual cortex. The mapping is nonuniform, with a proportionally larger representation of the central part of the visual field. In fact, about half of the neural mass both in the LGN and the cortex is involved in the processing of the foveal and near-foveal regions.

### 2.5.2 Principles of Cortical Architecture

#### Topographic Mapping and Foveal Magnification

As already noted, the visual field is topographically represented in the LGN and also in the primary visual cortex. Signals from ganglion cells with adjacent RFs terminate in adjacent location in the LGN which in turn project to neighboring cells within layer 4. The topographic or retinotopic mapping is highly nonuniform. A much larger part is devoted to the representation of the fovea compared to the periphery. The nonuniform representation can be formalized by a logpolar mapping.

The nonuniform mapping is predominantly determinated by the variations of ganglion cell density with retinal eccentricity. The mapping can be quantitatively described by the *cortical magnification factor*, which indicates the amount of cortex associated with one degree of visual field. The cortical magnification factor in monkey fovea is 100–300 $\mathrm{mm}^2/\mathrm{degree}^2$, which is 1000 times larger than in the periphery (Wässle et al., 1989, 1990).

### Vertical Layers and Vertical Circuitry

The human visual cortex is approximately 2 mm thick and has a layered architecture. Based on anatomical and morphological criteria, the visual cortex can be divided into six layers, which are numbered in increasing order from the pial surface to the white matter. Cells within each layer have specific afferents and efferents which at least partly determine their functional role. Layer 4 is the principal input layer which receives afferents from the LGN. Neurons in the superficial layers 2 and 3 receive input from layer 4 neurons. They send output to extrastriate cortical areas such as V2–V5 and MT, and make connections via the corpus callosum to reciprocal within the other cortical hemisphere. Neurons in layers 2 and 3 further make intracortical feedback connections to layer 5. The deep layers 5 and 6 receive input from layers 2, 3 and 4 and send feedback input to subcortical areas such as the LGN. The most superficial layer 1 is comprised primarily of axons and dendrites and can be considered as wiring layer of cell connections. The principal architecture of V1 is sketched in Fig. 2.12

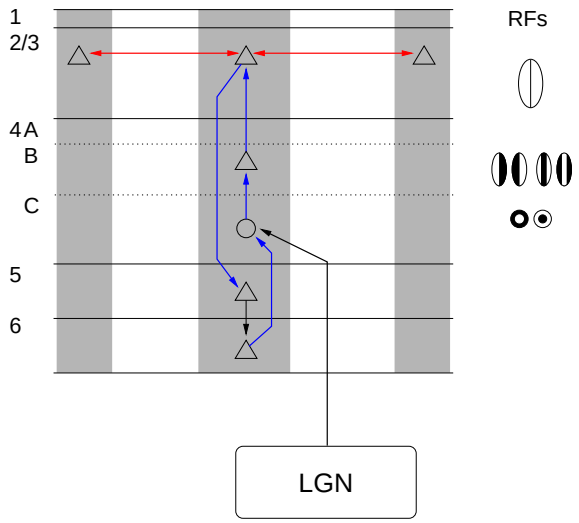


**Fig. 2.12:** Sketch of the principal architecture in V1.

In primates, the input layer 4 can be further subdivided into four sublayers or sublaminae, labeled A, B, C$\alpha$ and C$\beta$. LGN axonal input terminates in sublaminae C, with axons from the M pathway terminating in C$\alpha$ and axons from the P pathway terminating in layer C$\beta$.

Within V1, three major intracortical pathways exist. In an upward stream, the information flow goes from input layer 4 to layer 2 and 3 which projects to other cortical areas. Cells in layer 2 and 3 also take part in an intracortical excitatory circuit by projecting to layer 5. Cells in layer 5 project to layer 6 which completes the local excitatory circuit by projecting to inhibitory smooth stellate cells in layer 4. The inhibitory cells modulate the reponses of other excitatory cell within layer 4. Layer 6 cells also receive input from layer 4 and in turn make feedback connections to layer 4. The functional role of these various upward, downward and feedback connections is not clear. It has been suggested, e.g., that the recurrent intracortical interaction dynamically creates cells with more complex RF properties such as endstop cells (Bolz et al., 1989) or helps to sharpen orientation tuning of cells (e.g., Adorján et al., 1999; Somers et al., 1995).

To summarize, each layer makes specific input/output connections which determine its functional properties.

### Columnar Organization

**Orientation Columns and Receptive Fields in V1** In the retina and also in the LGN, cells have circular receptive fields (RF). In the primary visual cortex, a major qualitative change of RF structure occurs: the vast majority of cells have elongated RFs and respond best to bars, lines, or gratings of a preferred orientation. The pioneering work has been done by David Hubel and Torstein Wiesel who received the Nobel prize for medicine and physiology in 1981 (Hubel and Wiesel, 1962, 1968, 1974, 1977). Cells with circular RF, like in the LGN, are only found in input layer 4 and within specific regions of layer 2/3, the so-called blobs.

According to Hubel and Wiesel, the cells are categorized into two major groups, *simple cells* and *complex cells*. *Simple cells* have two or three separated, elongated subfields with alternating on and off responses and have specific axis of orientation. The most effective stimulus is a pattern of small dark and light patches that coincides with the RF subfields, such as a bar or line of a specific orientation. The same stimulus, but of orthogonal or oblique orientation evokes, no or only a small response. Such a stimulus is most effective for another simple cell with a RF at the position, but of orthogonal preferred orientation. Thus, any point in the visual field is analyzed in parallel by simple cells with different axes of orientation. The preferred orientation varies in discrete interval of about 10°, resulting in about 20 different orientations to be analyzed for the full range of 180°. Simple cell RFs are sketched in Fig. 2.13. As originally suggested by Hubel and Wiesel, the axis or orientation is predominantly determined by the input of LGN cells with properly aligned RFs (Reid and Alonso, 1995; Stryker et al., 1990).

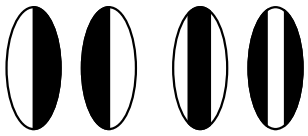

**Fig. 2.13:** Sketch of simple cell receptive fields. White regions indicate excitatory subfields, dark regions indicate inhibitory subfields. RFs of in vivo simple cells are less regular, but exhibit the same basic structure.

*Complex cells* have a preferred axis of orientation and larger receptive fields which, unlike simple cells, cannot be subdivided into distinct on and off subfields. The properties of complex cells have been suggested to result from pooling simple cells of different contrast polarity or slighty offset RF positions, but with the same preferred orientation.

Simple and complex cells decompose the outline of the visual image into short line segments of a certain orientation. This is probably the neural substrate for subsequent form analysis and object recognition, reminding of the primal sketch (Marr, 1982), the twodimensional representation of shape and contour.

Simple and complex cells are organized into vertical *orientations columns* of about 30–100 $\mu$m. Simple and complex cells within a column have the same preferred axis of orientation and RFs at almost the same position. Each orientation column by definition runs from the pial surface to the white matter and contains all cortical layers. Thus, a distinct orientation at a certain point in the visual field is analyzed at different levels of abstraction by simple and complex cells in different layers within one orientation column

**Ocular Dominance Columns and Blobs** Besides the orientation column, further vertical units exist in the primary visual cortex, *ocular dominance columns* and *blobs*.

Afferent fibers from different LGN layers associated with the ipsi- and contralateral eye terminate in separate stripes in layer 4. These *ocular dominance columns* have a width of 450 $\mu$m and are most prominent within layer 4. In the layers above and below layer 4, the information is combined for binocular processing, a prerequisite for further processing resulting in binocular depth perception.

The *blobs* are regions within layers 2 and 3, where cells have no orientation preference. Blobs receive direct input from the LGN and are concerned with the processing of color.

**Hypercolumns** As revealed by recordings of RFs at tangential penetrations, the preferred axis of an orientation column varies in an orderly manner. Adjacent orientation columns show a discrete shift in orientation preference of about 10°. Hubel and Wiesel introduced the concept of a *hypercolumn* which is a set of columns responsive to a particular region in space for all orientations and both eyes. Thus, a hypercolumn can perform a complete analysis (complete with regard to the properties analyzed within V1) for a particular region in the visual field. A hypercolumn has a size of about 1 mm$^2$. The average RF diameter within a hypercolumn and the inverse cortical magnification factor highly correlate, suggesting that the size of a hypercolumn does not vary with the eccentricity of RF position (Hubel and Wiesel, 1974).

### Horizontal Connections

Cells within different columns are linked by horizontal connections within a layer. Horizontal connections are most prominent between pyramidal cell in the superficial layers 2 and 3 (Gilbert and Wiesel, 1979; Rockland and Lund, 1983). The axons of layer 2/3 pyramide cell can extend up to 3 mm from the cell body and make clusters of axon colaterals at regular intervals that approximate the width of a hypercolumn. These long-range horizontal connections have a high functional specifity, i.e., only cells which share the same functional properties are connected, such as cells with the same orientation preference (Gilbert and Wiesel, 1989; Ts'o et al., 1986).

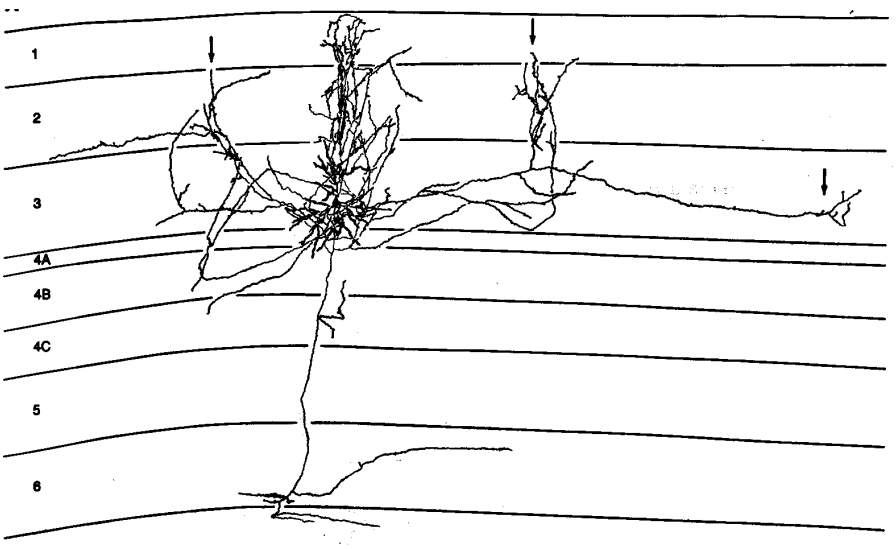


**Fig. 2.14:** Axonal arborization of a layer 2/3 pyramide cell show long-range horizontal connections. (From McGuire et al., 1991.)

By horizontal long-range connections, the response properties of a cell can be influenced by stimuli *outside* its classical RF. Horizontal long-range connections are involved in the processing of visual contrast and are consequently thought to be a neural substrate of psychophysical *context*

*effects*. For the processing of contour shape, e.g., it has been suggested that long-range connections implement the Gestalt principle of good continuation.

## 2.6 Higher Visual Areas

There is a large number of prestriate areas that are predominantly or exclusively concerned with vision (Maunsell and Newsome, 1987). Only some of them have been studied extensively, namely V2, V3, V4 and V5. Area V5 is also called middle temporal area (MT). The secondary visual area V2 contains like V1 all functional groups of cells. The other areas V3–V5 have more specialized functions as detailed below. Area V4, e.g., is primarily concerned with color. The most important principle of visual processing in higher visual areas is the parallel processing of different modalities in segregated, different area.

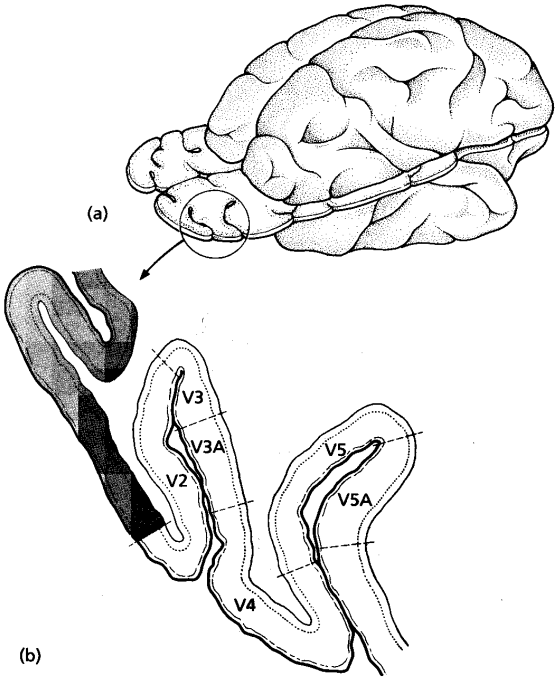


**Fig. 2.15:** Primary visual cortex V1 (underlayed in gray) and higher cortical areas V2–V5. (From Zeki, 1993.)

It has been suggested that the prestriate areas can be broadly grouped into two pathways (Mishkin et al., 1983): a ventral or *what* pathway that terminates in the temporal lobe and is concerned with objection recognition, and a second dorsal or *where* pathway that leads to the parietal lobe and is concerned with spatial vision. It has been hypothesized that the parvocellular and the magnocellular pathways exclusively contribute the what and where pathways, respectively. Recent studies, however, reveal that the doctrine of two pathways with separated input by the parvo- and magnocellular system is an oversimplification. At least three different pathways exist, concerned with *form*, *color*, and *motion*.

In this section we give a brief review of visual processing beyond the primary visual cortex. A more detailed description of the segregated, parallel processing with a particular focus on the processing of color is given by Zeki (1993).

### 2.6.1 Secondary Visual Cortex V2

Area V2 is an area that surrounds the primary visual cortex V1. Like V1, V2 has a topographic organization and contains a complete representation of all functional modalities of vision. V2 projects to the same specialized visual areas as does V1, i.e., V3, V4 and V5. There exists a topographically organized, point-to-point input from V1 to V2 (Cragg, 1969; Zeki, 1969).

V2 contains three major functionally different populations of cells, as revealed by physiology, which can be related to different patches in V2 as marked by cell staining techniques: Cells within *thick stripes* are *direction-selective* and concerned with the detection of motion and the three-dimensional organization of object, cells within *pale stripes* are *direction-selective* and concerned with the detection of form, cells within *pale stripes* are *wavelength-selective* and concerned with the perception of color (DeYoe and Van Essen, 1988; Livingstone and Hubel, 1987; Zeki, 1976). Thus, V2 further segregates the modalities as represented in V1. Cells with more specific response properties, such as cells responsive to illusory contours, have been found in V2, but not within V1 (von der Heydt et al., 1984). This suggests that in V2, the same modalities as in V1 are analyzed and represented, such as orientation of object contours, but on a higher, more abstract level.

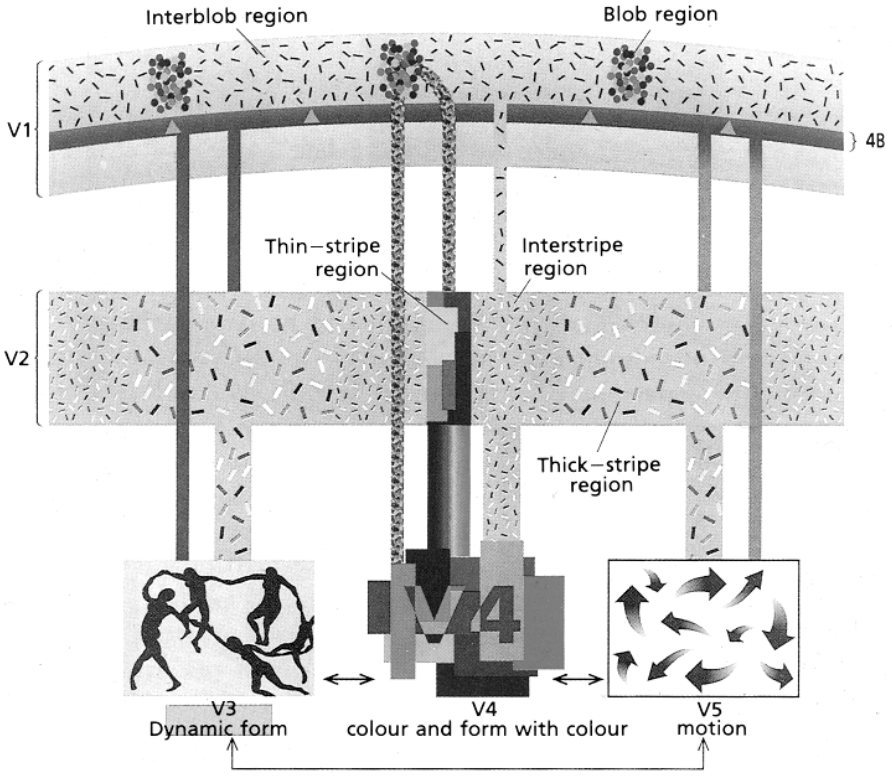


**Fig. 2.16:** Sketch of the perceptual pathways and their anatomical connections from V1 to the more specialized prestriate areas V2–V5. (From Zeki, 1993.)

### 2.6.2 Prestriate Visual Areas V3, V4 and V5

As already stated in the previous chapter, the different functional compartments within V2 can be related to different pathways which project to distinct cortical areas. An overview of the main pathways that project from V1 via V2 to higher prestriate areas V3–V5 is given in Fig. 2.16.

Area V3 is concerned with the processing of *dynamic form*. Area V3 receives input from direction and orientation selective cells in V1, both directly and through the thick stripes in V2. This pathway is driven by the M cell of the LGN.

Area V4 is concerned with the processing of *color* and *form with color*. The *color pathway* to V4 relays from the blobs in V1 via the thin stripes in V2 and also directly. The *form and color pathway* relays from the interblobs in V1 via the pale interstripes in V2 to V4 and also directly. Both pathway are principally driven by the P cell of the LGN.

Area V5 is concerned with the processing of *motion*. Area V5 receives direct input from the V1 layer 4B and also through the thick stripes in V2. This pathway is driven by M cells of the LGN.

To summarize, there are two form pathways, one concerned with dynamic form and driven by the M system, and another concerned with form and color, which is driven by the P system. Besides these two form pathways, there is a motion pathway which is principally driven by the M system and a color pathway principally driven by the P pathway. Signals from the P and M pathways can mix, so that the input to the specialized areas are not strictly separated according to these two pathways.

Chapter 3

# Contrast Processing

## 3.1 Introduction and Motivation

The robust detection of contrast is the first important step in the processing of visual stimuli for both natural as well as artificial vision systems. The term "contrast" here refers to *luminance contrast*, i.e., local changes in the luminance distribution of the input stimulus. More specifically, "contrast" denotes *oriented contrast*, i.e., small oriented segments of coherent luminance change such as lines or edges.

The detection of contrast is important, because contrast indicate a change in the physical world. The detection of such changes is a prerequisite for a large variety of basic vision functions, such as, e.g., motion estimation, depth perception or higher level functions such as object recognition.

A good contrast detector has to fulfill several criteria, such as a localized and unimodal response at contrast locationss and no response to homogeneous regions. Another important property is robustness, i.e., a contrast detector should not respond to noisy fluctuations within the input signal. Noise robustness is of particular importance, because common methods (Canny, 1986; Marr and Hildreth, 1980; Sobel, 1970) as well as our approach compute an approximation of the first or second order derivative of the image, a process by which noise is increased. Consequently, differentiation-based contrast detection is inherently sensitive to noise.

On the other hand, human and animal vision systems are able to extract the relevant information in noisy, cluttered scenes. In this chapter, we present a scheme for robust contrast detection which is motivated by early processing stages in the mammalian visual system. Our investigations are guided by computational considerations as well as physiological findings of basic architectural principles and wiring schemes found in vivo. The objective for the reference to biological vision systems if two-fold: First, the precise algorithmic description of basic response properties and processing stages of cells found in the early visual pathway allows for a deeper understanding of the computational role of these principles. Second, from a technical or engineering point of view, the construction of artificial vision systems can benefit from using computational principles encountered in natural vision systems, which are by far todays most successful and elaborated general purpose vision systems.

In particular, we have studied the wiring of isotropic LGN cells to simple cells subfields. In LGN cells as well as retinal ganglion cells, information is represented in two distinct, complementary domains, namely on and off cells. Based on physiological evidence, we propose a mechanism of *dominating opponent inhibition* (DOI), where a simple cell subfield is driven by both on and off domains, receiving more heavily weighted input from the opponent pathway. The DOI mechanism is integrated into an existing nonlinear simple cell model (Neumann and Pessoa, 1994; Neumann et al., 1999).

One motivation for the DOI mechanism was a study by Hammond and MacKay (1983) showing a strong decrease in simple cell responses to luminance gradient reversal. We show that the model with DOI successfully predicts the simple cell responses measured in the study. With

identical parameter settings, the model is applied to the processing of images. We show that DOI significantly reduces the response to noise. Moreover, this reduced response is largely invariant against variations of the noise level. This property may give a rationale for the strong inhibition measured physiologically and for the representation of contrast information in two complementary domains. Based on these findings we hypothesize that the visual system uses dominating opponent inhibition in order to robustly extract features in noisy environments.

This chapter is organized as follows: The first three sections give a more general overview of the addressed topic. In the first section, we give a brief review of biological findings regarding the very first stages of contrast processing in vivo. In this review we focus on the organization of LGN input to form simple cell subfields, where the proposed mechanism of DOI comes into play. In Sec. 3.3, we summarize other main approaches to model simple cell responses. In Sec. 3.4, a short review of traditional contrast detection methods in computer vision and more recent approaches is given. This section concludes the more general part of this chapter. In Sec. 3.5 the simple cell model together with the proposed mechanism of DOI gives is formally introduced. In Sec. 3.7 we show simulation results obtained with the new model. The results are compared to the nonlinear model without DOI and a basic linear model. The simulations cover a wide variety of topics, ranging from physiological studies on responses to contrast gradient reversal and orientation tuning, psychophysically measured responses to modified Glass dot pattern, to image processing. In Sec. 3.8, the response properties of DOI are evaluated. We present a stochastic analysis of the response to noisy homogeneous regions and a numerical evaluation of responses to homogeneous noisy and step edges. In Sec. 3.9 we present a sample pioneering application of DOI for object recognition. Section 3.10 concludes the chapter.

## 3.2 Empirical Findings

In this section, we first give an introduction covering basic simple cells properties. We shall then review major empirical findings mostly from physiology. Two direction of research can be distinguised, characterized by the focus on thalamic vs. cortical contributions to simple cell properties. More detailed reviews are given by Martin (1984) and Reid and Alonso (1996).

### 3.2.1 Introduction to Simple Cells

Processing of visual stimuli begins in the retina, where electromagnetic radiation within a certain frequency band is transformed into a neural code. The first major processing stage consists of retinal ganglion cells with circular, center-surround receptive fields (RFs) (Enroth-Cugell and Robson, 1966). Axons of ganglion cells form the optic nerve which terminates in a relay structure of the thalamus, the (LGN). The RFs of LGN cells have a circular, center-surround organization similar to that of retinal ganglion cells. LGN cells project upon the primary visual area V1 in the occipital lobe. In the visual cortex, to-called simple cells are found which have a RF structure that is considerably different from that of the ganglion cells. Simple cells have elongated RFs and respond best to bars of a certain orientation and position. The RF properties of simple cells have first been described by Hubel and Wiesel in cat (1962) and later in macaque (1968). Simple cell RFs have an alternating side-to-side arrangement of excitatory and inhibitory areas, the so-called on and off subfields. Simple cell RFs differ in their size and layout and can be tentatively grouped into even-symmetric and odd-symmetric cells based on the arrangement of on and off subfields with respect to the RF main axis. Common subfields as found by Hubel and Wiesel (1962) are sketched in Fig. 3.1.

In 1981, Hubel and Wiesel were awarded with the Nobel Prize for their research in the neurophysiology of the visual system.

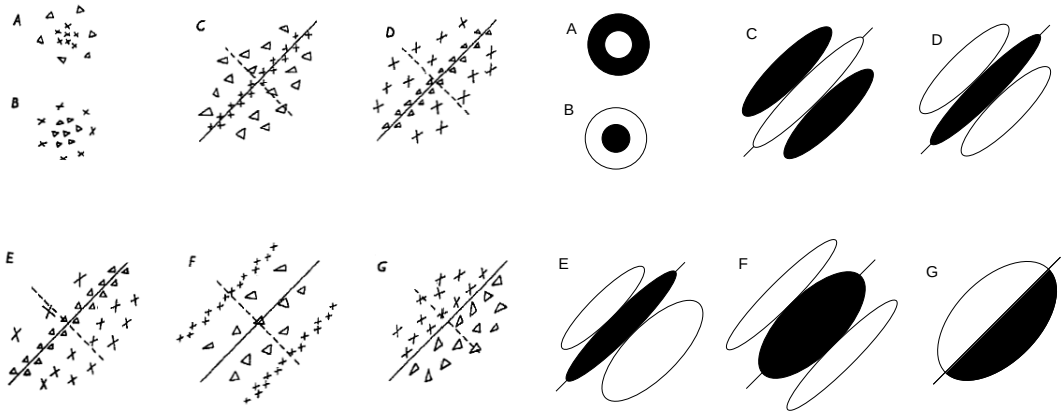


**Fig. 3.1:** Receptive fields of LGN cells and simple cells. *Left:* Sketch of various RFs; crosses denote areas of excitatory response, triangles denote areas of inhibitory response. (From Hubel and Wiesel, 1962.) *Right:* Corresponding RFs using an alternative display convention: white areas denote excitatory response, black areas denote inhibitory response. This convention is more intuitive, since the pattern of dark and light regions that make up the RF directly reflects the most effective stimulus for this type of cell. This conventions in used throughout this work.—*A and B:* RFs of LGN on cells *(A)* and off cells *(B)*. *C–G:* RF of simple cells with odd symmetry *(C–F)* and even symmetry *(G)*.

Compared to the transformation between retina and LGN, where RF properties of cells are virtually the same, the transformation of RF properties taking place from LGN to cortex is impressive: cortical cells have the fundamental new property of *orientation selectivity*. Since the days of Hubel and Wiesel, the origin of orientation selectivity has become one of the most thouroughly studied topics of the visual cortex. Various empirical studies have been undertaken and a plethora of models have been devised. Undoubtly, valuable new insight has been gained, but even after almost fourty years of intense research the underlying origin of orientation tuning has not yet been fully eludicated.

### 3.2.2 The Role of LGN Input

Having discovered the orientation selectivity of simple cells, Hubel and Wiesel (1962) also delineated a scheme for the origin of this property. This likewise elegant and straightforward scheme have become the forefather of whole class of simple cell models:

> To account for the spatial arrangements of excitatory and inhibitory regions of simple cortical fields we may imagine that upon each simpletype cell there converge fibres of geniculate origin having 'on' or 'off' centres situated in the appropriate retinal region. For example, a cortical cell with a receptive field of the type shown in Textfig. 2 C [Fig. 3.1 C in the present text] might receive projections from a group of lateral geniculate cells having 'on' field centres distributed throughout the long central narrow region designated in the figure by crosses. [...] For example, field type $G$ could be formed by having geniculate afferents with 'off' centres situated in the region below and to the right of the boundary, and 'on' centres above and to the left. (Hubel and Wiesel, 1962, pp. 141–142)

In other words, the orientation tuning is determined by the input of property aligned rows of LGN cells. The proposed simple cell model according to Hubel and Wiesel is shown in Fig. 3.2.

The proposal of Hubel and Wiesel is based, either implicitly or explicitly, on three assumptions with regard to the LGN input: (i) the input must come from "a large number of lateral geniculate cells" (Hubel and Wiesel, 1962), (ii) the input must be sufficiently strong and, most important,

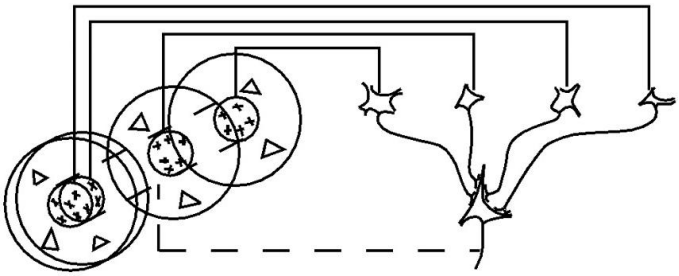

**Fig. 3.2:** Simple cell model of as originally proposed by Hubel-Wiesel's. The simple cell *(botttom right)* receives input from LGN on cells *(top right)*. The RFs of these on cells are properly aligned *(left circles)* and define the RF of the simple cell, which will have an elongated on center *(left, dashed lines)*. (Redrawn from Hubel and Wiesel, 1962.)

(iii) there must exist a precise and highly specific geometric wiring between properly aligned LGN cells and their cortical target cells. Empirical evidence for each of these assumption has been found which we shall detail in the following.

As reviewed by Martin (1988), a single synapse provides only a depolarization of about 0.2 mV. This is not enough to put the neuron above the firing threshold, which lies about 10–15 mV above resting level (Douglas et al., 1988; Martin, 1984). With respect to the proposal of Hubel and Wiesel the question arises whether there is a sufficiently large number of synapses of LGN afferents which contact a single cortical target cell. By evaluating HRP-filled LGN axons making synaptic contacts with Golgi-stained neurons, Martin and coworkers found that in most cases one afferent made only a single synaptic contact with the target cell. However, from the number of 400–800 axons of X cells that cover each point and the much more wider dendritic tree of a simple cell it can be concluded that "the availability of sufficient axons for each to provide but one synapse to a single target cell is not a problem" (Martin, 1988, p. 656).

It follows that the pure number of LGN afferents to simple cells can undoubtly be provided. But is the input also sufficiently strong? Evidence comes from analysing the firing patterns of pairs of thalamic and cortical cells (Tanaka, 1983). The two cell pairs were selected based on qualitative hand-plots of RF to have matching subfields, e.g., an on LGN cell overlapping with the on subfield of a simple cell. Using cross-correlation analysis it has been shown that the firing pattern of such a pair of thalamic and cortical cells are both strong and fast, i.e., a cortical neuron firing in the range of 1–4 ms following a thalamic spike. When a correlation was found, the thalamic spikes predicted up to 20% of the simple cell's spikes. This about ten times weaker than the correlation between cortical visual neuron (Toyama et al., 1981). Consequently, the thalamic input is presumably of particular importance for driving the cortical neuron.

The results of Tanaka (1983) have been confirmed and extended in a study of Reid and Alonso (1995). In this study, a quantitative technique of reverse correlation with spatio-temporal white-noise stimuli of random checker boards were used. This technique allows a precise mapping of RFs from simultaneously recorded LGN and cortical neurons. The RF maps of LGN and cortical cells were than fitted to a DoG function (Enroth-Cugell and Robson, 1966; Rodieck, 1965) or Gabor function (Jones and Palmer, 1987), respectively, to measure the degree of overlap of RF subfields. Reid and Alonso found a precise and specific connection pattern: There was a high probability of finding a monosynaptic connection between a geniculate afferent and a cortical simple cell when the center of the LGN RF was well superimposed on the simple cell subfield of the same sign (Fig. 3.3). This study provide evidence for the high degree of specific wiring of geniculate afferents as required by the model of Hubel and Wiesel.

So far, we have reviewed ample evidence for the Hubel-Wiesel model which is based on feedforward input to cortical cells from thalamic afferents. However, this is only part of the story, since simple

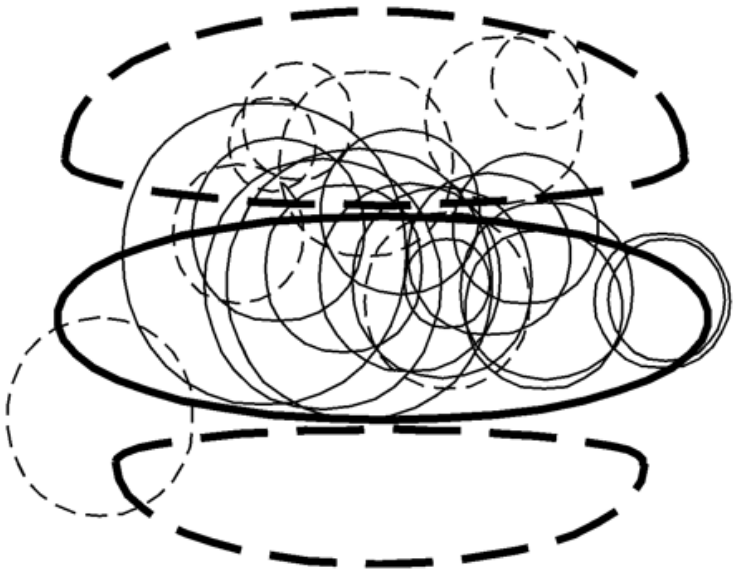

**Fig. 3.3:** Summary of the study of Reid and Alonso (1995). The *elliptic regions* are the central *(solid)* and flanking *(dashed)* subfields of a simple cell. The *circles* represent RF field centers of LGN cells which are monosynaptically connected to the simple cell. In almost every case, the RF center of the connected LGN overlapped the simple cell subfield of the same polarity. (From Reid and Alonso, 1996.)

cells also receive input from cortical neurons. The role of this intrinsic cortico-cortical connections in determining the orientation tuning of simple cells is reviewed in the following.

### 3.2.3 The Role of Cortical Input

The exclusive role of feedforward input as delineated in the Hubel-Wiesel model is questioned by the small proportion of LGN afferent to the total synaptic input to a simple cell. In fact, only 5–20% of the thalamic input to a layer 4 simple cell come from thalamic afferent, so that the vast majority of input is provided by cortical cells (Ahmed et al., 1994; Garey and Powell, 1971; LeVay and Gilbert, 1976; Peters and Payne, 1993). However, the plain quantity of afferents need not determine its efficacy. Reid and Alonso (1996) state that the sparseness of thalamic input can be compensated for by various mechanisms, such as synchronized input, among others.

Evidence for the role of cortical input in determining the orientation tuning comes from studies where the inhibitory cortical input is pharmalogically blocked by bicuculline, an antagonist of the inhibitory neurotransmitter GABA.

In a pioneering study, Sillito (1975) found the orientation selectivity of simple cells under iontophoresis of bicuculline to be reduced, but not eliminated. This supports the view that orientation selectivity is initially set up by the thalamic afferents and sharpend by intracortical inhibitory processes. Another study sought for a more complete elimination of inhibition by a parallel application of bicuculine and an inhibitor of GABA synthesis (3-mercaptoproprionic acid). In this case, orientation selectivity was completely lost in 2 out of 4 investigated cells (Tsumoto et al., 1979). Motivated by these findings, Sillito et al. (1980) reinvestigated the orientation tuning using a more potent GABA antagonist (N-methyl bicuculline, Nmb) for extracellular blockage of inhibiton. The results were impressive: All 13 investigated cells "showed a reversible and reproducible loss of orientation selectivity during Nmb application" (Sillito et al., 1980). From this study, orientation selectivity seems to be completely determined by intracortical inhibitory interactions.

More recent studies however provide information pointing in the opposite direction. In a study by Nelson et al. (1994), GABA was blocked *intracellularly* during whole cell recording. Although the

directional selectivity changed, all 18 out of 18 examined neurons remained selective for stimulus orientation. Ferster et al. (1996) eliminated the influence of all cortical input by cooling. During the cooling, virtually no cortical action potential were recorded, so that LGN cells provide the sole input to the simple cells. In this condition the simple cell responses were greatly attenuated, yet the orientation tuning was unchanged. Contradicting the findings of Sillito et al. (1980), the results of Nelson et al. (1994) and Ferster et al. (1996) imply that thalamic input is sufficient to provide the orientation tuning of simple cells.

### 3.2.4 Summary of Empirical Findings

By reviewing only the most important studies, an abundance of sometimes conflicting empirical findings has become apparent. The ultimate answer to the question of the origin of orientation tuning has still to be found. However, the emerging picture form the various findings is that orientation tuning cannot be attributed and explained by a single mechanisms. Most likely, the visual cortex has not been too particular in selecting the mechanisms by which robust orientation tuning can be provided: Both thalamic afferents as well as excitatory and inhibitory cortico-cortical inputs presumably play an important role. We tend to the position that LGN input provides the initial strong bias for orientation tuning, which is than modulated and sharpend by intracortical interactions.

## 3.3 Review of Simple Cell Models

Simple cell models can be divided into two major groups as implied by the supposed origin of simple cell properties: *feedforward models*, assuming a dominant role of LGN input, and *recurrent models*, assuming a dominant role of recurrent interactions. The recurrent interactions of the models reviewed in this chapter are limited to a localized population of cells whose RFs have different orientations but sense approximately the same location in the visual field, such as a cortical hypercolumn. This is to be distinguished from models that incorporate long-range horizontal interactions or recurrent interactions between cells of different layers or even cortical areas to describe the modulation of cell responses from outside its classical receptive field, as observed experimentally (Gilbert and Wiesel, 1990; Kapadia et al., 1995; Maffei and Fiorentini, 1976; Nelson and Frost, 1978, 1985; Zipser et al., 1996). Such influences of cell properties from outside its classical receptive field along with modeling approaches and possible functional roles are discussed in detail in the next chapter.

Other reviews with focus on particular properties of simple cells are given by von der Heydt (1987), Ferster and Koch (1987), Das (1996), Sompolinsky and Shapley (1997), Carandini et al. (1999), and Ferster and Miller (2000).

### 3.3.1 Feedforward Models

Feedforward models are basically in agreement with the classical proposal of Hubel and Wiesel (1962), in which simple cells responses originate from summing LGN cell responses of proper geometrical arrangement.

Feedforward models can be further subdivided into pure feedforward models and feedforward that incorporate inhibition.

### Pure Feedforward Models

Following the proposal of Hubel and Wiesel (1962), simple cells sample their input from LGN cell of proper geometrical layout. If the net LGN input is above some threshold, below which the response of the model simple cell is zero, the cell linearily responses to the input. The the feedforward models according to Hubel and Wiesel in given in Fig. 3.4.

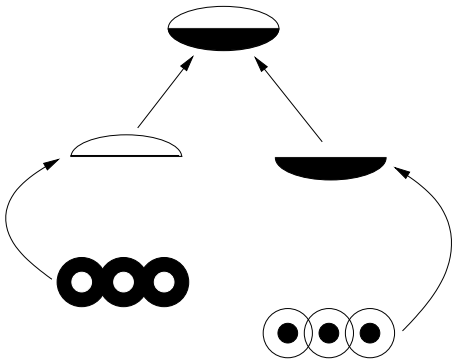

**Fig. 3.4:** Feedforward simple cell modell. An on subfields *(middle, left)* receives excitatory input from properly aligned LGN on cells *(bottom left)*; an off subfields *(middle, right)* receives excitatory input from properly aligned LGN off cells *(bottom right)*. Simple cell subfields are linearily integrated to form the simple cell response. *Arrows* denote excitatory input. For clearness of display, subfields are drawn separated.

Simple cells can than be modelled as linear spatial filters as reviewed by von der Heydt (1987). The weighted spatial summation of input as defined by the response profile of the RF can be described by differences of Gaussians with offset centers (Heggelund, 1981, 1986a,b; Heggelund et al., 1983; Parker and Hawken, 1988), by Gaussian derivatives (Koenderink and van Doorn, 1987, 1990, 1992; Stork and Wilson, 1990; Young, 1985), or by Gabor functions (Daugman, 1980, 1984, 1985; Jones and Palmer, 1987; Jones et al., 1987a,b; Marčelja, 1980; Pollen and Ronner, 1983; Webster and De Valois, 1985; Wilson et al., 1990). Gabor functions are motivated by the uncertainty principle as detailed below, while Gaussian derivatives exhibit optimal feature detection properties (Canny, 1986; Marr and Hildreth, 1980). The importance of both Gabor functions and derivatives of Gaussians as general purpose feature dectors in early vision is stressed by Michaelis (1997). Michaelis (1997) also investigated in detail the relationships of both classes of functions. It turns out that derivatives of Gaussians can be fitted quite good by Gabor functions. Especially the first order Gaussian derivative and higher order derivatives can be fit almost perfectly, while the largest $L^2$ error occured for the second order derivative (10%) and for the third order derivative (6%) (Michaelis, 1997, p. 46). From a qualitative point of view both Gabor functions and Gaussians yield quite similar results.

Gabor functions are sine or cosine modulated Gaussians. Gabor functions have been derived by Gabor (1946) on the basis of the uncertainty principle. The uncertainty principle states that a function cannot be localized in both the spatial and the frequency domain. The unique function that exhibits the best joint localization in both the space and frequency domain are Gabor functions.

The relevance of Gabor's work in the vision domain has first been stressed by Marčelja (1980) and Daugman (1980). Daugman (1985) then extended the uncertainty principle to 2D images and showed that the 2D Gabor filters exhibit the best possible joint localization in space and frequency. As pointed out by von der Heydt (1995), this reflects the duality of simple cells RFs for feature localization (Hubel and Wiesel, 1962, 1968; Livingstone and Hubel, 1987) and spatial frequency selectivity (De Valois and De Valois, 1988; De Valois et al., 1982; Maffei and Fiorentini, 1973; Wilson et al., 1990).

However, the importance of the uncertainty principle for early vision should not be overemphasized. The uncertainty principle is based on the assumption of a single measurement, which is contrasted to the large number of measurements at each spatial location that are available to visual cortex: Cells within a single cortical hypercolumn all sense virtually the same spatial location and sample orientated contrast approximately every 10 degree.

### Feedforward Models with Inhibition

The original proposal of Hubel and Wiesel is based on excitatory synapses. However, it has been suggested right from the beginning that "one should, however, consider the possibility of direct inhibitory connections." (Hubel and Wiesel, 1962, p. 142).

The inhibition is provided in a complementary arrangement compared to the excitation: Each on subfield is excitated by properly aligned geniculate on cells and also inhibited by properly aligned off cells. Similarily, an off subfield is excited by geniculate off cells and inhibited by geniculated on cells. The inhibition is mediated by a cortical interneuron. Thus each subfield receives *opponent inhibition* from cells of the opponent pathway (Ferster, 1987, 1989; Glezer et al., 1980; Palmer and Davis, 1981). A sketch of the assumend wiring scheme of direct excitation paralled by opponent inhibiton is shown in Fig. 3.5. Models using opponent inhibition have been termed "push-pull" models by von der Heydt (1987).

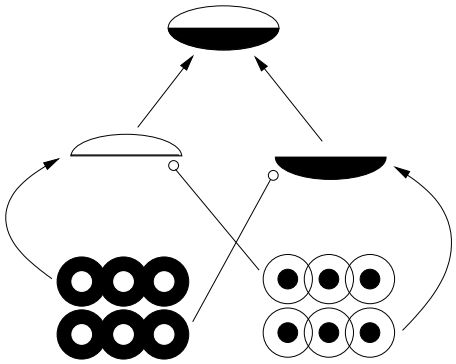

**Fig. 3.5:** Simple cell modell with opponent inhibition: An on subfield *(middle, left)* receives excitatory input from properly aligned LGN on cells *(bottom left)* and inhibitory input from LGN off cells *(bottom right)*. The reverse wiring pattern exists for the off subfields *(middle right)*. *Arrows* denote excitatory input, *circles* at the end of lines denote inhibitory input. For clearness of display, subfields are drawn separated; interneurons are omitted to simplify the sketch.

This wiring scheme is used in our model where the proposed mechanism of *dominating opponent inhibition* (DOI) assumes a stronger weighted input of the inhibitory component, as shall be detailed in Sec. 3.5.

Though strong evidence exists in favour for a *linear* simple cell model (Heeger, 1992, 1993), a pure linear model fails to account for all aspects of simple cell response properties. For example, a linear system obeys the principle of superposition. Consider a simple cell which responds vigourously to a grating of the preferred orientation but shows no response to a grating of oblique orientation. The principle of superposition would predict that the response to two gratings presented superimposed would be equal to the response evoked by the grating of preferred orientation alone. In fact, the response is strongly diminished, the so-called cross-orientation inhibition (Bonds, 1989). A detailed review of nonlinear properties of simple cells is given by Carandini et al. (1999).

Based on these findings a model has been devised which employes a normalization of cells responses by nonlinear, divisive interactions (Carandini and Heeger, 1994; Carandini et al., 1997) . The model consists of a linear stage where a linear filter samples output from LGN cells. The linear stage

is followed by a normalization stage, where the output of the linear stage divided by the pooled responses of a large number of cortical neurons (Fig. 3.6). Heeger et al. (1996) have proposed that a model with the same structure, i.e., a linear operation followed by normalization, can also account for responses in MT.

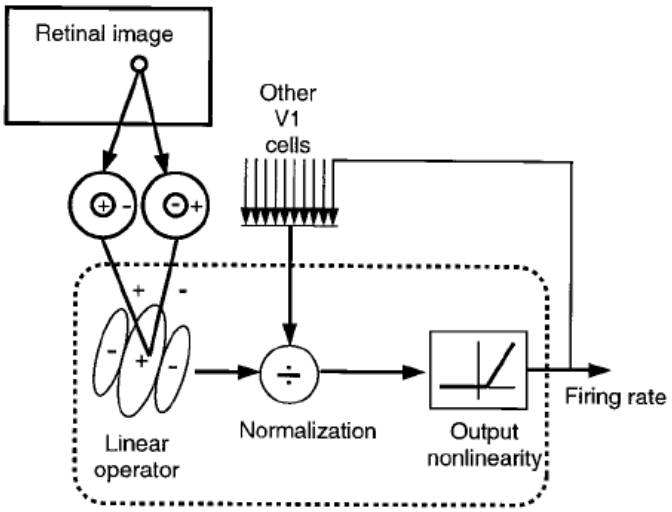


**Fig. 3.6:** Normalization model of simple cells. The retinal image is processed by isotropic on and off LGN cells. LGN input is sampled by a linear operator which is followed by divisive normalization and rectification as output nonlinearity. (From Heeger et al., 1996.)

Nonlinear interactions have also been employed in a model proposed by Neumann and Pessoa (1994). The computational goal is to amplify cell responses to luminance edges, where both subfields are simultaneously activated (soft AND-gate). The model is formalized using shunting equations to describe the disinhibition of subfield responses when both subfields are simultaneously activated. The resulting equation is a combination of an additive linear term and a multiplicative term, both subject to divisive self-normalization. The model is employed in a computational model of brightness perception (Neumann et al., 1998; Pessoa et al., 1995). A detailed investigation of its signal processing properties, in particular its scale-space behavior can be found in Neumann et al. (1999). This interaction scheme of simple cell subfiels is employed in the present work and combined with the new DOI scheme.

### 3.3.2 Recurrent Models

The feedforward models share the common assumption that the receptive field properties of simple cells are primarily determined by the spatial layout of LGN afferents in agreement with empirical data (Ferster et al., 1996; Reid and Alonso, 1995; Stryker et al., 1990, e.g.,). However, the contribution of intrinsic cortical connections to the receptive field properties of simple cells a likewise well founded by empirical studies (Ahmed et al., 1994; Sillito et al., 1980, e.g.,) Models that focus on the contribution of cortical input to the response properties of simple cells have been devised more recently (e.g., Adorján et al., 1999; Ben-Yishai et al., 1995; Douglas et al., 1995; Hansel and Sompolinsky, 1996; Somers et al., 1995). These models in particular focus on the emergence of orientation selectivity and its contrast invariance. We shall review the these class of models in the following.

**Somers and Coworkers.** Somers et al. (1995) have proposed a detailed model of recurrent cortical interactions using spiking neurons. Well shall review this model in detail because of its importance and that large range of empirical findings it covers.

A large body of anatomical and physiological data is incorporated in the model. The model has three layers, a layer of retinal ganglion cells (RGCs), a layer of geniculate cells and a cortical layer representing a full hypercolumn an neighboring columns of layer IV simple cells. Both retinal and geniculate layers were modeled as two arrays of $21 \times 21$ on and off cells. For the cortical layer, 21 orientation columns are modeled, each containing 84 excitatory and 21 inhibitory neurons, resulting in a total of 2205 cortical neurons. The preferred orientation varied monotonically between neighboring columns in steps of 15°, such that 12 columns span the full range of 189°.

Retinal ganglion cell responses are basically modelled by a spatio-temporal convolution with a DoG function and subject to half-wave rectification based on a previously proposed model (Wehmeier et al., 1989; Wörgötter and Koch, 1991). Firing rates of LGN cells are generated from RGC responses using a poisson process. Cortical neurons are modeled as single voltage compartments in separated populations for excitatory and inhibibitory neurons. The cortical neurons receive three kinds of input: excitatory LGN input, excitatory cortical input and inhibitory cortical input. The LGN input was broadly tuned for orientation to provide an orientation bias for each cortical cell. The aspect ration of subfields defined by LGN input varied between unoriented (1:1) and weakly tuned (1:3) with a mean ratio of 1:2.

An example for the connectivity of LGN afferents which provides the input to a cortical cell and the resulting postsynaptic potentials (PSPs) are shown in Fig. 3.7.

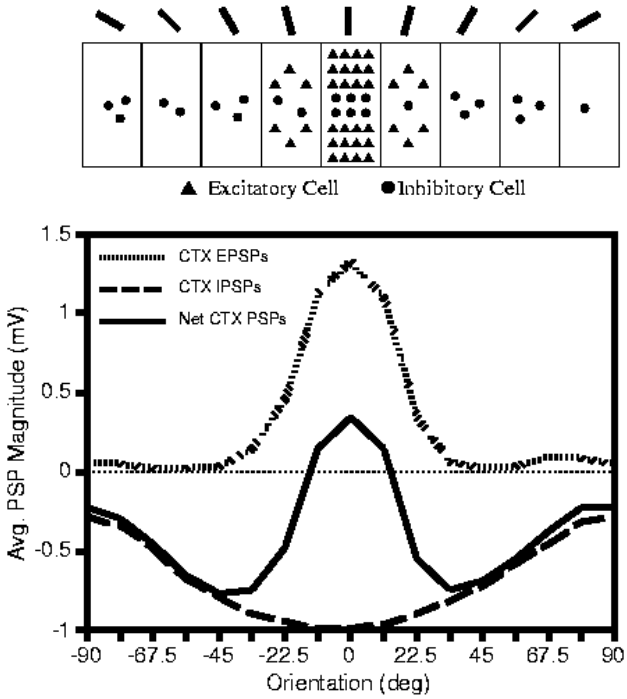


**Fig. 3.7:** Corticocortical connection in the model of Somers et al. (1995). *Top:* Distribution of excitatory and inhibitory connections; inhibitory connections have a broader tuning. *Bottom:* Average PSPs of a cortical cell. Cortical ESPS *(top curve)* evoked by narrowly tuned iso-orientation excitation, and IPSP *(bottom curve, dashed)* evoked by more broadly tuned iso-orientation inhibition results in a net PSP with a DoG like structure *(middle curve, solid).* (From Somers et al., 1995.)

Cortical input were defined by Gaussian distributions in the orientation domain, which defined the maximal connectivity within the cortical column both for excitatory and inhibitory input. Inhibitory input are provided from a more broadly tuned Gaussian (range of ±60°) than excitatory input (range ±15°). The net effect of cortico-cortical input thus resembles a DoG in the orientation domain. The cortical connectivity scheme is outlined in Fig. 3.7 (bottom).

The model by Somers et al. (1995) account for a broad range of findings. The models shows that sharp orientation tuning of physiologically plausible values in the range of 20° HWHH can be obtained from broadly tuned LGN input (about 55° HWHH) by recurrent cortico-cortical interactions. Tuning width were also contrast-invariant (Sclar and Freeman, 1982) and saturated at high contrasts. Somers et al. (1995) analyzed the fundamental requirements for orientation tuning and found the antagonistic center-surround organization of sharply tuned excitatory input and more broadly tuned inhibitory input to be essential. Broadly tuned inhibitory input alone also shar pened orientation tuning, but at the cost of reduced responsiveness.

The model was also used to simulate pharmacological blockage studies, where three different types have been investigated: (i) blockage of on retinal ganglion and LGN cells by injection of the glutamate antagonist 2-amino-4-phosphonobutyrate (ASP) which results in dramatic loss of response rates, buts does not effect orientation selectivity Horton and Sherk (1984); Schiller (1982), (ii) extracellular iontophoresis of GABA by bicuculline injection, which causes a disruption of orientation tuning along with a decrease in firig rates, and (iii) intracellular blockage of inhibition in single cortical neurons, which does not affect the orientation tuning. Impressingly, Somers et al. (1995) model acounts for all of these effects. In particular, the somewhat paradoxical results gained for extracelluar vs. intracellular blockage could be explained by the number of neurons that lost inhibitory input. Intracellular blockage of inhibition in a single does not affect the tuning of all other cells in the network, so that sharply tuned excitatory input of other cells can emerge and cause a sharp tuning in the blocked cell. In contrast, the extracelluar blockage of biculline injection results in a whitdrawl of inhibitory input for the whole population of cells, so that the sharp orientation tuning cannot emerge.

Somers et al. (1995) model focus on the cortico-cortical interactions and demonstrate the potency of recurrent interactions. This does, however, not negate the co-existence of simpler mechanisms. In particular, as pointed out by Somers et al. (1995, p. 5451), “it it quite possible that [the] cortex contains some neurons that receive thalamic inputs sufficient to support sharp tuning”. Furthe, inspite of this compelling competencies, the model cannot accout for the cooling experiments by Ferster et al. (1996), which claim to measure strong tuning in the absence of cortical input.

**A Simplification of the Model of Somers and Coworkers suggested by Carandini and Ringach.** More recently, Carandini and Ringach (1997) have shown that the detailed model of Somers et al. (1995) can be reduced to a single equation, namely a center-surround feedback filter in the orientational domain. Carandini and Ringach (1997) focussed on the generation of orientation tuning and showed that the findings of the full model could be quantitatively reproduced by the simplified model: orientation tuning was amplified and sharpend in response to bars and gratings (i.e., intrinsically 1D stimuli). The simplified model was then probed with intrinsically 2D stimuli containing multiple orientations, such as plaids or crosses. In this case, the model exhibits some peculiar properties: orientations closer than 45° could not be distinguished (attraction of orientation); orientation differences of more than 45° are overestimated (repulsion of orientation); three orientations separated by more than 60° were represented as a single equation. Moreover, spurious reponse at the orthogonal orientation occur when intrinsically 1D stimuli where corrupted with noise. While repulsion and attraction of orientation have observed at least qualitatively in psychophysical expriments (Blakemore et al., 1970; O’Toole and Wenderoth, 1977), a difference in response to gratings vs. plaids has not been found (Gizzi et al., 1990), and noise has a suppressive rather than enhancing/generating effect on response (Burr et al., 1981; Carandini, 1996). Carandini and Ringach (1997) showed that these effect could not be eliminated by varying parameters, such as the width of the inhibitory contribution. Thus, the observed effects are conclude to be intrinsic properties of a recurrent network with center-surround interactions in the orientation domain. It was suggested that center-surround interactions with multiple scales or integration of information from a broader spatial context (“long-range interactions”) might be used to detect multiple orientations.

**Douglas and Coworkers.** Douglas et al. (1995) describe the recurrent excitation and inhibition between neuron populations in analogy to electric circuits. Recurrent excitation and inhibition are described as separated network conductancies which are controlled by the firing frequency of the neuron. The recurrent interaction can generate robust responses from noisy input: Inhibitory interneuron generate an active threshold causing some neurons to cease firing. This results in an increase of the relative amount of excitatory coupling between the surving cells, and their gains remain high or even increase. The increased output of the survivors causes a increase of the inhibitory threshold, thus closing the recurrent loop. The network is shown to generate robust orientation tuning when probed with a single, noisy orientation. The response to intrinsically 2D signals has not been addressed.

**Kinetic Framework of Knight and Coworkers.** Knight and coworkers have developed an alternative framework to describe neural dynamics in terms of interacting sub-populations using a probabilistic formulation (Knight, 2000; Omurtag et al., 2000b). The resulting kinetic equations allow for to simulate the dynamics of a large population of neurons with a speed-up of several orders of magnitude. Omurtag et al. (2000a) applied this method to simulate the dynamics of orientation tuning. First, the model's response to time-independent input was successfully confirmed by replicating the findings of Somers et al. (1995). The model was then utilized to simulate the data-intensive dynamical experiments of Ringach et al. (1997), where good qualitative agreement was obtained.

All models outlined so far rely on *numerical* simulations of the specific algorithm. In an alternative approach, the proposed model equations are solved *analytically*.

**Ben-Yishai and Coworkers.** Ben-Yishai et al. (1995) have proposed a simple model of cortical hypercolumn to analytically investigate the relative contributions of LGN vs. cortical input in generation contrast-invariant orientation tuning. The model uses a deterministic mean field equation which models LGN input to, and cortical interactions within, a single hypercolumn. The angular cortical interactions have a Mexican hat like structure which is described by a cosine function: Cells of orientations not more than $\pm\pi/4$ appart from the target orientation have an excitatory influence, else an inhibitory influence. A similar equations has been derived by Carandini and Ringach (1997) as a simplified version of Somers et al. (1995).

Ben-Yishai et al. (1995) show that contrast-invariant orientation tuning can evolve in a so-called "marginal phase" in the absence of significantly anisotropic LGN input. In this scenario, a continuum of stable states with identical activity profiles but different peak location exist. Minimimal external input is sufficient to select the state which peak at the appropriate location. Further, the "marginal phase" scenario predicts a "virtual rotation" in response to a changing stimulus orientation. In the virtual rotation, a change in stimulus orientation causes dynamical "response sweep", where all cells with preferred orientation between the two presented orientations are successively activated with equal strength. This has not been observed experimentally. The view of the cortex that is implied in the model of Ben-Yishai et al. has been critizised by Ferster and Miller (2000): The recurrent processing defines a fixed set of stereotypical response pattern formalized as the attractors or stable states of a simple equation. This seems to be too limited to account for the variabilty observed in the cortical responses (Ferster and Miller, 2000).

The model also lumbs together excitatory and inhibitory reponses, which results in a smaller activtation range as pointed out by Li and Dayan (1999). Li and Dayan showed that a asymmetric network with distinct excitatory and inhibitory cell populations can achieve higher selective amplification than symmetric models. Extentions with distinct excitatory and inhibitory populations have been proposed by Wennekers (2001).

### 3.3.3 Summary of the Review of Simple Cell Models

The large body of empirical data that have been cumulated with respect to simple cell anatomy and physiology is paralled by a plethora of models that have been proposed over the past time period of almost fourty years. In the early models, simple cells are described as linear filters. Derivatives of Gaussians or Gabor filters are frequently applied and show little difference in their qualitative behaviour. Though basic simple cell properties can be expressed by linear operations, the restriction to linear mechanism is too severe. Nonlinear mechanims of amplification and normalization have been successfully applied to account for a broader variety of simple cell properties. More recently, the contribution of recurrent cortical processing has been incorporated into simple cell models. Center-surround processing in the orientation domain formalized by Mexican hat shaped functions (such as, e.g., a DoG or cosine function) is the central computational element in these models.

Each class of models describe essential computational principles found in simple cells. A model aimed for maximum comprehensiveness should consequently build on (at least) these principles. On the other hand, focussing on or stressing a single mechanism is useful in order to highlight the computational competencies of the particular principle.

## 3.4 Contrast Detection in Computer Vision

Contrast and edge detection can be undoubtly regarded as one of *the* topics in computer vision from the early days until present times. The notion of Marr (1982, p. 5), that "for the subject of vision, there *is* no single equation or view that explains everything" is in particular true for the subject of edge detection. A myriad of papers, book chapters and monographs has been published on this topic.[1] Consequently, the review given in this section can by no means be exhaustive. Nevertheless, we try to cover important steps and approaches. The review focuses on edge-detection algorithms based on filtering. This is for two reasons: Filtering-based algorithms comprise a widely used class of edge detection algorithms, and they can be linked to biologically motivated models as described in the previous section. In particular, the various filters proposed for edge detection can be compared to, and are sometimes motivated by, the receptive field structure of cells in the early visual pathway.

Edge detection is a problem of fundamental importance in image analysis. Introductions to edge detection can be found in virtually every textbook on computer vision, e.g., Levine (1985), Haralick and Shapiro (1992), Gonzalez and Woods (1993), Nalwa (1993) or Jähne (1997).

### 3.4.1 Introduction

Edges are associated with changes in the luminance distribution of the input image. These changes are caused by a change of at least one physical scene parameter, namely reflectance, illumination, surface normal or depth (Lee, 1990; Marr, 1982). The detection of these changes or edge features allow for a representation of image primitives in a flexible way such as the raw primal sketch in the sense of Marr (1982). Since edges often characterize object boundaries, edge features are useful for a multitude of higher level functions such as figure-round segregation, segmentation, or object registration and identification.

[1] For example, not less than 21 different edge-detection algorithms have been proposed in the relatively short period of three years (1993 through 1995) in three main computer vision journals, namely IEEE Transactions on Pattern Analysis and Machine Intelligence, Computer Vision and Image Understanding (named "Computer Vision, Graphics, and Image Processing: Image Understanding" before 1995) and Pattern Recognition, as reviewed by Heath et al. (1997).

Edges are the curves separating lighter and darker regions of an image, defining points of high change rate in the gray value of the image. The detection of edges thus requires a mechanism which selectively responds to such changes, while suppressing regions of constant gray values. The *differentiation* of the image is an operation suitable for this detection task, and builds the nucleus of most edge detection algorithms. The basic idea is sketched in Fig. 3.8: The derivation of a noisy step edge leads to an extremum in the first-order derivative and a zero-crossing in the second-order derivative (Fig. 3.8, second and third row). Based on the order of derivatives that are employed for an edge detection methods, one can distinguish between gradient-based methods (Sec. 3.4.2) and Laplacian-based methods Sec. 3.4.4.

In the following, we shall examine the basic methods as well as more sophisticated edge detection algorithms in more detail.

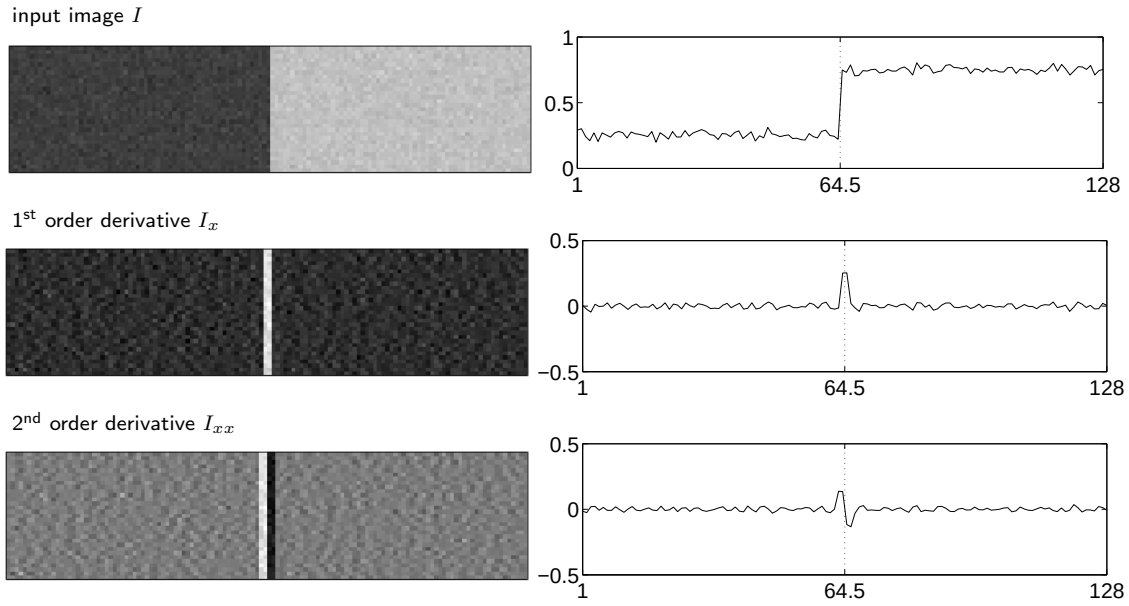


**Fig. 3.8:** Noisy step edge corrupted with 5% Gaussian noise, and derivatives. *Left column:* Gray value images. *Right colum:* Corresponding horizontal cross-sections. The *dashed vertical line* in the cross-sections indicate the position of the edge, i.e., the dark-light transition in the input image. *Top to bottom:* Input image, first-order derivative in $x$-direction, second-order derivative in $x$-direction. The edge leads to a maximum in the first-order derivative and a zero-crossing in the second-order derivative. The derivatives are approximated by central differences. The size of the images is $128 \times 32$ pixels.

### 3.4.2 Gradient-Based Edge Detection

Edge locations are characterized by strong, abrupt changes in the luminance distribution of the image. As stated above, the detection of these changes can be formalized by differentiation of the image, where an edge leads to a local maxima in the first derivative. Horizontal and vertical edges in an image $I$ can be detected by considering the directional derivatives in $x$- and $y$-direction, $I_x := \partial_x I$ and $I_y := \partial_y I$. For twodimensional images, edges are not limited to horizontal or vertical orientation, but can occur with arbitrary slope. Edges of any orientation can be characterized by the gradient

$$\nabla I = \begin{pmatrix} I_x \\ I_y \end{pmatrix} \quad .$$

The strength and orientation of the edge can be calculated from the gradient magnitude and orientation, respectively. The gradient magnitude is invariant to changes of the coordinate system

and is defined as the root sum of squares of the derivatives:

$$|\nabla I| = \left(I_x^2 + I_y^2\right)^{(1/2)} \quad .$$

The gradient orientation is defined as the arctangent of the ratio of the derivatives:

$$\angle \nabla I = \operatorname{atan}\left(\frac{I_y}{I_x}\right) \quad .$$

Edge detection based on computing the gradient is a widely used method. The individual approaches differ in the ways to compute the gradient image and the kind of postprocessing that is applied to identify edge points in the gradient image.

### Basic Difference Operators

In practice, images are given as a discrete set of pixels. The simplest approach to compute the gradient vector is to approximate the differentiation by a finite differences scheme. Three difference schemes can be used:

$$\begin{aligned}
&\text{backward:} && {}^{-}I_x \approx \frac{I(x,y) - I(x-h,y)}{h} \\
&\text{forward:} && {}^{+}I_x \approx \frac{I(x+h,y) - I(x,y)}{h} \\
&\text{symmetric:} && I_x \approx \frac{I(x+h,y) - I(x-h,y)}{2h} \quad .
\end{aligned}$$

For a grid distance of $h = 1$, the corresponding filter masks are given by

$$\begin{aligned}
&\text{backward:} && {}^{-}D_x = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} \\
&\text{forward:} && {}^{+}D_x = \begin{bmatrix} 0 & -1 & 1 \end{bmatrix} \\
&\text{symmetric:} && D_x = \tfrac{1}{2}\begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad .
\end{aligned}$$

Symmetric differences are also called central differences. The filter masks for the $y$-direction are given by transposing the masks for the $x$-direction.

The forward and backward filters are not suitable for computing the gradient, since the convolution with these filters results in an displacement of half the grid distance. Further, these simple gradient filters have rather poor performance measuring the magnitude and the direction of the gradient. The deviation of the true direction is zero only for the main axis and for the diagonals ($0°, 45°, 90° \ldots$) and approximately $\pm 10°$ for the directions in between ($22.5°$, $67.5°$, $\ldots$), as pointed out by Jähne (1997).

### Roberts Cross Detector

The smallest possible filter that does not result in an displacement of the measured edge position is given by the Roberts cross detector (Roberts, 1965):

$$R_{45°} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad R_{135°} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad .$$

The Roberts cross detection measures a gradient rotated by $45°$. The performance with respect errors in gradient magnitude and orientation is only slightly better than for the symmetric differences (Jähne, 1997).

**Prewitt and Sobel Operator**

Both the Prewitt and Sobel operator combine differentiation using symmetric differences with a smoothing orthogonal to the direction of the derivation (Prewitt, 1970; Sobel, 1970). The Prewitt operator uses a box filter $X = \frac{1}{3}\begin{bmatrix}1 & 1 & 1\end{bmatrix}$ for smoothing:

$$P_x = X^{\mathrm{T}} * D_x = \frac{1}{6}\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix} \quad \text{and} \quad P_y = P_x^{\mathrm{T}} \ .$$

The Sobel filter is build using a binomial mask $B = \frac{1}{4}\begin{bmatrix}1 & 2 & 1\end{bmatrix}$. Binomial masks are discrete approximation of Gaussians. The Sobel filter is defined as

$$S_x = B^{\mathrm{T}} * D_x = \frac{1}{8}\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix} \quad \text{and} \quad S_y = S_x^{\mathrm{T}} \ .$$

The errors of the Sobel filter in measuring the magnitude and gradient are only half the size of the simple difference operator. i.e., the angular error has a maximal value of 5°. Am optimized version of the Sobel operator with a much reduced angular error has been proposed by Scharr et al. (1997), see also Jähne et al. (1999):

$$F_x = \frac{1}{32}\begin{bmatrix}-3 & 0 & 3\\ -10 & 0 & 10\\ -3 & 0 & 3\end{bmatrix} \quad \text{and} \quad F_y = F_x^{\mathrm{T}} \ .$$

### 3.4.3 Derivatives of Gaussians

The small derivative operators described above are inherently sensitive to noise. The Sobel and the Prewitt operator combine differentiation with smoothing, resulting in a reduced sensitivity to noise. More generally, one can compute the derivatives on a smoothed image, or alternatively, use a smoothed derivative operator to be applied to the image (Torre and Poggio, 1980). The most general class of such smoothed derivative operators are comprised by derivatives of Gaussians.

**Motivation**

The importance of the Gaussian and its derivatives can be reasoned on a number of grounds. Gaussian smoothing is equivalent with linear scale-space (e.g., Lindeberg and ter Haar Romeny, 1994a; Weickert, 1998), as reviewed in detail in Sec. B.3.1. In particular, the Gaussians fulfills the semi-group property, i.e., the convolution of two Gaussians with different standard deviations $\sigma_1$ and $\sigma_2$ is again a Gaussian:

$$G_{\sigma_1} * G_{\sigma_2} = G_{\sqrt{\sigma_1^2+\sigma_2^2}} \ . \tag{3.1}$$

The Fourier transform of a Gaussian is also a Gaussian, which results in equally good localization properties in both the space and the frequency domain. More precisely, the Gaussian in the only *real* kernel that gives equality in the uncertainty equation, which states a lower bound for the product of the second-moments $\Delta x$ and $\Delta\omega$ in the spatial and frequency domain:

$$\Delta x \Delta\omega \geq \tfrac{1}{2} \ .$$

In other words, the Gaussian exhibits the best possible joint localization in both domains (Lindeberg and ter Haar Romeny, 1994a). As pointed out by Michaelis (1997), complex Gabor functions have the same optimal joint localization. This is because complex Gabor function results

from shifting the Gaussian in frequency space, an operation which does not affect the second-moments $\Delta x$ and $\Delta\omega$.

The Gaussian is also the density function of the normal distribution. The importance of the normal distribution in statistics is reflected by the central limit theorem, which states that the distribution of a large sum of stochastic variables (fulfilling rather general requirements) converges to a normal distribution. When transferred to image processing, the central limit theorem states that the repeated convolution of any *positive-valued* kernel with itself converges to a Gaussian (Forsyth and Ponce, 2000).

The Gaussian is infinitely differentiable, allowing the construction of differentiation operators of any desired order. The $n$-th order derivation of an image $I$ can then be computed by convolution with the $n$-th order Gaussian derivative operator. The convolution with the Gaussian is an integral transform with strong regularization property (Lindeberg and ter Haar Romeny, 1994a): Even if the image is not differentiable of any order or not continuous, the result of applying the Gaussian derivative operator is always well-defined. The only requirement is that the image is bound by some polynomial.

Using Gaussian derivative operators up to some order $n$, a set of spatial derivatives of the image can be obtained which allows to characterized the image intensity function in terms of the Taylor expansion up to order $n$. Such a *multi-scale n-jet signal representation* has been introduced by Koenderink and van Doorn (1987, 1992). The use of this $n$-jet representation is not limited to edge detection. In fact, it is a rather general framework which can be applied to modeling various operations in early vision, such as blob or junction detection (Lindeberg and ter Haar Romeny, 1994b). Bridging the gap between computational vision and biological models, the set of Gaussian derivatives have been used to describe receptive field properties in the early visual stages (Koenderink and van Doorn, 1990) and have been suggested as basis functions used by the visual system (Young, 1985).

Finally, derivatives of Gaussians can be motivated by their optimal feature detection properties (Canny, 1986; Marr and Hildreth, 1980; Perona and Malik, 1990a).

To sum up, the Gaussian and its derivatives are optimal filters in early vision regarding the most general criteria, i.e., for the low-level processing of in an uncommited visual system (Lindeberg and ter Haar Romeny, 1994a; Michaelis, 1997).

### Equations

The 2D-Gaussian $G(x, y)$ has the technically useful property of being separable, i.e., the 2D-Gaussian can be expressed as the product of two 1D-Gaussians

$$G(x, y) = g(x)g(y) \ ,$$

where the 1D-Gaussian with standard deviation $\sigma$ is given by

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \ .$$

The separability of the 2D-Gaussian allows for an computationally efficient implementation of filtering an image by Gaussian convolution: Instead of filtering with the 2D kernel, two convolution with the corresponding 1D kernels are applied. For an image of size $N$ and a 2D kernel of size $M$, the computational complexity is reduced from $O(MN)$ to $O(\sqrt{M}N)$.

Computing the derivatives of Gaussians can also benefit from separability. For example, the second order derivative in $x$-direction of the 2D-Gaussian can be expressed in terms of 1D-Gaussians:

$$G_{xx}(x, y) = g_{xx}(x)\, g(y) \ .$$

The general equation reads

$$G_{x^n y^m}(x, y) = g_{x^n}(x)\, g_{y^m}(y) \ .$$

Consequently, the derivatives of the 1D Gaussians are sufficient to compute the corresponding derivatives of the 2D Gaussian. The 1D derivatives of first and second order are given by

$$g_x(x) = -\frac{x}{\sigma^2}\, g(x)$$
$$g_{xx}(x) = \frac{1}{\sigma^2}\Big(\frac{x^2}{\sigma^2} - 1\Big)\, g(x) \ .$$

Derivatives of higher order and a general expression in terms of Hermite polynomials can be found in Sec. A.1.2.

#### Discretization

In practice, discrete approximations of the continuous functions have to be used. Sampling the continuous function is often sufficient, but leads to wrong results in some cases. For example, when the standard deviation of the Gaussian is small compared to the distance between grid points (usually one), the value of the central pixels are too high. The ad hoc solution to compensate for this effect is to normalize the filter coefficients. However, the discrete kernels obtained this way violate the semi-group property of scale-space (Eq. 3.1), as pointed out by Lindeberg (1990). If the standard deviation of the Gaussian is of the form $\sigma = \sqrt{p}/2$, $p \in \mathbb{N}$, binomial kernels of order $p$ can be used. For arbitrary values of the standard deviation, discrete approximation of the Gaussian and its derivatives in terms of modified Bessel functions have been suggested by Lindeberg (1990, 1994). This discrete Gaussian kernel can also be regarded as the limit case resulting from successive application of generalized binomial kernels (Lindeberg and ter Haar Romeny, 1994a):

$$g_\sigma \approx \lim_{n\to\infty} \begin{bmatrix} b & 1-2b & b \end{bmatrix}^n, \quad \text{with } b = \frac{\sigma^2}{2n} \ .$$

The operator notation $B^n$ denotes the $n$-times application of the operator $B$, in this case the $n$-times convolution of the generalized binomial kernel $B = \begin{bmatrix} b & 1-2b & b \end{bmatrix}$ with itself:

$$B^n := \underbrace{B * B * \ldots * B}_{n\text{-times}} \ .$$

A number of other discretization schemes have been proposed in the literature (e.g., Davis, 1987; Hashimoto and Sklansky, 1987; Wilson and Bhalerao, 1992).

### 3.4.4 Laplacian-Based Edge Detection

An edge is characterized by a local maximum of the first derivative or, alternatively, by a zero-crossing of the second order derivative. Instead of second-order directional derivatives, the Laplacian $\nabla^2$ is usually employed, since it is the lowest-order *isotropic* differential operator. The class of Laplacian-based edge detectors compute edge locations as follows: First, a discrete approximation of the Laplacian is computed, followed by marking the zero-crossings in the Laplacian-filtered image. The various methods essentially differ only in the way the Laplacian is approximated.

#### Basic Discretization Scheme

The Laplacian of a 2D image $I$ is defined as

$$\Delta I \equiv \nabla^2 I = I_{xx} + I_{yy} \ .$$

A straight forward approximation of the second order derivative combines two discrete first-order derivative filters, namely forward and backward differences:

$$D_{xx} = {}^{-}D_x * {}^{+}D_x = \begin{bmatrix}-1 & 1 & 0\end{bmatrix} * \begin{bmatrix}0 & -1 & 1\end{bmatrix} = \begin{bmatrix}1 & -2 & 1\end{bmatrix} \ .$$

The Laplacian filter $L$ is then given by the mask

$$L = D_{xx} + D_{yy} = \begin{bmatrix}1 & -2 & 1\end{bmatrix} + \begin{bmatrix}1\\-2\\1\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\1 & -4 & 1\\0 & 1 & 0\end{bmatrix} \ .$$

Other discrete approximations of the Laplacian can be found in Sec. A.2.

### Laplacian of a Gaussian

The approximations of the Laplacian based on finite differences are extremely sensitive to noise. In general, the sensitivity to noise increases with the order of derivation. To make the computation of the Laplacian more robust, the Laplacian can be applied to a smoothed image. As reviewed above (Sec. 3.4.3), the appropriate function for smoothing is the Gaussian. Combining both operations in a single filter leads to the Laplacian of Gaussian (LoG) filter.

The Laplacian of a 2D Gaussian is given by

$$\begin{aligned}\Delta G(x,y) &\equiv \nabla^2 G(x,y)\\ &= G_{xx}(x,y) + G_{yy}(x,y) = g_{xx}(x)g(y) + g_{yy}(y)g(x)\\ &= G(x,y)\Big(\frac{x^2+y^2}{\sigma^2} - 2\Big)\frac{1}{\sigma^2} \ .\end{aligned}$$

The LoG operator can be approximated by taking the difference of two Gaussians with different standard deviations, which leads to the so-called difference of Gaussians filter (DoG). The smallest difference of both operators is given if the two standard deviations of the center and surround Gaussians of the DoG have the ratio $\sigma_{\text{center}} : \sigma_{\text{surround}} = 1 : 1.6$ (Marr and Hildreth, 1980).

The method of detecting edges as zero-crossings of a LoG-filtered image was originally proposed by Marr and Hildreth (1980). The idea was to use an isotropic derivative operator which can be tuned to different scales. The need for multiple scales is based on the notion that a single operator of fixed size cannot detect all edges from the finest detail to the coarsest structures. The use of the second derivative instead of, e.g., the first order derivative was to avoid a subsequent thresholding operation. Thresholding is commonly applied to select sufficiently large maxima resulting from the first derivative operation at an edge location. To obtain good results, an "appropriate" value of the threshold has to be found, which strongly depends on the image and the amount of structure that needs to be detected. The zero-crossings of the second-order derivative seemed advantageous since they can be determined without the need of any thresholding operation. Based on these considerations, the isotropic Laplacian of Gaussian has been identified as the most appropriate operator. The application of LoG or DoG filters was also motivated based on biological grounds, since the receptive fields of retinal ganglion cells are commonly modeled by DoG filters.

The Marr-Hildreth edge detector is appealing because of its simple and well-motivated assumptions. In practice, however, the method is flawed by several intrinsic disadvantages. As already mentioned, the use of the second derivative makes the operator more sensitive to noise than a first-order derivative operator. Smoothing can diminish this effects, but a smoothed second-order derivative is clearly still more sensitive to noise than a equally smoothed first-order derivative operator. Using large smoothing kernels also leads to a dislocation of edges, a phenomenon well-known in the theory of scale-space (e.g., Witkin, 1983). In this case, edges found at larger scales have to be tracked down to smaller scales where they can be localized with higher precision. Further,

the second derivative tiles the image plane in region of positive and negative sign, separated by edge locations, the zero-crossings. Consequently, the edges always form closed contours, which is not appropriate to describe a large class of edge configurations. In particular, the zero-crossings of the LoG have incorrect responses at junctions (Berzins, 1984). For example, the zero-crossings cannot correctly divide three regions defined by three joining edges at a so-called 3-junction (e.g., Lindeberg, 1994). Finally, the presumed advantage of zero-crossings as a threshold-free methods turns out to become disadvantageous when a measure of the edge strength is needed to distinguish between strong and week responses.

### 3.4.5 Beyond Basic Edge Detection Methods

The edge detection method reviewed so far rely on a linear filtering of the input image followed by simple postprocessing methods such as thresholding or detection of zero-crossings. In the following, we shall review slightly more complicated approaches distinguished by more sophisticated postprocessing methods (such as the Canny edge detector) or the explicit use of nonlinear operations (such as Zucker's and coworkers logical-linear framework).

#### Canny Edge Detector

The algorithm devised by Canny (1986) was a major contribution and can still be regarded as one of the standard methods in edge detection. The approach is characterized by the use of mathematical optimization methods to derive an optimal filter kernel and the number of rather complex postprocessing operations following the application of the filter kernel.

Canny based the design of the edge detector on three criteria: good detection, precise localization, and a unique, single response to an edge. Good detection is given if the operator has a low probability of both failing to respond to an edge (false negative responses), and a low probability to respond in the absence of an edge (false positive responses), equivalent to a high signal-to-noise ratio. These criteria are mathematically formalized and are used to derive an optimal operator for 1D step edges by numerical optimization. The resulting filter can be efficiently approximated by a first order Gaussian derivative.

In 2D, first order derivatives of isotropic Gaussian are used to compute the gradient of the image. If the edges are locally straight, a set of highly directional operators results in a better performance than a simple circular operator. The directional masks are computed by summing the output of a number of circular masks along the preferred axis of orientation. A set of directional masks together with the profile of the masks taken both parallel and perpendicular to its preferred direction is shown in Fig. 3.9. Canny has shown that both the signal-to-noise ratio and the localization improve with the square $\sqrt{l}$ of the scaling $l$ parallel to the edge. Canny further points out that the use of directional operators has to be restricted to cases where they are applicable, i.e, where the image structure fits to a linear edge model. Testing for applicability is done by a simple goodness-of-fit measure, where the squared output of the directional mask is compared to the variance of its individual masks. The use of the directional operator is limited to those cases where the variance is smaller than some fraction of the squared output.

A thinning process of nonmaximum suppression is then used to derive contours of one pixel width from the gradient response. In nonmaximum suppression, the strength of the candidate edge pixel is compared to its two neighboring points along the gradient direction, and suppressed, i.e. set to zero, if it is not larger than both neighboring points. Neighboring points *orthogonal* to the gradient, i.e., along the edge, are not considered, since the continuity of the edge has to be preserved. On a discrete grid the values of neighboring points in the gradient direction have to be interpolated, e.g., by bilinear interpolation. Nonmaximum suppression corresponds to detect the zero-crossings in the 2D directional derivative along the gradient direction.

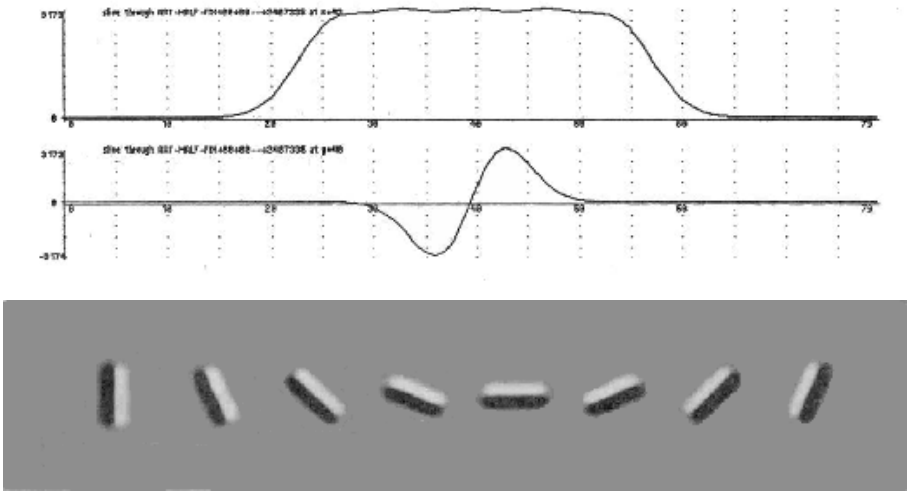

**Fig. 3.9:** Canny's directional step edge masks. *Top two graphs:* Cross sections of a directional mask parallel and perpendicular to the edge direction. *Bottom:* 2D weighting functions of eight directional masks. (From Canny, 1986.)

Points that survive the nonmaximum suppression are then subject to thresholding using hysteresis. In thresholding with hysteresis, two thresholds $\theta_{\text{low}}$ and $\theta_{\text{high}}$ are used. Values below $\theta_{\text{low}}$ are discarded, those above $\theta_{\text{high}}$ are preserved. The interesting points are those of intermediate value between $\theta_{\text{low}}$ and $\theta_{\text{high}}$. These points are preserved if they form a connected segment which is linked to a point above $\theta_{\text{high}}$. The use of hysteresis reduces the probability of streaking, i.e., the break-up of edges because of fluctuations in the operator response above and below a single threshold. The values of $\theta_{\text{low}}$ and $\theta_{\text{high}}$ can be determined by a global histogram estimate.

Canny also stresses the use of different operator sizes to cope with different signal-to-noise ratios. A fine-to-coarse heuristic using a feature synthesis approach is suggested to integrate edge information at several scales.

To sum up, the Canny edge detection methods combines a number of techniques to achieve considerably better results than obtained with simpler filtering approaches. However, the design of the filter is 1D-based, and the case of more than one edge at a single location is not considered. Consequently, failures such as small gaps occur where two ore more edges join.

From the viewpoint of simple cell models, the directional masks introduced for the detection of elongated 2D structures are of particular interest, since they resemble the RFs of simple cells. Moreover, Canny (1986, p. 695) states that "an efficient way of forming long directional masks is to sample the output of nonelongated masks with the same direction", which is basically the Hubel and Wiesel model for the origin of directional selectivity of simple cells. The importance of a goodness-of-fit measurement is also interesting in this perspective, since it motivates the need of adaptive RF properties from a computational point of view.

### Iverson and Zucker's Logical/Linear Operator

The edge detection methods reviewed so far, including the Canny operator, rely on linear convolutions followed by nonlinear operations such as thresholding or detection of zero-crossings. Further, these approaches assume either explicitly or implicitly that only a single structural change occurs within the support of the operator, thus ignoring the occurrence of junctions, where two or more edges or lines coincide. Iverson and Zucker (1995) point out that the assumption of traditional linear operators leads to a number of false-positive responses. To avoid such false-positive responses,

three criteria are distinguished. First, the operator has to exhibit a predictable behavior at junctions and should properly resolve and separate different features within its support. Second, the operator needs to preserve discontinuities and line ends. Third, the operator should not confuse lines and edges.

Iverson and Zucker (1995) devised a so-called *logical/linear* framework for the proper design of such operators. The basic idea is to decompose a linear model into a number of components, i.e., assertions for the existence of a tested feature. A boolean (or *logical*) combination of these assertions ensures that the operator responds only when *all* of these assertions are met. In this case, the response is a *linear* combination of the individual components, hence the name logical/linear operator.

Edges and lines are considered as *images curves*, which are described by their tangential and normal structure. The tangential or curvilinear structure is the same for lines and edges and is given by a local continuity condition. The normal structure is used to differentiate lines and edges. Logical/linear operators are designed to test both tangential and normal conditions. The normal conditions are expressed in terms of the differential structure across the curve. Extending traditional approaches, Iverson and Zucker make use of higher order derivatives. For example, an edge is characterized as a maximum of the first derivative and minimum of the third derivative. These conditions are formulated in a number of preconditions whose logical/linear conjunction comprise the edge operator. Derivatives of Gaussians are used to compute the various derivatives along the normal direction. The resulting operator is more selective than an traditional operator based on the linear combination of only a subset of the conditions used. The operators for the normal conditions are finally combined with the tangential operator, yielding logical/linear image curve operators. Three different operators for edges and lines (which are further subdivided into positive and negative contrast lines) are given.

The approach by Iverson and Zucker (1995) differs by various aspects from standard methods in edge detection. First, it stresses the need for *nonlinear* mechanism expressed in the logical combination of preconditions. Second, the need to consider the intrinsic 1D structures such as lines and edges as 2D image curves are explicitely addressed by verifying the local continuity along the tangential direction. Third, a rich representation of multiple features (edges, and positive and negative contrast lines) in multiple orientations at each spatial location is computed. Such a so-called "discrete image trace" is suggested to be refined by relaxation labeling techniques to extract a higher level description (Iverson, 1994). Finally, higher order derivatives are evaluated. To sum up, the approach makes a number of fruitful suggestions to overcome the shortcomings of traditional edge detection algorithms.

### 3.4.6 Summary of Contrast Detection in Computer Vision

We have reviewed basic and more advanced filter based approaches to contrast detection. All algorithms are based on the detection of luminance changes by computing an approximation of the first or higher order derivatives of the image. Gaussian smoothing is used to regularize the computation of the derivatives and to reduce the effects of noise. Edges are not a pure local event, but have a certain tangential extension along their axes. The integration of activity along this tangential or curvilinear direction results in a more robust response. Considering this tangential structure is expressed in the rather basic orthogonal smoothing performed by the Sobel operator, in the elongated directional masks proposed by Canny, and finally in the tangential logical/linear operator of Iverson and Zucker. Nonlinear interactions are important to suppress false-positive responses. The nonlinear operations applied range from simple thresholding and thresholding with hysteresis, via nonmaximum suppression to the logical/linear combination of elementary image features. From these considerations, a list of useful operations for contrast detection can be gathered, which—without claiming to be comprehensive—comprises the following

items: (i) approximation of the derivative, (ii) Gaussian smoothing, (iii) sampling of tangential support, and (iv) nonlinear suppression of responses.

In a broader perspective, the seemingly simple task of contrast detection in computer vision turns out to demand an unexpected high amount of computational effort. The almost effortless detection of lines and edges by the human visual system disguises the complexity of the underlying operations, all of which happen unconsciously. The fact that the operations performed by humans are not accessible to introspection and thus cannot be made explicit complicates both a realistic estimation of the computational power needed and the specification of proper methods. Insights into operations used by biological vision systems as revealed by empirical studies are thus an important counterpart to theoretically and computationally motivated approaches.

## 3.5 The Model

In this section a formal description of the model is given. The model consists of a hierarchical organization of two main processing stages, namely the extraction of contrast signals, followed by a simple cell circuit. In all equations, capital Roman letters denote the 2D maps of activity distributions at the various stages; Greek letters denote positively-valued model parameters.

### Contrast Signals

Contrast signals are generated from the initial luminance distribution of the input stimulus. Contrast signals occur at luminance differences and are intended to model integrated responses of retinal ganglion cells or LGN cells.

To model contrast signals, the initial luminance distribution is first processed by a center-surround mechanism similar to retinal ganglion cells. Center and surround responses are modeled separately by filtering the initial luminance distribution $I$ with isotropic Gaussians of different standard deviations $\sigma_\mathrm{c} = 1$ and $\sigma_\mathrm{s} = 3$:

$$\begin{aligned} I_\mathrm{c} &= I * G_{\sigma_\mathrm{c}} \\ I_\mathrm{s} &= I * G_{\sigma_\mathrm{s}} \quad , \end{aligned}$$

where $*$ is the spatial convolution operator. The Gaussians are sampled within a $3\sigma$ interval, resulting in a filter mask of size $7 \times 7$ for the center and $19 \times 19$ for the surround ($3\sigma \times 2 + 1$).

Center and surround responses provide the input to a shunting mechanism (Furman, 1965; Grossberg, 1970; Hodgkin, 1964). Shunting mechanisms yield a bounded activity and cause a compression of high amplitude activity following the Weber-Fechner law (Fechner, 1889; Weber, 1846):

$$\partial_t X = -\alpha X + (\beta - X)\mathrm{net}^+ - (\gamma + X)\mathrm{net}^- \quad .$$

In the above equation, $\alpha = 0.5$ is the activity decay and $\gamma = 0.1$, $\beta = 1$ denote the upper and lower bound of the activity, respectively.

The shunting equation is assumed to quickly reach steady state and is solved at equilibrium. The equilibrium solution is given for $\partial_t X = 0$ and can be written as a function $X$ of two input variables, an excitatory contribution $\mathrm{net}^+$ and an inhibitory contribution $\mathrm{net}^-$:

$$X(\mathrm{net}^+, \mathrm{net}^-) = \frac{\beta\mathrm{net}^+ - \gamma\mathrm{net}^-}{\alpha + \mathrm{net}^+ + \mathrm{net}^-} \quad . \tag{3.2}$$

The shunting interaction is modeled for two domains, namely on and off contrast signals. For the on domain, the excitatory input $\mathrm{net}^+$ is provided by the center filtered input $I_\mathrm{c}$, whereas the inhibitory input $\mathrm{net}^-$ is provided by the surround filtered input $I_\mathrm{c}$. The reverse holds true for the

off domain. Using the equilibrium function $X(\text{net}^+, \text{net}^-)$, on and off contrast signals $X_{\text{on}}$ and $X_{\text{off}}$ are thus modeled as

$$\begin{aligned} X_{\text{on}} &= X(I_{\text{c}}, I_{\text{s}}) \\ X_{\text{off}} &= X(I_{\text{s}}, I_{\text{c}}) \ . \end{aligned} \tag{3.3}$$

These shunting contrast signals exhibit nonzero response to homogeneous regions. As detailed by Neumann (1996), shunting contrast signals can be segregated into a pure contrast signal without any activity to homogeneous regions (i.e., a signal with "zero DC level") and a luminance signal given by a low-pass filtered copy of the input activity. Following Neumann (1996), zero DC level contrast signals $K_{\text{on}}$ and $K_{\text{off}}$ result from mutual inhibition of opposite domains:

$$\begin{aligned} K_{\text{on}} &= [X_{\text{on}} - X_{\text{off}}]^+ \\ K_{\text{off}} &= [X_{\text{off}} - X_{\text{on}}]^+ \ , \end{aligned} \tag{3.4}$$

where $[x]^+ := \max\{x, 0\}$ denotes half-wave-rectification.

### 3.5.1 Simple Cells

The next processing stage deals with simple cells, which are modeled for eight discrete orientations $\theta = 0°, 22.5°, 45°, \ldots, 157.5°$ and for two opposite contrast polarities, namely light-dark and dark-light.

**Simple cell subfields with dominating opponent inhibition.**

A simple cell has two adjacent subfields, an on subfield sensitive to light increments and an off subfield sensitive to light decrements. Simple cell subfields are defined by elongated, oriented weighting functions $G_\theta$. The weighting function $G_\theta$ is modeled with five isotropic Gaussians with standard deviation $\sigma = 2$, which are properly aligned along the preferred axis of orientation $\theta$ and spaced within a distance of two standard deviations. A sample weighting function for $\theta = 0°$ is depicted in Fig. 3.10. The modeling of the weighting function for the simple cell subfields results in a plateau-like RF which is $29/19 \approx 1.5$ times larger than the RF of the on and off cells. Generally, for $N$ Gaussians with a standard deviation $\sigma$, the length of the filter mask is given by $(N-1)2\sigma + 2 \cdot 3\sigma + 1$.

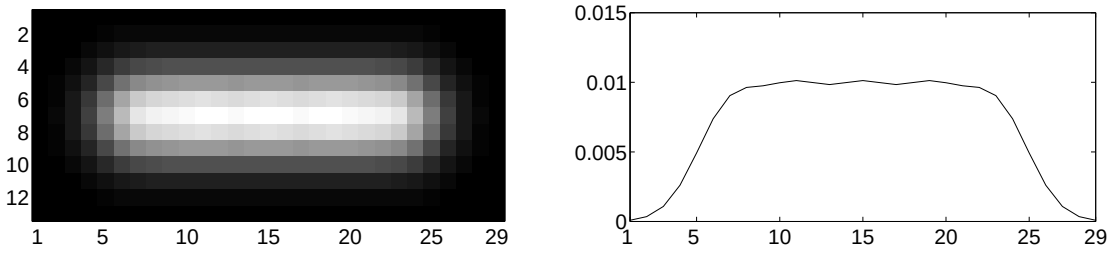


**Fig. 3.10:** *Left:* Filter mask for a simple cell subfield of orientation 0°. *Right:* The corresponding horizontal cross-section taken at the center of the mask.

Before integration, contrast activity of different polarity competes at each spatial location. Input activation for both on and off subfields $R_{\text{on}}$ and $R_{\text{off}}$ with a preferred orientation $\theta$ is computed by convolution of the weighted difference of unoriented LGN responses $K_{\text{on}}$ and $K_{\text{off}}$ with the subfield mask $G_\theta$ of the same orientation preference:

$$\begin{aligned} R_{\text{on},\theta} &= \big[(K_{\text{on}} - \xi K_{\text{off}}) * G_\theta\big]^+ \ , \\ R_{\text{off},\theta} &= \big[(K_{\text{off}} - \xi K_{\text{on}}) * G_\theta\big]^+ \ . \end{aligned} \tag{3.5}$$

A sketch of the interaction scheme which defines the subfields is given in Fig. 3.11. The case of equally weighted on and off inputs occurs for $\xi = 1$. The newly proposed scheme of dominating opponent inhibition (DOI) introduces a weighting parameter $\xi > 1$ which scales up the opponent contribution. This introduces a "one against many" situation, where, e.g., an on subfield only receives input if the contribution of the on channel $K_{\text{on}}$ is $\xi$ times larger than the contribution of the opponent off channel $K_{\text{off}}$. The subfield interaction with DOI is a special case of the push-pull models as reviewed in Sec. 3.3.1.

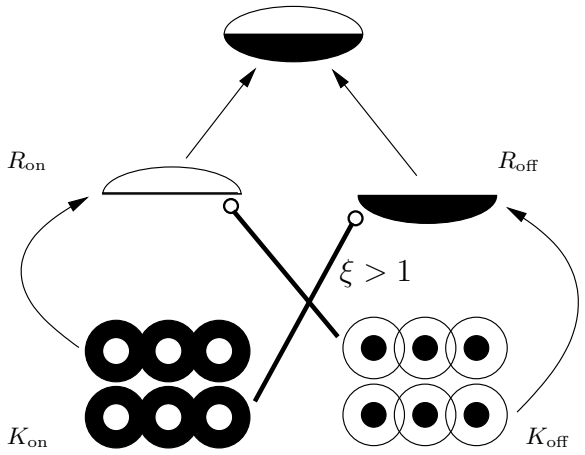


**Fig. 3.11:** Simple cell modell with dominating opponent inhibition: An on subfield $R_{\text{on}}$ receives excitatory input from properly aligned LGN on cells $K_{\text{on}}$ and inhibitory input from LGN off cells $K_{\text{off}}$. The scheme of dominating opponent inhibition proposes a stronger weighting of the inhibitory input with $\xi > 1$, as indicated by the thicker lines. The reverse wiring pattern exists for the off subfields $R_{\text{off}}$. *Arrows* denote excitatory input, *circles* at the end of lines denote inhibitory input. For clearness of display, subfields are drawn separated; interneurons are omitted to simplify the sketch.

The mechanism may also be interpreted in terms of voting, where excitatory and inhibitory inputs represent voting in favor or against a decision, i.e., whether the subfield responds or not. For balanced inhibition, a simple majority of 50% votes in favor results in a subfield response. For dominating inhibition, a majority greater than 50% of votes in favor are required to cause a response. More precisely, for a weighting of the opponent inhibition with $\xi > 1$, the excitatory drive has to comprise a fraction of $\xi/(\xi+1)$ of the total input to drive the cell. In terms of voting, this means that for setting, say, $\xi = 2$, a 2/3 majority is required to result in a response of the subfield.

DOI processing has important effects on the behavior of the model: It is the key feature for simulating data in a physiological study on luminance gradient reversal (Sec. 3.7.1), and it makes the model more robust to noise (Sec. 3.7.4 and 3.8). As stated above, DOI relies on strong inhibition. The assumption of strong inhibitory input to a simple cell that can overwhelm excitatory contributions is supported by many physiological studies. Evidence comes from both extracellular (Heggelund, 1981; Palmer and Davis, 1981) and intracellular recordings (Borg-Graham et al., 1998; Ferster, 1988; Hirsch et al., 1998).

**Nonlinear simple cell circuit.**

On and off subfields interact via a disinhibition circuit that boosts activities for spatially juxtaposed on and off contrast configurations. Such juxtaposed on and off contrasts occur at step edges, thus the simple cell model exhibits significantly higher responses for this configuration than for shallow luminance gradients, for example.

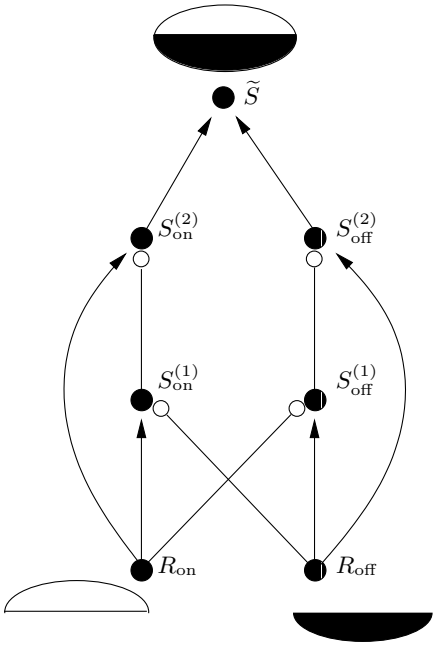


**Fig. 3.12:** Sketch of the simple cell circuit. *Arrows* denote excitatory input, *circles* at the end of lines denote inhibitory input. Index $\theta$ is omitted to simplify notation.

The circuit which defines the simple cell model comprises three intermediate stages, namely $S^{(1)}, S^{(2)}$, and $\widetilde{S}$ (Fig. 3.12). The various connections and their different computational roles are explained in the following. A comprehensive description and a detailed motivation of the nonlinear simple cell circuit are given by Neumann et al.. The basic circuitry is given by the excitatory $R_{\mathrm{on/off}} \rightarrow S^{(2)}_{\mathrm{on/off}} \rightarrow \widetilde{S}$ connections, which define the excitatory input to the simple cell from its two subfields $R_{\mathrm{on}}$ and $R_{\mathrm{off}}$. A model having only this basic circuitry results in a simple cell that linearly sums its input. To make the model more selective for juxtaposed on and off contrasts, additional connections are introduced. The on-channel path $R_{\mathrm{on}} \rightarrow S^{(1)}_{\mathrm{on}} \multimap S^{(2)}_{\mathrm{on}}$ implements a self-normalization by inhibition of $S^{(2)}_{\mathrm{on}}$, which prevents arbitrarily large activity of the cell. The same holds true for the off channel. The key connections of the model are the cross-channel inhibitory connections $R_{\mathrm{on}} \multimap S^{(1)}_{\mathrm{off}}$ and $R_{\mathrm{off}} \multimap S^{(1)}_{\mathrm{on}}$. By disinhibition, i.e., inhibiting the inhibition of $S^{(1)}$, the simple cell response is nonlinearly amplified if both subfields are active simultaneously.

The first two stages are steady-state solutions of inhibitory shunting interactions. The equations for the on channel read

$$S^{(1)}_{\mathrm{on}} = \frac{R_{\mathrm{on}}}{\alpha_S + \beta_S R_{\mathrm{off}}} \ ,$$
$$S^{(2)}_{\mathrm{on}} = \frac{R_{\mathrm{on}}}{\gamma_S + \delta_S S^{(1)}_{\mathrm{on}}} \ .$$

The corresponding equations for the off channel are obtained by interchanging on and off. Here and in the remainder of this section, variables occur for all discrete orientations. The index $\theta$ is omitted to simplify notation. The activity of the third stage $\widetilde{S}$ results from pooling the contributions of the on and off channel

$$\widetilde{S} = S^{(2)}_{\mathrm{on}} + S^{(2)}_{\mathrm{off}} \ .$$

Combining these equations and assuming a symmetric relation between the two channels by setting $\delta_S = \beta_S \gamma_S$ yields a more concise equation. The resulting simple cell activity consists of a linear

and a nonlinear, i.e., multiplicative, term

$$\widetilde{S} = \frac{\alpha_S(R_{\text{on}} + R_{\text{off}}) + 2\beta_S(R_{\text{on}}R_{\text{off}})}{\alpha_S\gamma_S + \beta_S\gamma_S(R_{\text{on}} + R_{\text{off}})} \quad . \tag{3.6}$$

The parameters are set to $\alpha_S = 1.0$, $\beta_S = 10\,000.0$, and $\gamma_S = 0.01$. Their specific choice is not critical as long as the linear components scaled by $\alpha_S$ and $\gamma_S$ are small compared to the nonlinear component scaled by $\beta_S$.

In the basic linear simple cell model, subfield responses $R_{\text{on}}$ and $R_{\text{off}}$ are simply added. In this case, Eq. 3.6 is replaced by

$$\widetilde{S}_{\text{lin}} = R_{\text{on}} + R_{\text{off}} \quad . \tag{3.7}$$

**Opposite Contrast Polarity Inhibition**

Simple cells sensitive to opposite contrast polarity, i.e., light-dark and dark-light, finally undergo mutual inhibition which sharpens the activity profile. The final simple cell responses are thus computed as

$$\begin{aligned} S_{\text{ld}} &= \left[\widetilde{S}_{\text{ld}} - \widetilde{S}_{\text{dl}}\right]^+ \quad , \\ S_{\text{dl}} &= \left[\widetilde{S}_{\text{dl}} - \widetilde{S}_{\text{ld}}\right]^+ \quad . \end{aligned} \tag{3.8}$$

Light-dark and dark-light simple cells are obtained by sampling the subfield activity with different offsets $\pm\tau = 3$ orthogonal to the axis of orientation of the simple cell: A light-dark cell has an on subfield with an offset to the left and an off subfield with an offset to the right. For a dark-light simple cell, left and right offsets are interchanged.

To sum up, the present simple cell model comprises two mechanisms with complementary functionality: DOI serves to suppress undesired spurious activity to noisy inputs, while the nonlinear simple cell circuit sharpens and amplifies desired responses to edges.

### 3.5.2 Complex Cells

Complex cell responses are insensitive to contrast polarity and are modeled by pooling responses of simple cells with opposite contrast polarity

$$C_\theta = S_{\text{ld},\theta} + S_{\text{dl},\theta} \quad . \tag{3.9}$$

Pooled complex cell responses $C_{\text{pool}}$ result from summing complex cell responses for all orientations

$$C_{\text{pool}} = \sum_\theta C_\theta \quad . \tag{3.10}$$

Pooled complex cells responses are a convenient way to display the lumbed responses of the whole model within a single image. For this purpose pooled complex cell responses are used, e.g., in Sec. 3.7.4.

## 3.6 Population Coding of Orientation

In the present model, the orientation of a stimulus is represented using a population code of complex cell responses. In signal processing, population coding is also known as multi-channel or diversity encoding. In population coding, a single quantity such as orientation is represented by a

number or population of neurons. The neurons have overlapping sensitivity or tuning curves, such that a particular signal (e.g., a bar of a certain orientation) results in the activity of many cells. Population codings allows for a more efficient transmission of information given a fixed number of channels with limited capacity, as reviewed by Mallot (2000).

Population coding involves two spaces: An explicit space consisting of the cell activities (such as the complex cell activities $C_\theta$), and an implicit space (Zemel and Hinton, 1994) of the underlying quantity of the visual world that is encoded (such as orientation). The transformations between these spaces are referred to as *encoding* and *decoding*. The encoding describes how a quantity of the visual world is expressed or encoded in terms of cell activities. The decoding describes the reverse process by which information about the quantity of the visual world can be extracted from these cell activities. The model presented above describes the process of encoding. In the following, we describe how a single quantity such as orientation can be decoded. An explicit decoding mechanism is useful to extract the information content that is present. This should not imply that the precise mechanism given below or even explicit decoding at all is used by the neural system (Lehky and Sejnowsky, 1998; Oram et al., 1998; Pouget and Zhang, 1997).

Decoding the measured local orientation $\theta_{\mathrm{dec}}$ from the complex cell responses $C_\theta$ can be formalized by vector addition, where an orientation vector $\mathbf{o}_\theta$ is assigned to each complex cell response (Jähne, 1997). The magnitude of the vector is given by the corresponding complex cell response. Since the orientation is limited by definition to the range $[0;\pi]$, the angle of the vector is given by the doubled orientation to cover the full range of $[0;2\pi]$. The equation for the orientation vector $\mathbf{o}_\theta$ associated with a complex cell $C_\theta$ is given by

$$\mathbf{o}_\theta = C_\theta \begin{pmatrix} \cos 2\theta \\ \sin 2\theta \end{pmatrix} \quad .$$

Expressing the 2D vector $(x,y)^{\mathrm{T}}$ as a complex number $z = x + \mathrm{i}y$ and using Euler's famous equation $\exp \mathrm{i}\theta = \cos\theta + \mathrm{i}\sin\theta$ yields the following more concise expression for the orientation

$$o_\theta = C_\theta \exp(2\mathrm{i}\theta) \quad .$$

The overall orientation vector is then given by the sum of the individual vectors

$$o = \sum_\theta o_\theta = \sum_\theta C_\theta \exp(2\mathrm{i}\theta) \quad .$$

The decoded orientation $\theta_{\mathrm{dec}}$ can be computed as arctangent of the overall orientation vector

$$\theta_{\mathrm{dec}} = 0.5 \operatorname{atan}\left(\frac{\mathrm{Re}(o)}{\mathrm{Im}(o)}\right) \quad .$$

Population coding is a general principle in neural systems. Examples can be found in such diverse fields as, e.g., the representation of limb movements by neurons in the motor cortex (Georgopoulos et al., 1986), representation of self-location by hippocampal place cells in rat (O'Keefe and Dostrovsky, 1971), color coding by different types of cones, or retinal position and hyperacuity by overlapping receptive fields of retinal ganglion cells (Fahle and Poggio, 1981). The benefits of population coding of visual orientation by striate cells has been studied in a model of Vogels (1990). The model can account for the highly accurate orientation estimation in the presence of cells which are only broadly tuned for orientation. Lüdtke et al. (2000) have devised a model for feature localization and tangent field estimation based on population coding of stimulus orientation by the response moduli of a bank of complex Gabor filters. Lüdtke et al. showed that the population coding approach can resolve finer details of the tangent field than other methods for orientation measurement (Leite and Hancock, 1997; Zucker et al., 1988). From a theoretical point of view, Bayesian inference has been proposed as an optimal decoding method for population codes (Oram et al., 1998; Zemel et al., 1998).

# 3.7 Simulations

## 3.7.1 Hammond and MacKay Study

In order to demonstrate the physiological plausibility and relevance of the proposed model, basic properties of simple cells found in vivo are simulated. In particular, we simulate a study of (Hammond and MacKay, 1983), who investigated the response of simple cells in cat to optimally oriented bars. This study is challenging for any model of simple cells because it shows classical effects like linear contrast summation up to saturation and strong, possibly nonlinear, suppressive effects.

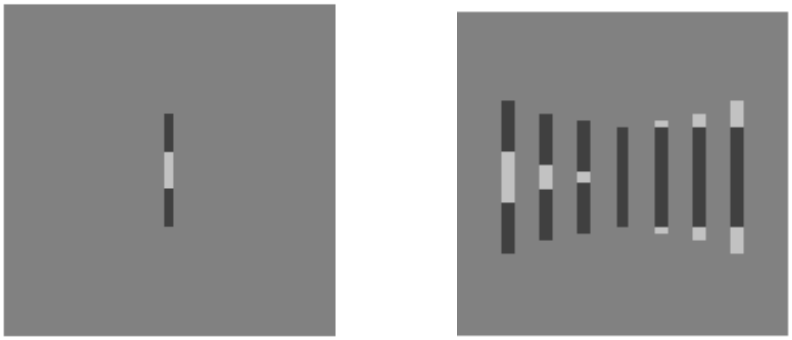

**Fig. 3.13:** *Left:* Example of stimulus used. *Right:* A set of stimuli for a fixed length of the dark bar.

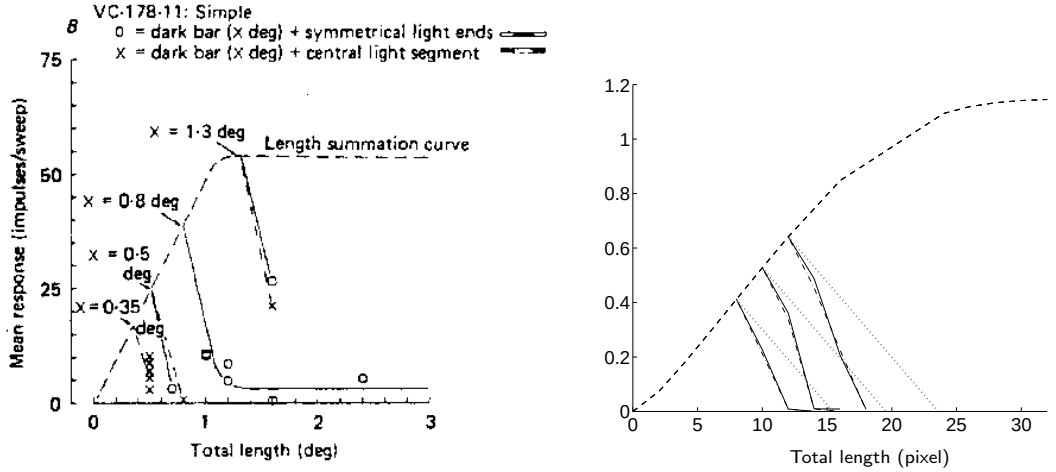


**Fig. 3.14:** *Left:* Results of physiological recordings (reprinted from Hammond and MacKay (1983) with permission of the publisher). *Right:* Simulation result. Both plots show the length-summation curve *(dashed)* and responses to LDL and DLD bars (*solid* and *dash-dotted*, respectively). For comparison, predictions by linear contrast summation *(dotted)* are shown in the simulation plot.

In their study, Hammond and MacKay recorded simple cell responses to three types of bar stimuli: Dark bars, dark bars with light segments added in the middle (DLD) and dark bars with light segments added at both ends (LDL). Figure 3.13 depicts the single stimulus used and a sample of the whole stimulus set. A main result of their work is shown in Fig. 3.14. For bar stimuli, linear response up to saturation is observed ("length-summation curve"). When light segments are added to the dark bars (DLD and LDL), the average response decrement is much larger than predicted from linear contrast summation. Linear summation would suggest that the slopes of the length-summation curve and of the LDL and DLD curves are the same.

Our model predicts that simple cell responses as observed by Hammond and MacKay can be generated on the basis of the proposed DOI scheme. Results are shown in Fig. 3.14 (right). The same model parameters as for the processing of images in Sec. 3.7.4 are employed. The declining slopes of the curves for both DLD and LDL stimuli are much steeper than the ascending slope of the length-summation curve, as reported by Hammond and MacKay. In summary, a good qualitative fit with the physiological data is obtained. Note that for the non-dominant case, i.e., setting the DOI parameter $\xi = 1$, no strong suppression occurs, but the responses for LDL and DLD bar stimuli lie on the dotted line as predicted by linear contrast summation. To rule out effects of the nonlinear simple cell, the circuit is replaced by a linear model. For the linear model, subfield responses are simply added: $\widetilde{S} = R_{\text{on}} + R_{\text{off}}$, instead of the nonlinear interaction in Eq. 3.6. For the linear model with DOI, the results obtained are qualitatively the same.

In their paper, Hammond and MacKay speculated that nonlinear or shunting suppression might cause the observed behavior. Considering the results mentioned here this notion cannot be rejected, but our model shows that a linear mechanism is also sufficient to explain the data.

In this section we have shown that DOI processing can account for both classical linear response up to saturation as well as strong depression effects as measured by Hammond and MacKay. In the next sections we investigate the influence on DOI on further properties of simple cells. In particular, we study the effects of inhibition on orientation tuning curves (Sec. 3.7.2) and on the perception of Glass pattern (Sec. 3.7.3).

### 3.7.2 Contrast-Invariant Orientation Tuning

The most prominent receptive field property of simple cells is their orientation selectivity. One can plot orientation tuning curves of simple cells by measuring the mean firing rate of a simple cell for stimulus items of different orientations. Orientation tuning curves of simple cells have a Gaussian shape which peaks at the prefered orientation. Remarkably, the width of these orientation tuning curves remains constant, even when the contrast of the stimulus is changed (Sclar and Freeman, 1982; Skottun et al., 1987). As pointed out by Ferster and Miller (2000), this property is difficult to explain in a simple feedforward model, because the responses of both retinal ganglion cells and LGN cells strongly depend on stimulus contrast (Cheng et al., 1995; Troy and Enroth-Cugell, 1993).

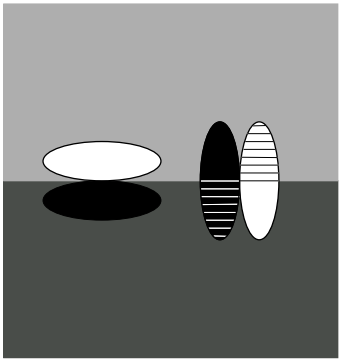

**Fig. 3.15:** Simple cell RFs with optimal (*top*) and orthogonal (*bottom*) orientation for a vertical dark-light transition. Even at the orthogonal orientation, the simple cell receives a certain input, as sketched by the dashed regions.

To understand this difficulty, consider a dark-light transition with orthogonal orientation to the cells preferred orientation (Fig. 3.15, bottom cell). Here, the off subfield receives a certain input generated from off cells responses in the dark part near the transition. Likewise, the on subfield receives input from on cell responses generated in the bright part near the transition. If the contrast

is sufficiently large, this localized input would drive the cell, even at non-preferred orientations. Further, this cell activation grows with contrast. As a results, a stimulus of high contrast would evoke a response for a cell of nonoptimal orientation, which could be of the same size or even higher than the response of the cell to a stimulus of optimal orientation at low contrasts.

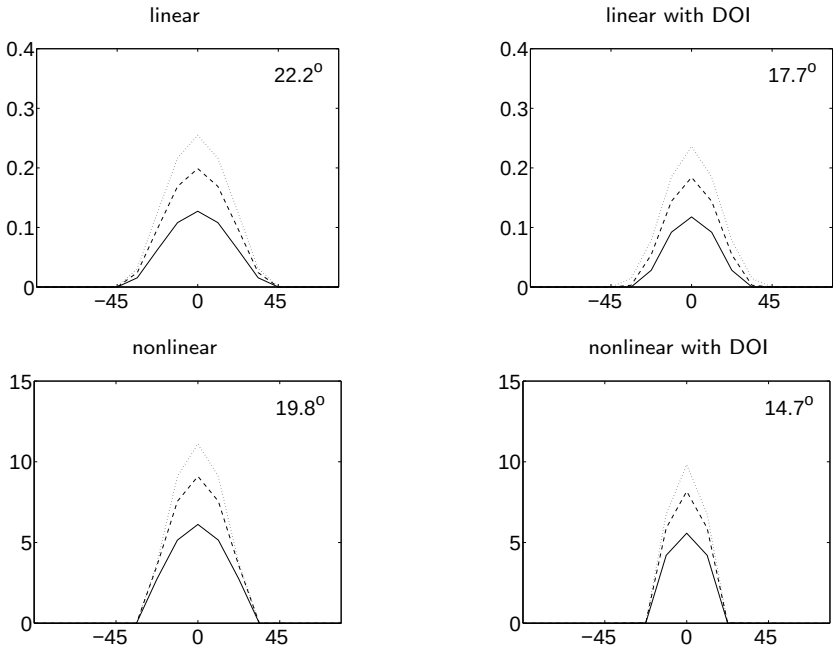


**Fig. 3.16:** Orientation tuning curves for the linear model and the nonlinear model with and without DOI. The abscissa denotes orientation in degree, the ordinate denotes the simple cell response at a fixed spatial location. The three curve in each plot corresponds to different contrast, namely 0.25, 0.5 and 0.8 marked with *solid, dashed* and *dotted* lines, respectively. *Top right inset* denotes mean HWHH. The orientation tuning curves show contrast invariant orientation tuning for all four models and a sharpening of orientation tuning by DOI.

As reviewed in Sec. 3.2, the wiring scheme as outlined in the simple feedforward model, where an on subfield gets input from on-center LGN cells only, has been outdated by physiological findings (e.g., Ferster, 1988; Heggelund, 1981; Hirsch et al., 1998; Palmer and Davis, 1981). Instead, simple cells receive strong push-pull inhibition, i.e., strong inhibition of the off-center responses in their on subfields and strong on inhibition in their off subfield. Troyer et al. (1998) have proposed a feedforward model using strong inhibition that accounts for contrast-invariant orientation tuning. This model is similar to the DOI scheme as it uses dominant rather than balanced inhibition. However, in Troyer et al. model, inhibition occurs between simple cells rather than at the level of LGN afferent of simple cell subfields.

In order to study the effects of DOI and the subsequent linear vs. nonlinear processing on the orientation selectivity of model simple cells, we probe the circuit with a sinusoidal grating stimulus of different contrasts and record orientation tuning curves. The orientation tuning curves are obtained from simple cells responses of $O_{\max} = 16$ orientations at the same fixed location, which is equivalent to probing a single simple cell with stimulus patterns of different orientations, as done in physiological experiments. The location is chosen which results in maximal response for a linear light-dark simple cell and is kept fixed for all parameter variations.

In a first study, we compare the effect of DOI on orientation tuning curves for the linear and the nonlinear model (Fig. 3.16). As a basic result, all curves show contrast invariant orientation tuning. Next, for both the linear and the nonlinear model, DOI, i.e, increasing the inhibition at the level of LGN cells, results in a sharper orientation tuning. Note that DOI has only small effects on the response magnitude of the optimally tuned cell.

The orientation tuning curves of the linear and the nonlinear models exhibit two general differences. First, the orientation tuning curves sharply fall off for the nonlinear model and more smoothly roll of for the nonlinear model. Second, the width of the curves are smaller for the nonlinear than for the linear model. To sum up, both DOI and nonlinear processing has the effect of decreasing the width of the tuning curves, with nonlinear processing additionally sharpens the shape of the curves. Consequently, the best results are obtained with the nonlinear model with DOI.

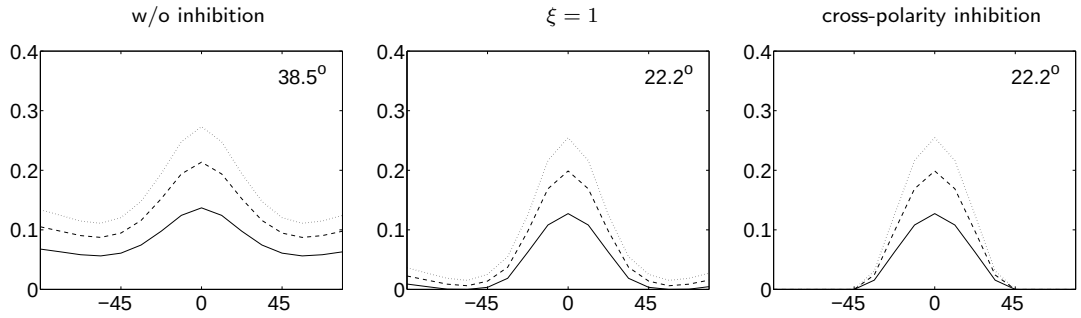


**Fig. 3.17:** Effects of inhibition on orientation tuning. Simple cell response versus orientation in degree for gratings of different contrasts (0.25, 0.5 and 0.8 marked with *solid, dashed* and *dotted* lines, respectively). *Left to right:* Linear simple cell model without any inhibition, inhibition added at the subfield level ($\xi = 1$), inhibition added at the simple cell level (cross-polarity inhibition). *Top right inset* denotes mean HWHH. Without inhibition, the model does not exhibit contrast invariant orientation tuning. Adding either kind of (balanced) inhibition suffices to a large extend to generate invariant tuning.

Since all models exhibit contrast invariant orientation tuning, especially the linear model with balanced inhitibion, we investigate the origin of this invariance property. Two possible candidates can be identified, namely the (balanced) inhibition at the subfield level and the cross-polarity inhibition at the simple cell level. Therefore, a plain linear model without any of the two kinds of the abovementioned inhitibions is simulated, and inhibition on either the subfield or the simple cell level is added. The results are depicted in Fig. 3.17. The plain model without any inhibiton corresponds two the simple cell model a suggested by Hubel and Wiesel. Clearly, this model does not show contrast invariant orientation tuning. Rather, cells of all orientations have non-zero responses. This can be circumvented by thresholding, but the resulting widths of orientation tuning curves would still remarkebly vary with contrast. By introducing inhibition, contrast independent tuning ca be obtained, where the inhibition at the simple cell level results in higher invariance.

The width of orientation tuning of simple cells in vivo exhibits a certain variation. The half width at half height (HWHH) between orientation tuning curves in monkey varies between 5° and 50°, with a most frequent tuning of 20° (Schiller et al., 1976). More recently, Carandini and Ferster (2000) have found an average HWHH for the spike responses of simple cells of 23°. The values obtained for the models are thus in good agreement with the physiological experiments.

To sum up, all models with any kind of inhibition have contrast tuning curves which are largely invariant to contrast. We conclude that balanced inhibition suffices to a large extend to provide contrast invariant orientation tuning. Therefore, contrary to the notion of Troyer et al. and Ferster and Miller, we propose that that the sharpening of orientation tuning cannot be considered the primarily functional role of dominant inhibition. Rather we suggest that DOI makes the simple cells more robust to noise, as further explicated.

Since all model at least balanced inhibition at the subfield level, the undesired behavior as sketched for a simple feedforward model without push-pull inhibition does show up.

### 3.7.3 Glass Dot Patterns

Glass dot patterns are composed of arrangements of individual dots evoking the perception of certain patterns, e.g., radial or concentric lines, depending on the spatial relationsship between individual dots (Glass, 1969). Glass pattern can be constructed from a random dot pattern superimposed by a transformed, e.g, rotated or translated, copy of the random pattern. A local dot arrangement is defined by a dot and its transformed counterpart. Depending on the kind of local arrangements, different global patterns are perceived. A radial pattern, for example, as shown in Fig. 3.18 (left) results from scaling, which aligns the dot pairs along radial paths. The perception

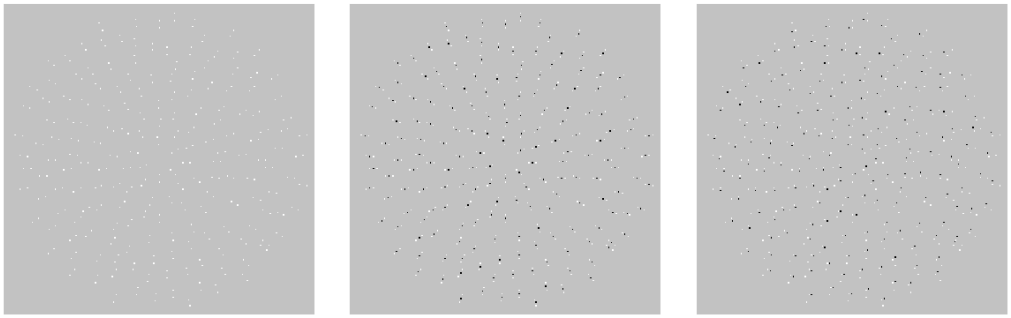

**Fig. 3.18:** Radial glass dot pattern (*left*) and modifications used in the study of Brookes and Stevens (1991). In the modified Glass pattern, a black dot is added between the white dots (*middle*) or adjacent to (*right*). Brookes and Stevens showed that in the dot-adjacent condition, a radial pattern is perceived, similar to the standard Glass pattern, while in the dot-between condition the pattern is more likely to be perceived as concentric. Due to reduced resolution and additional discretization, the perceptual effect may be substantially reduced

of the structure generated by a Glass pattern depends on two types of interaction: a measurement of the local orientation defined by the dot pairs and a more global long-range integration of oriented items (Sagi and Kovács, 1993). Simple cells have been proposed as responsible for the local orientation measurement: a dot pair would be roughly equivalent to a continuous line of equal total energy presuming a linear summation within the receptive field. As shown by Hammond and MacKay, a light segment of opposite contrast polarity placed within the receptive field significantly attenuates the cell response.

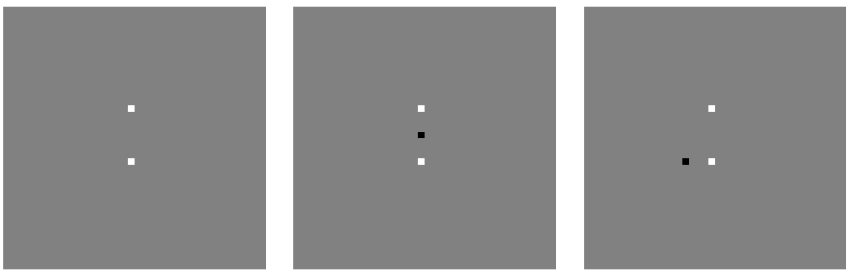

**Fig. 3.19:** Individual dot items used for the simulations. Local dot item of two white dots to generate the standard Glass pattern (*left*) and modification introduced by Brookes and Stevens (1991). In the modifications, a black dot is added either between (*middle*) or adjacent to (*right*) the white dot pair resulting in categorial changes in the perception of the Glass pattern constructed from this items

To further confirm the contribution of simple cells to the perception of Glass pattern, Brookes and Stevens investigated the perception of modified Glass pattern. They generated classical Glass pattern of oriented white dot pair items with radial or concentric orientation relative to the pattern center and added a black dot of opposite contrast polarity to each of the white dot pairs. There

were two possible positions for the opposite-contrast dot. In the "dot-between" case, the black dot was placed in the middle of the two white dots, in analogue to the LDL stimuli used by Hammond and MacKay. In the "dot-adjacent" case, which serves as a control, the black dot was placed adjacent to the pair of white dots at a distance of half the spacing between the pair. If the perception of structure in Glass pattern depends on the measurement of local contrast direction, the perception should be significantly affected in the "dot-between" case, since the black dot disrupts the response for the white dot pair item. Moreover, the "dot-between" case should give rise to the perception of an orthogonal organization, such that a radial arrangement is more likely to be perceived as concentric, and vice versa. These predictions were confirmed in a psychophysical experiment using alternative forced-choice judgement.

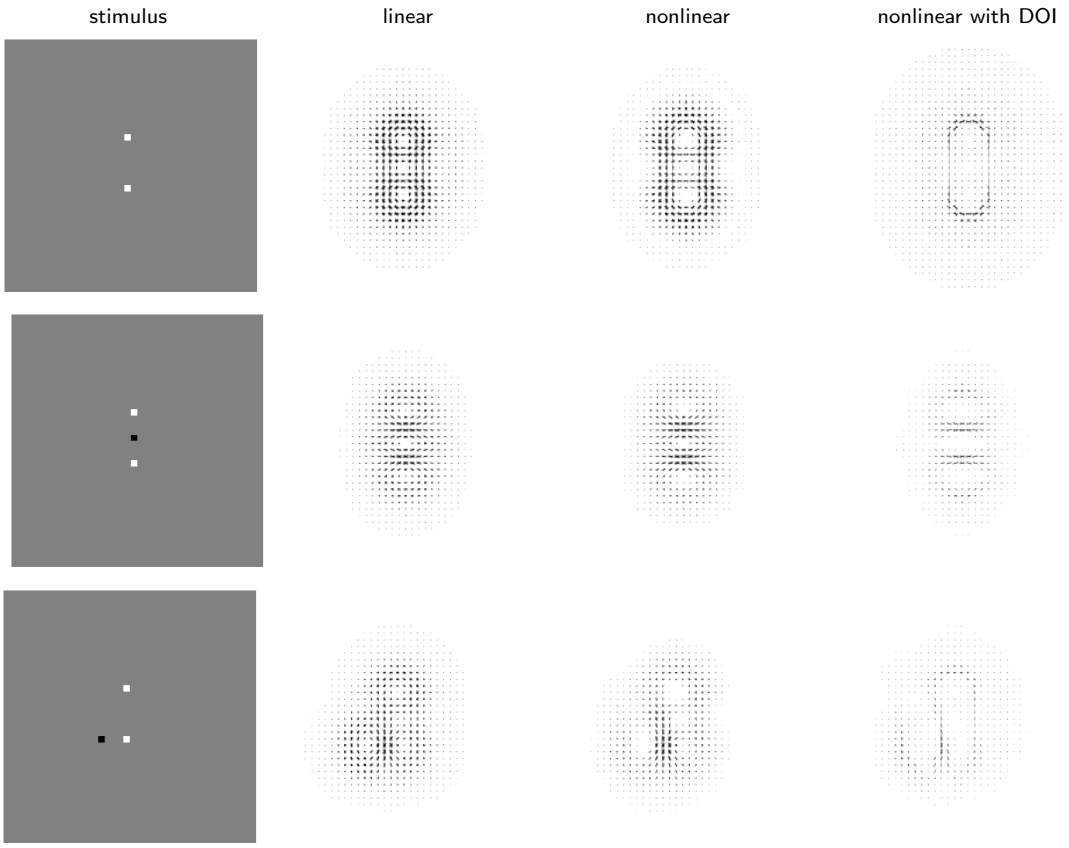


**Fig. 3.20:** Results of processing the local dot items by the simple cell circuit. For each of the three different item in Fig. 3.19, namely standard, dot-between and dot-adjacent (top to bottom row), simulation results of the linear (*left*), nonlinear (*middle*) and nonlinear simple cell model with DOI (*right*) are shown. The orientation plots show for each position the activity of the $O_{\max} = 8$ orientations. The length of each oriented line encodes the relative magnitude of response with respect to the maximum response in that display

The experiment of Brookes and Stevens further confirms the assumption that measurement of local contrast configurations, such as provided by simple cells, is essential for the perception of Glass pattern. Motivated by this study, we probe our simple cell circuit and its modifications with the local dot patterns used by Brookes and Stevens. Using the local dot pattern instead of the whole Glass dot pattern is valid, since simple cell provide a local contrast measurement. We use the individual dot items as depicted in Fig. 3.19 as input, (standard white dots, dot-between and dot-adjacent modification) and compare the model output of three different simple cell models, namely, linear, nonlinear and nonlinear with DOI. All three models exhibit a significantly different pattern of activation for the dot-between than for the dot-adjacent condition, as reported by Brookes and

Stevens: in the dot-adjacent condition, the predominant activation of simple cells is parallel to the axis defined by the two white dots (similar to the standard condition of two white dots), while in the dot-between condition, the main activity is orthogonal to this axis. The three models differ in the fuzziness of response: the nonlinear model shows a sharper and more localized response than the linear model, and DOI further sharpens the response. Such sharper responses presumably leads to higher responses of a subsequent more global grouping mechanism that integrates local items into larger structures.

### 3.7.4 Processing of Images

In this section, we show the performance of the model on synthetic and on natural images. The values of the model parameters are as described in Sec. 3.5 and are the same in all simulations. In the simulations we compare the new mechanism of DOI (setting $\xi = 2$) to a linear simple cell model and to the nonlinear model without DOI ($\xi = 1$). Recall that for the linear model subfield responses are simply added to replace the nonlinear interaction of Eq. 3.6. The linear model approximates filtering with a first order Gaussian derivative (Neumann et al., 1999). The edge images show pooled complex cell responses (Eq. 3.10), which are obtained by summing up simple cell responses of both contrast polarities for all orientations. The edge images are reverted, such that dark values indicate high responses.

#### Synthetic Images

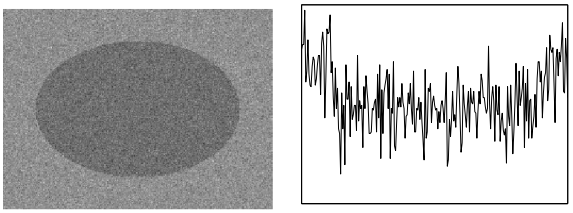

**Fig. 3.21:** Noisy ellipse *(left)* and corresponding horizontal cross-section *(right)* taken at the center of the image. The size of the ellipse image is $253 \times 189$ pixels.

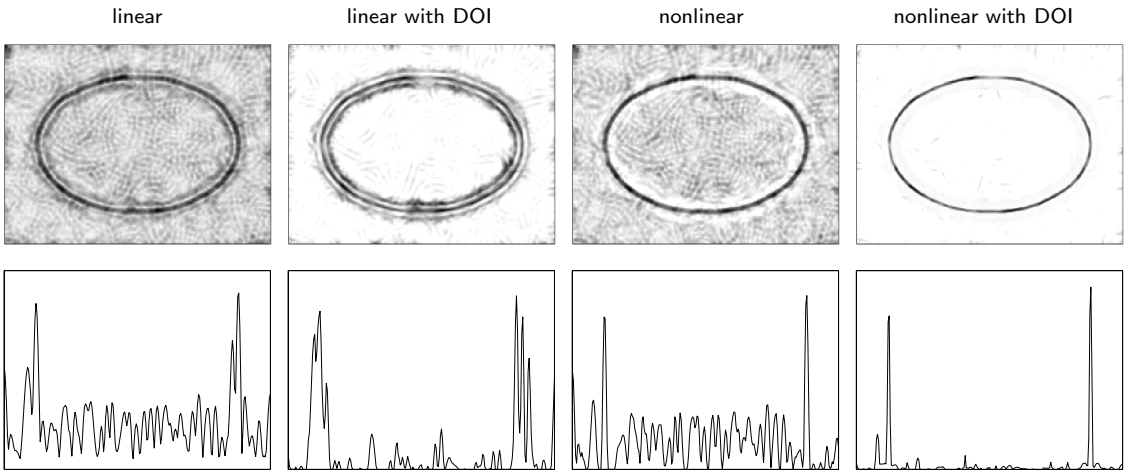


**Fig. 3.22:** *Top row:* Simulation results for the noisy ellipse stimulus. *Bottom row:* The corresponding horizontal cross-sections taken at the center of the images. The size of the grayscale images is $253 \times 189$ pixels.

In the first study we employ a synthetic image of a dark ellipse on a lighter background, corrupted with 50% additive Gaussian noise. Figure 3.21 shows the input stimulus together with a horizontal cross-section taken at the center of the image. For this stimulus, four results are generated by combining two interaction at the subfield level (standard and DOI) with both the linear and the nonlinear simple cell model. The simulation results are shown in Fig. 3.22. All models show pronounced responses at the edge locations, but only the nonlinear models have a unimodal response to an edge. Moreover, the results show that the models with DOI are considerably less sensitive to noise. The simulation results for this stimulus exemplify the complementary properties of DOI and of the nonlinear simple cell circuit: DOI serves to suppress noisy inputs, while the nonlinear interaction sharpens the responses to edges.

### Natural Images

A further challenge to the model is posed by processing of natural images. We use the tree image shown in Fig. Fig. 3.23 together with the simulations results.

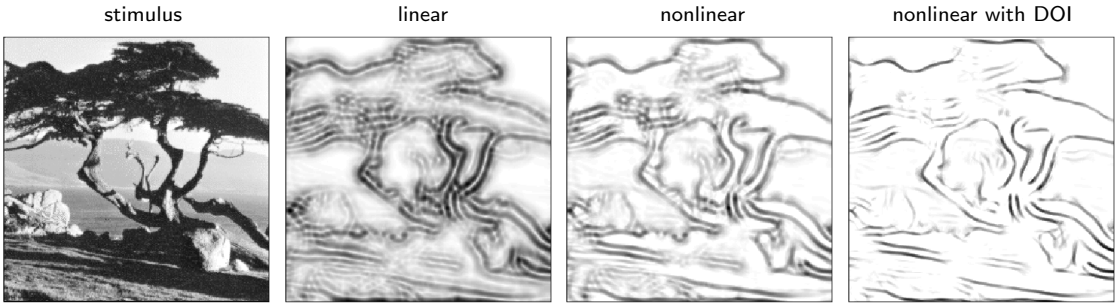


**Fig. 3.23:** Natural image of a tree and simulation results. The size of the images is $255 \times 256$ pixels.

For the DOI processing, responses to the lawn are largely suppressed, while responses to the contour of the tree and to the shadow are enhanced. We also employ an image of a 3D laboratory scene as input stimulus (Fig. 3.24). Here, the contours of the cube are sharper and the spurious responses at the floor vanish for DOI processing.

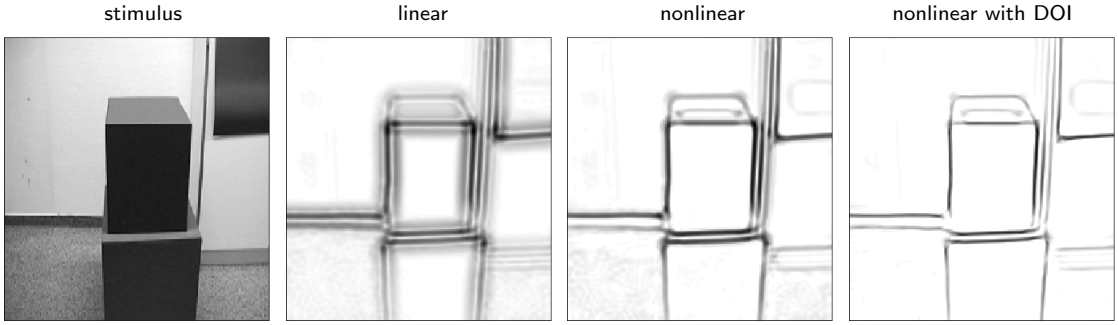


**Fig. 3.24:** Image of a laboratory scene and simulation results. The size of the images is $230 \times 246$ pixels.

In two further simulations, we employ images from a set used in an evaluation study of edge detection algorithms (Heath et al., 1997). We use a larger scale to show the simulation results to compensate for the approximately doubled image size compared to the previously shown images.

For the traffic cone, the DOI processing results in sharp, pronounced responses the the shape outline of the cone and the car, while spurious responses to the leaves are successfully suppressed. Similar results are obtained for the second image of a golf cart, where the edges of the cart are

reliably detected, while responses to small noisy structures such as the lawn and the leaves are suppressed by DOI. For better visualization of the simulation results, we employed a gamma correction with $\gamma = 0.6$ because of the high contrast variations in this image.

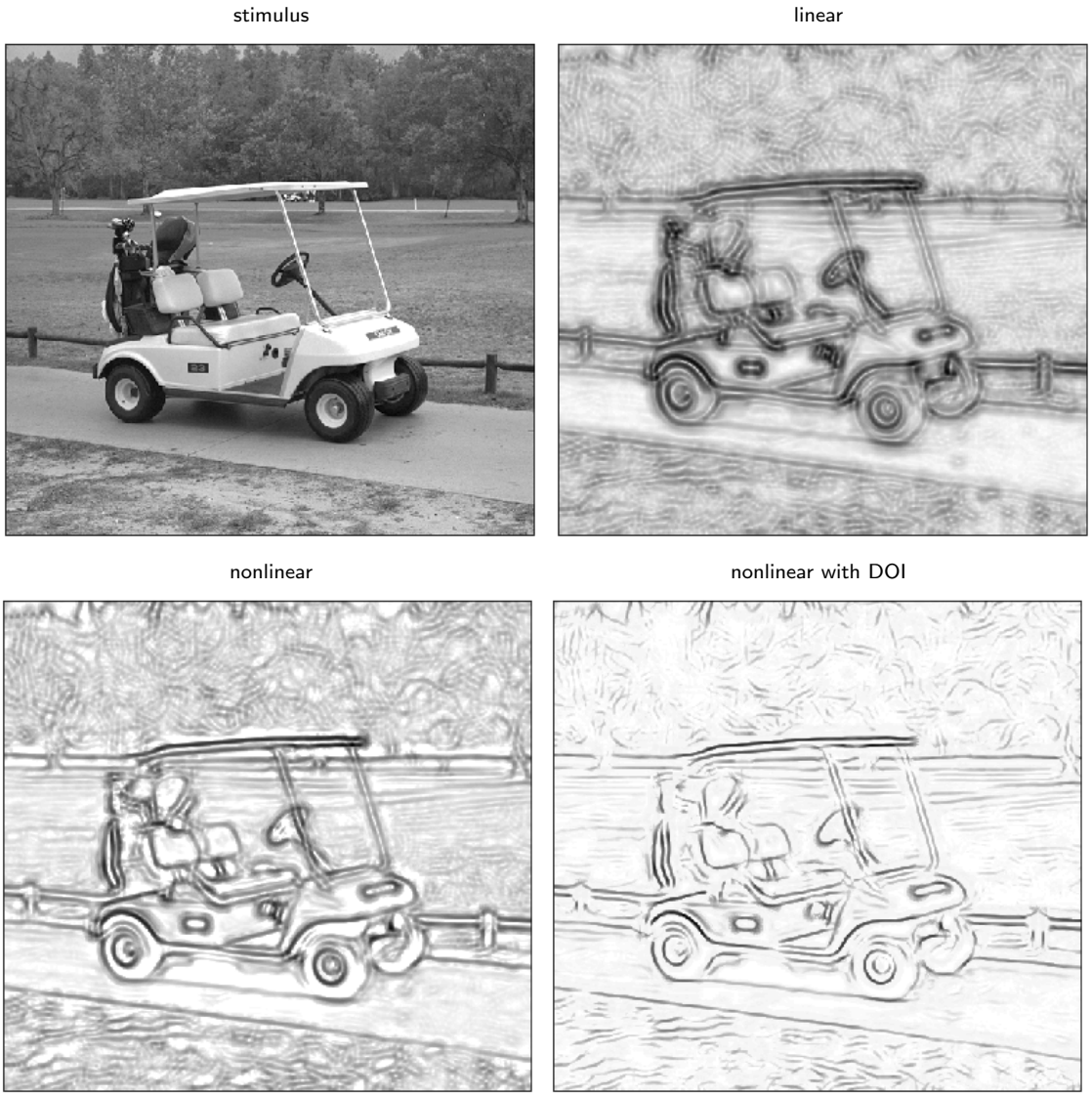


**Fig. 3.25:** Golf cart stimulus and simulation results. The edge images are gamma corrected with $\gamma = 0.6$. The size of the images is $548 \times 509$ pixels.

In a concluding simulation study, we employ natural images obtained online from Kodak (Kodak, 2001). The simulation results further exemplify the benefits of DOI processing: Responses to noisy, irregular texture-like structures such as the sand in the geyser image or the fur in the koala image are suppressed, while responses to salient edge structures are retained. In some of the simulation results, the DOI processing seems to suppress also responses to low contrast edges. This effect is studied in detail in 3.8.2. It turns out that the small responses are also retained for DOI processing, but their visibility is reduced due to the rendering of the simulation results.

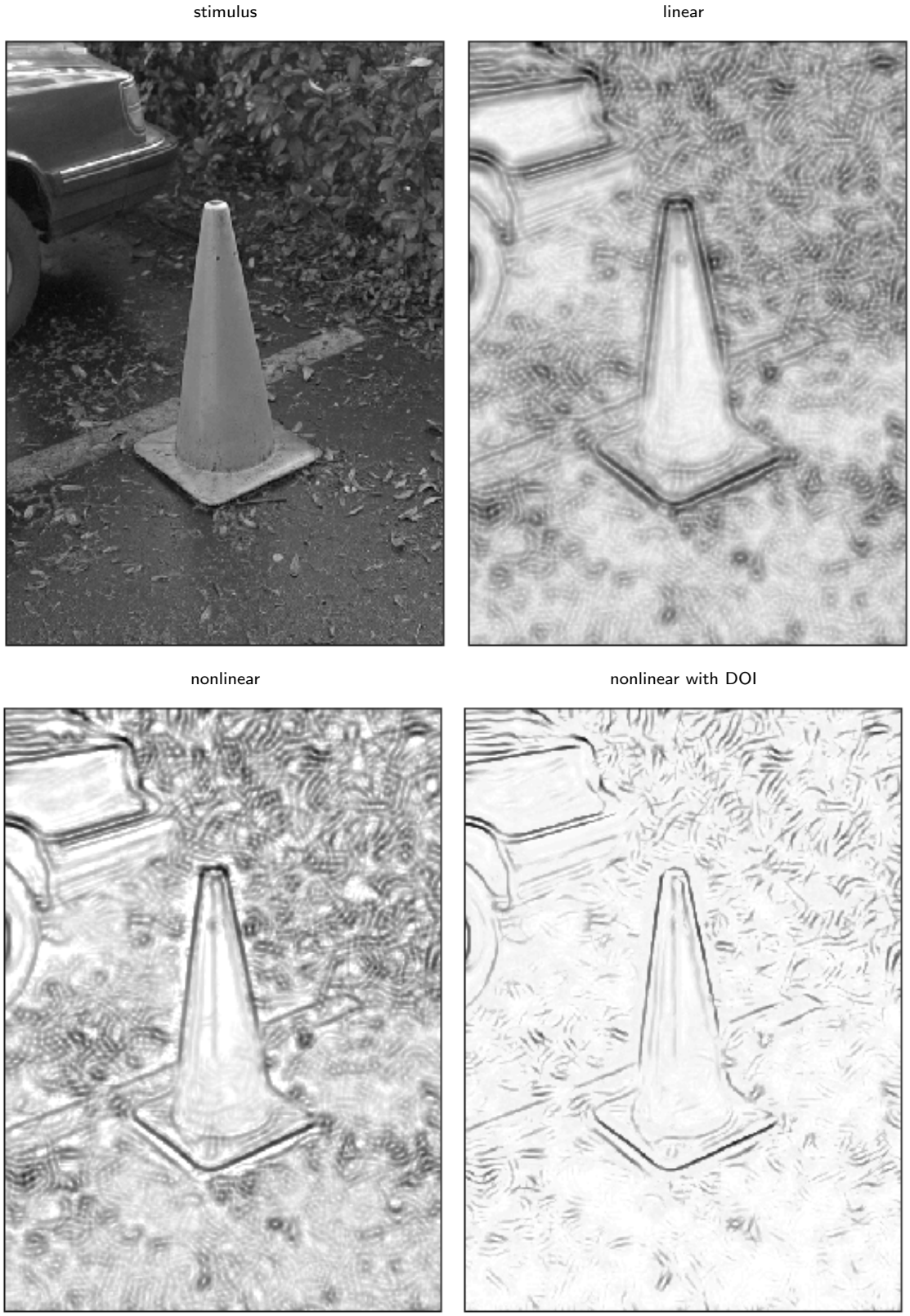


**Fig. 3.26:** Traffic cone stimulus and simulation results. The size of the images is $437 \times 604$ pixels.

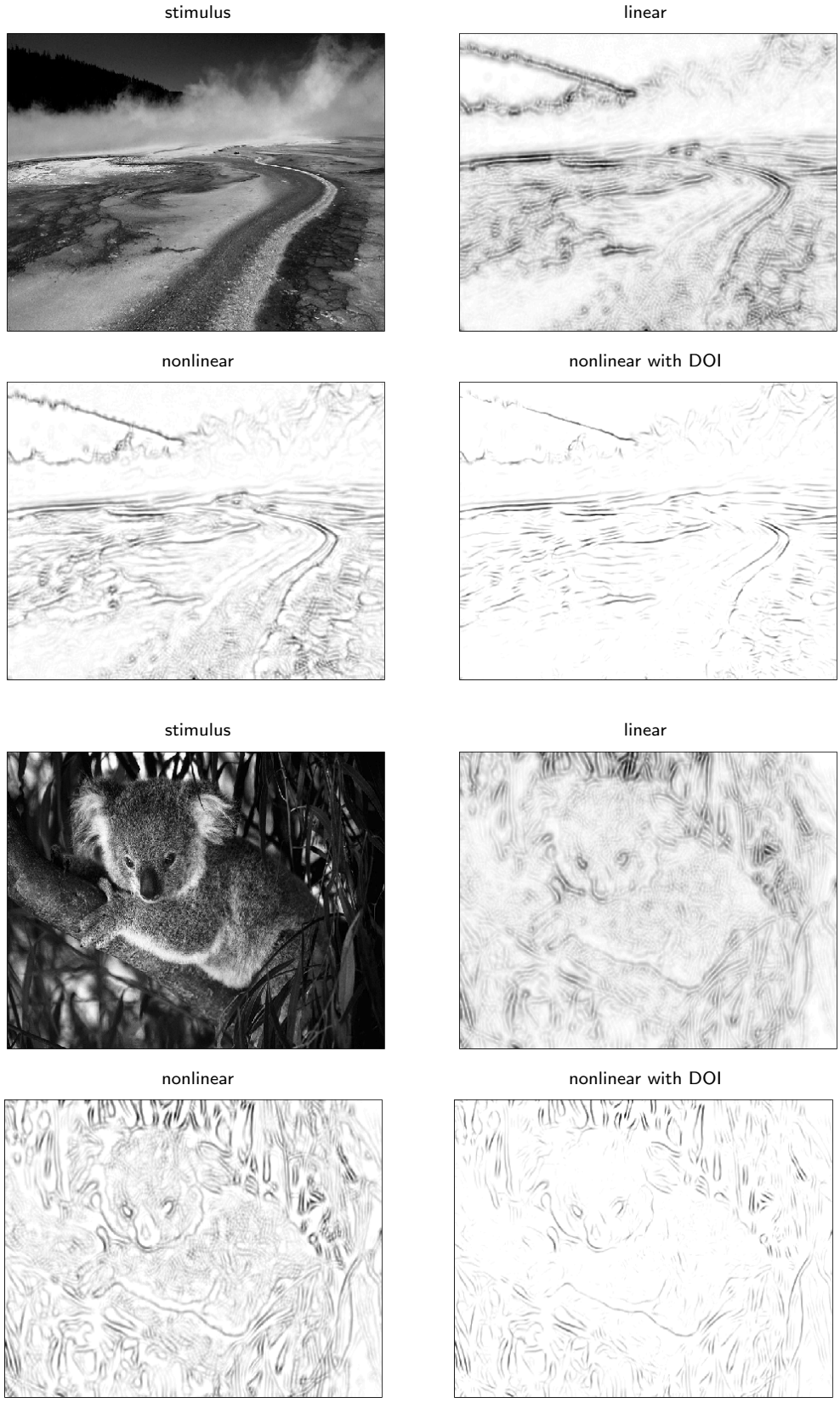


**Fig. 3.27:** Geyser and koala stimulus and simulation results. The size of the geyer images is $732 \times 490$ pixels, the size of the koala images is $737 \times 493$ pixels.

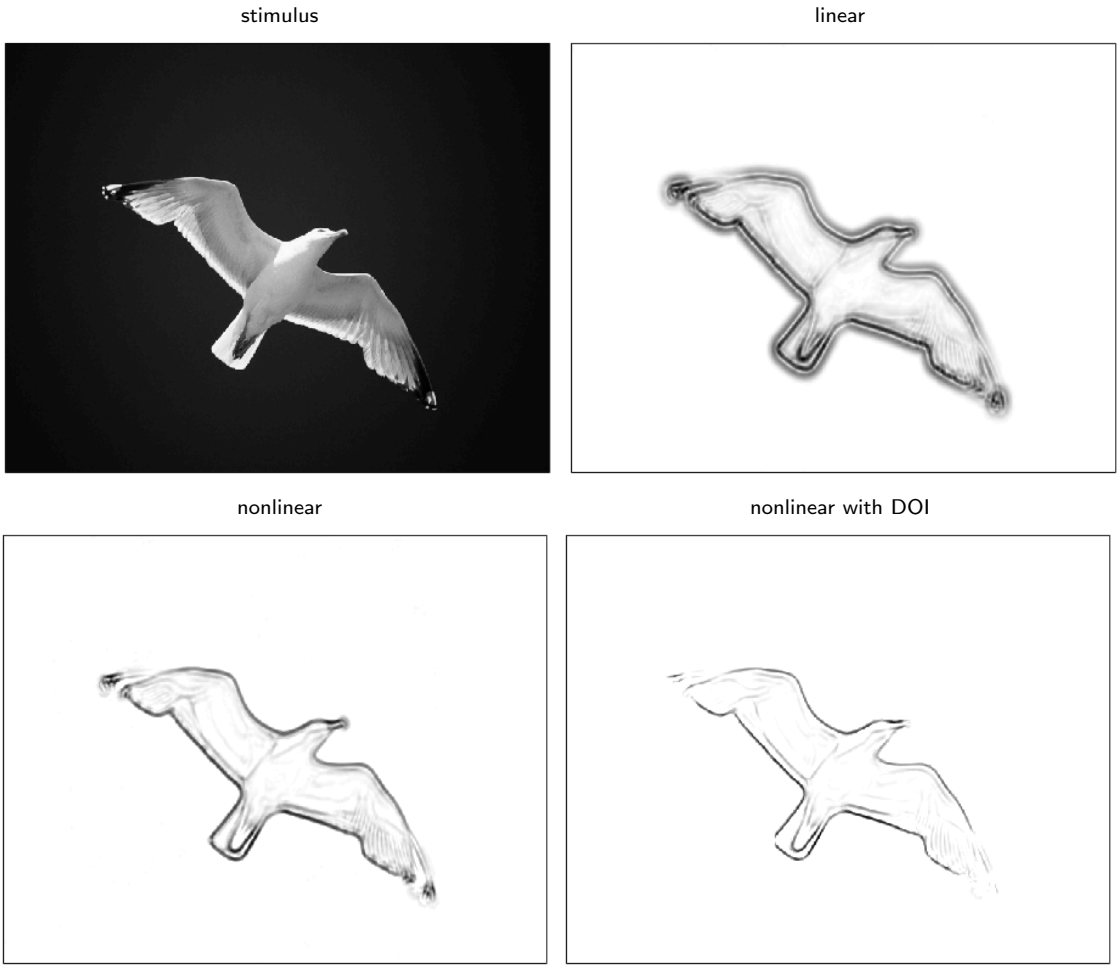


**Fig. 3.28:** Seagull stimulus and simulation results. The size of the images is $735 \times 486$ pixels.

## 3.8 Evaluation of DOI Properties

In the previous section, we have shown the properties of DOI qualitatively. In this section, we clarify the properties by a stochastic analysis and by numerical simulations. The stochastic analysis shows that DOI processing introduces an adaptive threshold, using the noise level itself to determine the amount of suppression. As a result, good noise suppression for various noise levels is achieved. The numerical evaluation is subdivided into two parts. In the first part, subfield responses $R_{\text{on}}$ are evaluated to determine the value of the DOI parameter $\xi$. The DOI parameter is determined to match two conflicting requirements as well as possible, namely suppression of noise and responsiveness to edges. In the second part, complex cell responses $C_{\text{pool}}$ are evaluated for the three simple cell models (linear, nonlinear and nonlinear with DOI) to clarify the response of DOI to small contrast changes in the presence of noise. Also, the suppression of background noise irrespectively of the amount of noise is evaluated for the complex cell responses.

### 3.8.1 Stochastic Analysis

In this section, we conduct a stochastic analysis of the response properties of DOI to noisy homogeneous regions. Let $I$ be an input image of homogeneous intensity, corrupted by an additive Gaussian noise process, and let $X$ be the result of a contrast extraction operation, approximated by the application of a DoG filter to $I$. Because DoG filtering is a linear operation, the resulting $X$ is a Gaussian process as well (Papoulis, 1965). At each spatial position, $X$ can be described by a random variable $\mathbf{x}$ with a density $f_{\mathbf{x}}(x)$ that is defined by a Gaussian distribution $g_\sigma(x)$. The DOI interaction in Eq. 3.5 can be written as

$$\begin{aligned}\mathbf{y} &= \mathbf{x}_{\text{on}} - \xi\,\mathbf{x}_{\text{off}} \\ &= [\mathbf{x}]^+ - \xi[-\mathbf{x}]^+ = \begin{cases}\mathbf{x} & \text{if } x \geq 0 \\ \xi\mathbf{x} & \text{if } x < 0\end{cases}.\end{aligned}$$

The density $f_{\mathbf{y}}(y)$ can be determined with the fundamental theorem on transformations of densities (Papoulis, 1965):

$$f_{\mathbf{y}}(y) = \begin{cases} f_{\mathbf{x}}(y) = g_\sigma(y) & \text{if } y \geq 0 \\ 1/\xi\; f_{\mathbf{x}}(1/\xi\; y) = g_{\sigma\xi}(y) & \text{if } y < 0\end{cases}.$$

From the above one can see that for values of $\xi \neq 1$ the resulting density is discontinuous at zero and described by Gaussians with different standard deviations for the positive and negative part of the $y$-axis. Fig. 3.29 shows density plots for different values of $\xi$.

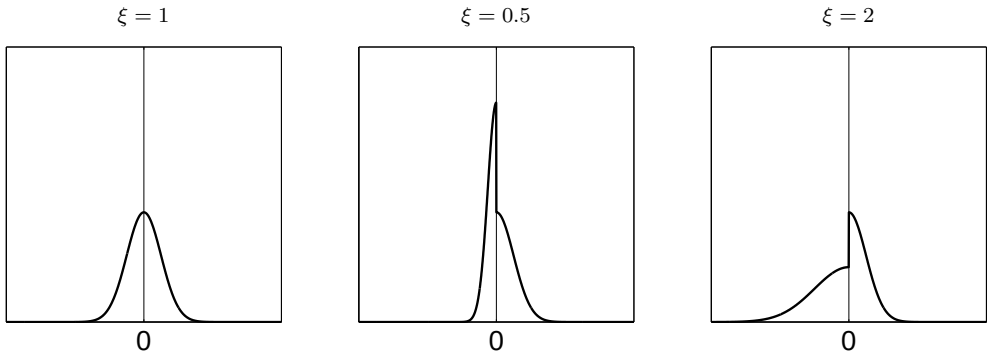


**Fig. 3.29:** Density plots for constant $\sigma$ and different values of $\xi$.

The next processing step is the convolution of $\mathbf{y}$ with the subfield mask $G_\theta$, which realizes a weighted average over a limited spatial neighborhood. Under the general assumption of an ergodic

process (Papoulis, 1965), the ensemble (or spatial) average in homogeneous regions corresponds to the mean of the individual units. The mean of $\mathbf{y}$ is given by

$$\begin{aligned} E\{\mathbf{y}\} &= \int_{-\infty}^{\infty} y f_{\mathbf{y}}(y)\ \mathrm{d}\, y = \int_{-\infty}^{0} y g_{\sigma\xi}(y)\ \mathrm{d}\, y + \int_{0}^{\infty} y g_{\sigma}(y)\ \mathrm{d}\, y \\ &= -\frac{1}{\sqrt{2\pi}}\,\sigma\xi + \frac{1}{\sqrt{2\pi}}\,\sigma \\ &= -\frac{1}{\sqrt{2\pi}}\,\sigma(\xi - 1)\ . \end{aligned} \tag{3.11}$$

The result shows that the mean is i) negative for $\xi > 1$ and ii) proportional to the DOI parameter $\xi - 1$ and to the noise level $\sigma$. The expression for the mean of $\mathbf{y}$ (Eq. 3.11) explains the suppression of noise largely independently of the noise level. As the noise level increases, the mean proportionally shifts to more negative values, so that the amount of positive activity left after the subsequent rectification (Eq. 3.5) remains small. This theoretical finding is confirmed by numerical simulations in the next section (cf. Fig. 3.30).

### 3.8.2 Numerical Evaluation

#### Determination of the DOI Parameter

In order to determine the optimal strength of inhibition which is controlled by the DOI parameter $\xi$ and to investigate the circuit's response properties to noisy inputs, we measure mean subfield responses $R_{\mathrm{on}}$ (Eq. 3.5). We vary the noise level and the value of the DOI parameter $\xi$. Simulations are done for two basic situations, namely noisy a homogeneous regions and a noisy step edge. In both cases, Gaussian noise is added to the ideal stimulus.

**Noisy Homogeneous Region.** In the first study, a stimulus of homogeneous intensity is corrupted by additive Gaussian noise. We measure the mean response over all spatial positions of an on subfield $\bar{R}_{\mathrm{on},\theta}$. Since noisy homogeneous regions do not have any preferred orientation, the choice of the orientation of the subfield $\theta$ is irrelevant. For the simulations, $\theta = 90°$ is chosen.

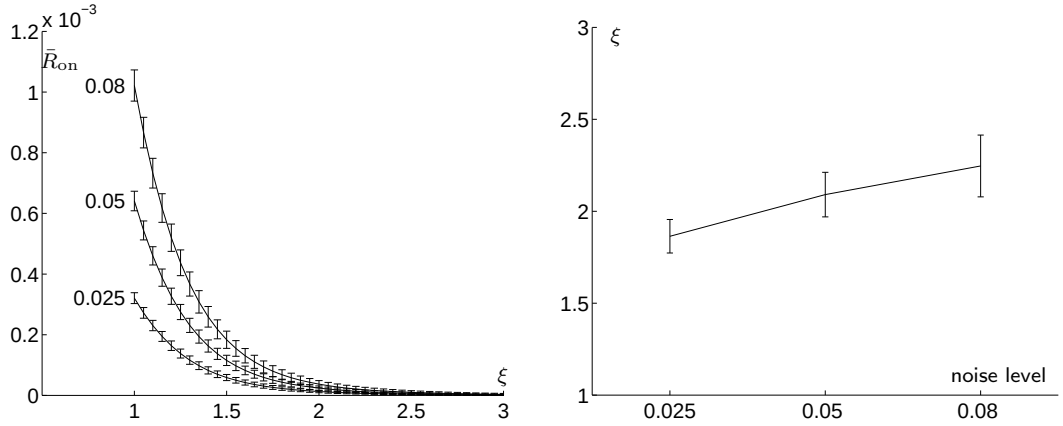


**Fig. 3.30:** *Left:* Mean subfield responses $\bar{R}_{\mathrm{on}}$ to homogeneous regions. *Right:* Minimal $\xi$ for which the mean subfield response of the respective noise level falls below $2 \cdot 10^{-5}$. For the three noise levels, the respective mean is given by 1.86, 2.09, 2.25. For both plots, three different noise levels (standard deviation 0.025, 0.05, 0.08) are employed. Responses are averaged over 100 realizations of the respective noise leve; error bars denote $\pm 1$ standard deviation. The curves show that for dominating opponent inhibition with $\xi > \approx 2$, noise is suppressed largely independently of the noise level.

The results are depicted in Fig. 3.30 (left). The curves correspond to Gaussian noise of decreasing standard deviations (top to bottom). We observe that the mean subfield response decreases as $\xi$ gets larger and is almost zero for $\xi > \approx 2$. For a more quantitative evaluation, the value of $\xi$ is determined for which the mean subfield response of the respective noise level falls below a certain threshold of $2 \cdot 10^{-5}$ (Fig. 3.30, right). For the highest noise level, $\xi$ has the value 2.25. The curves show that suppression occurs for values of $\xi$ that are significantly larger than 1, a value which corresponds to balanced excitation and inhibition. Further, suppression is largely independently of the noise level, as only a slight decrease of $\xi$ with the noise level can be observed.

Insight into this adaptive behavior can be gained by analyzing Eq. 3.5 in more detail. Using Eq. 3.4 and the equality $[x]^+ - [-x]^+ = x$, we can rewrite Eq. 3.5 as

$$\begin{aligned} R_{\text{on},\theta} &= \big[(K_{\text{on}} - \xi K_{\text{off}}) * G_\theta\big]^+ \\ &= \big[(K_{\text{on}} - K_{\text{off}}) * G_\theta - (\xi - 1) K_{\text{off}} * G_\theta\big]^+ \\ &= \big[K * G_\theta - \underbrace{(\xi - 1) K_{\text{off}} * G_\theta}_{\text{dynamic threshold}}\big]^+ \ . \end{aligned} \tag{3.12}$$

This shows that DOI interaction introduces a dynamic threshold that is proportional to $\xi$ and depends on the strength of the signal in the opponent pathway. Note that in the non-dominating case for $\xi = 1$, Eq. 3.12 reduces to

$$R_{\text{on},\theta}|_{\xi=1} = \big[K * G_\theta\big]^+ \ .$$

To summarize, a noisy stimulus generates responses in both on and off pathways. DOI interaction introduces a dynamic threshold by scaling up the contribution of the opponent pathway, which causes a decrease of response proportional to the noise level.

The simulations suggest that any value of $\xi > \approx 2.25$ would be appropriate to suppress responses to noisy homogeneous regions. However, the effect of large $\xi$ to suppress desired responses to signals like edges, for example, also needs to be clarified. This issue is addressed in the next section.

**Noisy Step Edge.** In a complementary experiment, an ideal is corrupted with Gaussian noise of the same standard deviations as in the case of noisy homogeneous regions in the previous section. For a luminance difference at the edge of 0.1, this results in 25%, 50% and 80%, Gaussian noise (i.e., noise with a standard deviation of 25%, 50%, and 80% of the luminance difference at the edge). Two kinds of responses are distinguished: Response of an optimally oriented subfield $R_{\text{on},90^\circ}$ with an orientation parallel to the edge, and responses of nonoptimally oriented subfields not parallel to the edge. Since all the examined non-optimal orientations of 0°, 45°, and 135° yield comparably large responses, we choose one representative, namely the orthogonal orientation $R_{\text{on},0^\circ}$. For both the optimal and the nonoptimal orientation, we measure the mean response of $R_{\text{on},\theta}$ along a line parallel to the edge. The horizontal position of the line is analytically determined as the position of maximal response of an optimal subfield to an ideal edge.

Simulation results are shown in Fig. 3.31. For the optimal orientation (top left), responses are a decreasing function of $\xi$, indicating that $\xi$ cannot be chosen arbitrarily large. For the nonoptimal orientation (bottom left), responses are almost zero for $\xi > \approx 2$. For a more quantitative evaluation, the value of $\xi$ is determined for which the mean nonoptimal response of the respective noise level falls below a threshold of zero (Fig. 3.31, bottom right). Here, $\xi$ depends more on the noise level than in the homogeneous case. For the highest noise level of 80%, $\xi$ has the value of 2.01 (Fig. 3.31, bottom right). Because the optimal response is a decreasing function for $\xi$ and the threshold is set to zero, the respective values of $\xi$ also determine the maxima of the signal-to-noise ratios, i.e., the ratios of the optimal and the nonoptimal response $\bar{R}_{\text{on},90^\circ} / \bar{R}_{\text{on},0^\circ}$ for each noise level.

These results provide criteria for the choice of $\xi$. A value of $\xi \approx 2$ yields the maximal signal-to-noise ratio for the highest noise level. Since the optimal responses decrease more slowly for

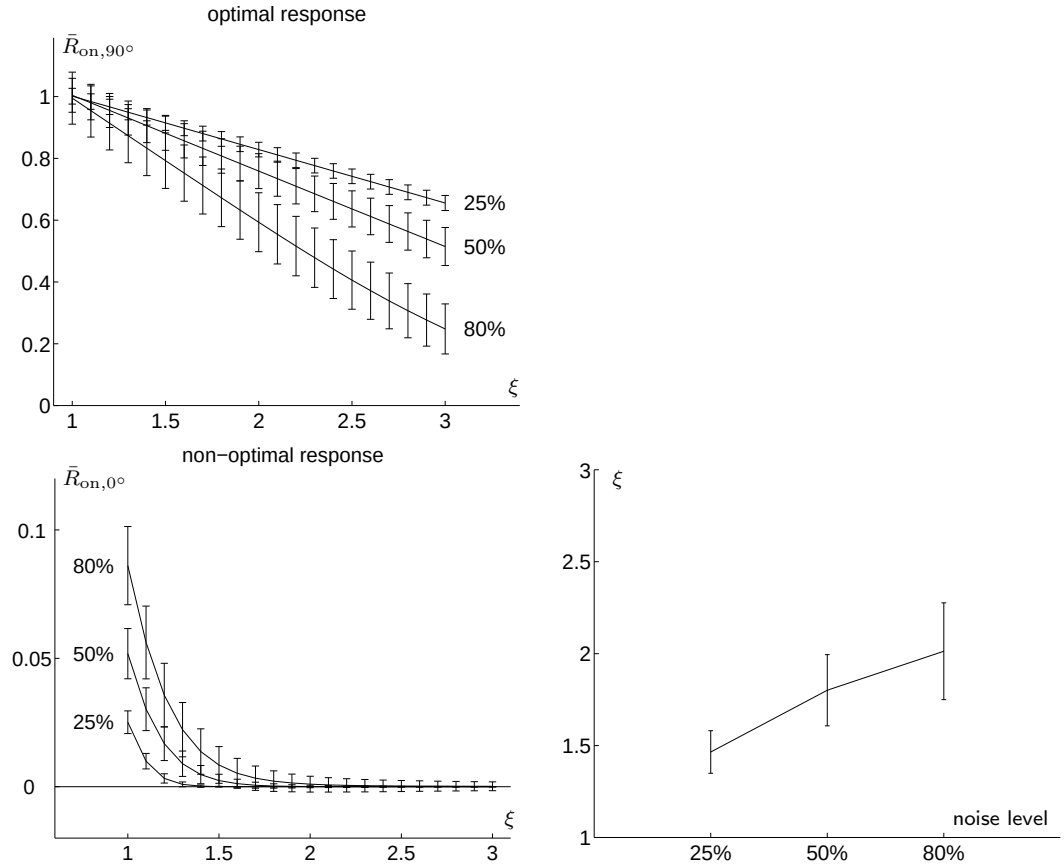


**Fig. 3.31:** The mean subfield responses to a noisy step edge, corrupted with 25%, 50%, and 80% additive Gaussian noise for various values of the DOI parameter $\xi$. Responses are averaged over 100 different realizations of the respective noise level; error bars denote $\pm 1$ standard deviation. Responses are normalized to allow for better comparison. *Top left:* The mean response of an optimally oriented subfield. *Bottom left:* The mean response of nonoptimally oriented subfield. *Bottom right:* The minimal $\xi$ for which the mean subfield response of the respective noise level is zero. For the three noise levels, the respective means are given by 1.47, 1.80, 2.01. Results show that for $\xi \approx 2$ nonoptimal responses are almost zero, while the optimal responses are still considerably large.

small noise levels, this value also results in considerably large signal-to-noise ratios for small noise levels. From the evaluation of the responses to homogeneous noise, values of $\xi$ lie in the range $\approx [1.85; 2.25]$, depending on the noise level. Since the signal-to-noise ratios for high noise levels decrease considerably for $\xi > 2$, a value of $\xi = 2$ is chosen for all the simulations of the nonlinear model with DOI presented in Sec. 3.7.

### Response to Small Contrasts

In the previous section, the response properties of DOI for various noise levels and fixed contrast have been evaluated. In the section, the contrast is varied to evaluate the response properties of DOI to small contrasts in the presence of noise.

Recall that the DOI equation can be rewritten as the standard interaction $\pm(K_{\text{on}} - K_{\text{off}})$ minus a space-variant threshold:

$$R_{\text{on},\theta} = \left[(K_{\text{on}} - K_{\text{off}} - (\xi - 1)K_{\text{off}}) * G_\theta\right]^+$$
$$R_{\text{off},\theta} = \left[(K_{\text{off}} - K_{\text{on}} - (\xi - 1)K_{\text{on}}) * G_\theta\right]^+ \quad .$$

This relation has been derived in Eq. 3.12 on page 77. With respect to this thresholding property of DOI it needs to be clarified whether, and if so to which degree, the DOI threshold has a suppressive effect on contrast responses. To address this question the response to small contrast changes in the presence of high level noise is evaluated for three simple cells models, namely linear, nonlinear and nonlinear with DOI.

For the simulations a synthetic test stimulus of alternating on-off and off-on vertical step edges of increasing contrast is employed. The on-off contrasts vary from 0.01 to 0.1 in steps of 0.01 and are centered around a mean luminance level of 0.5. For each on-off contrast, the stimulus contains an individual substimulus of size $128 \times 256$ pixels, resulting in a total stimulus size of $1280 \times 256$ pixels. The stimulus together with a horizontal cross-section is depicted in Fig. 3.32.

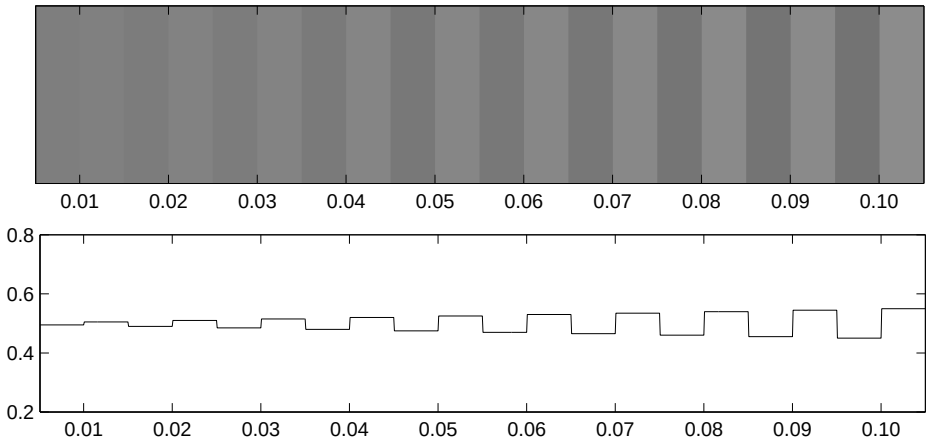


**Fig. 3.32:** *Top row:* Test stimulus for small contrast responses. *Bottom row:* The corresponding horizontal cross-sections. The values at the abscissa denote the contrast of the step edge at this position.

In a pioneering study, the test stimulus is corrupted with Gaussian noise of standard deviation 0.05 which is equivalent to 500% noise for the smallest contrast and 50% for the largest contrast. The input stimulus and the simulation results for the three simple cell models are depicted in Fig. 3.33. Like all simulations in this section, the results are obtained from the complex cell responses which are pooled over all orientations. Most obviously, the result gained by the model with DOI differs from the results gained by the models without DOI in the response to noise. With DOI, responses to noise are absent or at least largely suppressed. At the same time, upon first visual inspection, the model with DOI indeed seems to suppress responses to small contrast. For example, compare the response to the contrasts of 0.04 for the three models: While both the linear and nonlinear model without DOI yield a visible response, the DOI response appears fragmented and faint.

To further analyze this behavior, the sum over each column is computed for the three models (Fig. 3.34). Apart from a scaling factor, the column sum shows the mean of the response along the vertical axis. The plots show that all models yield a vertical mean response above the mean noise level for contrast of about 0.03 and higher, corresponding to noise of 166% or less. This can be seen also in the corresponding vertical profiles taken at a sample contrast location of 0.4 (Fig. 3.35). For all three models, the profile along the contrast edge lies well above a profile taken at the noisy background.

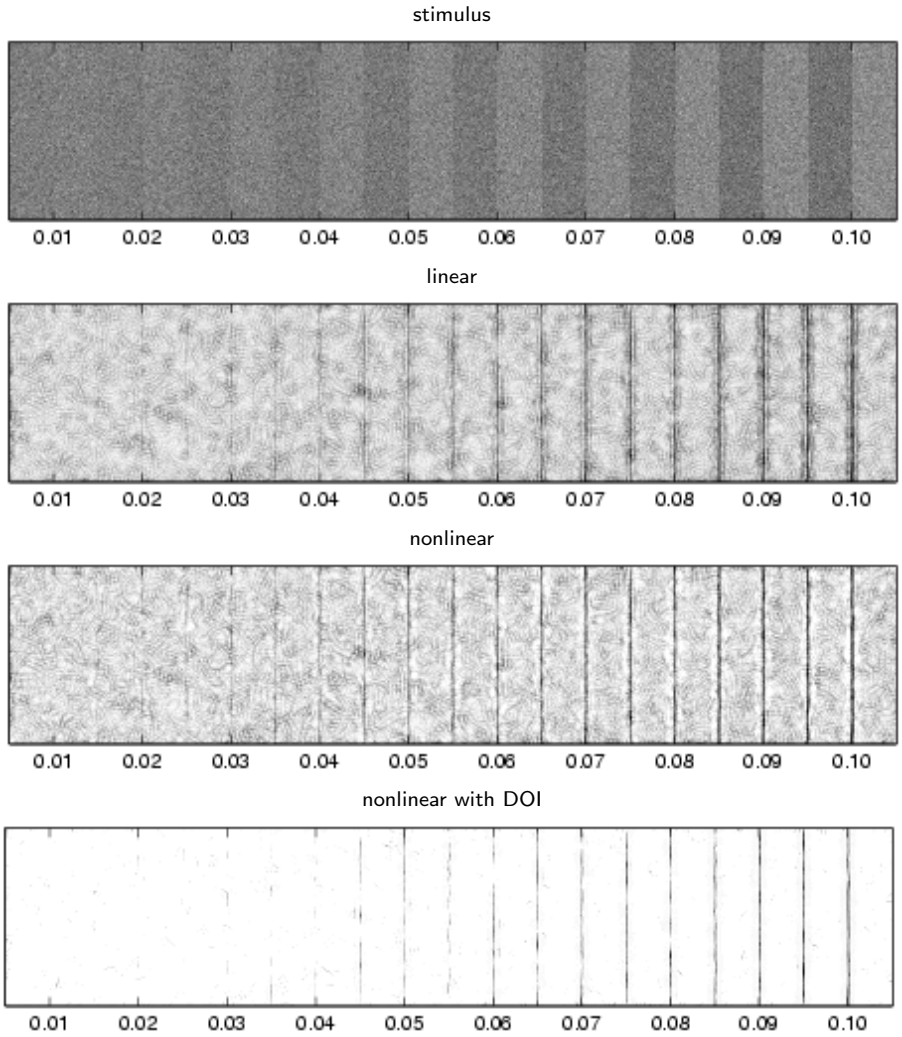


**Fig. 3.33:** Test stimulus corrupted with noise of standard deviation 0.05 and corresponding simulation results.

The reason why these responses are visible for the two models without DOI but seem to fade for the model with DOI are caused by the properties of the graphical representation. For each

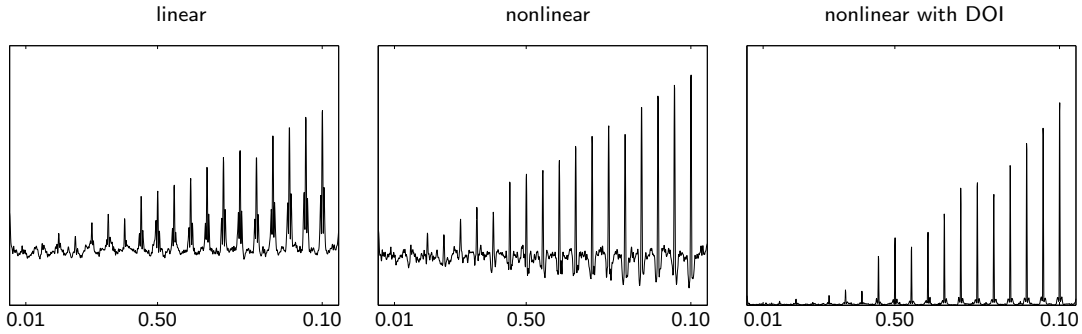


**Fig. 3.34:** Column sum of simulation results shown in Fig. 3.33, using a test stimulus which is corrupted with noise of standard deviation 0.05. The values at the abscissa denote the contrast of the step edge at this position.

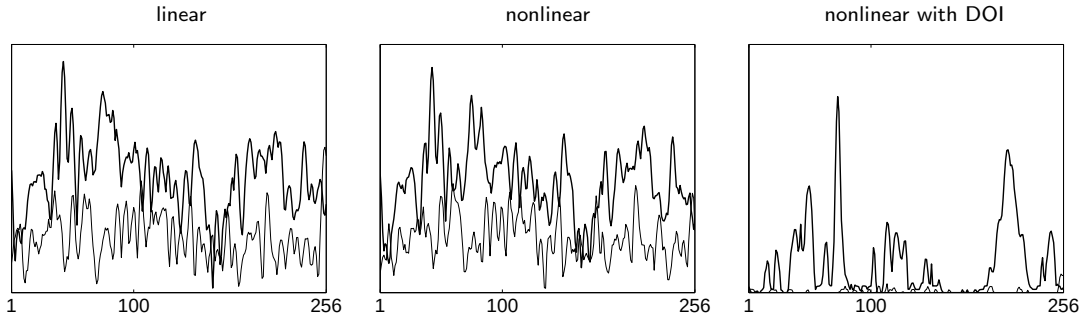


**Fig. 3.35:** Vertical cross-section at 0.04 contrast *(bold, upper curve)* compared to a vertical cross-section of the noisy background *(normal, lower curve)* for the three models.

model the resulting image as shown in Fig. 3.33 is scaled individually, mapping the lowest value to white and the highest to black. For the model without DOI, the responses to the edges add on a significant noise level, causing, e.g., a mean response to 0.4 contrast at about 30% of the maximum response. For DOI, on the contrary, the noise level is virtually zero, and the mean response to 0.4 contrast is only at about 5% of the maximum response. To sum up, the almost vanishing response to noise for the model with DOI renders the response to small contrast less visible, though this response is still present.

The vertical profiles shown in Fig. 3.35 indicate that the responses for all three models at 0.04 contrast and noise of standard deviation 0.05, i.e., 125% noise, differ from a vertical response profile taken at a background location. The question to be addressed in the following is whether these differences are significant. Therefore, the mean responses and standard deviations at each dark-light contrast step (signal) are computed and compared to the mean response and standard deviation at the background (noise). Results for the three simple cells models are shown in Fig. 3.36.

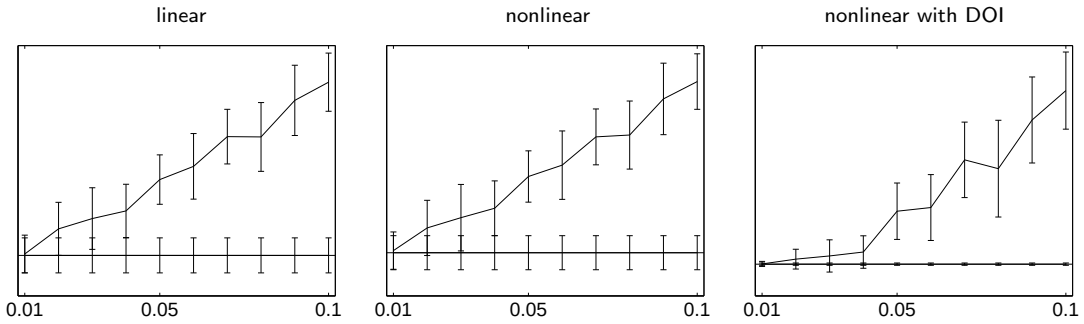


**Fig. 3.36:** Mean response at dark-light contrast edges *(upper line)* compared to mean response at the background *(lower horizontal line)* for the three models. Error bars denote ± standard deviation.

The plots allow to determine the contrast which yields a significant response. Visually, a significant response is to be found at contrast locations where the error bars of signal and noise do not overlap. Formaly, a significant response is defined by a mean response to an edge minus its standard deviation with is larger than the response to the noise background plus the standard deviation of this noise, i.e.,

$$\overline{C}_{\text{edge}} - \text{std}(C_{\text{edge}}) \; > \; \overline{C}_{\text{background}} + \text{std}(C_{\text{background}}) \; .$$

The results show that the amount of contrast which is necessary to yield a significant response is the same for all models, in this case 0.05 contrast.

So far, we have addressed the significant response for one noise level and a single realization of this noise level. In a concluding study on small contrast responses, the considerations of the previous

paragraph are quantitatively evaluated for a great variety of parameter changes. In this study a modified version of the test stimulus is used with dark-light contrasts increasing in steps of 0.005 instead of 0.01 to allow for a finer resolution of the significant contrast. The test stimulus is corrupted with noise of ten different standard deviations ranging from 0.01 to 0.1 in steps of 0.01. For each noise level and each simple cell model, the smallest contrast is determined which yields a significant response. The results are averaged over 100 realizations of the noise level, resulting in a total of $3 \times 10 \times 100$ simulation runs.

The resulting significant contrast response vs. noise level is depicted in Fig. 3.37. The curves show that the model with DOI (dotted line) tends to rely on slightly larger contrasts (maximum 0.01 larger) to yield a significant response compared to the models without DOI (linear, solid line and nonlinear, dashed line). However, this need for larger contrasts is not significant. Further, the models without DOI show almost identical responses. Taken together, the result gained from a single simulation (Fig. 3.37) has been validated for a large variety of parameter settings, namely, that the amount of contrast which is necessary to yield a significant response is the same for all three simple cell models.

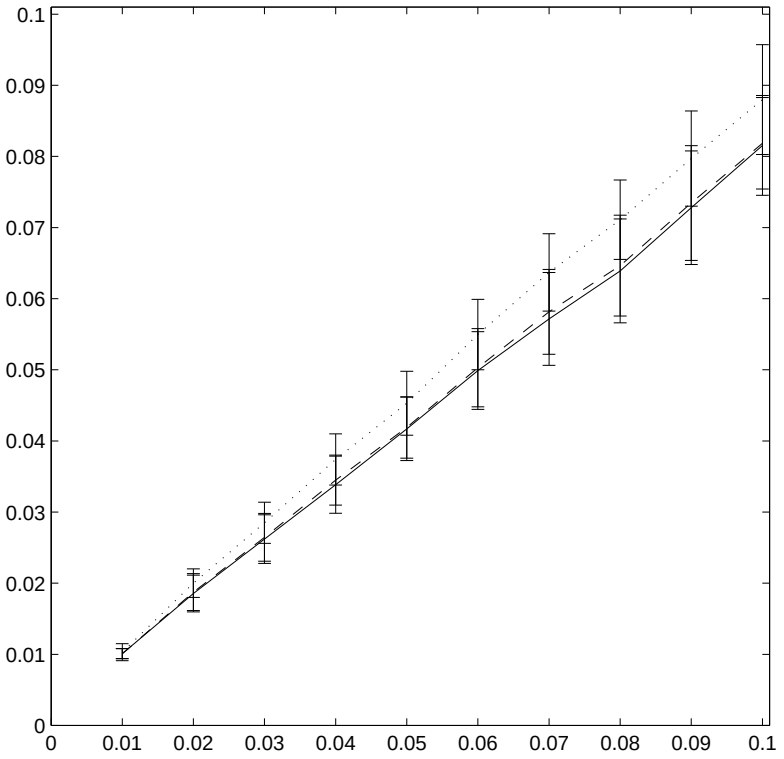


**Fig. 3.37:** Contrast which yields a significant response *(ordinate)* for different noise levels with standard deviation $\sigma_{\text{noise}} \in \{0.01, 0.02, \ldots, 0.1\}$ *(abscissa)* for the three simple cell models linear *(solid)*, nonlinear *(dashed)*, nonlinear with DOI *(dotted)*. Response are averaged over 100 realizations of each noise level. The curves show that the model with DOI tends to rely on larger contrasts to yield a significant response, but this difference to the other models is not significant.

**Response to Background Noise** The simulations with noise of standard deviation 0.05 in Fig. 3.34 show that the response to noisy homogeneous background is virtually zero for the DOI model. Furthermore, this is true also for other noise values of 0.01 and 0.1 as shown in Fig. 3.38. The remarkable property has already been shown to arise at the subfield level (Sec. 3.8.2), and is here shown to hold true also at the level of complex cell responses. This robustness and invariance against noise of different levels is an outstanding property of DOI processing. Both the linear

and the nonlinear model without DOI exhibit nonvanishing responses to noisy background. This response is not constant but depends on the noise level: As larger the noise level, as larger becomes the response to noise. This can be seen in Fig. 3.38 and also Fig. 3.34.

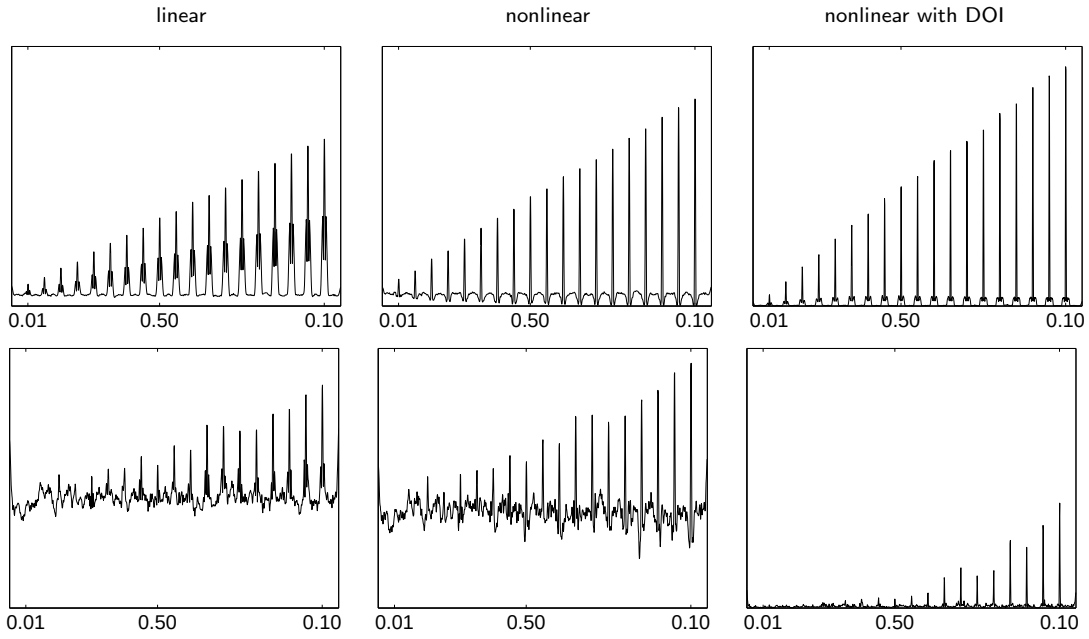


**Fig. 3.38:** Column sum of simulation results for the test stimulus corrupted with noise of standard variation 0.01 *(top row)* and standard variation 0.1 *(bottom row)*. The scale of the ordinates are equal within models and are also equal to the scale used for the simulations with 0.05 noise shown in Fig. 3.34. While for all three models the responses at the contrast locations scale down as the noise increases, the models differ considerably in their response to the noisy background. Both the linear and the nonlinear model without DOI exhibit an increasing response to the noisy background for higher standard deviation of the noise. The nonlinear model with DOI, on the contrary, shows almost zero response to the noisy background irrespectively of the amount of noise.

### Summary of Results Regarding the Response to Small Contrasts

The results of this section can be interpreted in two perspectives. First, with regard to the response properties of DOI to small contrasts it is shown that DOI yields responses which are as significant as the responses without DOI. The assumption that the good suppression of noise for the model with DOI might be accompanied by the undesired feature of suppressing responses to small contrast changes can thus be rejected. Second, on the other hand, the data show that the DOI model does not outperform the other models with respect to contrast sensitivity, i.e., the lower response to noise does not cause a significantly better response at lower contrast compared to the other models. The basic advantage of DOI is thus the suppression of noise irrespectively of the noise level. A non-zero response gained with the model with DOI is a strong indication of a contrast change in the original signal. For the models without DOI, a non-zero response might be caused by a contrast change or, alternatively, by noisy fluctuations within the signal. In this case a threshold might be applied to distinguish signal from noise. The determination of this threshold is not trivial, since the threshold is not constant but depends on the noise level which must somehow be determined from the response signal by a separate mechanism. With DOI, this is achieved implicitly by an elegant mechanism.

## 3.9 Application to Object Recognition

In a further study, we applied the model as feature detector for object recognition. In cooperation with the classification group at our department (Steffen Simon and Hans Kestler), we provided the feature vector, while the actual classification and evaluation of classification performance has been pursued by the classification group.

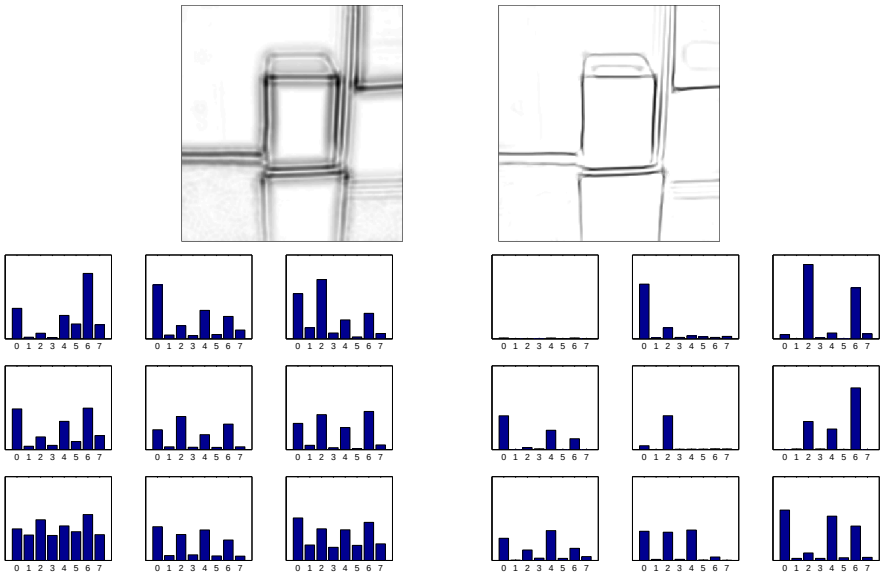

**Fig. 3.39:** Edge images and corresponding orientation histogram for a cube image obtained using the linear simple cell model *(left)* and the nonlinear model with DOI *(right)*.

We used a simple preprocessing scheme based on orientation histograms to compute the characterizing feature vector (Freeman and Roth, 1995). The image is subdivided into $n \times n$ non-overlapping square regions. For each region, an orientation histogram of $O_{\max} = 8$ orientations is computed. The orientation histogram consists of 8 bins of equal width $180°/8 = 22.5°$, each counting the number of respective orientations within the region. The histogram values for each each region are concatenated to form the resulting feature vector with dimension $O_{\max} n^2$. Figure 3.39 show the resulting histograms for the processing of a cube image within a laboratory environment. The histogram resulting from the nonlinear simple cell model with DOI is compared to the linear simple cell model. The histograms show that the edges are more reliably detected using the nonlinear simple cell model.

For the classification task, a hierarchical neural network architecture is employed. This architecture consists of a first level network which provides a coarse classification, followed by a set of networks each recursively refining the discrimination patterns (Kestler et al., 1999).

**Fig. 3.40:** Sample images for the classification task.

The classification performance has been tested on a set of real world images taken from objects in an laboratory environment. The images are partly taken under poor lighting condition and

with distracting shadows. The data base contains 606 images of six objects: bottle (standing and lying), cylinder (standing and lying), bucket and ball. Figure 3.40 shows examples images.

To evaluate the classification performance, three different simple cell models are used for feature detection, namely, linear, nonlinear, and nonlinear with DOI. Table 3.1 shows the mean error of tree cross-validation runs both on the training set and the test. The cross-validation runs differ in the random permutation of the data. Within each $n$-fold cross-validation run, the data is divided into $n = 7$ disjoint sets, and $n$ training runs are performed with $n - 1$ sets. After each run, the classification accuracy is evaluated on the remaining (test) set (Kestler et al., 1999).

**Table 3.1.** Cumulated results of three seven-fold cross-validation runs on the training set and the test set for three different feature dectectors (linear, nonlinear, nonlinear with DOI). The nonlinear model with DOI shows a better performance on both sets than the other models.

| | cross-validation 1 | | cross-validation 2 | | cross-validation 3 | |
|---|---|---|---|---|---|---|
| mean error | training | test | training | test | training | test |
| linear | 12.5 | 27.9 | 13.3 | 27.9 | 12.7 | 26.9 |
| nonlinear | 12.6 | 27.2 | 13.2 | 25.3 | 13.1 | 27.2 |
| nonlinear with DOI | 11.1 | 25.4 | 11.1 | 25.2 | 11.3 | 22.4 |

The results show that the nonlinear model with DOI slightly performes better than the other models. This indicates that the performance of the classification system can benefit from the more robust detection capacity of the proposed model.

## 3.10 Discussion and Conclusion

In this chapter, we propose a simple cell model with dominating opponent inhibition (DOI). DOI is integrated into a push-pull interaction defining the simple cell subfield responses. In push-pull interaction, a subfield (e.g., on) receives both excitatory input from the like domain (i.e., on) and inhibitory input from the opposite domain (i.e., off). DOI proposes a stronger weighting of the opponent input, resulting in a more selective response. The subfields are combined using a nonlinear simple cell model (Neumann et al., 1999).

The proposed model can account for a number of empirical findings. The model reproduces the physiological data of simple cell responses to luminance gradient reversal (Hammond and MacKay, 1983), showing a large decrease of activity if small patches of opposite contrast polarity are added to an optimal bar stimulus. We have further investigated the generation of contrast invariant orientation tuning. Inhibition at either the subfield level or between simple cells of opposite contrast polarity turns out to be sufficient to generate contrast invariant orientation tuning in a feedforward regime. The effect of DOI is to sharpen the tuning curves. The model can also generate responses to modified Glass pattern which are consistent with psychophysical findings. After simulating empirical findings from both physiology and psychophysics, the model is used with identical parameter settings to process noisy synthetic and natural images. The results show that the robustness of the response increases for the model with DOI. The model with DOI has a lower probability of false-positive responses, while the sensitivity to salient edges is preserved. Finally, we conduct a stochastic analysis and detailed numerical simulations to clarify the role of DOI. We particular focus on the strength of the DOI parameter and the response to small contrasts. We determine an optimal parameter value of $\xi = 2$ which is used in all simulations, and show that the DOI mechanims remains sensitive to small contrasts. The numerical evaluation further shows a unique property of DOI, namely the *adaptive* suppression of noise. In adaptive suppression, the amout of suppression scales with the noise level which is signalled in our model by high activity in the opponent path. Postprocessing operations such as thresholding can benefit

from this adaptive suppression property. As an outlook to a sample application we present results of a pioneering study on object recognition.

Dominating inhibition is also used in a detailed physiological model by Troyer et al. to explain contrast invariant orientation tuning of simple cells. In contrast to our nonlinear model, Troyer et al. use linear Gabor filters to model simple cells. Strong "anti-phase" inhibition occurs between Gabor filters of phase shift 180°, i.e., opposite contrast polarity, while we employ inhibition between isotropic on and off responses. With the current physiological knowledge, evidence for both models can be found. In our model, contrast-invariant orientation tuning can be generated to a large extent by balanced inhibition at the subfield level ($\xi = 1$, Eq. 3.5) and by balanced inhibition between simple cells of opposite contrast polarity (Eq. 3.8). DOI sharpens the orientation tuning and slightly increases the contrast invariance, but cannot be ascribed the primary role in generating contrast-invariant orientation tuning, as in the model of Troyer et al. Provided that intracortical recurrent interaction may also play a significant role in generating contrast-invariant orientation tuning (cf. Sec. 3.3.2), we suggest a different functional role of DOI. On the basis of our findings we suggest that the visual system mainly uses dominating opponent inhibition to robustly extract oriented contrast features in noisy environments.

Chapter 4

# Contour Grouping

## 4.1 Introduction and Motivation

The response properties of neurons in the early visual stages are classically characterized by their receptive field (RF) properties (Hubel and Wiesel, 1962). The classical RF is shaped by the feedforward connections of afferent fibers. More recent studies have shown that the response properties of neurons are substantially influenced by stimuli outside their classical RFs (e.g., Gilbert, 1992; Gilbert and Wiesel, 1990; Kapadia et al., 1995; Lamme, 1995). These contextual influences are mediated by lateral horizontal interactions and feedback connections. The precise functional role of both horizontal and feedback connections is unclear. It has been hypothesized that the more global integration of information from distant areas neurophysiologically "highlights" perceptually salient image features (Lamme and Spekreijse, 2000). The contextual influences seem to be related to low- and mid-level task such as contour and feature linking (Allman et al., 1985; Grossberg and Mingolla, 1985b; Li, 1998) or preattentive texture segmentation (Li, 1999b) as well as higher level processes like perceptual organization, attention, and visual awareness (Lamme and Spekreijse, 2000).

In this chapter, we focus on the role of recurrent long-range processing for the enhancement of contours. The robust and reliable extraction of contours is an important task in early visual processing. However, initial contrast measurements which define the first processing stage in the computation of contour signals are often noisy and fragmented. Therefore, the salient or prominent contours have to be determined out of an array of noisy, cluttered contrast responses.

How can this task be accomplished? We suggest a computational framework involving *long-range connections*, *feedback*, and *recurrent interactions*. The task of contour extraction cannot be solved solely on the basis on the incoming data alone, but requests for additional constraints and assumptions on the shape of frequently occurring contours. An important principle of salient contours is colinearity, as expressed in the Gestalt law of good continuation. It has been suggested that horizontal *long-range connections* found in the superficial layers of early visual areas like V1 and V2 provide a neural implementation of the law of good continuation (Schmidt et al., 1997). The assumptions or a priori information such as expressed in the law of good continuation have to be carefully matched against the incoming data. We suggest that *feedback* plays a central role in this matching process by selectively enhancing those feedforward input signal which are consistent with the assumptions. The interaction between feedforward data and feedback assumptions requires certain time steps. In each step the result of the interactions is recursively fed into the same matching process. Such a process of *recurrent interaction* might be used by the brain to determine the most stable and consistent representation depending on both the assumptions and the given input data.

Motivated by empirical findings we present a model of *recurrent long-range interaction in the primary visual cortex* for contour processing.

The chapter is organized as follows. In Sec. 4.2 we shall review empirical findings regarding lateral long-range connections and recurrent processing in early vision. Next, in Sec. 4.3, we give an

overview of other models of contour integration that have been proposed. We shall then detail in Sec. 4.4 our model of recurrent interaction for contour enhancement. Simulations and an evaluation of the model's competencies are given in Sec. 4.5. In Sec. 4.6, a variant of the model using early feedback is introduced and its specific capabilities are demonstrated. Section 4.7 concludes the chapter.

## 4.2 Empirical Findings

The present section is subdivided into two parts: In Sec. 4.2.1 we give a review of empirical findings regarding lateral processing and horizontal integration in early vision. Next, in Sec. 4.2.1, a framework for describing properties of recurrent processing is suggested prior to a brief summary of empirical data.—A review covering both lateral and recurrent elements in visual processing is given by Lamme et al. (1998).

### 4.2.1 Lateral Long-Range Processing

In this section we review empirical evidence for lateral or horizontal long-range processing in early vision. Evidence comes from a large variety of methods, ranging from anatomy via physiology to psychophysics. Often, different methods are combined (such as anatomical cell staining techniques and optical imaging) or parallel studies are conducted (such as physiological recordings in monkey and psychophysical experiments in man using the same or similar stimuli) to account for a more complete characterization of the properties of long-range processing, linking dendritic morphology and axonal projections to functional properties of individual cells (Bosking et al., 1997; Gilbert, 1993) or to identify possible neural correlates of higher level perceptual or cognitive processes such as contour saliency, segmentation or binding (Kapadia et al., 1995, 2000).

In physiology, lateral long-range interactions are intimately linked to the notion of the "nonclassical" RF of a neuron (Maffei and Fiorentini, 1976) which has been introduced to characterize modulatory facilitatory influences on the response properties of a neuron from outside its classical RF. A number of functional roles have been proposed regarding long-range interactions and nonclassical RF mechanisms, such as colinear contour integration (Grossberg and Mingolla, 1985a,b; Li, 1998; Neumann and Sepp, 1999; Schmidt et al., 1997), perceptual grouping (Grossberg et al., 1997; Grossberg and Raizada, 2000; Ross et al., 2000), illusory contour generation (Heitger et al., 1998), contrast gain control (Heeger, 1992), perceptual filling-in (Gilbert, 1992; Gilbert et al., 1996), and sparse coding and response decorrelation (Vinje and Gallant, 2000).

The present review focus on lateral long-range interaction in early visual areas and also briefly covers the characteristics of short-range lateral connections in the superficial layers 2/3 of V1. While we focus on vision, it should be noted that long-range connections are found in all areas in the cortical hierarchy and seems to be a general principle of cortical organization (Amir et al., 1993). For an introductory description of horizontal connections the reader is referred to Mason and Kandel (1991). More detailed reviews can be found in a series of papers by Charles Gilbert and coworkers (Gilbert, 1992, 1993, 1998; Gilbert et al., 1996) as well as Miikkulainen and Sirosh (1996), Kovács (1996), Polat (1999), and Fitzpatrick (2000).

We shall begin by reviewing anatomical data which has been provided first evidence for the existence of horizontal long-range connections.

#### Anatomy

Cell labeling techniques, such as retrograde labeling using dye injection of HRP have provided the prime example for horizontal long-range connections (Gilbert and Wiesel, 1979, 1983, 1989; Martin

and Whitteridge, 1984; Rockland and Lund, 1982, 1983). The studies revealed a characteristic pattern of axonal terminals of layer 2/3 pyramidal cell. Two type of *short range* and *long-range connections* can be distinguished (Fig. 4.1).

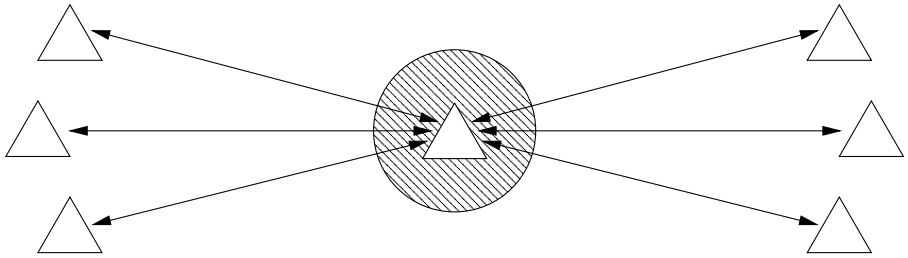

**Fig. 4.1:** Schematic view of lateral short- and long-range interactions. Pyramidal cells *(triangles)* are reciprocally linked by long-range connections *(double-arrow lines)* to pyramidal cells with coaligned, cooriented RFs, and form unspecific short-range connections *(hatched circle)*.

Short-range or local connections are confined to a distance of about 500 $\mu$m from the cell center and form a dense, isotropic plexus around the cell. Comparing the location of presynaptic terminal boutons of these short-range connections with the orientation preference map showed that short-range connections are rather unspecific for orientation (Amir et al., 1993; Bosking et al., 1997; Malach et al., 1993). Most likely, the short range connections belong to an inhibitory system mediated by smooth-dendritic GABAergic interneurons. Inhibitory connections are known to extend for shorter distances, are uniformly distributed and rather unspecific for orientation (Albus and Wahle, 1994; Albus et al., 1991; Kisvarday and Eysel, 1993; Kisvarday et al., 1994; Matsubara and Boyd, 1992). There is one kind of inhibitory cell known as basket cell with more long-range axonal spreading up to 1.5 mm (Kisvarday and Eysel, 1992), but the connections are also rather unspecific for a particular orientation (Kisvarday et al., 1994).

Long-range connections, on the other hand, span large distances of 6–8 mm and branch off in a regularly spaced pattern of clustered axon collaterals separated by 0.5–1 mm. These clustered terminals of long-range connections tend to be anisotropic, i.e., form distributions that are elongated along a particular axis. This anisotropy may partly reflect the corresponding anisotropy in the visual map (Amir et al., 1993; Malach et al., 1993; McGuire et al., 1991) but cannot account for the full degree of anisotropy (Gilbert and Wiesel, 1983, 1989; Kisvarday and Eysel, 1992). Long-range connections between cells are mostly reciprocal (Kisvarday and Eysel, 1992). Thus, the target cells of output fibers (as indicated by the locations of presynaptic axonal terminal boutons) most likely provide input to cells.

The extend and clustering of horizontal long-range connections can be compared to the size and distance of cortical hypercolumns (Gilbert et al., 1996). The distance of two hypercolumns is roughly 1.5 mm, which is the minimal distance of cells with nonoverlapping RFs. Thus, long-range connections can integrate response from cell which are several RF diameters apart. In the tree shrew, e.g., long-range connections extend up to 4 mm (corresponding to 20°), compared to a classical RF size at this eccentricity of only 5° (Bosking et al., 1997). Further, the distance between clustered axon collaterals in the range of 0.5–1 mm is approximately the width of an individual hypercolumn. The similarity of distances suggests a close relationship between the spreading of horizontal connections and the pattern formed by orientation columns.

The functional characteristics of these relationships can be revealed by anatomical methods. In principle, both the orientation columns and the axonal long-range arborization have to be determined, overlayed and related to the representation of the visual map. Difficulties arise from the requirement of precise alignment and the distorted representation of the visual space in the maps of most mammals. By using an approach as sketched above, the pattern of horizontal connections has been revealed by retrograde labeling with extracellularly applied tracers, and the orientation columns have been visualized by autoradiography or optical imaging. Studies both in cat (Gilbert

and Wiesel, 1989) and monkey (Malach et al., 1993) have shown that horizontal connections link patches of the same orientation preference.

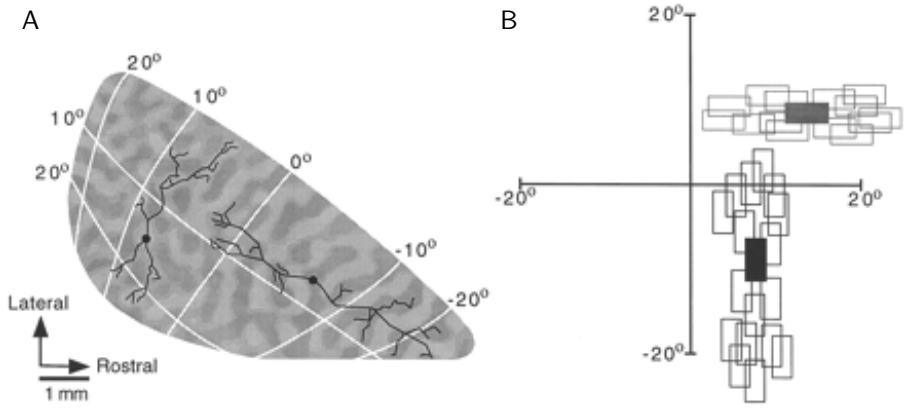


**Fig. 4.2:** Specificity of horizontal connections of layer 2/3 cells in V1 of tree shrew. *(A)* Overlay of axonal long-range arborization of two cells *(black)* over a map of orientation preference *(gray)* partitioned by a grid of the visual space *(white lines)*. The upper border (labeled from 20° to −20°) represents the vertical meridian, the left border (labeled from 10° to 20° represents the horizontal meridian). *Light gray* denotes region with 0° orientation preference, *dark gray* denotes regions with 90° orientation preference.—The *left neuron* has an orientation preference of 0° and lies in a light gray regions. The axonal terminals lie in regions of the same orientation preference, i.e., light gray regions, and lie along a horizontal line in visual space. The *right neuron* has an orientation preference of 90° and makes axonal connections with cells of 90° preference lying along a vertically oriented line. *(B)* Corresponding RFs of both cells *(dark solid rectangles)* and RFs of cells that provide input to the cells *(open rectangles)*. (From Bosking et al., 1997.)

Based on computer simulations, it has been argued that the simple rule "like connects to like" cannot explain the patchy connections and should be accompanied by a second rule which further restricts the connections between like-oriented cells to those cells with coaxial aligned RFs (Mitchison and Crick, 1982). Exactly this pattern of connections has been confirmed by anatomical studies in squirrel monkey (Sincich and Blasdel, 1995), tree shrew (Bosking et al., 1997; Fitzpatrick, 1996) and cat (Schmidt et al., 1997). An example for the high degree of specification of the lateral connections summarizing the results of Bosking et al. (1997) is given in Fig. 4.2.

### Physiology

The anatomical findings on long-range connections have their physiological counterpart in the notion of excitatory influences from outside the classical RF. Before reviewing some studies regarding these influences, we give a brief overview of the terminology used and point to some studies regarding inhibitory influences.

**Classical and nonclassical RFs** Generally speaking, the RF of a cell is the area in the visual space where stimuli can influence the response property of the cell. These influences can be either excitatory or inhibitory, and involve both driving or generating effects near the RF center as well as modulatory surround effects. Though the terminology is somewhat lousy, the RF of a cell can tentatively be divided into the three zones: An excitatory center, an inhibitory surrounding region (both defining the "classical" RF), and more distant fields having modulatory, facilitatory influences (the "nonclassical" RF).

The RF center can be measured using simple bar stimuli. The size of the RF center is given as the visual space where an increase in stimuli length causes an increased response.

**Inhibitory influences within the classical RF** The existence of inhibitory surrounding regions both orthogonal (side-inhibition) and along (end-inhibition or end-stopped) the preferred orientation of the cell is well known since the early studies of Hubel and Wiesel (1962) and have been confirmed in a number of more recent ones (Bishop et al., 1973; Gilbert, 1977; Li and Li, 1994; Maffei and Fiorentini, 1976). Inhibitory influences expressed in the notion of so-called end-stopped cells have been implicated in processes such as curvature detection (Dobbins et al., 1987), illusory contour generation (von der Heydt and Peterhans, 1989), texture segmentation (Knierim and Van Essen, 1992) and perceptual pop-out (Lamme, 1995; Zipser et al., 1996).

**Influences from outside the classical RF** The functional properties of horizontal long-range connections have been examined in a number of studies using cross-correlation analysis of cells spaced within the distance of long-range axon collaterals. In cross-correlation analysis, the spike pattern of two cells are recorded and a histogram of their differences in spike time is constructed. Cells which are connected or receive a common input generate a peak in the histogram (Gilbert, 1993). Cross-correlation analysis revealed that cells separated by about 1 mm, which is the distance of the axonal collaterals, fire simultaneously in response to stimuli with the same orientation. This suggests that horizontal long-range connections primarily link cells with the same orientation preference (Ts'o and Gilbert, 1988; Ts'o et al., 1986).

The relation of excitatory, long-range influences in V1 to contour integration and grouping have been revealed more recently in a number of studies. These studies show that the response to an optimal stimulus in the classical RF can be greatly enhanced by the presentation of coaligned, co-orientated bars in a larger neighborhood (Kapadia et al., 1995; Nelson and Frost, 1985; Polat et al., 1998).

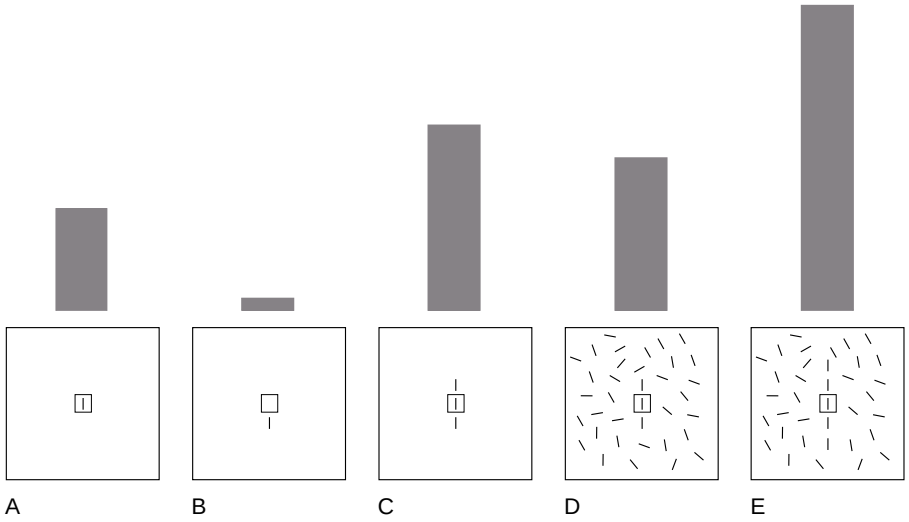


**Fig. 4.3:** Contextual influences of bar stimuli. *Bottom, large squares:* Stimulus configurations used; *small central squares* indicate the extend of the classical RF. *Top, gray bars:* Sketch of the corresponding activation. *(A)* Optimal activation of the RF by a single bar stimulus within the classical RF evoke a moderate response. *(B)* The same bar stimulus outside the classical RF evoke no or only a weak response. *(C)* Presenting the flanking bars together with a bar within the classical RF cause an increase of response. *(D)* Randomly oriented bars have a suppressive effect. *(E)* Further adding coaligned, cooriented bars results in an increase of response. (Adapted from Kapadia et al., 1995.)

Kapadia et al. (1995) have studied the contextual influences of bar stimuli placed outside the RF in the superficial layers of V1 in alert monkey. The results are summarized in Fig. 4.3. The response to the central bar can be enhanced by coaxially aligned bars placed outside the classical RF, and grows with the number of properly aligned bars. The bars alone evoke no or only a weak response. An orthogonal line placed between the flanking lines eliminates the facilitatory effects.

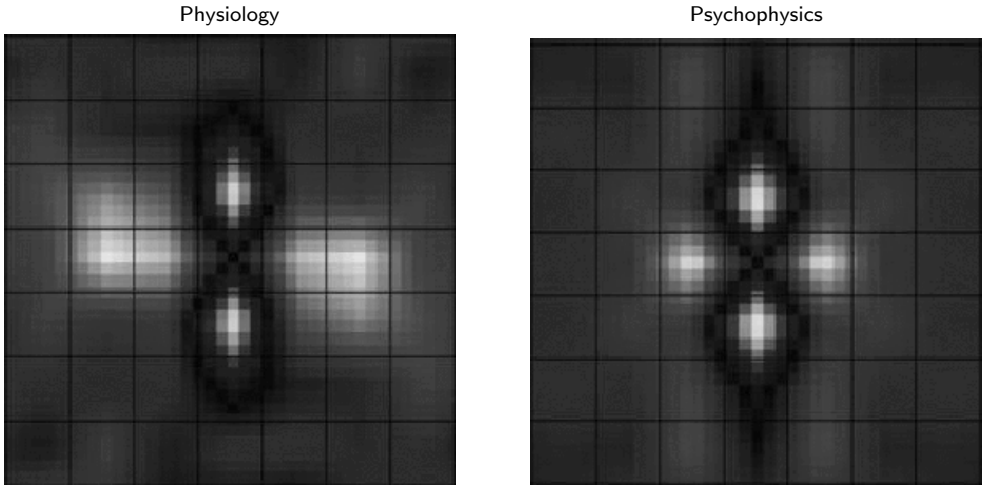


**Fig. 4.4:** Comparison of physiologically and psychophysically obtained maps of long-range interactions. Interactions are shown for a vertical orientation presented in the central square. Light bipole-shaped areas above and below the central square indicate excitatory interactions, light areas to the left and right indicate inhibitory interactions. Note the striking similarity between both patterns. (Adapted from Kapadia et al., 2000.)

Randomly oriented bars in the neighborhood have a suppressive effect. Rotating a few of this randomly oriented bars toward colinear orientation eliminates the inhibition and even caused in some cells an increase in response compared to the central bar stimulus alone.

In a more recent study, this results have been confirmed and extended (Kapadia et al., 2000). In this studies in alert monkey, the flanking stimuli were systematically varied across position, and detailed maps indicating positions of facilitatory and suppressive responses were obtained. The stimuli consist of two mirror-symmetric flanking bars of the same orientation as the central bar. Reproducing the findings of the earlier study by Kapadia et al. (1995), it was shown that colinear flanks caused a more than three-fold increase in response, thought the flanks alone elicited little response. Further, the side-by-side placement of the flanks revealed significant nonlinearities: The two flanking stimuli alone caused an increasing response, while the response in conjunction with the central bar decreased. From the response to the whole stimulus set, 2D maps of the RF and the surround effects have been constructed. A so-called "context map" was created by subtracting the response to the central bar alone from the response evoked by the three-bar condition. In the context map, facilitatory regions are located along the colinear axis, while suppressive regions are located at the sides, i.e., with an offset orthogonal to the preferred orientation of the cell (Fig. 4.4, left).

In another set of experiments, the effect of *contrast* on the contextual influences have been studied. First, the contrast of both the central target and the flanking bars was varied simultaneously. In this case, strong colinear facilitation and weak lateral inhibition was found at low contrast, and the opposite effect on the interaction strength at hight contrast. Second, the contrast of flank and target were varied independently. For a low contrast central bar, the position of maximal facilitation shifts toward the periphery. Evidence for the potential limitation of colinear facilitation to low contrast stimuli is found in rat visual cortex, where long-range excitation is evoked only by weak stimulation (Shao and Burkhalter, 1996; Shao et al., 1996).

In addition to the effect of contrast, the facilitatory influence of contextual lines is also modulated by attention (Ito and Gilbert, 1999). The influence of attention on the processing of visual information in early visual areas has been recently reviewed in a number of papers (Lamme and Spekreijse, 2000; Posner and Gilbert, 1999; Treue, 2001).

Long-range horizontal connections have also been implicated to play a role in context-dependent synchronization or binding of cells with nonoverlapping RFs (Engel et al., 1990, 1991; Gray et al., 1989; Livingstone, 1996).

Taken together, colinear facilitation from beyond the classical RF turns out to be a characteristic feature of processing in V1. Further, the surround can be partitioned into bipole-shaped excitatory and inhibitory regions. The facilitatory effects depends not only on the spatial arrangement of contextual lines, but is further modulated by contrast and attention. The physiological findings are paralleled by results of psychophysical studies in human observers, as shall be reviewed below.

### Psychophysics

In psychology, the effects of context on the perception of objects have a long tradition. In particular, the work of the Gestalt school of perception (Max Wertheimer and his students Wolfgang Köhler and Kurt Koffka) at the beginning of the 20th century stressed the relevance of whole arrangement or form of individual items (the "Gestalt") for the perception of a scene (Koffka, 1935; Köhler, 1929; Wertheimer, 1938). The main idea is reflected in the central tenet stating that "the whole is different from the sum of its parts".

The Gestalt psychologists devised a set of criteria or rules, the "Gestalt principles", to describe how the grouping of individual items leads to the emergence of stable figures to be perceived. Important elements that influence the grouping process are proximity, similarity, good continuation, and closure (Fig. 4.5).

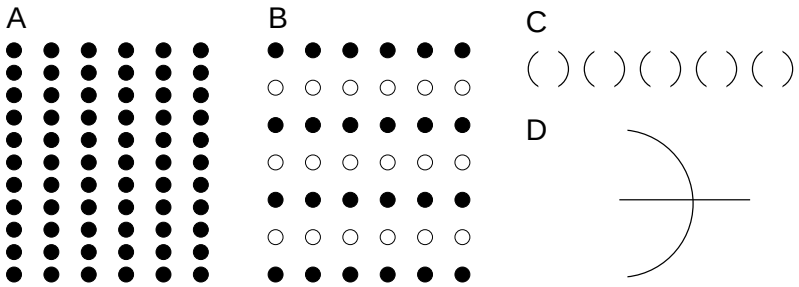


**Fig. 4.5:** Gestalt principles of perceptual organization. Grouping by *(A)* proximity, *(B)* similarity, *(C)* closure, *(D)* good continuation. (Partly adapted from Wertheimer, 1923 and Rock and Palmer, 1990.)

Results of more recent studies on contextual "long-range" effects reflect basic ideas which can be tracted back to the school of Gestalt perception. The results obtained with these psychophysical studies are also closely related to physiological results, suggesting that the neural substrate of perceptual grouping can be found in the early stages of visual processing.

In accordance with the Gestalt principle of good continuation and in agreement with comparable physiological studies, it has been shown that the contrast detection of a target stimulus is significantly improved by the presence of colinear, superthreshold context stimuli (Dresp, 1993; Kapadia et al., 1995, 2000; Polat and Sagi, 1993, 1994). Similar results have been reported for illusory contours, were the colinear arrangement of inducers facilitated the response of a target line. Interestingly, facilitation occurs independent of the contrast polarity of the inducers and the target line (Dresp and Bonnet, 1995).

The stimuli used in the contrast detection experiments usually consists of a central target given by a low contrast Gabor signal, and two high-contrast flanking masks. The dependence of detection threshold on spatial distance, orientation and spatial frequency between target and mask has been studied by Polat and Sagi (1993). Varying the distance between target and mask reveals a small suppressive region and a five-times larger facilitatory region. The spatial extend of the regions depends on the spatial frequency of the Gabor elements: As the frequency decrease, the

spatial distance increases. A decrease of both suppressive and facilitatory effects was observed if either the orientation or the spatial frequency of the flankers deviate from the target. In both cases the decrease was more strongly for the facilitatory effect. Overall, the results revealed a characteristic pattern of lateral interaction consisting of a small region of rather unspecific suppression surrounded by a much larger, more selective facilitatory region. A similar structure of antagonistic center-surround interactions has also been found for the detection of orientation pop-up at texture singularities (Sagi, 1990)

In a psychophysical study complementary to their physiological experiments (see above), Kapadia et al. (2000) have obtained a map of contextual interactions with separated regions of excitatory (attractive) and inhibitory (repulsive) influences based on the tilt illusion. In the experiment, the effect of two mirror-symmetric, tilted flanks on the perceived orientation of a vertical target line was monitored. Depending on the position of the flanks, regions of attractive influences were found in regions colinear to the target line, while repulsive influences were located orthogonal to the target line (Fig. 4.4, right).

Another paradigm used to study the spatio-orientational mechanisms underlying perceptual grouping in contour detection is based on so-called "path-finder" displays (Beck, 1983). These displays consists of contour elements embedded in an array of randomly oriented background elements. The individual elements or "micropatterns" (Dakin and Hess, 1998) are the same for the contour and the background, and a contour is defined only by the relative alignment of neighboring elements. The individual elements are most often patches of Gabor functions, which are thought to selectively activate the elements involved in the process of contour grouping (Field et al., 1993). Similar displays consisting of lines as individual elements have been used for the study of texture segregation (Beck et al., 1989; Nothdurft, 1991). In the experiments, two pathfinder displays are successively shown, and the subject has to decide in a two-alternative forced choice procedure (2AFC) which display contains the target contour.

Field et al. (1993) have used this paradigm to investigate the different rules that govern contour grouping. Consistent with the studies of contrast detection, they found that the path could be more reliably identified if the individual elements are coaligned or can be connected by a smooth path without any inflection points (cf. Kellman and Shipley, 1991). Detection performance was not impaired by increasing the distance between the individual elements. The distance in visual space was larger than the distance covered by an individual orientation column, thus suggesting the involvement of long-range connections between cells in different orientation columns. Substantial impairment in the detection task occurred if the orientation of the elements deviated from the orientation of the path or if the elements were oriented orthogonal to the path. The results are summarized in the notion of an so-called "association field" which integrates information of neighboring elements of similar orientation.

Dakin and Hess (1998) have studied the spatial-frequency tuning of contour integration, using a variante of the path-finder display described above. In this variante, two Gabor elements of different spatial frequency were employed for both the target contour and the background. Dakin and Hess (1998) showed that colinear paths can be detected with high reliability even if the individual elements differ in spatial frequency up to 1.3 octaves. However, if the elements defining the contour deviate from colinearity, the integration is narrowed to spatial-frequency differences of 0.7 octaves. In other words, the tolerable difference of spatial frequency between individual elements to be integrated into a single contour decreases with increased curvature of the contour.

The pattern of connections defined by the association field conforms to the Gestalt law of good continuation. The role of other Gestalt principles (such as closure or common fate) for contour detection have also been investigated systematically. Closed paths could be more reliably detected for larger distances between the individual elements than open paths (Kovács and Julesz, 1993, 1994; Pettet et al., 1998), but sharp corners result in impaired detectability (Pettet et al., 1998). The results have been qualitatively simulated in a single layer feedback network with mutually facilitatory interactions depending on distance, curvature and change in curvature (Pettet et al.,

1998). More recently, the effect of common fate have been shown to synergistically interact with good continuation in the detection of contours against a cluttered background (Lee and Blake, 2001).

### Statistics

Complementary to the empirical results concerning the "inner" mechanisms for contour processing and integration, the "outer" structure of frequently occurring contour patterns in the natural world can be statistically evaluated. This approach has been pursued in two recent studies (Geisler et al., 2001; Sigman et al., 2001). Both studies employ the same hierarchical processing scheme with two stages, but differ in the particular methods used at each stage. The basic scheme consists of a first stage, where the images are preprocessed to extract the local edge orientation at every image point. This operation is followed by thresholding in order to reduce the number of edges to be considered. At the next stage, the joint statistics for the cooccurrence of two edges are evaluated. The statistics is evaluated for a fixed orientation of the reference element in two ways, namely (i) to find the most probable orientation at every position, and (ii) to find the most probable position for every orientation difference. Remarkably, the results obtained in both studies are the same, though the both studies differ in the particular methods and the set of evaluated images. Both studies found that (i) iso-oriented, or parallel edges are the edges of the most probable orientation at every spatial position, with the highest probability for colinear edges, and (ii) for every orientation difference, the most probable position is predicted by the cocirculary constraint, and a bipole-shaped co-occurrence plot emerged in the study of Geisler et al. (2001) (Fig. 4.6). The studies show that distribution of edges in natural scenes are consistent with the Gestalt principles of cocircularity, and colinearity. Geisler et al. (2001) further performed a Bayesian analysis using conditional probabilites to determine the likelihood ratio that a given pair of edges belong to the same vs. different physical contour. The likelihood function also shows that edges of smooth, cocircular orientation most likely belong to the same physical contour.

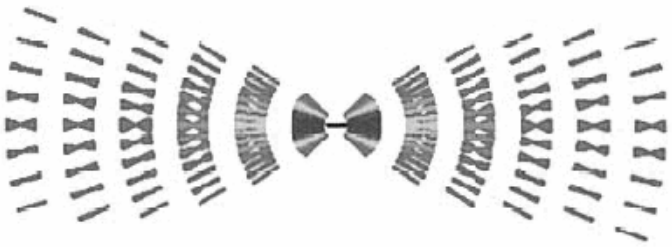

**Fig. 4.6:** Co-occurrence plot of the most frequently occurring edge directions (for each possible distance and orientation difference) for a horizontal reference edge. (From Geisler et al., 2001.)

### Summary of Lateral Long-Range Interactions

Based on a number of studies from different disciplines their is now ample evidence that Gestalt criteria of similarity, proximity, and smoothness, which describe the contextual interactions involved in, e.g., contour grouping, find their neural substrate in the long-range connections mediated by axon collaterals of layer 2/3 pyramidal cells in V1. These long-range connections (i) span large distances in the order of several classical RF diameters, and (ii) preferentially link neurons with colinear oriented RFs (i.e., neurons of the same orientation preference with an RF located along a colinear axis in the map of the visual space). Long-range connections between layer 2/3 cells are reciprocal, defining a single layer recurrent network of excitatory connected cells.

From a functional and computational viewpoint it is quite clear that such a system of excitatory coupled units needs to be paralleled by a second dual system of *inhibitory* connections to prevent

an increase of activity up to saturation or uncontrolled oscillations (cf. Crick and Koch, 1998). This inhibitory system is defined by short-range connections among layer 2/3 cells. Unlike long-range connections, the short-range connections have no strong orientation preference and provide unspecific input to neighboring cells.

The strength of the contextual interactions does not only depend on the spatio-orientational configuration of the individual items, but also on contrast. The facilitatory effect of properly aligned contour fragments on a target item is strongest at low contrasts. This finding is paralleled by recent studies, showing that the size of the classical RF, as determined by conventional measurements, is not fixed, but depends on stimulus contrast. An up to 4-fold increase of RF size ("classical", noteworthy), has been reported at low contrast (Kapadia et al., 1999; Sceniak et al., 1999).

The results suggests a novel view of neurons at the early visual stages such as V1: Rather than representing a bank of static filters, V1 neurons exhibit an adaptive, dynamic change of response properties and RF size depending on stimulus conditions such as contrast and context, and are thought to be involved in higher level processes such as contour grouping and perceptual saliency (cf. Lee et al., 1998).

### 4.2.2 Recurrent Processing

Recurrent processing is ubiquitous in the brain. A thorough review of empirical data concerning recurrent processing would be beyond the scope of this section. Instead, we focus on basic findings regarding recurrent interactions in early vision. Prior to this review, we shall define the terminology used to describe and distinguish intrinsic elements of recurrent interactions.

#### Elements of Recurrent Interaction

Recurrent circuits are defined by a number of local connections between units. To characterize recurrent interactions, it is thus useful to have a closer look at the basic properties of connections between individual neurons.

One basic property refers to the *polarity* of the connection, i.e., whether an activation of the connection has a depolarizing (or excitatory) or hyperpolarizing (or inhibitory) effect on the target cell. The polarity of a connection is the underlying physiological property to distinguish between cooperative and competitive interactions.

Besides the quantitative property of polarity, the connection between two neurons is characterized by a quantitative *efficacy* on the activation of the target neuron, given by the strength (or weight) of the connection. With respect to recurrent processing, a coarse but useful distinction between weak (or modulating) and strong (or driving) connections has been suggested by Crick and Koch (1998). A driving connection provides input that can generate a response in the target neuron, while modulating connections cannot cause a response, rather than a change of activity which is already present.

Another property is motivated by the organization of the cortex in a *hierarchy* of levels. Based on the level within the hierarchy, three different kinds of inputs can be distinguished: *(i) feedforward* or bottom-up input from neurons of a previous level, *(ii) lateral* or horizontal input from neurons of the same level, and *(iii) feedback* or top-down input from neurons of a higher level (e.g., Salin and Bullier, 1995). The different kinds of input are summarized in Fig. 4.7.

A necessary prerequisite for the distinction of feedforward, feedback and lateral connections is the proper determination of a hierarchy. Different methods can be used to determine the hierarchical level, based on anatomical or physiological methods. Anatomically, one may define the hierarchy by the largest number (or "longest path") of synaptic connections to an sensory receptor. Since such a measurement is difficult if not almost impossible for individual cortical neurons, anatomical

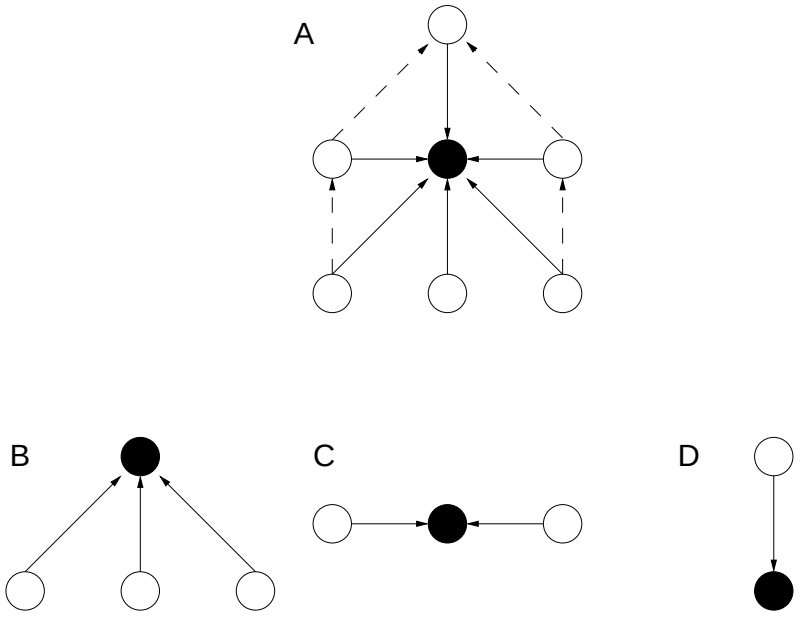


**Fig. 4.7:** *(A)* Different kinds of input *(solid lines)* to a target neuron *(filled black circle)* can be decomposed into *(B)* feedforward, *(C)* lateral, and *(D)* feedback input. The *dashed lines* in A are added to indicate a hierarchy of levels necessary for the decomposition into different types of input. (Adapted from Neumann and Mingolla, 2001.)

criteria based on the cortical layer of input and target units have been suggested to distinguish between feedforward, feedback, and lateral connections between two cortical areas (Rockland and Pandya, 1979). According to this scheme, feedforward connections terminate mainly in layer 4 and originate preferentially in the supragranular layer 2/3. Feedback connections, on the other hand, terminate mainly in layer 1 and layer 6 and originate preferentially in the infragranular layers 5 and 6. Lateral connections are more unspecific. The construction of a hierarchy of cortical areas by Felleman and Van Essen (1991) is based on these rules. Physiologically, the hierarchical level can be determined by a temporal criteria given by the response latency of a specific area. See Lamme and Roelfsema (2000) for an analysis of latencies as measured in different studies and a discussion of the (non-)correspondence to an anatomically defined hierarchy.

Besides the hierarchical organization, the cortex has a specific *structure* of layers, columns, and areas. According to this structural organization, three different types of connections can be distinguished: *(i) intralaminar* connections running within a layer, *(ii) interlaminar* connections which run between different layers and which are often confined to a single column, and *(iii) interareal* connections between different areas.

So far we have characterized the connections between two individual units. In the following, we shall consider the circuit properties that characterize recurrent interactions involving a number of units.

A *recurrent* interaction between occurs if a unit $A$ provides input to another unit $B$, and in turn receives input from unit $B$. If one describes the network by means of a directed graph (also called digraph), recurrent interactions define a cycle in the digraph.

Recurrent circuits can be characterized by the *number of units* that are involved in the interaction. A coarse distinction can be made between circuits of two units, which are bidirectionally (or reciprocally) linked, and circuits involving more than two units.

Similar to the structural distinction made for the connections between individual units, a recurrent circuit is characterized by the smallest structure to which the individual connections of the recurrent circuit are confined, i.e., *intralaminar*, *interlaminar*, or *interareal*.

We are know able to establish a framework which allows to describe elements of recurrent interactions by summarizing the above considerations. Such a framework may prove useful in disentangling different kinds of recurrent interactions as a prerequisite for a functional characterization of these interactions.

We suggest a differentiation of recurrent interactions according to five properties, subdivided into properties of the whole circuit and properties of individual connections.

*Circuit Properties*

1. *Structure*

   Motivated by the organization of cortical architecture into (hierarchical) *areas*, horizontal *layers* and vertical *columns*, three different structural types of recurrent connections can be distinguished:

   (a) *interareal*

   Interareal connections occur between different areas of the brain and can be further subdivided into *cortico-cortical* connections between different cortical areas such as V1 and V2, and *cortico-thalamic* connections between the cortex and the thalamus.

   (b) *interlaminar*

   Interlaminar connections occur between different layers within a particular area of the brain.

   (c) *intralaminar*

   Intralaminar connections occur between neurons *within* a particular layer, such as layer 2/3.

2. *Topology*

   With respect to the topology of connections, recurrent interactions can be either *direct* or *indirect*. *Direct connections* occur if neuron A provides input to B and vice versa. Direct connections are also termed reciprocal or bidirectional. *Indirect connections* occur if either or both the input from neuron A to B or from neuron B to A are mediated by a number of intermediate neurons.—The indirect recurrent interaction in the laminae of V1 has been termed "folded feedback" in the models by Grossberg and coworkers (Grossberg and Raizada, 2000; Ross et al., 2000).

*Connection Properties*

3. *Polarity*

   Connections can be either excitatory or inhibitory.

4. *Efficacy*

   Connections can provide either *driving* input which by itself is sufficient to generate a response in the target neuron, or *modulating* input which by itself cannot generate a response rather than has either a facilitatory or suppressive effect on the activity which is already present.

5. *Hierarchy*

   Connections can originate and terminate in levels at the same or at a different different position within the cortical hierarchy. Three different kinds of inputs can be distinguished:

*(i) feedforward* input from neurons of a previous level, *(ii) lateral* or horizontal input from neurons of the same level, and *(iii) feedback* input from neurons of a higher level.

According to this scheme, recurrent interactions can be characterized with respect to *structure*, *topology*, *polarity*, *efficacy*, and *hierarchy*. For example, the reciprocal horizontal long-range connections in V1 layer 2/3 as reviewed in Sec. 4.2.1 define recurrent interactions which are intralaminar, both direct and indirect, excitatory, presumably modulating, and lateral.

The differentiation according to structure seems to be of particular importance. Examples for each kind of structural interactions in the domain of early vision are *(i) interareal:* the recurrent cortico-thalamic interactions between LGN and V1 and the cortico-cortical interactions between V1 and V2, *(ii) interlaminar:* the "folded" feedback circuit within V1 from layer 2/3 via 5 and 6 and back via layer 4 to layer 2/3 and *(iii) intralaminar:* the reciprocal interactions between layer 2/3 pyramidal cell mediated by long-range connections. These prototypic interactions are sketched in Fig. 4.8.

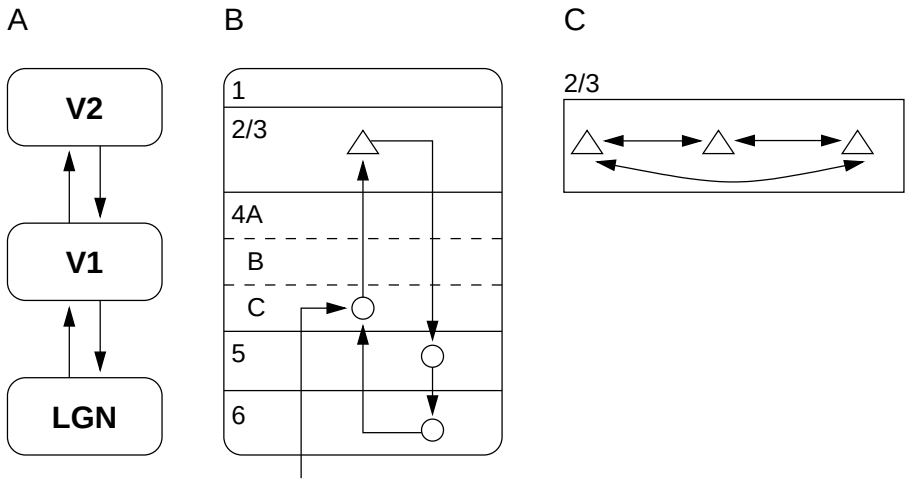


**Fig. 4.8:** Different structural types of recurrent connections. *(A)* Interareal recurrent connections between LGN and V1 (corticothalamic) and V1 and V2 (cortico-cortical). *(B)* Interlaminar recurrent connections. *(C)* Intralaminar recurrent interactions between V1 layer 2/3 pyramidal cells.

### Recurrent Interactions in Early Vision

Having outlined a framework for describing properties of recurrent interactions, we shall now briefly review empirical findings in the domain of early vision. As suggested by the considerations above, the data is presented according to the structure where the recurrent interaction takes place, i.e., interareal, interlaminar and intralaminar. Since the model presented in Sec. 4.4 below does not include interareal connections, we focus and inter- and intralaminar recurrent interactions. Findings regarding interareal interactions are only briefly covered for completeness.

Recurrent interactions mediated by top-down feedback have been suggested to be involved in a number of phenomena, such as synchronization of the firing of thalamic relay cells (Sillito et al., 1994), generation of end-stopped properties (Bolz et al., 1989; Rao and Ballard, 1991), perceptual pop-out (Kastner et al., 1997), figure-ground segregation (Hupé et al., 1998; Lamme et al., 1999), or attention (Treue, 2001).

**Interareal Recurrent Interactions** Recurrent interactions between areas are usually direct, i.e., an area which receives feedforward projections in turn makes feedback connection to the same area. This pattern of bidirectionally linked areas is a general principle of brain organization

(Felleman and Van Essen, 1991; Salin and Bullier, 1995). An example for this principle is given by the pattern of cortico-thalamic connections depicted in Fig. 4.9.

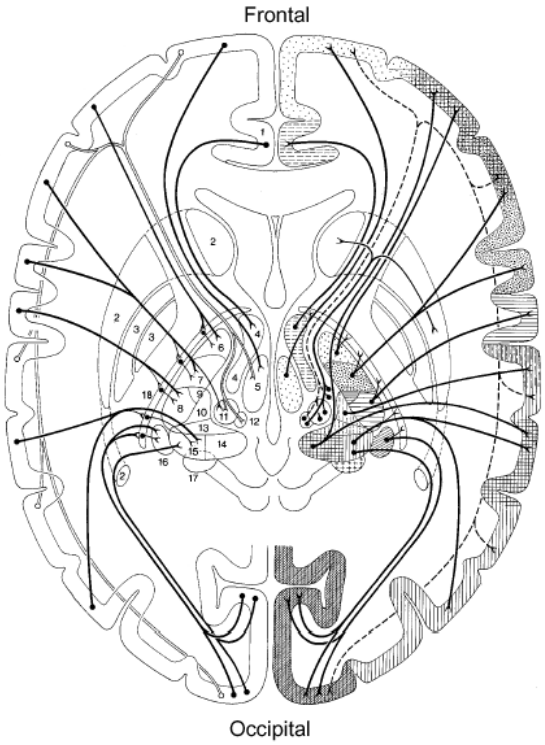


**Fig. 4.9:** Horizontal section of the human thalamocortical system. The almost perfect axial-symmetric arrangement of thalamocortical projections *(right hemisphere)* and corticothalamic projections *(left hemisphere)* exemplify a major principle of the wiring circuitry in the central nervous system, namely that most areas are bidirectionally linked by feedforward and feedback connections. *4–18* Thalamic nuclei, *16* LGN. (Adapted from Nieuwenhuys et al., 1988.)

With respect to cortico-cortical connections, we focus on the reciprocal connections between V1 and V2. The V1–V2 connections are largely guided by a point-to-point scheme. The pattern of feedforward projections preferably link patches of similar feature preference, as shown for orientation selective cells in V1 and V2 (Gilbert and Wiesel, 1989). The pattern of feedback projections show a precise retinotopic correspondence (Bullier et al., 1988). Such a retinotopic linkage has been suggested for the connections of cells in cytochrome oxidase blobs and bands (Livingstone and Hubel, 1984). However, the feedback connections diverge from V2 to multiple clusters in V1, which may reflect the convergence of information flow within V2 (Rockland and Virga, 1989). With respect to the polarity of V1–V2 feedback connections it has been shown that the connections have a on-center off-surround organization: topologically precise connections synapse on excitatory cells, while widespread connections preferentially synapse with inhibitory interneurons (Bullier et al., 1996; Salin and Bullier, 1995).

Burkhalter and coworkers have studied the corticortical feedback circuitry between area V1 and LM (secondary visual area lateromedial) in rat. The rodent visual cortex can serve as a model system for studying the visual system, since the basic principles of organization are the same as in other mammmals, such as in cat and monkey. The studies show that the feedback connections from LM to V1 (i) preferentially contact pyramidal cells in V1 and avoid GABAergic interneurons (Johnson and Burkhalter, 1996), (ii) produce monosynaptic excitatory responses (Shao and Burkhalter, 1996), and (iii) provide strong input to forward projecting neurons (Johnson and Burkhalter, 1997). The connections thus form direct loops of recurrent excitatory circuits.

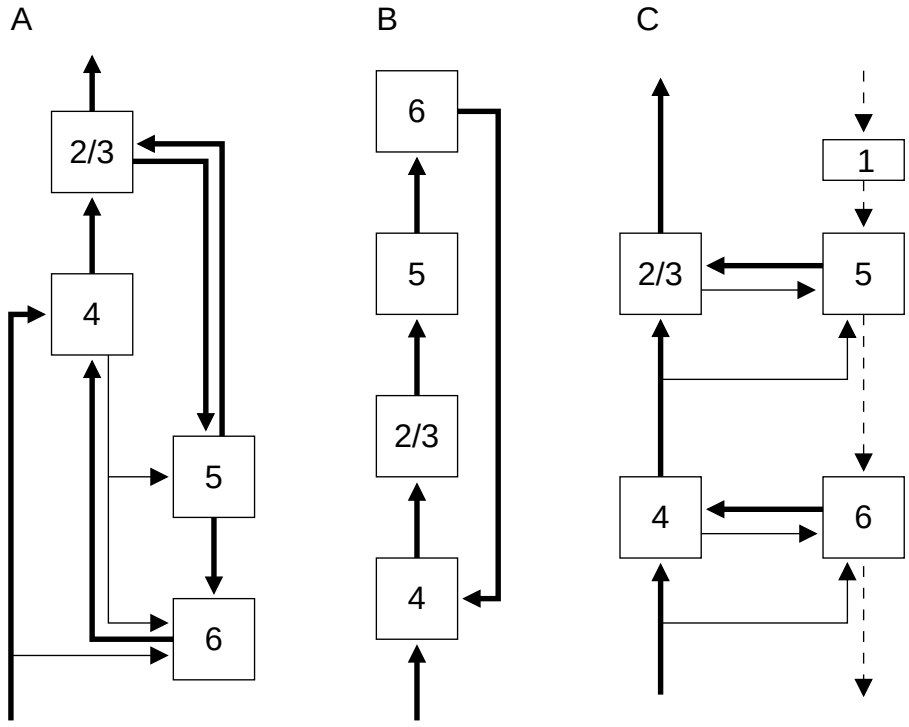


**Fig. 4.10:** Models of intralaminar circuitry in V1. *(A)* Sketch of the basic interlaminar circuitry. In this diagram, unlike in diagrams B and C, the position of the layers from top to bottom is chosen according to their ordering in the cortex. *(B)* Recurrent loop proposed by Bolz et al. (1989). *(C)* Two-level model originally proposed by Callaway (1998) and extended by an additional feedback path *(dashed arrows)*. *Thick lines* indicate strong connections, *thin lines* indicate weak connections. For simplicity, the numbering convention of the layers is the same in the three diagrams. A more precise labeling would change 4 to 4C and 2/3 to 2–4B in diagram C.

**Interlaminar Recurrent Interactions** The interlaminar connections within V1 is governed by a general, stereotyped layout (Fig. 4.8, B): The input from the LGN terminates in the granular layer 4 (sublamina C), layer 4 projects to the supragranular layer 2/3, layer 2/3 project to the infragranular layer 5. Layer 5 projects to layer 6, which in turn projects back to layer 4, thus closing the recurrent loop (Gilbert and Wiesel, 1979).

A possible functional role of this major recurrent loop has been proposed based on pharmalocical studies (Bolz and Gilbert, 1986; Bolz et al., 1989). In these studies, layer 6 has been reversely inactivated by GABA injection. During inactivation, layer 4 cells lost their end-inhibition property, which was recovered after layer 6 become active again. This effect could be generated by layer 6 cells with large RFs by providing surround inhibition to layer 4 cells, mediated by inhibitory interneurons. The model is supported by evidence from EM serial reconstruction (McGuire et al., 1984), showing that the majority of layer 6 target cells in layer 4 are smooth stellate cells, i.e., inhibitory interneurons.

The wiring of the intralaminar circuit described so far has been found in cat. A recent review of the local circuits in monkey is given by Callaway (1998). In monkey the laminar organization share the same basic properties. However, the dense projection from layer 5 to 6, which is present in cat, seems to be missing in monkey (Callaway and Wiser, 1996). Instead, layer 5 project mainly to layers 2–4B (analougous to layer 2/3 in cat), and only occasionally branch off into layer 6. Input from layer 5 to 6 can nevertheless be provided by class II neurons in layer 6, which have many dendritic branches in layer 5.

Making several simplifications of this circuitry (such as considering only excitatory primary output of a layer), a two-level model of local circuitry has been suggested by Callaway (1998). The model is depicted in Fig. 4.10, C. In this model, the layers are grouped into feedforward and feedback layers or "modules". Layer 4C (analogous to layer 4 in cat) and layer 2–4B are thought to be feedforward, while the deep infragranular layers 5 and 6 are feedback layers. Feedforward and feedback layers come in pairs of associated layers (namely 4C and 6, 2–4B and 5). The connections between the two paired layers have the same characteristic structure: The feedforward layer receive strong excitatory input from the next lower layer and in turn provides strong input to the next layer. The feedback layer receive weak input from two sources, the next lower layer and the feedforward layer, and sends feedback connection to the feedforward layer. Thus, the feedforward stream relays information directly to the next layer, while the feedback layers combine input from and to its associated feedforward layer to modulate the output of this feedforward layer.

The two-level model can be critizised since the feedback layers 5 and 6 receive no input from a higher layer, but from the previous layer and the associated feedforward layer only. We suggest to extend the two-level model by an additional main feedback path from V2 via layer 1, layer 5 and layer 6 back to the LGN in analogy to the main feedforward path. Such a main feedback path would provide additional information from higher layers or areas. The suggested feedback path is shown in dashed lines in Fig. 4.10, C.

The basic intralaminar connections and the two different interpretations as a recurrent loop (Bolz and Gilbert, 1986; Bolz et al., 1989) and a two-level model of bidirectionally connected layers are depicted in Fig. 4.10. Clearly, both layouts are simplifications of the actual circuitry within V1. The question to which degree the most important functional properties are captured by such simplifications remains an issue of further research.

**Intralaminar Recurrent Interaction** Recurrent intralaminar connections in V1 occur in layer 2/3 and are constituted by long-range connections between pyramidal cells with cooriented, coaligned RFs. See Sec. 4.2.1 for further details.

Cells with long lateral axonal arbors have also been found in layer 5 both in cat (Gilbert and Wiesel, 1979; Martin and Whitteridge, 1984) and monkey (Callaway and Wiser, 1996). Based on inactivation experiments similar to the studies of end-inhibition as described above it has been hypothezised that long-range connections in layer 5 allow for the construction of large RFs in layer 6 (Bolz and Gilbert, 1989; Bolz et al., 1989).

**Efficacy: Feedback is modulatory.** Besides the structural distinction of different levels of recurrent interaction, another important determinant is the *efficacy* of the connection. Several physiological studies indicate that feedback projections have a gating or modulating rather than generating effect on cell activities (Hirsch and Gilbert, 1991; Hupé et al., 1998; Salin and Bullier, 1995). Feedback alone is not sufficient to drive cell responses (Hirsch and Gilbert, 1991; Sandell and Schiller, 1982). In fact, the modulatory efficacy seem to be a hallmark of feedback connections (Callaway, 1998), which can thus aid to distinguish between feedforward and feedback connections.

### 4.2.3 Summary of Empirical Findings on Lateral and Recurrent Interactions

The input to a neuron is determined by feedforward, lateral and recurrent connections. Different functional roles may be subscribed to each type of these connections. Feedforward connections provide the information about the sensory data present in the visual stimulus. Lateral connections integrate activities from a larger part of the visual field in a modality specific way by forming the neural substrate of a priori assumptions or expectations about the nature of "good" sensory data.

Feedback connections from higher cortical areas modulate the responses based on higher level and more abstract, but still data-oriented representations, or convey the influence of cognitive factors such as attention. Local recurrent interactions within an area may serve to integrate all these inputs and steer the neuron's response to the best achievable compromise with respect to the available input.

While this description probably captures essential properties of cortical processing in agreement with recent empirical findings, one should be aware that it is still rather tentative and at best incomplete. The highly interconnected and recurrent nature of cortical wiring makes it hard to entangle the different influences. A number of neurons in, say, V1 may be viewed as providing the feedforward sensory input to V2, but at the same time the output of this neurons is itself involved in a plethora of recurrent interactions within lateral connections, between layers, and between other cortical and thalamic areas, all of which are ignored in the rather simplistic view of these neurons as providing feedforward sensory-driven input alone. With these limitations in mind, the assignment of different functional roles to different kinds of connections is a useful concept in a number of further steps to be taken aiming at unraveling the nature of the highly connected recurrent interactions in the brain.

## 4.3 Review of Contour Models

Between the visual stages of low-level feature extraction and high-level scene interpretation there is—among other potential processing stages—an intermediate stage of *grouping*, where elementary feature items are integrated or grouped into more meaningful chunks.

In this section we give an overview over the rich literature on computational approaches to perceptual grouping for contour integration. The approaches can be subdivided into two classes: computational *models* and computational *algorithms*. Computational models are motivated by empirical data and address the question how contour integration is realized in biological vision systems. Computational algorithms, on the other hand, are pure computer vision approaches where neither the internal computational scheme nor the class of problems these algorithms are applied to are explicitly motivated by empirical data. The proper class for a particular approach is most often naturally suggested by the main focus of the computational approach. However, there is some amount of mutual cross-fertilization between computer vision algorithms and biologically motivated models, so that the two classes cannot be strictly separated.

The core of this section is devoted to computational models (Sec. 4.3.2), since the model presented in this work in Sec. 4.4 belongs to this class. Computational algorithms are more marginally covered in Sec. 4.3.3. Prior to the review of computational approaches, we shall describe in Sec. 4.3.1 the computational elements underlying contour integration, and delineate an overall framework or classification scheme which allows to characterize and classify different computational approaches.

### 4.3.1 Elements of Contour Integration

Grouping mechanisms underlying contour integration are defined by selective interactions between local contour elements (guided by principles such as colinearity, cocircularity, smoothness, or proximity) in a three-dimensional space of position and angle. In the following we shall first describe the geometrical layout of the grouping interactions in this 3D spatio-angular feature space and contrast it to other common interaction schemes in early vision, such as center-surround processing. Further, the parameters defining the geometric relationship between contour elements and the notion of the bipole are introduced. We shall then describe the main processing stages of any computational scheme of contour integration and the basic mechanisms that are used to implement the grouping process. These considerations are summarized in a suggested framework

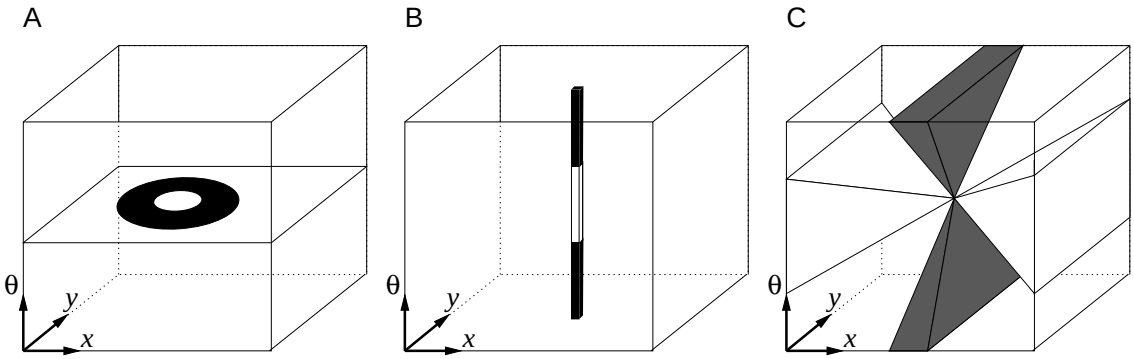


**Fig. 4.11:** Schematic diagram of the geometric layout of different spatio-angular interaction schemes. *(A)* Center surround-interaction in the spatial domain $\mathcal{S} = \langle x, y \rangle$, e.g., in retinal ganglion cells. *(B)* Center-surround interactions in the orientational domain $\mathcal{S} = \langle \theta \rangle$, e.g., within a cortical hypercolumn. *(C)* Spatio-angular interactions $\mathcal{S} = \langle x, y, \theta \rangle$, between different hypercolumns as mediated, e.g., by horizontal connections between layer 2/3 pyramidal cells. *White areas* indicate zones of potentially excitatory influences, *dark areas* indicate zones of potentially inhibitory influences.

for describing computational approaches to contour grouping. The considerations presented in this section are influenced and inspired by recent work of Neumann and Mingolla (2001).

### The Geometry of Spatio-Angular Interactions

**The Feature Space of Lateral Interactions** Contour integration involves interactions between elements in a three-dimensional space of position and orientation. Contour integration is thus a special case of local interactions between units or neurons in a high-dimensional feature space. The specific properties of contour integration may be elucidated by comparison to other patterns of local interactions. The first major processing stage in the primary visual pathway are center-surround interactions of retinal ganglion cells. In the present framework, these interactions take place in a 2D retinotopic map of spatial positions, $\mathcal{S} = \langle x, y \rangle$. Another principal scheme of interaction in early vision involve intracolumnar orientational interactions, i.e., within a cortical hypercolumn (e.g., Ben-Yishai et al., 1995; Somers et al., 1995). In this case, the feature space is given by the 1D space of orientations $\mathcal{S} = \langle \theta \rangle$ at a particular position. Finally, the interactions underlying contour integration involve intercolumnar interactions between cortical cells assemblies of adjacent or neighboring hypercolumns. The corresponding 3D feature space is given by position and angle, $\mathcal{S} = \langle x, y, \theta \rangle$.

To get an idea of the geometric layout of interactions underlying contour integration, the basic layout of these three schemes of interactions are visualized in the 3D cube of the spatio-angular feature space (Fig. 4.11).

**The Geometrical Relationship of Edge Elements** Above we have considered the feature space of contour integration as compared to other schemes of lateral interactions. In this paragraph we shall introduce the basic elements of contour integration and their geometrical relationship.

Rotated edge elements (or edgels, Guy and Medioni, 1996) are the basic elements underlying contour integration. An edge element is defined by a spatial quantity which defines its position or location $(x, y)$ and an angular quantity $\theta$ defining the slant of the edge element. The angular quantity can be formulated as a *direction* which has unique values for the full range $\theta \in [0; 2\pi]$, or as *orientation*, where the values are limited to the range $\theta \in [0; \pi]$ and suffices $\theta = \theta + \pi$. The difference between direction and orientation is illustrated in Fig. 4.12.

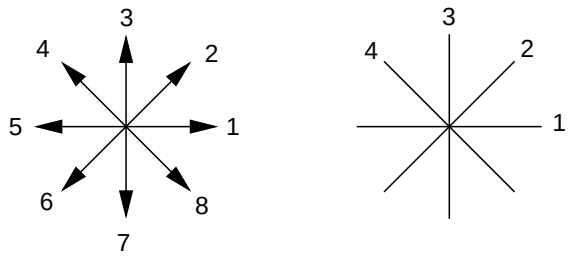


**Fig. 4.12:** Diagram illustrating the difference between direction and orientation. *Left:* Eight different directions $0°, 45°, \ldots, 315°$ are defined by the star-like pattern. *Right:* The same pattern defines four different orientations $0°, 45°, \ldots, 135°$ which are limited to the range $[0°; 180°]$.

Overall, an edge elements defines a point $(x, y, \theta)$ in the 3D spatio-angular feature space $\mathcal{S} = \langle x, y, \theta \rangle$. Contour integration then involves the mutual influence of two edge elements $\mathbf{e}_0 = (x_0, y_0, \theta_0)$ and $\mathbf{e}_1 = (x_1, y_1, \theta_1)$ within this feature space. The interaction between two elements can be described by a grouping function $f(\mathbf{e}_0, \mathbf{e}_1)$. The grouping function most often obeys two general properties:

1. *Euclidean Invariance*

   The grouping strength between two elements is invariant against translation and rotation of the whole ensemble of the two elements. Note that the *relative* position and orientation of the two elements remains the same under these transformations.

2. *Symmetry*

   For *oriented edges*, the grouping function is symmetric: the support an edge $\mathbf{e}_1$ receives from another edge $\mathbf{e}_2$ is the same as the support which $\mathbf{e}_2$ receives from $\mathbf{e}_1$, i.e., $f(\mathbf{e}_1, \mathbf{e}_2) = f(\mathbf{e}_2, \mathbf{e}_1)$.

   For *directed edges*, the grouping function in general does not exhibit a symmetry like for oriented edges, but shows another kind of symmetry. Let $\mathbf{e}'$ be the directed edge at the same location as an edge $\mathbf{e}$, but with flipped orientation $\theta' = \theta + \pi$. Then the support a directed edge $\mathbf{e}_1$ receives from another edge $\mathbf{e}_2$ is the same as the support which the flipped edge $\mathbf{e}'_2$ receives from $\mathbf{e}'_1$, i.e., $f(\mathbf{e}_1, \mathbf{e}_2) = f(\mathbf{e}'_2, \mathbf{e}'_1)$. This kind of symmetry has been termed "time-reversal symmetry" (Williams and Thornber, 1999) or simply "reversal symmetry" (Mahamud et al., 1999). See Fig. 4.13 for an illustration.


**Fig. 4.13:** "Time-reversal symmetry" between directed edges. The support for a reference edge $\mathbf{e}_1$ by a neighboring edge $\mathbf{e}_2$ *(left)* is the same as the support which the flipped edge $\mathbf{e}'_2$ receives from the neighboring edge $\mathbf{e}'_1$ *(right)*. *Solid black arrows* denote the respective target edge, *open arrows* denote the neighboring edge.

Because of the invariance against ensemble translation and rotation, the strength of interactions does not depend on the absolute values of position and angle, rather than on the relative values. Consequently, the interaction can be described by a function of only three parameters $f(\Delta x, \Delta y, \Delta\theta)$ with respect to a reference element. Since the spatial interactions are most often expressed in terms of distance and direction, it is more convenient to describe the spatial difference in polar coordinates of radius $r$ and azimuth $\varphi$. The geometric relationship between a pair of edge elements is thus described by a function of three parameters $f(r, \varphi, \theta)$, leaving the delta $\Delta$ for the benefit of a more concise notation. The geometric relationship between two edge

elements is depicted in Fig. 4.14. Here and in the following we will assume a gauge coordinate system where the axis of the target element defines the reference orientation.

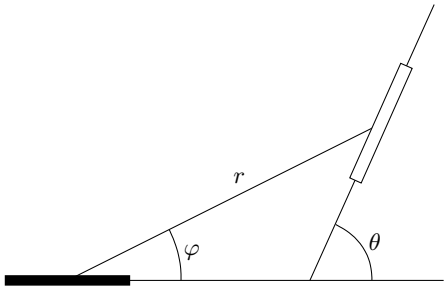


**Fig. 4.14:** Basic geometrical relations between a pair of edge elements, given by a reference element *(solid black rectangle)* and a neighboring element *(open rectangle)*. The relationship is defined by three parameters: the distance between the two elements $r$, the angular position $\varphi$ of the neighboring element with respect to the orientation of the reference element, and the difference $\theta$ in orientation between the two elements.

In some computational schemes, the grouping function can be subdivided into two components, given by a pure spatial function of location differences $f_{\text{spatial}}(r, \varphi)$ and a function of angular interactions between features $f_{\text{feature}}(\varphi, \theta)$. Further, the spatial function is often polar separable into an angular and a radial component, i.e., $f_{\text{spatial}}(r, \varphi) = f_{\text{radial}}(r) f_{\text{angular}}(\varphi)$.

The interaction between features is mostly specified based on elementary connection patterns of parallelity, radiality or cocircularity. Formaly, the three elementary patterns can be defined by the predicate function

$$f_{\text{feature}}(\varphi, \theta) = \begin{cases} 1 & \text{if the specific constraint is fulfilled} \\ 0 & \text{else} \end{cases},$$

with the specific constraint for the different connection patterns given by

| | |
|---|---|
| $\theta = 0$ | parallelism |
| $\theta = \varphi$ | radiality |
| $\theta = 2\varphi$ | cocircularity . |

Colinearity is the only pattern to be found in all three basic connection patterns and is given by the constraint

| | |
|---|---|
| $\theta = 0$ and $\varphi = 0$ | colinearity . |

The three basic connection patterns are depicted in Fig. 4.15.

In a concrete model implementations, a binary-valued predicated function defining the feature cooperation is replaced by a continuous-valued function. This function takes maximal values at the optimal orientation as defined by the particular basic connection pattern used, and smoothly decays with deviation from the optimal orientation.

**The Bipole: Describing the Spatial Extent of Interactions** In general, contour integration does not occur between elements of arbitrarily large distances or angular differences. Instead, the spatio-angular integration is confined to elements within a region of moderate distance in the feature space with respect to a reference element. The projection of this three-dimensional region to the $(x, y)$-plane has a characteristic shape of two symmetric lobes. We will refer to this shape confining the local region of spatio-angular integration as "bipole" (Grossberg and Mingolla, 1987).

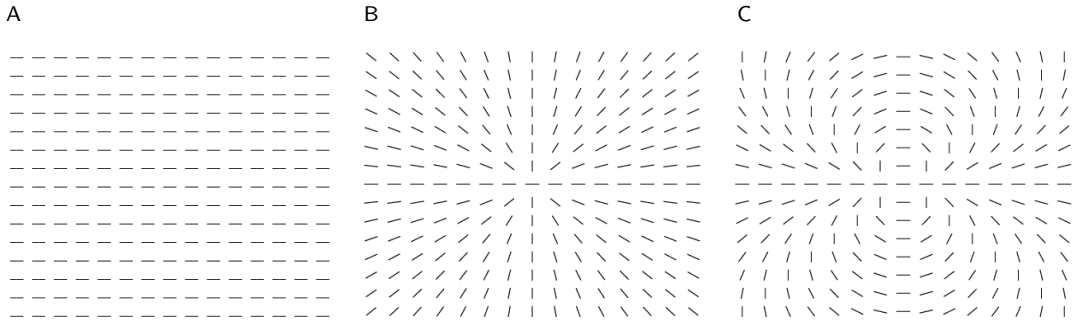


**Fig. 4.15:** Basic connection patterns of *(A)* parallelism, *(B)* radiality, and *(C)* cocircularity for a horizontal reference orientation on a discrete $15 \times 15$ grid.

The grouping function is generally chosen so that the bipole is symmetric. If specified for oriented edges, the bipole then has two axes of symmetry, namely the main axis defined by the orientation of the reference element and the orthogonal axis passing through the center of the reference element. If directed edges are considered, the bipole may reduce to a single lobe which is symmetric with respect to the main axis (cf. Williams and Thornber, 1999, Fig. 5).

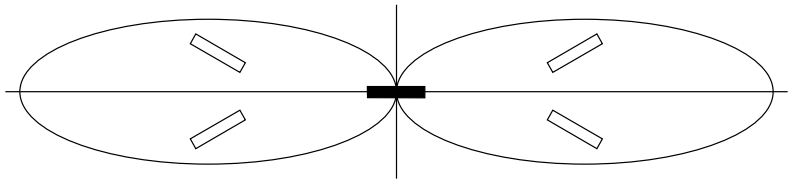

**Fig. 4.16:** The bipole icon *(two ellipses)* for a horizontal reference element *(solid black central rectangle)* for oriented edges. The bipole has two axes of symmetry *(straight horizontal and vertical lines)*. The *open rectangles* indicate neighboring elements which provide identical support due to the symmetrical shape of the bipole.

The symmetry of the bipole does not necessarily follow from the symmetry of the grouping function. If the geometric relation between two elements $\mathbf{e}_1$ and $\mathbf{e}_2$ is described by $(r, \varphi, \theta)$, the reverse relation between $\mathbf{e}_2$ and $\mathbf{e}_1$ is given by $(r, \pi-\theta+\varphi, \pi-\theta)$. For arbitrary values of $\varphi$ and $\theta$, these two edges are not symmetric. It can be shown that the additional constraint of either mirror- or point-symmetric relation results in a pattern of cocircular respectively parallel orientation (Sec. A.4).

The interactions between elements as summarized by the bipole can be interpreted in two dual perspectives. First, in terms of convergent or integrating interactions, the bipole describes the relative support of neighboring elements to a target element at the reference location. This interpretation relates to the implementation of the bipole as a 3D filter in the spatio-angular feature space. Second, in terms of divergent or distributing interactions, the bipole defines the relative support contributed by the element at the reference location to its neighboring elements. This interpretation relates to, e.g., the interactions mediated by the axonal spreading of in vivo neurons. These two interpretations have been termed "in-field" and "out-field" by Grossberg and Mingolla (1985b). Formaly, for a given grouping function $f(\mathbf{e}_1, \mathbf{e}_2)$ which specifies the relative support an edge $\mathbf{e}_1$ receives from another edge $\mathbf{e}_2$, the in-field is given by $f(\mathbf{o}, \mathbf{e})$ for all neighboring edges $\mathbf{e}$, whereas the out-field is given by $f(\mathbf{e}, \mathbf{o})$. Here $\mathbf{o} = (0, 0, 0)$ denotes the null vector. For a symmetric grouping function, both fields are the same. More precisely, both fields are the same if oriented edges are considered, and mirror-symmetric with respect to the axis orthogonal to the reference direction if directed edges are considered.

The pattern of bipole-shaped integration is found in a number of empirical studies, ranging from anatomy and physiology via psychophysics to statistics (Sec. 4.2.1). A large number of different terms have been assigned to the pattern of spatio-angular interactions such as, e.g., cooperative

cell (Grossberg and Mingolla, 1985a,b), bipole (Grossberg and Mingolla, 1987), projection field (Zucker et al., 1989), association field (Field et al., 1993), extension field (Guy and Medioni, 1996) or grouping field (Heitger et al., 1998).

### Processing Stages of Contour Integration

In this paragraph the main stages that comprise most computational schemes of contour integration are outlined. Further, the particular computational elements which are commonly used to specify the mechanisms at each stage are introduced. These considerations are then summarized in a framework to describe the elements of perceptual grouping for contour integration.

The computational approaches to contour grouping usually consists of two main processing stages, sometimes followed by a third stage of postprocessing (cf. Geisler et al., 2001). First, the elementary items or input elements for the grouping process are extracted from the raw image data. Second, these local input elements are grouped or bound together according to various constraints. This grouping stage comprise the core stage of any computational scheme for contour grouping. Finally, the output of the grouping stage may be used for the explicit extraction of a global contour description. To sum up, the three processing stages are:

1. Preprocessing stage: extraction of the input elements
2. Grouping stage: binding or grouping of elements

[3. Postprocessing stage: explicit formation of globally "salient" contours]

We will now have a closer look at the characteristic computational elements and properties of the two main stages, namely the preprocessing stage and the grouping stage, which are to be found in virtually every computational approach to contour grouping. The computational framework derived from these elements and properties allows to distinguish and classify different computational schemes as proposed in the literature (see Sec. 4.3.2 and Sec. 4.3.3).

**Preprocessing Stage: Extraction of Input Elements** The computational schemes for contour grouping usually involve a first stage of low-level image processing, where the basic elements of the grouping process are extracted from the raw image data. In some approaches which exclusively deal with synthetic image data, such as artificial psychophysical stimuli (e.g., Pettet et al., 1998), the basic elements may not be computed rather than provided as a priori available information. Whether prespecified or computed, different kinds of input elements may be used. In most approaches, the basic elements are oriented or directed edges, but other elements such as endpoints as the output of end-stopped operators, or unoriented features such as dots may also be used.

**Grouping Stage** The grouping stage is the core stage of any scheme of contour integration. The interactions at the grouping stage define the relative support of neighboring elements to an element at the target location. These interactions can be subdivided into a number of computational elements.

One fundamental element within these interaction is the *grouping function* or *net coupling strength* (Neumann and Mingolla, 2001) between two elements in the feature space. In general, the feature space is a high-dimensional space $\langle x, y, \mathbf{f} \rangle$ of spatial locations in a 2D retinotopic map, $(x, y)$, and an $n$-dimensional feature $\mathbf{f} = (f_1, f_2, \ldots, f_n)$. Elementary features are, e.g., motion or disparity. In the case of contour grouping, the elementary feature is mostly the one-dimensional orientation $\theta$ of the edge element. The grouping function then defines the mutual influences of elements in the feature space. i.e., the *relatability* (Kellman and Shipley, 1991) or *compatibility* (Parent and Zucker, 1989) or *affinity* (Williams and Thornber, 1999) between two edge elements.

The grouping functions can be subdivided into two classes depending on the specification used for the coupling strength within a *geometrical* or *analytical* (respectively *probabilistic*) framework. Grouping functions of the first class are based on geometrically defined constraints such as colinearity, cocircularity or a smooth connection without inflection point (Gove et al., 1995; Grossberg and Mingolla, 1985a,b; Kellman and Shipley, 1991; Parent and Zucker, 1989). Grouping function of the second class define a connection curve of least energy. Such curves of least energy or elastica have a long tradition and can be tracked back to Euler (1744), as pointed out by Mumford (1994a). Formaly, let $C$ be a curve with curvature $\kappa(s)$ at each point $s$ of the curve. Then the energy of the curve is defined by $E(C) = \int_C \alpha\kappa(s)^2 + \beta \,\mathrm{d}\, s$. (The parameter $\beta$ adds a contribution which is linearly proportional to the length of the curve, and $\alpha$ is a weighting parameter.) Curves of least energy can also be formulated in a Bayesian framework as the maximum likelihood of particles in stochastic motion (Mumford, 1994a). A number of computational approaches to perceptual grouping and contour completion employ curves of least energy or variants (Horn, 1983; Sha'ashua and Ullman, 1988; Thornber and Williams, 1996; Ullman, 1976; Williams and Jacobs, 1997a,b).

The coupling strength between two elements as defined by the grouping function usually takes greater values for elements at small distances which can be linked by straight lines or by smooth curves without any inflection points. The underlying principles that govern the construction of the grouping function are thus given by, e.g., proximity, colinearity, cocircularity, or smoothness. While at least some of these principles are realized in virtually every computational approach to contour grouping, the choose of a particular set of principles and the precise way of their implementation differs between approaches. The resulting bipole pattern may further differ in shape (e.g, ellipsoidal, fan-like, or spades-like), opening angle, and the weighting both along and orthogonal to the preferred axis of orientation.

The grouping function defines the relative support of a *single* element to a target or reference element. In general, there are a number of elements in the neighborhood of the reference element, each of which potentially supports the element at the reference location. The next step in the grouping schemes is thus the combination or aggregation of individual contributions from different elements in the neighborhood of the reference element by an *aggregation function* or *saliency function* (Williams and Thornber, 1999). This aggregation function is often evaluated separately for each of the two lobes of the bipole. In the simplest case, the aggregation is defined by the addition of the individual contributions. In this case, the aggregated activity for a single lobe of the bipole can be computed by a linear 3D-correlation of the spatio-angular kernel (as specified by the grouping function) with the input activations. More advanced schemes make use of a nonlinear, often multiplicative combination of the two lobes. A nonlinear, AND-gate like operation ensures that the target element is only supported if *both* lobes are simultaneously activated. This property is necessary to preserve tangential continuity (Williams and Thornber, 1999).

The separate aggregation of activities with respect to the two lobes of the bipole implies the next step in computation, which is the *combination of the left and right lobe*. Again, in the simplest case, this combination is linear. In this case, the aggregation of activity and the combination of lobes can be combined in a single stage of 3D-correlation. In general, however, the two lobes of the bipole are nonlinearly combined by, e.g., by multiplication or using a compressive nonlinearity such as shunting interaction. The nonlinear combination is often used to realize an AND-gating mechanism, such that the target cell only receives an input if both lobes are simultaneously activated.— The main axes of the two lobes of the bipole are usually coaligned to the orientation of the reference element (Fig. 4.16). Some schemes may also combined lobes which have a slanted orientation with respect to the reference element (e.g., Heitger et al., 1998).

So far, we have considered the elements to compute the facilitatory influences expressed in a *single bipole interaction*. For a particular reference location, a number of bipoles with different orientations or slants of the lobes may be computed, which undergo either facilitatory or inhibitory interactions. The final stage of computation is then constituted by the *cooperative-competitive interactions* between the result of bipole groupings at different angles. Similar to the preprocessing

stage, one may distinguish simple linear schemes from more complex ones which incorporate a number of nonlinear interactions.

The computational elements outlined so far relate to how the grouping is realized. To further characterize the grouping process, the *kind of the output*, i.e., what is computed, has also to be considered. The kind of output can be given by a number of oriented elements in a representation similar to the input stage, but with different individual strengths as determined by the grouping process. Alternatively, the output may incorporate new features such as depth relation of the elements, or an explicit assignment of a contour to a particula surface (so-called "border ownership" (Nakayama and Shimojo, 1992; Nakayama et al., 1989) or "boundary ownership" (Grossberg, 1997)).

**Computational Architecture** With respect to the overall *computational architecture* of the grouping scheme, one can distinguish between pure feedforward schemes and iterative or recurrent feedback schemes. In the feedforward schemes, the grouping interactions are computed in a one-step process, whereas feedback schemes use a number of recurrent interactions until the output of the grouping stages has converged to a stable result. Schemes of recurrent interactions can be further subdivided according to the number of layers involved in the recurrent loop. One may distinguish between (i) single-layer architectures, where lateral input to the bipole interaction is provided by horizontal integration of activity within a layer, and (ii) multi-layer architectures, where bottom-up input is provided by a convergent feedforward mechanism from a lower layer.

### A Framework to Describe Computational Approaches to Contour Grouping

The above considerations can be summarized in framework suggested to describe the core elements of computational approaches to contour grouping.

*Overall Computational Architecture*

1. *Structure*

   One can distinguish feedforward from recurrent processing schemes. Recurrent schemes can be further subdivided with respect to the number of layers involved in the recurrent interaction: *single-layer* recurrent network with lateral input and *multi-layer* recurrent network with bottom-up input.

*Input Stage*

2. *Image data*

   Different kinds of image data may be used to validate the grouping scheme, e.g, black and white line drawings, artificial gray scale images, or natural images.

3. *Extraction of input elements*

   The input elements may be prespecified or computed by some kind of low-level image processing operation.

4. *Kind of input elements*

   Different kinds of input elements may be used, such as, e.g., oriented edges, directed edges, or dots.

*Grouping Stage*

5. *Class of grouping function*

   The grouping function may be specified within a geometrical or analytical (respectively probabilistic) framework.

6. *Grouping principles*

   Grouping principles are often based on Gestalt principles like proximity and good continuation. These principles can be used to guide the definition of the spatial component of the interaction by proximity and (near) colinearity, as well as the featural component using the elementary connections patterns of parallelism, radiality or cocircularity.

7. *Aggregation of activity*

8. *Combination of lobes*

9. *Cooperative-competitive interactions (CC interactions)*

   The grouping at a particular location and orientation may be followed by a number of cooperative-competitive interactions between the result of other grouping operations at different locations or angles.

10. *Kind of output*

    The result of the grouping process may be represented like the input activations, e.g., by a number of oriented features. Alternatively, additional features such as depth relations of boundary ownership may be computed.

### 4.3.2 Computational Models

In this section we give an overview of computational models for contour integration. A recent review focusing on neural models for perceptual grouping is given by Neumann and Mingolla (2001).

#### Boundary Contour System by Grossberg and Coworkers

Grossberg and coworkers were among the first to make use of long-range contour grouping schemes (Grossberg and Mingolla, 1985a,b) and have continuously advanced the early proposed schemes (Gove et al., 1995; Grossberg and Mingolla, 1987) by a detailed modeling of interactions between the laminar circuits in V1 and V2 (Grossberg, 1999; Grossberg et al., 1997; Grossberg and Raizada, 2000; Grossberg and Williamson, 2001; Raizada and Grossberg, 2001; Ross et al., 2000). The fundamental property all these models share is the use of recurrent feedback processing. Further characteristic mechanisms used include center-surround connection schemes and nonlinear, divisive normalization resulting from shunting inhibition.

In the early work, which already constitutes an impressively developed and complex model, the grouping interactions are used in the BCS/FCS framework of a boundary contour system (BCS) and feature contour system (FCS) for the purpose of scene segmentation of textured images (Grossberg and Mingolla, 1985b). The grouping interaction occur within the boundary system and is formalized as a recurrent feedback loop between short-range competitive interactions and long-range cooperative interactions.

The grouping function at the cooperative stage is specified and evaluated separately for the left and right subfield $B_{\mathrm{sf}}$ of the bipole. For each subfield, the grouping function is the half-wave rectification of the product of three functions

$$B_{\mathrm{sf}} = \left[ f_{\mathrm{radial}} f_{\mathrm{angular}} f_{\mathrm{feature}} \right]^{+} \quad .$$

The radial and angular function define the spatial weighting of the bipole. The radial function decays in a Gaussian fashion with the distance $r$, and the angular function decays in a cosine fashion with deviation from colinearity. Similarly, the feature function decays in a cosine way with the deviation from radiality, given by the orientation of the virtual line passing through the

reference location and the neighboring point, i.e., the angular position $\varphi$. Using the notation as introduced in Sec. 4.3.1, the equations for a gauge coordinate system of zero reference orientation write

$$\begin{aligned} f_{\text{radial}}(r) &= \exp(-2(r/\alpha_1 - 1)^2) \\ f_{\text{angular}}(\varphi) &= \cos(\varphi)^{\alpha_2} \\ f_{\text{feature}}(\varphi, \theta) &= |\cos(\varphi - \theta)|^{\alpha_3} \quad , \end{aligned}$$

where $\alpha_1, \alpha_2$ and $\alpha_3$ denote positively valued model parameters. Variants of this kernels have been used in other model of Grossberg and coworkers (e.g., Gove et al., 1995; Ross et al., 2000).

The grouping function specifies the receptive field or *infield* of a cell at the cooperative stage, or likewise, in a complementary view, the projection field or *outfield* of the cell. In other words, the infield describes the weighted input from neighboring cells to a target reference cell, whereas the outfield describes the weighted output from a source reference cell to its neighboring cells. For a particular parameter choice of $\alpha_1 = 5, \alpha_2 = 55, \alpha_3 = 5$, the corresponding infield and outfield is depicted in Fig. 4.17. Note that both fields are different, which implies a non-symmetric grouping function, as outlined in Sec. 4.3.1.

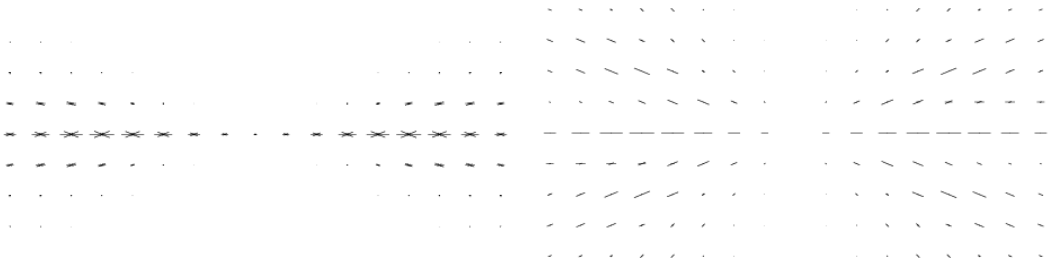

**Fig. 4.17:** Infield *(left)* and corresponding outfield *(right)* of a cooperative cell. The lengths of lines encode the relative weighting of a cell at the particular position and orientation. (Adapted from Grossberg and Mingolla, 1985b.)

Within each subfield of the bipole, the individual contributions according to the grouping function are added. The contributions of the left and right subfield are nonlinearly compressed and added, which results in a AND-gating of the individual lobes, as pointed out by Neumann and Mingolla (2001). The processing thus realizes inward completion without extrapolation, i.e., the grouping function is superthreshold only if *both* lobes of the bipole receive input. The term "bipole property" has been coined for this characteristic feature.

The output of the subfield combination is fed into a stage of shunting on-center off-surround interaction in the spatial domain to sharpen the activities. The output of this center-surround stage generates the final output of the grouping stage, which is fed back to an earlier stage of feedforward on-center off-surround interaction.

In more recent models, the early ideas have been extended and mapped to the laminar structure of visual areas V1 and V2. In contrast to the older models, the new models have implemented the bipole property together with *analog sensitivity*. The term analog sensitivity refers to the strength of the illusory contours and thus the result of the underlying grouping function, which should take continuous "analog" values ranging from weak to strong rather than binary values of "all-or-none" contour completion (Grossberg et al., 1997; Ross et al., 2000).

In the more recent models, the two lobes of the bipole are not evaluated separately and then nonlinearly combined. Instead, the excitatory input to the grouping stage results from spatial correlation with a bipole kernel which samples colinear activity. Such a kernel alone would result in an outward spreading of activity, since only activation of a single lobe is sufficient to generate a response, i.e., it would violate the bipole property. To realize the bipole property, an additional

**Table 4.1.** Summary of properties of two models by Grossberg and coworkers (Grossberg and Mingolla, 1985b; Ross et al., 2000).

| Paper | **Grossberg/Mingolla (1985b)** | **Ross et al. (2000)** |
|---|---|---|
| *Computational Architecture* | | |
| structure | multi-layer recurrent network | |
| *Preprocessing Stage* | | |
| image data | artificial images (black/white line drawings) | artificial images (black/white, grayscale) |
| extraction | prespecified | oriented filters |
| input elements | oriented edges | |
| *Grouping Stage* | | |
| class | geometric | |
| principles (spatial) | proximity, colinearity | |
| (feature) | radiality | cocircularity |
| aggregation | additive, squashing-type compression | additive |
| combination of lobes | additive (functional AND-gating) | additive |
| CC interactions | on-center off-surround for each orientation | coaxial long-range inhibition, on-center off-surround spatio-orientational interaction |
| output | oriented edges | |

system of long-range inhibitory interactions is introduced. This system causes an inhibition which shuts off all extrapolated excitatory activity resulting from a single lobe of the bipole (one-against-one), but prevents interpolated activity resulting from two lobes (two-against-one). The kernels used to model the long-range interactions are depicted in Fig. 4.18. The resulting activity of this cooperative-competitive interactions is fed back to engage a feedforward on-center off-surround network.

The whole model incorporates a large number of further computational mechanisms, such as short-range excitatory connections, recurrent spatio-orientational sharpening, top-down cortico-geniculate feedback, as well as cortico-cortical feedback from V2, modeled as a homologous replication of the V1 circuit at a larger scale.

Variants of this model have been proposed for the processing of real world gray scale images from synthetic aperture radar (SAR) (Grossberg et al., 1995; Mingolla et al., 1999).

**Fig. 4.18:** Spatio-orientational kernels (infields) to model long-range grouping interactions. *Top:* V1 kernel, *bottom:* V2 kernel. (From Ross et al., 2000.)

### Ortho/Para-Keypoint Grouping by Heitger and Coworkers

Heitger and coworkers have developed a feedforward model for contour grouping (Heitger and von der Heydt, 1993; Heitger et al., 1998). The grouping process is used to infer contours from occlusion features, which are provided by end-stopped filters. The grouping results in the generation of illusory contour both along and orthogonal to line ends. Heitger and coworkers have coined the term "para" and "ortho" grouping for these different types of grouping. Further, the model output incorporates a prediction of figure-ground relationship. A variant of this model have been proposed recently to account for physiological findings regarding the signaling of depth order (Peterhans and Heitger, 2001).

A key assumption of the model is that grouping processes are exclusively initiated by so-called "key-points", such as corners or line-ends. Consequently, such key-points are first extracted from the input stimulus using a hierarchical filter scheme of simple, complex and end-stopped filters (Heitger et al., 1992). End-stopped filters of orientations parallel ("para" grouping) and orthogonal ("ortho" grouping) to the preferred grouping direction are locally combined according to a cornerness measurement and provide the input to scheme of bipole integration. A single lobe of the bipole is defined by a spatial weighting function similar to that used in Grossberg and Mingolla (1985b). The bipole property is then realized by a multiplying the two lobes of the bipole. For each preferred orientation of the bipole, two modes of integration are realized to signal the two possible types of figure-ground direction.

**Table 4.2.** Summary of properties of the model by Heitger et al. (1998).

| Paper | **Heitger et al. (1998)** |
|---|---|
| *Computational Architecture* | |
| structure | feedforward |
| *Preprocessing Stage* | |
| image data | artificial and natural images |
| extraction | oriented filters |
| input elements | end-stopped points |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity, colinearity |
| (feature) | parallelism |
| aggregation | additive |
| combination of lobes | multiplicative |
| CC interactions | combination of different curvature classes, cross-orientation inhibition |
| output | contour points plus figure-ground relation |

In order to allow for curved illusory contours, the grouping response is then computed from the weighted average of bipole responses with orientations adjacent to the preferred orientation according to three different curvature classes (0°, 30° and 60° difference between the individual lobes of the bipole). Multiple orientation are suppressed by cross-orientation inhibition. The final result of combined illusory contours resulting from the grouping stage and real contours of the initial complex cell operator are further subject to non-maximum suppression. The capability of the model to successfully generate illusory contours from occlusion features is demonstrated for a variety of classical psychophysical stimuli such as the Kanizsa triangle as well as for images of natural scenes. The model not only computes illusory contours but further explicitely signals

figure-ground relationship along such contours. The properties of the model by Heitger et al. (1998) are summarized in Tab. 4.2.

As stated above, the model of Heitger et al. relies on the explicit computation of end-stopped points which provide the input for the grouping process. The use of such key-points is consistent with the theory of Kellman and Shipley (1991) based on earlier results results (Shipley and Kellman, 1990). Kellman and Shipley propose that first order discontinuities in the curve along the object boundary are fundamental for the unit formation process. However, counter-evidence for the exclusive significance of such discontinuities exist (Purghé and Russo, 1996). As an aside, the idea of an explicit filter-based computation of end-stopped operator has been challenged by physiological studies suggesting that end-stopped properties emerge from surround inhibition within a recurrent interlaminar loop (Bolz and Gilbert, 1986; Bolz et al., 1989).

Further, the feedforward scheme used in the model of Heitger et al. is different to reconcile with findings suggesting that feedback plays an important role in figure-ground segregation (Hupé et al., 1998; Lamme, 1995; Zipser et al., 1996).

### V1 Model of Excitatory-Inhibitory Coupled Oscillators by Li

Li has developed in a number of papers a neural model for contour integration and texture segmentation in primary visual cortex (Li, 1998, 1999a,b,c, 2001). The model is defined by two reciprocally connected recurrent populations of excitatory and inhibitory neurons. The recurrent interactions within each population is mediated by horizontal connectivity patterns. Two different patterns are used for the excitatory and inhibitory population, which either extend colinear to the reference orientation (for the excitatory population) or orthogonal to (for the inhibitory population). The kernels implementing the excitatory respectively inhibitory connection patterns are depicted in Fig. 4.19. The model equations are governed by linear or piece-wise linear interactions and do not employ multiplicative or shunting interactions. The model is thus more apted for a detailed mathematical analysis compared to models using a large variety of nonlinear interactions. The properties of the model by Li (1998) are summarized in Tab. 4.3.

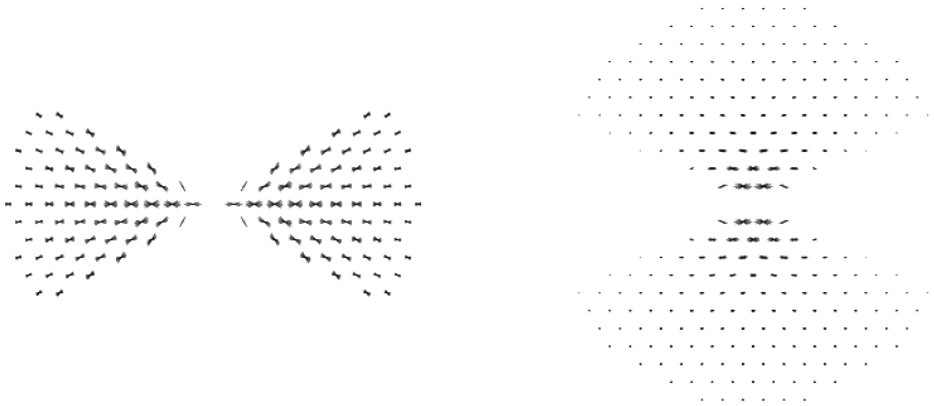

**Fig. 4.19:** Spatio-orientational kernels (infields) to model long-range grouping interactions. *Left:* excitatory kernel, *right:* inhibitory kernel, both for a horizontal reference orientation. (Adapted from Li (1998) using a reimplementation of the filter kernels courtesy of Axel Thielscher.)

The connection patterns are designed using mean field techniques and dynamic stability analysis to fulfill a number of criteria such as no spontaneous pattern generation or "no hallucination" (Li and

**Table 4.3.** Summary of properties of the model by Li (1998).

| Paper | **Li (1998)** |
|---|---|
| *Computational Architecture* | |
| structure | two-layer recurrent network (excitatory and inhibitory layer) |
| *Preprocessing Stage* | |
| image data | artificial images, (one natural image) |
| extraction | prespecified, (oriented filters) |
| input elements | oriented edges |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity, colinearity |
| (feature) | cocircularity |
| aggregation | additive |
| combination of lobes | additive |
| CC interactions | transaxial long-range inhibition, orientational inhibition, normalization |
| output | oriented edges |

Dayan, 1999), higher responses at region borders, and enhancement of smooth contours (including filling-in of small gaps and no increase in contour width or length). Despite the claim that the model fulfills all criteria, homogeneous regions nevertheless may give rise to higher responses (e.g., Fig. 6B in Li, 1999a). Such obvious artefacts are then rationalized as a medial axis representation (Lee et al., 1998) tracing object skeletons (Blum, 1973). The significance of both the physiological data and the model (see also Li, 2000) is questionable, since higher responses at the medial axis are only measured respectively generated for objects of a specific size. A robust skeleton representation, on the other hand, would be invariant against scale.

The model has been successfully applied to a number of different tasks such as pre-attentive texture segmentation (Li, 1999a,b), and contour enhancement (Li, 1998) and also accounts for asymmetrical pop-out of targets among distractors in psychophysical search paradigms (Li, 1999c). The dynamical properties of the employed network structure of reciprocally connected excitatory and inhibitory neurons (E-I network) have been analyzed in detail (Li, 1998, 2001; Li and Dayan, 1999).

### Model of Cortico-Cortical Recurrent Interactions by Neumann and Sepp

Neumann and Sepp (1999) have developed a model for the recurrent interaction between two reciprocally connected cortical areas V1 and V2. In this model of cortico-cortical V1–V2 interaction, the lower area V1 serves as a stage of feature measurement and signal detection, whose activity is propagated in the feedforward path to the higher area V2. The higher area in turn represents expectations about visual structural entities and context information to be matched against the incoming data carried by the ascending pathway. The matching process generates a pattern of activation which is propagated backwards via the descending feedback pathway. This activation pattern serves as a signature for the degree of match between the data and possible boundary outlines. The activation is used to selectively enhance those signal patterns that are consistent with the model expectations. A gain control mechanism, that is accompanied by competitive interactions in an on-center/off-surround scheme, realizes a "soft gating" mechanism that selectively filters salient input activations while suppressing spurious and inconsistent signals. As a result

the primary functional role of the feedback pathway realizes a gain control mechanism driven by top-down model or expectation information (Mumford, 1991; Ullman, 1995), similar to ideas of adaptive resonance theory (Carpenter and Grossberg, 1988; Grossberg, 1980).

The model information is stored in bipole-shaped "curvature templates" which represent shape segments of varying curvature derived from a cocircularity constraint. An additional template or connection pattern of parallel and near parallel orientations is used to model inhibitory contributions (Fig. 4.20).

**Fig. 4.20:** Spatio-orientational kernels modeling the excitatory *(top)* and inhibitory *(bottom)* subfields of a contour cell. (Adapted from Neumann and Sepp (1999) using kernel weightings provided courtesy of Wolfgang Sepp.)

The templates are matched against the feedforward measurements of local oriented contrast using spatio-orientational correlation of the two lobes of the bipole. The two lobes are then multiplicatively combined using a micro-circuit realizing an AND-gate. The resulting activities undergo a spatio-orientational center-surround interactions realizing a compressive normalization by divisive inhibition. The properties of the model by Neumann and Sepp (1999) are summarized in Tab. 4.4.

**Table 4.4.** Summary of properties of the model by Neumann and Sepp (1999).

| Paper | **Neumann and Sepp (1999)** |
|---|---|
| *Computational Architecture* | |
| structure | multi-layer recurrent network |
| *Preprocessing Stage* | |
| image data | artificial images, (one natural image) |
| extraction | oriented filters |
| input elements | oriented edges |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity, colinearity |
| (feature) | cocircularity |
| aggregation | additive |
| combination of lobes | multiplicative (AND-gate micro-circuit) |
| CC interactions | coaxial long-range inhibition, spatio-orientational divisive inhibition |
| output | oriented edges |

The model accounts for a number of different phenomena such as illusory contour generation of abbuting gratings, correctly predicts illusory contour strength for different Kanizsa and Varin figures and accounts for different effects of an overlay of dense vs. coarse line patterns on the

visibility of an underlying line of the same orientation. The model further performs contour enhancement and grouping of noisy fragmented shape in both artificial and natural images.

### Other Approaches

Kellman and Shipley (1991) have outlined a theory of visual interpolation and unit formation. Central to their theory is the idea that two basic completion phenomena, namely completion of partly occluded objects and illusory contour formation (or amodal and modal completion), can be described by the same unit formation process. The unit formation process is formalized in the concept of "relatability": To edges are relatable, if their linear extensions intersect, and their angle of intersection is obstuse or 90°. In this case, the two edges can be connected by a first-order continuous curve. The proposed scheme is consistent with a number of psychophysical data. Note that relatability is a binary predicate function or all-or-none criterium, which does not account for different degrees of the perceptual strength of completion. Also, the distance between the two edges is not taken into account. It might be instructive to compare the relatability definition to an algorithm of endpoint linking based on shared, simple occluders devised by August et al. (1999), which is based on complementary cues: Whereas the former scheme is defined using the orientation of endpoint tangents and ignores the distance between endpoints, the latter ignores the tangents and crucially depends on the distance between endpoints. The algorithm of August et al. (1999) is discussed in more detail in Sec. 4.3.3 on page 122.—Unlike all other approaches reviewed in this section, the theory of Kellman and Shipley is not implemented in a computer model.

Yen and Finkel (1998) have developed a noniterative model to compute perceptual saliency and preattentive pop-out of contours. The input elements are a measure of the local orientation and are computed as the sum of a quadrature pair of steerable basis filters (Freeman and Adelson, 1991). The input elements are grouped by model horizontal connections according to a cocircularity constraint. In addition, facilitatory input is also provided from parallel oriented elements in a narrow region orthogonal to the orientation of the target element (trans-axial). These orthogonal connections are supported by anatomical evidence (Fitzpatrick, 1996; Lund et al., 1985; Mitchison and Crick, 1982; Rockland and Lund, 1982, 1983) and allow to simulate facilitatory effects of parallel aligned contour elements, as has been observed psychophysically (Field et al., 1993; Polat and Sagi, 1994). The output of the grouping stage undergoes a dynamic thresholding, followed by competitive winner-takes-all interactions between the co-axial and trans-axial lobes. The most distinguishing feature of the model is the final stage, where the superthreshold activity is fed into a network of homogeneous coupled neural oscillators. The saliency of a contour is then determined by the activity of synchronized elements. Such temporal synchronization has been proposed as a general mechanism for feature binding and allows for the flexible representation of context-depending grouping (Eckhorn, 1994; Singer and Gray, 1995).

Pettet et al. (1998) have developed a single layer recurrent network for contour detection. The model is designed to explain detection performance in a psychophysical search task of contours among a field of distractors. The model utilizes a long-range interaction between model neurons which minimizes distance, curvature and change in curvature. In the model, a higher saliency of closed curves as has been found experimentally (e.g., Kovács and Julesz, 1993) is an emergent feature, resulting from mutual reverbarating excitation among model neurons.

Dakin (1997) has proposed a noniterative model that can account for the psychophysically measured performance at contour detection in Glass patterns. The model is based on the local selection of the most active filter across orientations.

With respect to contour integration in the periphery, a model for the detection of colinear or near-colinear contours has been proposed that involves no linking or grouping of individual elements at all (Hess and Dakin, 1997, 1999). In this system the processing of near-straight contours simply results from the independent output of multiple coarse-scale oriented filters.

Williams and Jacobs (1997a,b) have developed a model for contour linking and illusory contour generation within a probabilistic framework. The generation of a connecting contour between two inducing contour fragments (the so-called source and the sink) is formulated using prior probability distributions of completion shapes. These prior probability distributions are modeled as paths followed by particles traveling in stochastic motion from the source to the sink. The magnitude of the stochastic completion field is the probability for a path joining the two contour fragments. The probability density functions underlying the stochastic completion field evolve according to a Fokker-Plank equation (Mumford, 1994a). Recently, a new method for computing the completion field has been proposed with the unique feature of Euclidean invariance. Using a basis of so-called shiftable-twistable functions, the computations are invariant under arbitrary rotations and translations on a discrete grid (Zweck and Williams, 2000).

Ritter and coworkers have proposed an energy-based approach to feature binding and segmentation within the framework of recurrent neural networks, the so-called competitive layer model CLM (Wersing, 2000; Wersing and Ritter, 1999; Wersing et al., 2001). The CLM consists of a number of layers, and grouping of features is represented by a common activity of these features within a particular layer. The CLM makes use of lateral interaction between compatible features, a vertical dynamical winner-take all circuit between layers and a self-inhibitory interaction. These terms are combined into a standard additive activity dynamics. The CLM dynamics can be efficiently simulated by a sequential asynchronous update rule, where the equilibrium solution is computed independently for a randomly chosen node of the network. In the context of contour grouping, the lateral interaction incorporate a cooperative interaction between cocircular features and a short-range competitive interaction which is insensitive to orientation. The model performance is demonstrated on a number of natural images, where the model highlights the salient contours. Further, the model allows to distinguish between contours belonging to different objects, which are represented in different layers. A major drawback of the model seems to be the fixed number of layers for the representation of different contours. A fixed number is too large for a small number of objects, so that the provided neuronal capacity is unused, and conversely too small for a large number of objects. Also, the use of layers for the representation of the same feature to represent the belonging to different objects has not been found in vivo. Instead, the different layers or laminae found in the cortex seem to represent different features (such as sublaminae 4C$\alpha$ and 4C$\beta$) or different contextual influences mediated by different kinds of intralaminar, cortico-cortical or cortico-thalamic recurrent interactions (Blasdel and Fitzpatrick, 1984; Callaway, 1998).

### 4.3.3 Computational Algorithms

#### Relaxation Labeling Approach by Zucker and Coworkers

Zucker and coworkers have proposed a scheme to refine an initial noisy estimate of a local orientation field by iterative relaxation labeling (Iverson, 1994; Parent and Zucker, 1989; Zucker, 1985; Zucker et al., 1988, 1989). The relaxation scheme maximizes a local cocircularity constraint within a neighborhood which is partitioned into distinct curvature classes of same arc length.

Relaxation labeling is a general method to derive a globally consistent interpretation from an initial ambiguous and noisy measurement. Given a set of objects, a set of labels associated with each object, and a constraint relation over pairs (or more generally $n$-tuples) of objects, the goal is to find a *consistent labeling*, i.e., to assign a label to each object which is consistent with the constraint relation which defines the degree of compatibility between neighboring labels (Hummel and Zucker, 1983). In the domain of contour grouping, the objects are given by spatial positions, usually on a discrete grid, and the labels are given by a set of orientations. The constraints implement a priori assumptions on the nature of desirable contours, such as smoothness, and define the support a particular orientation at a particular location receives from its spatio-orientational neighborhood. At each iterative relaxation step, the individual labels are updated until convergence.

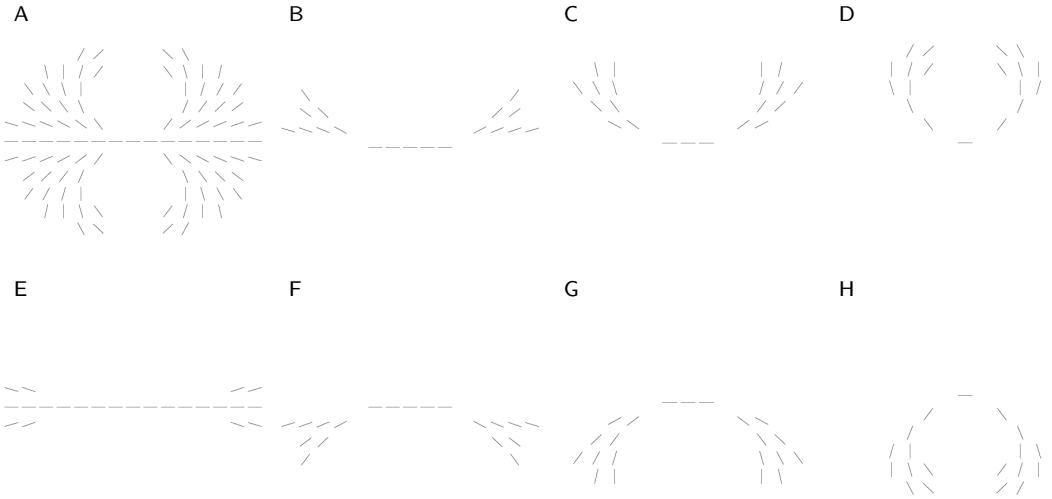


**Fig. 4.21:** Equilength $15 \times 15$ neighborhood *(A)* for a horizontal reference orientation and its partioning into seven curvature classes *(B–H)*. (Adapted from Parent and Zucker, 1989.)

In the scheme proposed by Parent and Zucker (1989), initial estimates of the orientation at each position are provided by linear convolution with a Gaussian line detector. A global normalization is necessary to map the orientation estimates to the range $[0; 1]$. The local support for a particular orientation is provided by cocircular neighboring orientations. The spatial neighborhood is partitioned into distinct curvature classes covering all osculating circles whose radii lie in the limits of the respective class (Fig. 4.21). The extend of the spatial neighborhood is defined such that each class has the same maximal arc length. An additional intrapixel length correction compensates for discretization effects. A further comparison of local curvature classes results in consistency of curvatures. Finally, lateral maxima are selected to extract curves of one-pixel width.

**Table 4.5.** Summary of properties of the algorithm by Parent and Zucker (1989).

| Paper | **Parent and Zucker (1989)** |
|---|---|
| *Computational Architecture* | |
| structure | recurrent relaxation network |
| *Preprocessing Stage* | |
| image data | artificial and natural images |
| extraction | oriented filters |
| input elements | oriented, normalized edges |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity |
| (feature) | cocircularity |
| aggregation | additive |
| combination of lobes | additive |
| CC interactions | curvature classes, curvature consistency, nonmaximum suppression |
| output | oriented edges |

Several application of the algorithm to a variety of images show that the relaxation labeling results in a robust inference of contour orientation from noisy estimates. The properties of the algorithm by Parent and Zucker are summarized in Tab. 4.5.

### Noniterative Tensor-Based Approach by Medioni and Coworkers

Medioni and coworkers have developed a noniterative scheme for the enhancement of salient contours within a tensor-based framework. A first description of the approach is given by Guy and Medioni (1996). A more rigorous definition of the tensor-based interaction scheme along with an extension to higher dimensions is given in Medioni et al. (2000).

The input to the grouping process is provided by either nonoriented points or oriented edges. The interaction between input points if defined by a bipole-shaped "extension field" (Fig. 4.22). This the support from the extension field is maximal for cocircular points and decreases with distance and higher curvature. The input to a particular location is defined by gathering the votes of neighboring points which strength is defined by the extension field. This interaction between points is defined in a tensor-based approach, where the input and output is represented as a $2\times2$ tensor in 2D. The output tensor can be rotated toward its eigensystem along its principal axis, such that the results at each location can be visualized by a rotated ellipse parameterized by the minimal and maximal eigenvalues of the corresponding tensor, i.e., by the triplet $(\lambda_{\max}, \lambda_{\min}, \theta)$. An "enhanced" saliency measure is defined which is maximal for long, elongated ellipses, i.e., for a high value of the maximal eigenvalue $\lambda_{\max}$ and a low eccentricity $1 - \lambda_{\min}/\lambda_{\max}$. Junction points can be inferred as points with high value of the minimal eigenvalue $\lambda_{\min}$.

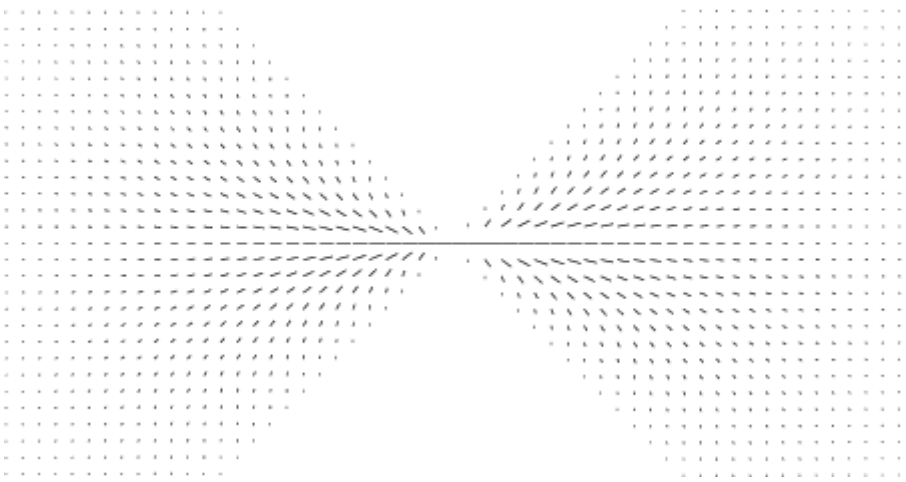

**Fig. 4.22:** Spatio-orientational kernels implementing the "extension field" (outfield) in the algorithm of Guy and Medioni (1996). The field is defined for a $51 \times 29$ neighborhood. (Adapted from Guy and Medioni, 1996.)

The algorithm is evaluated for simple synthetic images only, where it shows good performance. The properties of the algorithm by Guy and Medioni are summarized in Tab. 4.6.

The most distinguishing feature of the approach by Medioni and coworkers is the use of tensors for the coherent encoding of nonoriented and oriented input data and the final interpretation of the result using tensor decomposition (Medioni et al., 2000, p. 64). The advantage of a tensor representation is rationalized in contrast to a vector representation: a vector can represent a directed edge, with the length coding the strength of the edge, but fails to represent nonoriented features of different strength. A tensor, on the other hand, which can be visualized as an ellipse in 2D can encode the *type* of information, i.e., point vs. line, by the shape of the ellipse, and

**Table 4.6.** Summary of properties of the algorithm by Guy and Medioni (1996).

| Paper | **Guy and Medioni (1996)** |
|---|---|
| *Computational Architecture* | |
| structure | feedforward |
| *Preprocessing Stage* | |
| image data | artificial images |
| extraction | (not specified) |
| input elements | tensors: nonoriented points or nonoriented edges |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity, colinearity |
| (feature) | cocircularity |
| aggregation | additive |
| combination of lobes | additive |
| CC interactions | (none) |
| output | oriented edges, (junctions) |

the *strength* or saliency of the information by its size. Clearly, a representation using three values (describing a tensor in 2D) instead of only two (describing a vector) allows for a richer description. On the other hand, almost all discussed schemes use a representation with a number of orientations at each position. In contrast to a tensor-based representation, this allows to encode, e.g, corners and junctions with non-orthogonal intersection angles or more than two intersecting lines. Consequently, a representation with more than two discrete orientations at each positions is superior to a tensor representation.

### Other Approaches

Sha'ashua and Ullman (1988) have devised an iterative approach to extract the most salient curves out an array of cluttered line segments. The saliency measure is based on contour length and curvature and favors long smooth curves with few gaps. While computing the saliency the networks also smoothes the contours and completes or closes gaps. The saliency networks has been extensively analyzed by Alter and Basri (1996). They found that the approach is efficient, i.e., the network converges linearly in the size of the image and the overall complexity is quadratic in size of the network, namely $O(p^2o^2)$ for $p$ pixels and $o$ orientations at every position. However, the analysis revealed also a number of weaknesses, such that the most salient element may not lie on the most salient curve, or a scale-dependent performance (e.g., a line is more salient than a circle at small scale, but vice versa at larger scale). Further, the saliency map prefers closed curves with a single large gap over closed curves containing a number of small gaps of same total size (which is at odds with experimental findings reported by Elder and Zucker (1994), and has problems with identifying salient curves other than the most salient one.

August et al. (1999) have addressed the problem of long-distance grouping of contour fragments. In contrast to other schemes discussed so far, the algorithm of August et al. only specifies which contour fragments should be linked without computing or interpolating an estimate of this linking contour. The approach is motivated by the notion that two contour endpoints should be linked if and only if they results from the same simple occluder (so-called "shared, simple occlusion" or SSO constraint). To fix ideas, disks are suggested as the only generic shape of a simple occluder, and two endpoints are linked if the "gap disk" adjacent to both endpoints contains no other

contour points. The approach is formalized in the framework of skeletons by using a fraction of the skeleton representation of the background (the so-called "gap skeleton") to decide which endpoints are to be linked. Overall, the scheme constitutes an original approach and performs well on simple grouping problems such as a single fragmented line (Fig. 13 in August et al., 1999). On a more complex artificial image of fragmented B's however (Fig. 17), the algorithm fails to detect a number of links while likewise suggesting a number of wrong links. We conjecture that these failures are due to the complete ignorance of the orientation of the contour tangents at the endpoints. While one might admit that the influence of tangent orientation decreases with the distance of endpoints, the orientation may still provide a significant bias for the grouping process. In fact, the vast majority of grouping schemes, in particular the well-known concept of "relatable" contours by Kellman and Shipley (1991), are based on the orientation of endpoint tangents.

Gigus and Malik (1991) have proposed a noniterative feedforward mechanism that is based on an adaptive tuning of local filters to circular image structures.

Hérault and Horaud (1993) have proposed an algorithm which assigns a binary label to segment a set of oriented edges into figure and ground. The interaction of oriented edges is based on the Gestalt principles of cocircularity, smoothness and proximity and is used to define a quadratic cost function to be minimized by simulated annealing.

### 4.3.4 Discussion of Contour Models and Algorithms

A diversity of computational approaches for contour grouping have been proposed, based on a number of different design principles. In the previous section we have characterized a selection of prominent approaches and pointed to relative merits and drawbacks of particular schemes. In the following, we shall discuss some of the different properties according to the suggested framework presented in Sec. 4.3.1 within a more broader context.

**Feedforward vs. Feedback Schemes** With respect to the overall computational architecture, feedforward and feedback approaches have been proposed. While feedforward models are usually faster and more easier to design, since no care has to be taken to ensure stable processing, we favour feedback models for a number of reasons. First, as pointed out by Neumann and Sepp (1999), feedback model provide a "rough and ready" computation within the first feedforward sweep, similar to pure feedforward models, and offer the possibility of subsequent refinement of the result within the recurrent loop. Second, the drawback of a restriction of a fixed spatial entension of integration kernel is less severe for feedback models, since the resulting spreaded activity of the previous cycle is fed back, allowing an integration within a range of distances. Consequently, feedback schemes are more suitable for scale-invariant grouping, i.e., the integration of contour elements at different spatial scales (cf. Dakin and Hess, 1998; Gigus and Malik, 1991). Last but not least, feedback is a common principle in visual processing and has been suggested to be involved in figure-ground segregation (Hupé et al., 1998; Lamme, 1995; Zipser et al., 1996).

**Oriented vs. Directed Edges and the Bipole Property** Intrinsically 1D elements as input for the grouping stage can be either sensitiv or not sensitive to contrast polarity, i.e., being either directed or oriented. The perception of contours, e.g., in the reverse-contrast Kanizsa square suggests polarity insensitive grouping using oriented edges (cf. Grossberg, 2000). On the contrary, Williams and Thornber (1999) argued based on the evaluation of different algorithms that directed edges ensure tangential continuity and allow for a better grouping result compared to oriented input elements. These different findings may be reconciled in the notion of the bipole property, i.e., the splitting of the grouping kernel in two lobes and the demand of simultaneous activation of the two lobes in order to trigger grouping.

**Grouping Principles** Their is large agreement concerning the grouping principles underlying the design of the grouping function, such as proximity, colinearity and cocircularity. However, their are substantial differences in the precise implementation. Based on a cocircularity principle, the kernels may allow grouping interaction only between very low curvature (e.g., Ross et al. (2000), as depicted in Fig. 4.18 or high high curvature (e.g., Parent and Zucker (1989), Fig. 4.21). Cocircularity constraints are incorporated in a number of models and can be motivated by differential geometry considerations, arguing that inverse radius of the osculating circle approximates curvature, i.e., the second derivative of curve, whose change and local maxima mark interesting points (cf. Parent and Zucker, 1989). Further cocircular connection schemes have been recently revealed in the second order statistics of contour cooccurrences in natural scenes (Geisler et al., 2001; Sigman et al., 2001). Despite this plausible motivations, a neural implementation of a cocircularity constraint such as in the axonal arborization of pyramidal cells, has not been found. Instead, only parallel respectively colinear connection schemes have a direct biological correlate. Interestingly, none of the proposed models make exclusive use of such biologically most plausible of colinear respectively near colinear kernels. A model using this type of grouping kernels will be presented in the next section.

**Design and Partitioning of the Grouping Kernel** In the simplest case, the grouping kernel may additively gather the contributions of the neighboring positions. For a cocircularity constraint, three different configurations that give rise to the same support, namely (i) cocircular orientations, which is the desired support, (ii) orientations on half-circles joining with an inflection point at the reference location (violating the principle of relatability, Kellman and Shipley 1991), and (iii) orientation on half-circles forming a cusp and joining with a curvature singularity at the reference location (violating the principle of tangential continuity). The three configurations are depicted in Fig. 4.23. To ensure only a support from cocircular orientation (case (i)), one has to introduce the bipole property to eliminate case (ii) and a horizontal split of the bipole in two curvature classes of up and down opening circles. While both the bipole property and curvature classes can be found in different models and algorithms, the joint use of both principles is not to be found in any scheme.

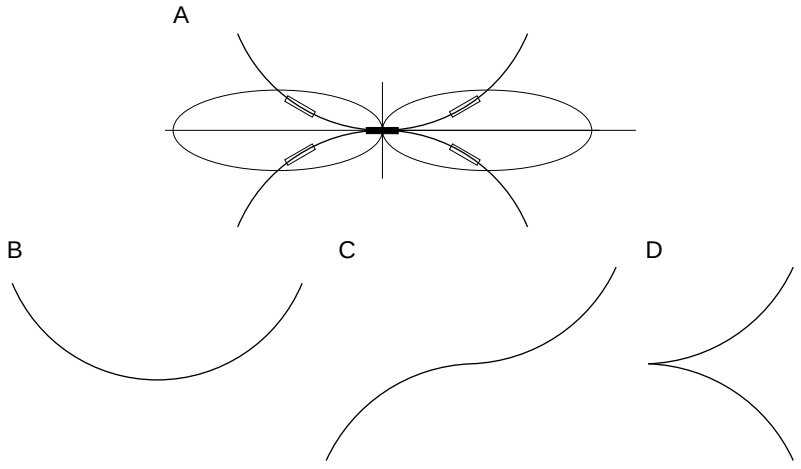


**Fig. 4.23:** Different configurations giving equal support in a cocircular condition. *A:* Sketch of four neighboring elements cocircular to the horizontal reference element. *B–D:* Three configurations that would give equal support if the bipole interaction is implemented by spatio-orientational correlation of a single kernel. However, only configuration *B* fulfills the cocircularity condition.

**Evaluating the Approaches** The above considerations summarize some crucial properties of grouping schemes. However, a key questions remains open: What computational scheme has the

"best" performance?—A rigorous answer to this and related questions would involve a thorough evaluation of different approaches on a range of images, both synthetic and natural. Such an evaluation would be faced with a number of practical problems, e.g., the proper selection of images, the accurate reimplementation (respectively obtainment of the authors' implementations), the selection of parameters, the proper adaption to different scales used by various schemes (e.g., $15 \times 15$ kernels by Parent and Zucker (1989) vs. $29 \times 51$ kernels by Guy and Medioni (1996)), and so on. However challenging this approach may be, a number of valuable insights would be gained from such an extensive evaluation.

Instead of evaluating the whole model, a more trackable approach is to compare a particular subset of model features. In the only existing comparative study (to our knowledge), Williams and Thornber (1998, 1999) have pursued this approach. They compared different methods of aggregating the individual contribution of neighboring sites (so-called "saliency measures") for a number of schemes using the same preprocessing and the same grouping function (so-called "affinity values"). The validy of the approach is clearly based on the assumption that no dependencies exists between the grouping function and the subsequent aggregation. Williams and Thornber choose an affinity function as proposed for their own model and find that their algorithm outperforms the other tested schemes. They conjecture that two properties cause the superior performance of their model, namely the represention of directed instead of oriented edges and the explicit enforcement of closed contours. The performance on images containing open fragmented contours has not been evaluated.

## 4.4 The Model

In this section we propose a biological plausible model of recurrent long-range interaction in the primary visual cortex for contour processing. We shall first describe the computational principles used and point to their biological plausibility and relevance. We shall then give a rigorous mathematical definition of the model.

### 4.4.1 Model Overview

The computational model uses localized receptive fields for oriented contrast processing and incorporates feedforward and feedback processing, cooperative horizontal long-range integration and lateral competitive interactions.

The key properties of the model are motivated by empirical findings.

- *Horizontal long-range connections*

  The grouping of aligned contours require a mechanism that links cells of proper orientation over larger distances. Horizontal long-range connection found in the superficial layers of V1 may provide such a mechanism: They span large distances (Gilbert and Wiesel, 1983; Rockland and Lund, 1983) and selectively link cells with similar feature preference (Gilbert and Wiesel, 1989) and colinear aligned RFs (Bosking et al., 1997; Schmidt et al., 1997).

- *Short-range connections*

  Short-range connections are rather unspecific for a particular orientation (Amir et al., 1993; Bosking et al., 1997; DeAngelis et al., 1994) and most likely belong to an inhibitory system (Kisvarday et al., 1994).

- *Modulating feedback*

  Several physiological studies indicate that feedback projections have a modulation or gating rather than generating effect on cell activities (Hirsch and Gilbert, 1991; Hupé et al., 1998;

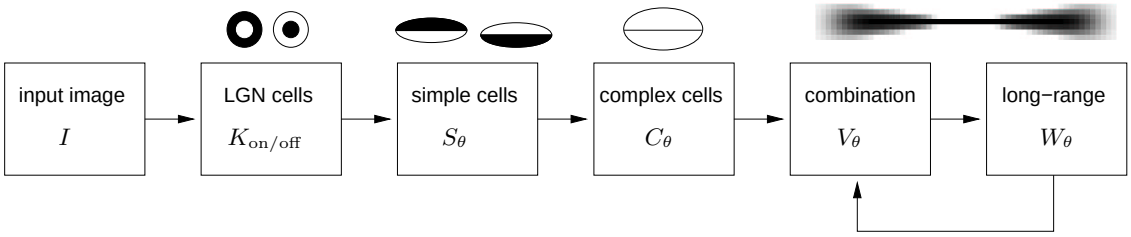


**Fig. 4.24:** Overview of model stages together with a sketch of the sample receptive fields of cells at each stage for 0° orientation. For the long-range stage, the spatial weighting function of the bipole filter is shown.

Salin and Bullier, 1995). Feedback alone is not sufficient to drive cell responses (Sandell and Schiller, 1982) and initial bottom-up activity is necessary to generate activity.

The model architecture is defined by a sequence of preprocessing stages and a recurrent loop of long-range interaction. We propose a functional architecture for the processing within the recurrent loop (Hansen et al., 2001; Neumann and Sepp, 1999). In this architecture of two interacting regions, let them be cortical layers or areas, each region has a distinctive purpose. The lower region serves as a stage of feature measurement and signal detection. The higher region represents expectations about visual structural entities and context information to be matched against the incoming data carried by the feedforward pathway.

Overall, the model realizes a simplified architecture of V1 (Gilbert, 1993) and is depicted in Fig. 4.24.

### 4.4.2 Feedforward Preprocessing

In the feedforward path, the initial luminance distribution is processed by isotropic LGN-cells, followed by orientation-selective simple and complex cells. The interactions in the feedforward path are governed by basic linear equations to keep the processing in the feedforward path relatively simple and to focus on the contribution of the recurrent interaction. A more elaborated processing in the feedforward path would make use of, e.g., nonlinear processing at the level of LGN cells and simple cells (Hansen et al., 2001; Neumann et al., 1999). In our model, complex cell responses $C_\theta$ as output of the feedforward path (cf. Fig. 4.24) provide an initial local estimate of contour strength, position and orientation which is used as bottom-up input for the recurrent loop.

#### LGN On- and Off-Cells

Responses of isotropic LGN-cells are modeled by correlation of the initial input stimulus $I$ with a difference-of-Gaussians (DoG) operator. Two types of LGN cells are modeled, namely on and off, which generate rectified output responses $K_{\text{on/off}}$

$$\begin{aligned} K &= \text{DoG}_{\sigma_\text{c},\sigma_\text{s}} \star I \\ K_\text{on} &= [K]^+ \\ K_\text{off} &= [-K]^+ \ , \end{aligned} \tag{4.1}$$

where $\star$ is the spatial correlation operator and $[x]^+ := \max\{x, 0\}$ denotes half-wave-rectification. The DoG is parameterized by the standard deviation of the center and surround Gaussian ($\sigma_\text{c} = 1, \sigma_\text{s} = 3$), respectively.

#### Simple Cells

Simple cells in V1 have elongated subfields (on and off) which sample the input of appropriately aligned LGN responses. Input sampling is modeled by correlation with rotated, anisotropic Gaussians. The Gaussians are shifted perpendicularly to their main axis by $\pm\tau = 3$ to model left and right subfields of an odd-symmetric simple cell. Thus, e.g., for the on-channel, the equations read

$$\begin{aligned} R_{\text{on,left},\theta} &= K_{\text{on}} \star G_{\sigma_x,\sigma_y,0,-\tau,\theta} \\ R_{\text{on,right},\theta} &= K_{\text{on}} \star G_{\sigma_x,\sigma_y,0,\tau,\theta} \ . \end{aligned} \tag{4.2}$$

The activations of the off-channel are computed analogously.

Simple cells are modeled for two polarities (dark-light and light dark) in $O_{\max} = 4$ orientations ($\theta = 0,\ \pi/O_{\max}, \ldots, (O_{\max} - 1)\,\pi/O_{\max}$). The standard deviations of the anisotropic Gaussians are set to $\sigma_y = 1, \sigma_x = 3\,\sigma_y$. For each orientation, the simple cells activity is computed by pooling the two subfield responses. The equation for light-dark (ld) and dark-light (dl) simple cells read

$$\begin{aligned} S_{\text{ld},\theta} &= R_{\text{on,left},\theta} + R_{\text{off,right},\theta} \\ S_{\text{dl},\theta} &= R_{\text{off,left},\theta} + R_{\text{on,right},\theta} \ . \end{aligned} \tag{4.3}$$

#### Complex Cells

Cortical complex cells are polarity insensitive. Their response is generated by pooling simple cells of opposite polarities. Before pooling, simple cells of opposite polarities compete and are spatially blurred. The corresponding equations read

$$\begin{aligned} \widetilde{S}_{\text{ld},\theta} &= \left[ (S_{\text{ld},\theta} - S_{\text{dl},\theta}) \star G_{\sigma_x,\sigma_y,0,0,\theta} \right]^+ \\ C_\theta &= \widetilde{S}_{\text{ld},\theta} + \widetilde{S}_{\text{dl},\theta} \ . \end{aligned} \tag{4.4}$$

### 4.4.3 Recurrent Long-Range Interaction

The output of the feedforward preprocessing defines the input to the recurrent loop which has two stages, namely a combination stage where bottom-up and top-down inputs are fused, and a stage of long-range interaction.

#### Combination Stage

At the combination stage, feedforward complex cell responses and feedback long-range responses are combined. Feedforward inputs $C_\theta$ and feedback inputs $W_\theta$ are added and subject to shunting interaction

$$\begin{aligned} &\partial_t V_\theta = -\alpha_V V_\theta + (\beta_V - V_\theta)\ \text{net}_\theta \ , \qquad \text{where} \\ &\text{net}_\theta = C_\theta + \delta_V W_\theta \ . \end{aligned}$$

Solving the equation at equilibrium $\partial_t V_\theta = 0$ results in a normalization of activity

$$V_\theta = \beta_V \frac{\text{net}_\theta}{\alpha_V + \text{net}_\theta} \ . \tag{4.5}$$

The weighting parameter $\delta_V = 2$ is chosen so that dimensions of $C_\theta$ and $W_\theta$ are approximately equal, the decay parameter $\alpha_V = 0.2$ is chosen small compared to $\text{net}_\theta$ and $\beta_V = 10$ scales the activity to be sufficiently large for the subsequent long-range interaction. For the first iteration step, feedback responses $W_\theta$ are set to $C_\theta$.

### Long-Range Interaction

At the long-range stage, the contextual influences on cell responses are modeled. Directional sensitive long-range connections provide the excitatory input. The inhibitory input is given by undirected interactions in both the spatial and orientational domain. Long-range connections are modeled by a bipole filter (Grossberg and Mingolla, 1985a). The spatial weighting function of the bipole filter is narrowly tuned to the preferred orientation, reflecting the highly significant anisotropies of long-range fibers in visual cortex (Bosking et al., 1997; Schmidt et al., 1997) (see Fig. 4.24, top right). The size of the bipole is about four times the size of the RF of a complex cell.

Essentially, excitatory input is provided by correlation of the feedforward input with the bipole filter $B_\theta$. A cross-orientation inhibition prevents the integration of cells responses at positions where responses for the orthogonal orientation also exist. The excitatory input is governed by

$$\mathrm{net}_\theta^+ = \left[V_\theta - V_{\theta\perp}\right]^+ \star B_\theta \ ,$$

where $\star$ denotes spatial correlation and $[x]^+ = \max\{x, 0\}$ denotes half-wave rectification.

The profile of the bipole filter is defined by a directional term $D_\varphi$ and a proximity term generated by an isotropically blurred circle $C_r \star G_\sigma$ where $r = 25$, $\sigma = 3$. The detailed equations read

$$\begin{aligned} B_{\theta,\alpha,r,\sigma}(x,y) &= D_\varphi \cdot C_r \star G_\sigma \\ D_\varphi &= \begin{cases} \cos(\frac{\pi/2}{\alpha}\varphi) & \text{if } \varphi < \alpha \\ 0 & \text{otherwise} \end{cases} , \end{aligned}$$

where $\varphi$ is defined as $\mathrm{atan2}\,(|y_\theta|, |x_\theta|)$ and $(x_\theta,\, y_\theta)^{\mathrm{T}}$ denotes the vector $(x, y)^{\mathrm{T}}$ rotated by $\theta$. The operator $\cdot$ denotes point-wise multiplication of two filter kernels or 2D matrices. The parameter $\alpha = 10°$ defines the opening angle of $2\alpha$ of the bipole. The factor $\frac{\pi/2}{\alpha}$ maps the angle $\varphi$ from the range $[-\alpha; \alpha]$ to the domain $[-\pi/2; \pi/2]$ of the cosine function with positive range. A plot of the bipole filter for a reference orientation of 0° is depicted in Fig. 4.25.

**Fig. 4.25:** Spatial weighting function for the long-range interaction for a reference orientation of 0°.

Responses which are not salient in the sense that nearby cells of similar orientation preference also show strong activity should be diminished. Thus an inhibitory term is introduced which samples activity from both orientational $\widetilde{g}_{\sigma_o,\theta}$, $\sigma_o = 0.5$, and spatial neighborhood $G_{\sigma_{\mathrm{sur}}}$, $\sigma_{\mathrm{sur}} = 8$,

$$\mathrm{net}_\theta^- = \mathrm{net}_\theta^+ \circledast \widetilde{g}_{\sigma_o,\theta} \star G_{\sigma_{\mathrm{sur}}} \ ,$$

where $\circledast$ denotes correlation in the orientation domain.

The orientational weighting function $\widetilde{g}_{\sigma_o,\theta}$ is implemented by a 1D Gaussian $g_{\sigma_o}$, discretized on a zero-centered grid of size $O_{\max}$, normalized, and circularly shifted so that the maximum value is at the position corresponding to $\theta$. Fig. 4.26 shows a sample plot of $\widetilde{g}$ for $O_{\max} = 4$, $\sigma_o = 0.5$ and $\theta = 2\,\pi/O_{\max} = \pi/2$.

The spatially inhibitory interactions reflect the extend of short-range connections. The parameterization of this inhibitory interaction results in an effective spatial extension of about half the size of the excitatory long-range interaction implemented by the bipole filter.

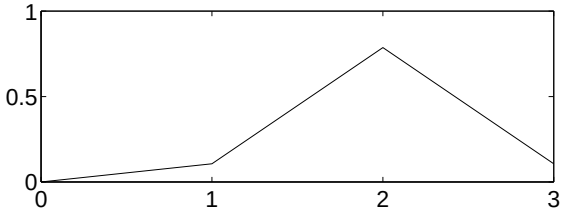


**Fig. 4.26:** Orientational weighting function.

Excitatory and inhibitory terms combine through shunting interaction

$$\partial_t W_\theta = -\alpha_W W_\theta + \beta_W V_\theta \left(1 + \eta^+ \operatorname{net}_\theta^+\right) - \eta^- W_\theta \operatorname{net}_\theta^- \ .$$

The equation is solved at equilibrium, resulting in a nonlinear, divisive interaction

$$W_\theta = \beta_W \frac{V_\theta \left(1 + \eta^+ \operatorname{net}_\theta^+\right)}{\alpha_W + \eta^- \operatorname{net}_\theta^-} \ . \tag{4.6}$$

where $\alpha_W = 0.2$ is the decay parameter and $\eta^+ = 5$, $\eta^- = 2$, and $\beta_W = 0.001$ are scale factors.

The multiplicative contribution of $V_\theta$ ensures that long-range connections have a modulating rather than generating effect on cell activities (Hirsch and Gilbert, 1991; Hupé et al., 1998). The result of the long-range stage is fed back and combined with the feedforward complex cell responses, thus closing the recurrent loop. The shunting interactions governing both the long-range interactions and the combination of feedback and feedforward input ensure a saturation of activities after a few recurrent cycles.

**Table 4.7.** Summary of properties of the model proposed in this work.

| Paper | **Hansen and Neumann (1999d)** |
|---|---|
| *Computational Architecture* | |
| structure | feedback |
| *Preprocessing Stage* | |
| image data | artificial and natural images |
| extraction | oriented filters |
| input elements | oriented edges |
| *Grouping Stage* | |
| class | geometric |
| principles (spatial) | proximity, colinearity |
| (feature) | parallelism |
| aggregation | additive |
| combination of lobes | additive |
| CC interactions | spatio-orientational divisive inhibition |
| output | oriented edges |

The model is robust against parameter changes which is mainly caused by the compressive transformation equations employed. For the combination of responses (Eq. 4.5), however, it is crucial to have activities in both streams of similar order of magnitude. Also the relative RF sizes must not be substantially altered. The current parameter setting results in relative RF sizes of complex cells : isotropic short-range filter : long-range interaction of about $1 : 2.5 : 4$, assuming a cut-off of the Gaussians at $2\sigma$ (or 95% of the total energy).

For comparison to the models of contour grouping discussed in Sec. 4.3, the properties of the model are summarized in Tab. 4.7.

## 4.5 Simulations

In this section we show the competencies of the model in a number of simulations. We shall begin with an artificial image which allows to highlight and evaluate various properties of the proposed model. We shall then demonstrate the performance of the model on a number of real world images. The resulting edge images for both the complex cell stage and the long-range stage show pooled responses which are obtained by summing over all orientations.

The values of the model parameters as specified in Sec. 4.4 are employed in all simulations. The model response saturates after a few recurrent cycles. Unless denoted otherwise, a number of 12 recurrent cycles is employed. The resulting edge images for both the complex cell stage and the long-range stage show pooled responses which are obtained by summing over all orientations.

### 4.5.1 Processing of Noisy Artificial Images

In a first simulation a synthetic stimulus of a noisy square is employed. The image is heavily corrupted by high-amplitude additive Gaussian noise of standard deviation equal to 100% of the luminance difference at the edge (so-called 100% Gaussian noise). Figure 4.27 demonstrates the functionality of lateral long-range interaction for the enhancement of coherent structure. Outline contrasts are detected and subsequently enhanced such that the activities of salient contrast as well as orientation significance is optimized.

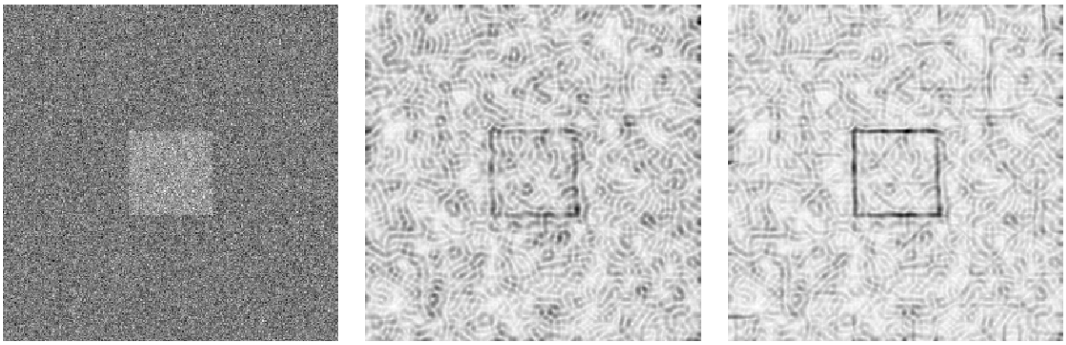

**Fig. 4.27:** Processing of a square pattern with additive high amplitude noise. The size of the images is $256 \times 256$ pixels. *Left to right:* Input image, initial complex cell responses, and the result of the recurrent long-range processing.

The capabilities of the model can be further assessed by a close-up of the top left corner of the square (Fig. 4.28). The simulations demonstrate three important properties of the recurrent long-range interaction:

1. *Contour enhancement:* The weak initial orientation estimates of the square contour are enhanced.
2. *Noise suppression:* Spurious noisy activities in the background are suppressed.
3. *Preservation of multiple activities:* At corners, the significant orientations are preserved.

The preservation of multiple activities near corners and junctions provide an implicit signature or labeling of such points which can be used to define a corner detection scheme, as shall be detailed

in the next chapter. These higher order features play a significant role in object recognition and depth segregation (e.g., Biederman, 1985).

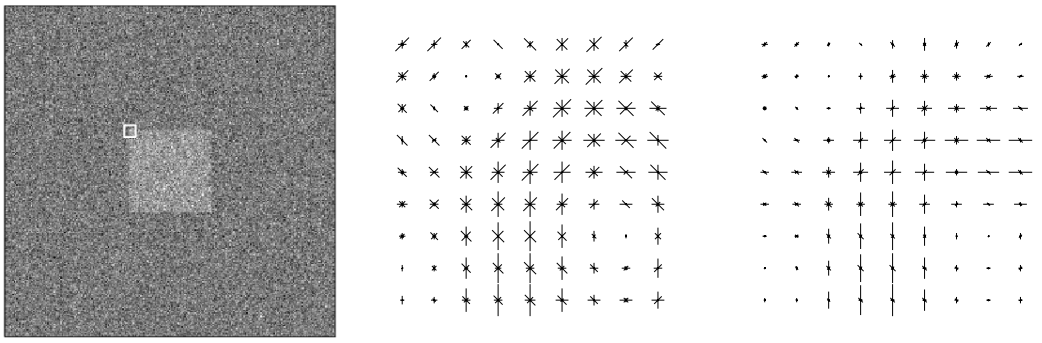

**Fig. 4.28:** Orientation plot of a close-up of the top left corner of the noisy square. The size of the close-up images is $9 \times 9$ pixels. *Left to right:* Input image and close-up of the top left corner (white square inset in the input image) for complex cell responses and long-range responses. In the close-ups, three important properties of the long-range interaction can be seen: i) enhancement of the orientation coaligned to the contour, ii) suppression of noisy activity in the background, and iii) preservation of the significant orientations at corners.

Overall, the proposed model circuit is thus capable to perform main tasks of low and med-level vision, namely contour enhancement, noise suppression and junction detection within a single architecture. The model thus defines a computational framework suggesting how these tasks can be accomplished by the neural machinery.

### 4.5.2 Quantitative Evaluation

In the following, we shall define two measurements of *contour saliency* and *orientation significance* which allow for a quantitative, more rigorous assessment of the capabilities of the model.

#### Contour Saliency

One core property of the proposed circuit is contour enhancement: The activity at contour locations should be increased compared to background locations. This property has already been demonstrated in the initial simulation of the noise square (Fig. 4.27) and becomes even more prominent in a close-up of the top border of the square (Fig. 4.29). Besides the strengthening of the contour activity, the results also demonstrate that the model circuit can close gaps (as in the right part of the contour), as long as some amount of initial bottom-up activity is present.

To quantify the contour enhancement, we use a saliency measurement as suggested by Li (1999a,c). Li defines the net saliency $S$ at each position as the response of the maximally activated orientation

$$S = \max_{\theta} X_\theta \quad .$$

The relative enhancement of contour activity is then defined as the ratio of the mean saliency along the contour and the mean saliency measured over all positions:

$$r = \frac{\overline{S}_{\text{contour}}}{\overline{S}_{\text{all}}} \quad .$$

A second measurement compares the standard deviation of the saliencies at all positions $\sigma_{\text{all}}$ with the difference of the mean saliencies:

$$z = \frac{\overline{S}_{\text{contour}} - \overline{S}_{\text{all}}}{\sigma_{\text{all}}} \quad .$$

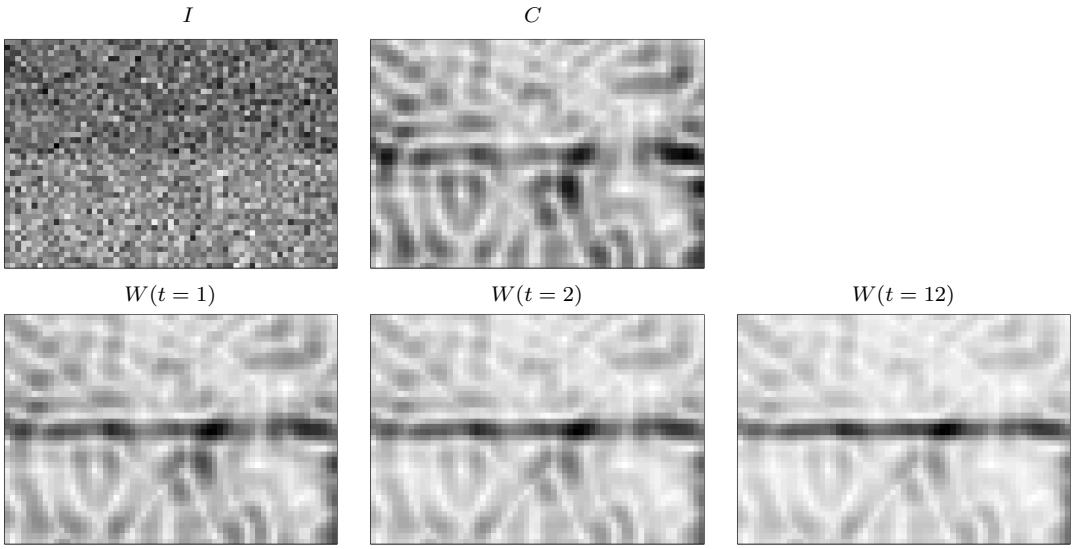


**Fig. 4.29:** Close-up of the processing results obtained for the top contour of the noisy square. The size of the close-up images is $61 \times 42$ pixels. *Top row, left to right:* Input image and the initial complex cell responses. *Bottom row, left to right:* Result of the recurrent long-range processing at three discrete time steps ($t = 1, 2, 12$).

A salient contour is characterized by high values of $r$ and $z$.

We can now evaluate the contour enhancement property of the proposed model in terms of the defined saliency measure. A plot of the temporal evolution of both saliency values $(r, z)$ for the noisy square stimulus is given in Fig. 4.30. For the contour values, we considered positions defined by a stripes of two pixel width along each of the four sides of the square. Mean saliency values at these positions are then compared against mean saliency values measured over all pixels of the whole input stimulus. The curves show an increase in saliency for both values during the recurrent processing: initial saliency values of $(r, z) = (2.3, 2.9)$ as obtained at the complex cell level are increased to $(r, z) = (5.7, 7.0)$ after 12 recurrent cycles. The simulation results show that the model successfully enhances the contour saliency of a feedforward input of cluttered responses with low contour saliency.

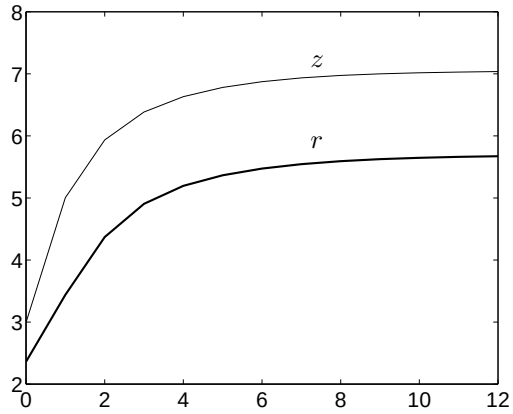


**Fig. 4.30:** Temporal evolution of contour saliency for the noisy square. The ordinate denotes the saliency, the abscissa denotes the discrete time steps ($t = 0, 1, \ldots, 12$). *Top solid curve* denotes the temporal evolution of the $z$ value, the *bottom bold curve* denotes the $r$ value. The initial values $(r, z) = (2.3, 2.9)$ monotonically increase during the recurrent interaction and level off at about 12 recurrent cycles at values $(r, z) = (5.7, 7.0)$.

We have also investigated modifications of the saliency definition as given above. In these modifications, we defined the net saliency as the sum across all orientations at a given position. Further, the mean value at the contour has been compared against the mean value of the background instead of comparing it against the mean value of all positions (including the contour). For all these modifications qualitatively similar runs of saliency curves have been obtained, so we decided to keep the original definition.

Finally, we checked the effect of scale variations on the saliency results. For a fixed scale of the preprocessing stages we modified the parameters of the long-range interaction. In particular, we varied the effective radius $r$ of the long-range filter, and the standard deviation $\sigma_{\text{sur}}$ of the Gaussian used for the short-range inhibition. Recall that the parameters are set to $r = 25$ and $\sigma_{\text{sur}} = 8$ in the original model as specified in Sec. 4.4. This setting results in a relative RF size of complex cells : isotropic short-range filter : long-range interaction of about $1 : 2.5 : 4$. For the scale variations, we simulated three different scenarios, with decreasing relative RF size of the long-range filter, namely 3, 2, and 1.5 times the filter size used at the complex cell level. The relative size of the short range filter is kept fixed at approximately $2.5/4 = 0.625$ the size of the long-range filter. The parameter settings for the three scenarios are as follows:

1. scale 3, relative RF sizes $1 : 1.875 : 3$, ($r = 19$, $\sigma_{\text{sur}} = 6$)
2. scale 2, relative RF sizes $1 : 1.25 : 2$, ($r = 13$, $\sigma_{\text{sur}} = 4$)
3. scale 1.5, relative RF sizes $1 : 1 : 1.5$, ($r = 9$, $\sigma_{\text{sur}} = 3$)

The resulting temporal evolution of the saliency values are depicted in Fig. 4.31. The plots show that the saliency drops to considerably smaller values if the range of interactions at the long-range stage is decreased.

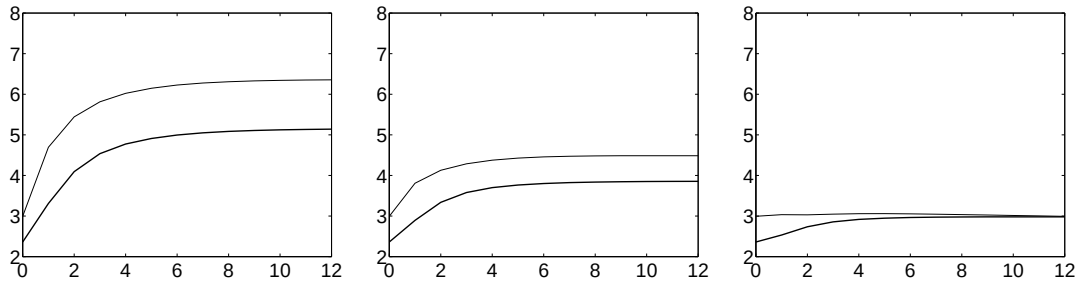


**Fig. 4.31:** Temporal evolution of contour saliency for the noisy square under variations of the scale of long-range interactions. The ordinate denotes the saliency, the abscissa denotes the discrete time steps ($t = 0, 1, \ldots, 12$). In each plot, the *top solid curve* denotes the temporal evolution of the $z$ value, the *bottom bold curve* denotes the $r$ value. *Left to right:* Saliency values for relative RF sizes of $1 : 1.875 : 3$, $1 : 1.25 : 2$ and $1 : 1 : 1.5$. The increase of the initial values $(r, z) = (2.3, 2.9)$ is considerably smaller if the range of the long-range interaction decreases: the final significance values are $(r, z) = (5.1, 6.4)$ for the scale 3, $(r, z) = (3.9, 4.5)$ for the scale 2 and $(r, z) = (3.0, 3.0)$ for scale 1.5, as compared to $(r, z) = (5.7, 7.0)$ for the original parameter setting (see Fig. 4.30).

### Orientation Significance

The saliency measure as introduced above allows to compare the response at different positions, but does not allow to quantify the strength of a particular response relative to other responses of different orientations at the same position.

In order quantify this relative enhancement of contour response across orientations, a measure of orientation significance is defined. Orientation significance is a measure of orientation bandwidth which is bounded between zero and one. Cells not tuned for orientations have a zero orientation significance. Cells that are very sharply tuned have orientation significance values close to one.

The orientation significance is defined as the length of the vector resulting from summing all orientations normalized by the absolute sum of orientations (Batschelet, 1981; Ringach et al., 1997). The formal definition reads

$$\mathrm{osgnf}(W) := \frac{|\sum_\theta W_\theta \exp(2\mathrm{i}\theta)|}{\sum_\theta W_\theta} \quad . \tag{4.7}$$

The factor of 2 in the argument of the exponential function stretches the range of orientations $\theta \in [0; \pi]$ to the full turn $[0; 2\pi]$, such that circular statistics can be applied.

In a pilot study, the changes of orientation significance during recurrent long-range interaction are examined for the image of the noisy square. Figure 4.32 depicts the temporal evolution of the mean orientation significance for two patches of size of 2×40 pixels. One patch is placed at the contrast boundary (upper solid line) and the other at the background (lower solid line). The curves for the two patches show that the recurrent interaction increases orientation significance only at the borders while leaving the significance at the background almost unchanged.

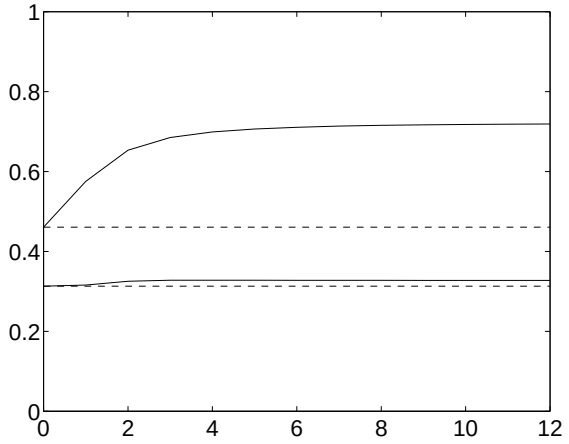


**Fig. 4.32:** Temporal evolution of mean orientation significance for a patch placed at the contrast boundary *(upper solid line)*, and a second patch placed at the background *(lower solid line)*. The *dashed horizontal lines* indicate the corresponding mean initial significance obtained for the complex cell responses. The abscissa denotes the discrete time steps ($t = 0, 1, \ldots, 12$), the ordinate denotes mean orientation significance. The curves show a pronounced increase in orientation significance at the borders (increasing from 0.46 to 0.72), while the significance at the background remains almost unchanged (increasing only slightly from 0.31 to 0.33). This corresponds to an increase in the ratio of the orientation significance at the border and at the background from 1.47 to 2.20 during the recurrent processing.

While the two curves depicting the temporal evolution of the orientation significance differ considerable for the background and the contour, a number of issues may need to be clarified. For example, one may intuitively presume a decrease of orientation significance at background locations. Further, one may wonder about the values assumed by the orientation significance function: Why do the values at the border level off at about 0.8 instead of approaching 1, and why, on the other hand, does the background significance assumes values about 0.33 instead of lower values near zero? In order to address these points and to develop some intuition about the values assumed by the orientation significance function, it is instructive to consider the orientation significance for a synthetic orientation distribution. A typical distribution resulting from the preprocessing scheme has a maximum response at the orientation of the local image structure $\theta_{\mathrm{opt}}$, a near zero response at the orthogonal orientation $\theta_{\mathrm{nonopt}}$, and some residual responses at all other orientations. For the purpose of analysis we consider an idealized synthetic orientation response distribution $W_{\mathrm{synth}}$ having unit-valued residual responses, zero response at the orthogonal direction and

a response $w_{\text{opt}} > 1$ at the preferred orientation:

$$W_{\text{synth}} = \begin{cases} w_{\text{opt}} > 1 & \text{if } \theta = \theta_{\text{opt}} \\ 0 & \text{if } \theta = \theta_{\text{nonopt}} = \theta_{\text{opt}} + \pi/2 \\ 1 & \text{else .} \end{cases}$$

For an even number of orientations $O_{\max} \mod 2 = 0$, the responses can be grouped in pairs of mutually orthogonal orientations $\theta$ and $\theta_\perp$, whose responses $w_\theta$ and $w_{\theta_\perp}$ provide antagonistic contributions to the orientation significance:

$$\begin{aligned} w_\theta \exp(2\mathrm{i}\theta) + w_{\theta_\perp} \exp(2\mathrm{i}(\theta + \pi/2)) &= w_\theta \exp(2\mathrm{i}\theta) + w_{\theta_\perp} \exp(2\mathrm{i}\theta) \exp(\mathrm{i}\pi) \\ &= (w_\theta - w_{\theta_\perp}) \exp(2\mathrm{i}\theta) \ . \end{aligned}$$

The residual responses obey $w_\theta = w_{\theta_\perp} = 1$, such that these responses cancel each other. The only remaining term in the numerator of the significance function is then given by the pair of optimal and nonoptimal responses. The denominator is the sum of all responses. The orientation significance for the synthetic orientation response distribution $W_{\text{synth}}$ thus reads

$$\text{osgnf}(W_{\text{synth}}) = \frac{|(w_{\text{opt}} - w_{\text{nonopt}}) \exp(2\mathrm{i}\theta_{\text{opt}})|}{w_{\text{opt}} + (O_{\max} - 2)} = \frac{w_{\text{opt}}}{w_{\text{opt}} + O_{\max} - 2} \ .$$

For a number of $O_{\max} = 4$ as used in the simulations, a sample plot of the different values assumed by the orientation significance function for varying strength of the optimal orientation $w_{\text{opt}} = 1, 2, \ldots 10$ is depicted in Fig. 4.33. In the case of the minimal and maximal value used ($w_{\text{opt}} = 1$ and $w_{\text{opt}} = 10$), the orientation significance assumes values of 0.33 and 0.83, respectively. These values are in good agreement with the orientation significance at the background and the final orientation significance at the contour as measured in the simulation of the noisy square stimulus (compare Fig. 4.32). Increasing the number of orientations results in a shift of the orientation significance curve toward zero and a slightly steeper sloop of the curve. For $O_{\max} = 8$, e.g., the significance values for $w_{\text{opt}} = 1$ and $w_{\text{opt}} = 10$ are given by 0.14 and 0.63, respectively.

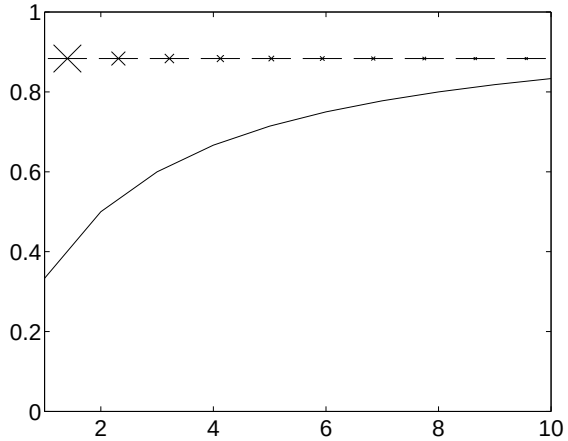


**Fig. 4.33:** Evaluation of orientation significance for a synthetic orientation response distribution. The abscissa denotes the response strength of the preferred orientation, $w_{\text{opt}} = 1, 2, \ldots 10$, the ordinate denotes the corresponding orientation significance. The 10 orientation plots inset a the top depict the corresponding orientation distribution.

The above analysis has clarified that the orientation significance cannot decrease below 0.33 (given a zero response at the orthogonal orientation), and that an orientation significance of about 0.8 results from a tenfold higher response at the preferred orientation compared to the residual responses. Consequently, values significantly below 0.33 can only arise for equal responses along all

orientations. On noisy artificial as well as on natural images, however, the local image structure always has a small orientation bias, leading to higher responses at this orientation and vanishing or near zero responses at the orthogonal orientation. A pronounced decrease of orientation significance during long-range processing would require an increase or even generation of activity orthogonal to the local image structure, thus violating the idea of modulating feedback.

So far, we have shown the competency of the model to increase orientation significance only for a single noisy image, i.e, for a fixed contrast and noise level and a particular realization of the noise process. For a more complete assessment of the competency of the model, the temporal evolution of mean orientation significance is evaluated for a square image using a larger variation of the input parameters (contrast and noise levels) and for a number of different realizations of each noise level (Fig. 4.34). The term "contrast" here denotes the contrast amplitude, i.e., the difference between maximal and minimal luminance values in the input image. We have simulated a square of 0.1 contrast, corrupted with 10%, 20%, 50% and 100% Gaussian noise (Fig. 4.34, left) and a square of higher contrast of 0.2, corrupted with 5%, 10%, 20% and 40% Gaussian noise (Fig. 4.34, right). The different and smaller noise values for the high-contrast stimulus are chosen to guarantee input luminance values in the range $[0; 1]$. The simulations results are averaged over 100 different realizations of the noise process to exclude effects resulting from a particular realization. The results show that the capability of the model circuit to increase orientation significance along contours is robust against changes of both the input contrast and the noise level.

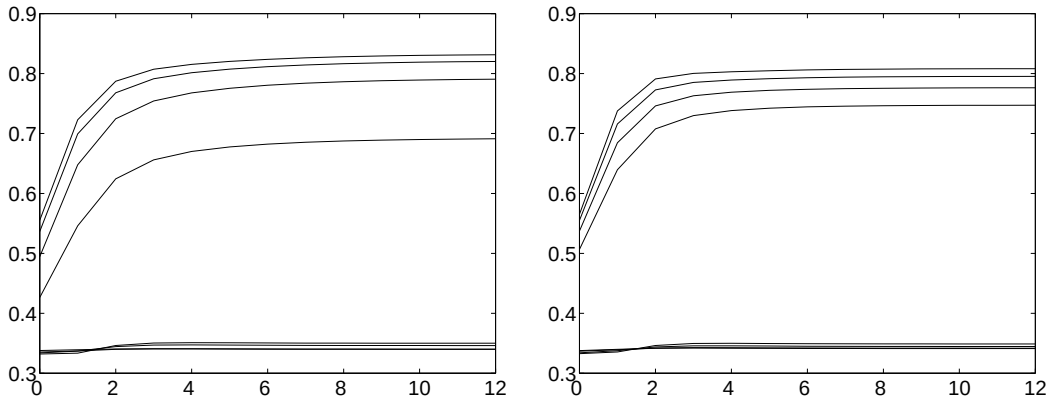


**Fig. 4.34:** Temporal evolution of mean orientation significance under variation of the input contrast and noise level. Results are averaged over 100 different realizations of the noise process. *Left:* Contrast of 0.1, corrupted with 10%, 20%, 50% and 100% Gaussian noise. *Right:* Contrast of 0.2, corrupted with 5%, 10%, 20% and 40% Gaussian noise. The *top four curves* in each plot show the mean orientation significance along the contours of the square for increasing noise levels *(top to bottom)*. The *bottom curves* show the corresponding orientation significance at the background. The results show an enhancement of orientation significance along the contour, while at the background the orientation significance remains almost the same.

### 4.5.3 Response to Curved Patterns

The bipole-shaped weighting function used in the present model is tuned for colinear respectively near-colinear contrast configurations in accordance with anatomical and psychophysical findings. However, many objects in both in man-made and natural environments have a curved shape outline. How does the model perform when processing such stimuli? To address this question, we conducted a set of simulations using circular segments of different radii, which are corrupted which additive Gaussian noise. The contrast value is kept fixed at 0.1 in this experiment. Sample plots of the input stimuli used are depicted in Fig. 4.35.

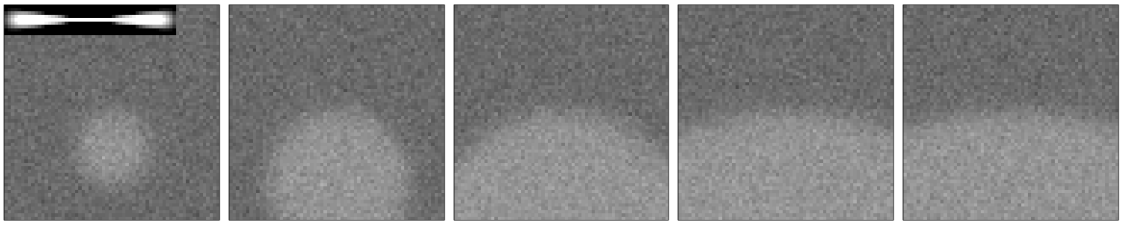

**Fig. 4.35:** *Left to right:* Circles of 0.1 contrast and radius 10, 20, 40, 80, and 100 pixels, corrupted with 20% Gaussian noise. The size of the images is $64 \times 64$ pixels. The inset in the first image shows the relative size of the bipole filter.

For each stimulus, the temporal evolution of orientation significance tangential to the circular segment is measured. In order to compensate for effects resulting from a particular realization of the noise process, results are averaged over 100 trails. The resulting plot of the temporal evolution of the orientation significance for circles of different radii is shown in Fig. 4.36.

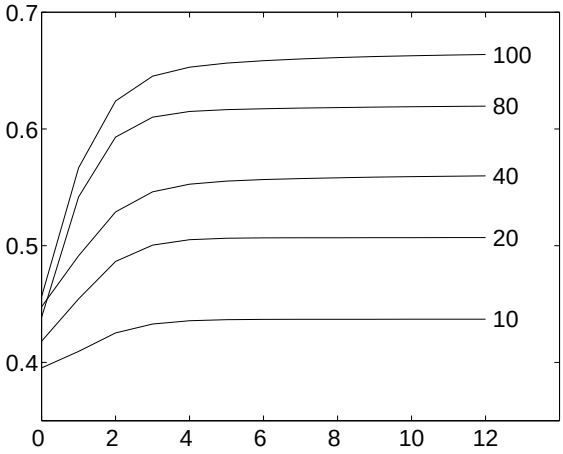


**Fig. 4.36:** Temporal evolution of orientation significance for noisy circles of 0.1 contrast and radius 10, 20, 40, 80, and 100 pixels, corrupted with 20% Gaussian noise. The abscissa denotes the discrete time steps ($t = 0, 1, \ldots, 12$), the ordinate denotes the orientation significance. The labels at the different curves denote the radius of the input circle. The curves shows for all radii a continuous increase in orientation significance during recurrent processing.

The results show for circles of all radii a constant increase in orientation significance by the recurrent long-range processing. The final value of the orientation significance depends on the radius: the larger the radius, the higher the orientation significance. Thus, as the pattern approaches a colinear configuration, orientation significance successively increases. We conclude that the model circuit can successfully process curved shape fragment to a certain degree which depends on the curvature of the input pattern.

### 4.5.4 Processing of Natural Images

To further examine the model, natural stimuli are used as input. We first employ a cell image depicted in Fig. 4.37. The results show that the outlines of the cells are enhanced by the recurrent long-range interaction. In particular, week initial estimates and low contrast measurements are enhanced, e.g., the longer colinear structures at the top border of the rightmost cell or smaller salient structures like the nuclei. This demonstrates how the proposed circuit can accentuate meaningful structures like object boundaries and may thus serve as a prerequisite of figure-ground segregation.

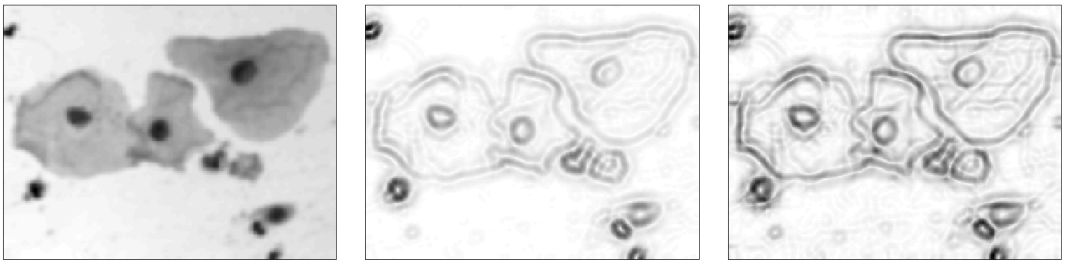

**Fig. 4.37:** Processing of a cell stimulus. The size of the images is $256 \times 195$ pixels. *Left to right:* Input image, initial complex cell responses, and V1 cell responses as the result of the recurrent long-range processing.

We have also processed a 3D image of a laboratory scene (Fig. 4.38). Those locations with high contrast complex cell responses and low orientation uncertainty are further stabilized by the recurrent loop. In addition, weak initial estimates such as the pedestal of the cube are enhanced. Also, weak spurious responses to the ground are suppressed.

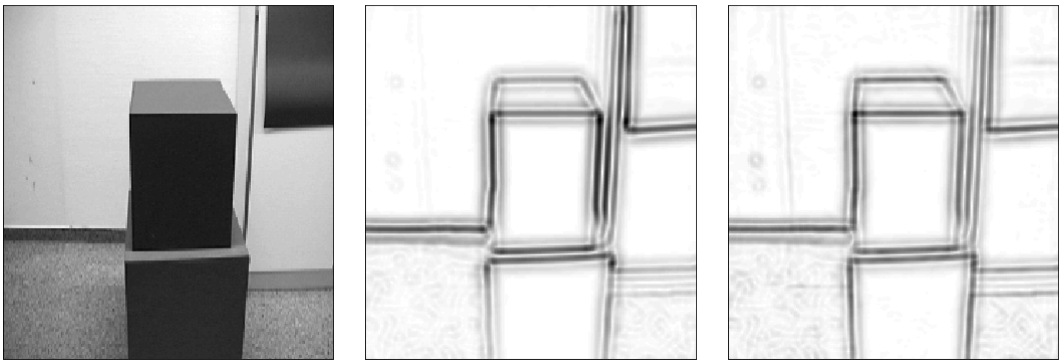

**Fig. 4.38:** Processing of a laboratory scene. The size of the images is $230 \times 246$ pixels. *Left to right:* Input image, initial complex cell responses, and V1 cell responses as the result of the recurrent long-range processing.

In a final set of simulations we used images from a collection of fruit and vegetable images (Williams, 2001). For the banana image, a closeup of the orientation distribution is shown for the complex cell responses and the long-range responses. Wrong orientation responses from noisy initial estimates at the complex cell stage are suppressed by the long-range interaction, and only the salient responses along the contour remains, which are coherent within a more broader context (Fig. 4.39).

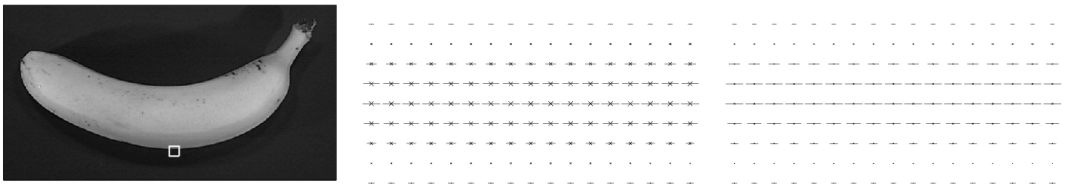

**Fig. 4.39:** Close-up of the processing results obtained for the banana image. The size of the close-up images is $17 \times 9$ pixels. *Left to right:* Input image and close-up of the lower contour (white square inset in the input image) for complex cell responses and long-range responses. Only those orientation responses which are coaligned along the contour are enhanced by the recurrent long-range processing relative to other spurious responses.

In a last simulation on natural images we used an image of a sweet potato. Simulations results for the potato image are shown in Fig. 4.40.

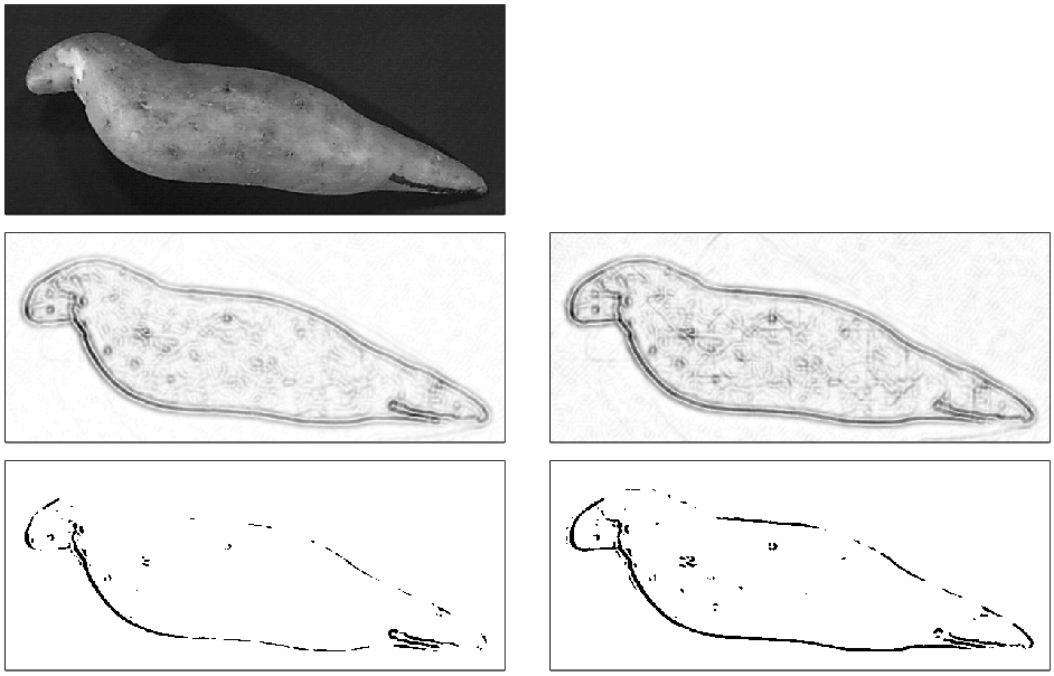

**Fig. 4.40:** Processing of a sweet potato image. The size of the images is $610 \times 256$ pixels. *Top row:* Input image. *Middle row, left to right:* Initial complex cell responses, and V1 cell responses as the result of the recurrent long-range processing. *Bottom row:* Corresponding thresholded images, thresholded at 30% of the relative maximum value for each image.

The initial complex cells responses to this stimulus already provide a rather good representation of the contours, but of considerable amplitude differences. These differences are compensated for by the long-range processing, yielding a more equal contour activity. To demonstrate this property, each image is thresholded at 30% of its relative maximum value (Fig. 4.40, bottom row). In the thresholded image after long-range interaction, longer parts of the object boundary are visible due to long-range processing. Some gaps however remain in the thresholded contour, where the initial response amplitude is considerably lower compared to other parts of the contour. These locations are not enhanced such that values as strong as at other locations of initially higher responses would be reached. The model responses thus can assume a range of values instead of "all-or-none" responses only, in accordance with the idea of analog sensitivity (Grossberg et al., 1997). The selective equalization of amplitude differences may ease the tasks of subsequent processing stages of, e.g., figure-ground determination and object recognition.

The simulations of natural images demonstrate core competencies of the model. Initial complex cell activations generated for localized high contrast contours are further stabilized. Initially weak activations in coherent spatial arrangements are enhanced. The results show that noisy low contrast arrangements can be significantly enhanced to form elementary items of smooth contour segments.

### 4.5.5 Simulation of Empirical Data

In order to address the empirical relevance of the model, the model is probed with fragmented contour patterns and texture stimuli such as used in the study of Kapadia et al. (1995). In particular, we investigated the relative contributions of surround inhibition by randomly oriented bars and long-range excitation from colinear flankers on the activity of a central bar element (Kapadia et al., 1995; Knierim and Van Essen, 1992). The simulation results are depicted in Fig. 4.41.

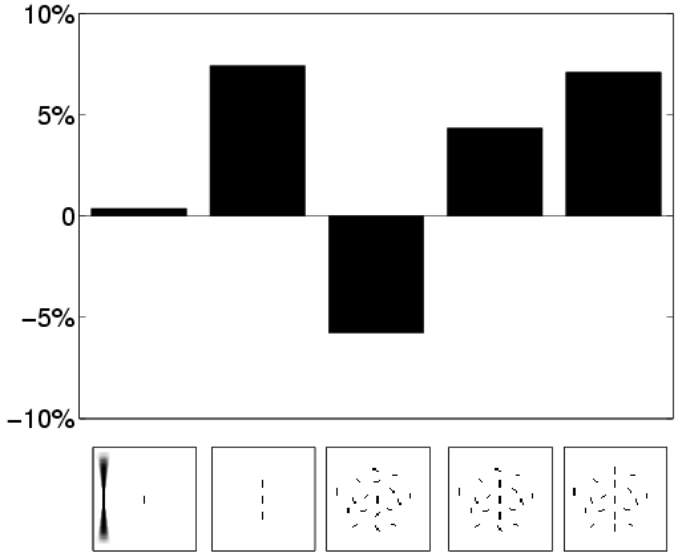


**Fig. 4.41:** Model response to generic contour patterns as use in an empirical study by Kapadia et al. (1995). The results depict the percent change of response *(bar plot, top)* for different input stimuli *(bottom)* relative to the response to a single bar *(bottom, leftmost stimulus)*. The left inset in the leftmost stimulus indicates the relative size of the bipole filter. Adding colinear flankers to a central bar element results in a response increase, adding a texture of randomly oriented bars results in a response decrease. The net effect of combined flankers and texture is a response increase, which strength depends on the strength of supporting activity from colinear flanking bars.

We first simulated the response to a single central bar element which serves as the reference activity. Adding two colinear flanking bars results in a increase of activity. The responses of feedforward complex cells to the flanking bars are integrated by the long-range filter, resulting in higher activation at the long-range stage. In the feedback loop, this higher activity selectively enhances the response to the central bar element. In another experiment, the central bar is embedded into a texture of randomly oriented bars. This results in a response decrease compared to the central bar alone. Here, the short-range inhibitory filter which samples activity summed across all orientations is activated by the randomly oriented bars. We finally simulate the response to the combined pattern of colinear flankers and a texture of randomly oriented bars. Here, the contributions of surround inhibition and long-range excitation result in a net effect of excitatory feedback, which is weaker than the response without the textured surround. Adding two more colinear flankers further increases the excitatory feedback and causes a higher activity of the central bar. Overall the model exhibits basic response characteristics of surround inhibition and long-range excitation in good agreement with empirical data.

## 4.6 Model Variant Using Early Feedback

In the model proposed above, the feedback signal is combined with the feedforward signal at a relatively high level, namely with the feedforward signal as generated by the complex cells (see Fig. 4.24 in Sec. 4.4). This circuit models the intralaminar recurrent interaction of layer 2/3 pyramidal cells. Physiological and anatomical studies indicate the existence of another interlaminar recurrent cycle in V1. This recurrent loop runs from layer 4 to layer 2/3 and feeds back via layer 5 and 6 to layer 4 (Bolz et al., 1989), were the feedback signal is combined with layer 4 simple cells. In this section we shall introduce a modification the model using early feedback, i.e., the feedback signal influences simple cell responses. Note that layer 5 and 6 are thought to integrate extrastriate cortico-cortical and thalamic input, which are not considered in the present V1 model. The core model architecture using early feedback is depicted in Fig. 4.42.

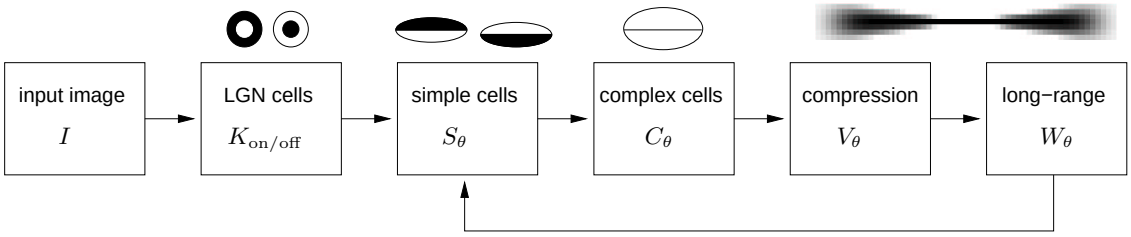


**Fig. 4.42:** Overview of model stages of the new model using early feedback. Instead of terminating at the combination stage (cf. Fig. 4.24), the feedback signal now feeds into the simple cell stage. The stages are shown together with a sketch of the sample receptive fields of cells 0° orientation. For the long-range stage, the spatial weighting function of the bipole filter is shown.

So far we have only motivated the new termination of the feedback connection within the simple cell layer. How should this combination be specified? Recall that we have introduced a mechanism of dominating opponent inhibition (DOI) in Chap. 3.DOI has been shown to reduce the noise sensitivity of the simple cells by introducing an adaptive threshold controlled by the activity in the opponent path. This allows to suppress activity at the background, but not at the contrast locations. However, in the simple cell model of Chap. 3, the contrast locations are only determined locally in a single feedforward sweep. The DOI mechanism could be made even more robust by integrating more global contour information such as provided by the long-range interactions. The idea is to use DOI only at the background and reduce the level of inhibition at contour locations signaled by high orientation significance provided by long-range feedback signals. These ideas are made rigorous in the following section, were the precise equations of the model variant with early feedback are provided.

### 4.6.1 Modification of the Model Equations

The model equations as introduced in Sec. 4.4 are modified at two stages, namely at the simple cell stage and the combination stage.

**Simple Cells** The simple cells combine input from different subfields. In the new model, the subfield responses $R$ are generated by a feedback controlled DOI signal $\Xi$. The equations Eq. 4.2 are replaced by

$$\begin{aligned} R_{\text{on,left},\theta} &= \left[(K_{\text{on}} - \Xi \cdot K_{\text{off}}) \star G_{\sigma_x,\sigma_y,0,-\tau,\theta}\right]^+ \\ R_{\text{on,right},\theta} &= \left[(K_{\text{on}} - \Xi \cdot K_{\text{off}}) \star G_{\sigma_x,\sigma_y,0,\tau,\theta}\right]^+ \quad , \end{aligned} \tag{4.8}$$

where the feedback control is defined as

$$\Xi = \xi - \text{osgnf}(W) \ . \tag{4.9}$$

The DOI parameter $\xi$ is set to 2, as determined in Sec. 3.8. In Eq. 4.9, high orientation significance along contours leads to a decrease of inhibition of the opponent channel. Low significance $\text{osgnf}(W) \approx 0$ results in dominating inhibition with $\Xi \approx \xi$. In the first iteration, $W$ is set to zero. The activations of the off-channel are computed analogously.

The old model results from setting $\Xi = 0$, the simple cell model with DOI as defined in Sec. 3.5 uses $\Xi = \xi = 2$.

The use of the orientation significance!for feedback control to control the DOI signal $\Xi$ is motivated mainly computationally: The orientation significance allows to robustly distinguish between oriented and non-oriented structures, is bound between $[0; 1]$ and grows during long-range processing. From a biological point of view, the computation of orientation significance requires only local computations within a hypercolumn. While the use of orientation significance for feedback control has not been shown empirically and hence is speculative, such a signal could principally be computed by the neural machinery.

**Compression Stage** At the compression stage (the former combination stage), the signal is compressed by using the complex cell responses $C_\theta$ instead of the feedback signal $W$ at every iteration step:

$$\begin{aligned} \text{net}_\theta &= C_\theta + \delta_V C = (1 + \delta_V)\, C \\ V_\theta &= \beta_V \frac{\text{net}_\theta}{\alpha_V + \text{net}_\theta} \ . \end{aligned} \tag{4.10}$$

### 4.6.2 Simulation Results Using Early Feedback

In this section we show simulation results for the model with early feedback. We start with a simulation of the noisy square image used in the simulations with the standard model. The simulation results are depicted in Fig. 4.43. The recurrent processing with the new model generated a slightly more pronounced contour activity (Fig. 4.43, top row as compared to simulation results for the standard model in Fig. 4.27), while the orientation plot of the top right corner appears to be nearly the same (Fig. 4.43, bottom row as compared to simulation results for the standard model in Fig. 4.28).

For a rigorous comparison of the two models, more qualitative measurements are needed. Such measurements have been defined above as contour saliency and orientation significance.

In a first experiment, we compare the contour saliency of the two models Fig. 4.44. The results show a stronger increase in saliency for the new model, especially in the relative enhancement as expressed by the $r$ value.

We have also redone the extensive simulations measuring the temporal evolution of orientation significance for a variety of input parameters (noise and contrast) and different realizations of each respective noise level Fig. 4.45. The results show a slightly better performance of the old model. In particular, the orientation significance of background locations is increased less for the old model. Due to the opponent inhibition at the simple cell level, the new model has a sharper orientation tuning, resulting in a slightly higher initial orientation significance at the the background (about 3.6 as compared to 3.3 in Fig. 4.34, left) as well as a slightly higher final significance value (about 0.37 as compared to 3.4 in Fig. 4.34, left). The small decrease of the orientation significance at the background after the first processing step is statistically not significant.

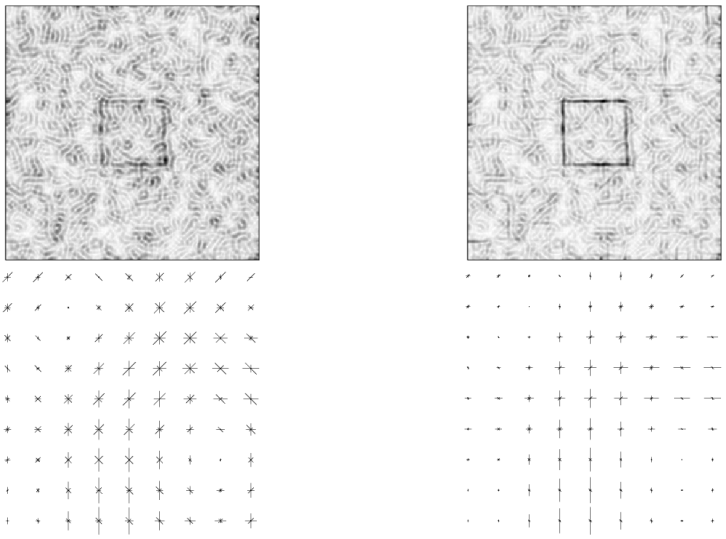

**Fig. 4.43:** Processing of a square pattern with additive high amplitude noise by the model with early feedback. The size of the images is $256 \times 256$ pixels. *Top row, left to right:* Initial complex cell responses and the result of the recurrent long-range processing. *Bottom row, left to right:* Orientation plot of a close-up of the top right corner of the noisy square for the complex cells and the long-range stage. The size of the close-up images is $9 \times 9$ pixels.

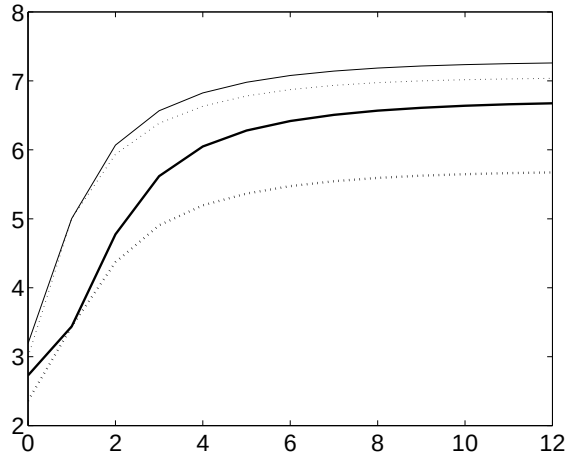


**Fig. 4.44:** Temporal evolution of contour saliency for the noisy square generated by the model with early feedback. The ordinate denotes the saliency, the abscissa denotes the discrete time steps ($t = 0, 1, \ldots, 12$). *Top solid curve* denotes the temporal evolution of the $z$ value, the *bottom bold curve* denotes the $r$ value. The *dashed lines* indicate the respective curves of the standard model. The saliency values monotonically increase during the recurrent interaction and level off at about 12 recurrent cycles at values $(r, z) = (6.7, 7.3)$ as compared to $(r, z) = (5.7, 7.0)$ for the standard model. The increase of the $z$ value is nearly the same for both models (4.0 for the standard model and 4.1 for the new model), but the increase of the $r$ value is considerably stronger for the model variant with early feedback (3.3 compared to 4.0).

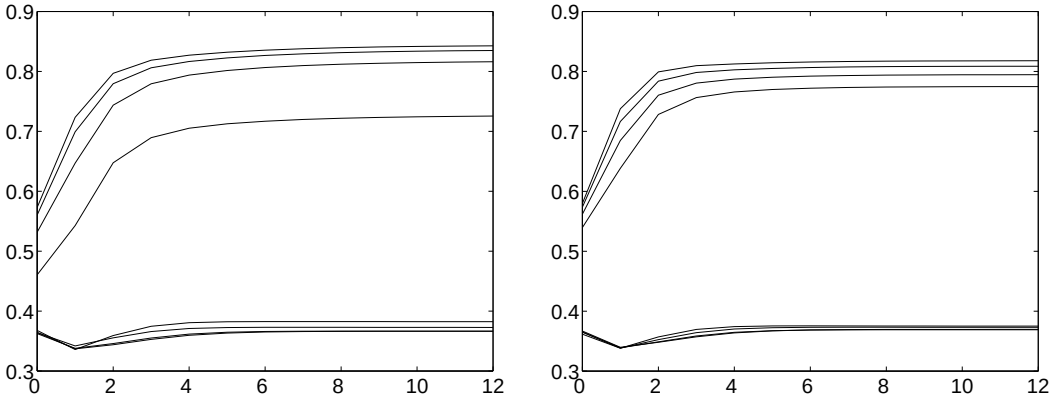


**Fig. 4.45:** Temporal evolution of mean orientation significance under variation of the input contrast and noise level for the new model with early feedback. Results are averaged over 100 different realizations of the noise process. The corresponding results for the standard model are depicted in Fig. 4.34. *Left:* Contrast of 0.1, corrupted with 10%, 20%, 50% and 100% Gaussian noise. *Right:* Contrast of 0.2, corrupted with 5%, 10%, 20% and 40% Gaussian noise. The *top four curves* in each plot show the mean orientation significance along the contours of the square for increasing noise levels *(top to bottom)*. The *bottom curves* show the corresponding orientation significance at the background. The results show an enhancement of orientation significance along the contour, while at the background the orientation significance increases only slightly.

The results can be interpreted in terms of complementary roles of the two kinds of feedback as employed in the model variants: while early feedback results in a more salient contour response, late feedback results in higher orientation significance. The results may indicate a functional role for the different kinds of feedback loops as observed in vivo.

## 4.7 Discussion and Conclusion

In this section we have proposed a model circuit of V1 contour processing utilizing long-range interactions and recurrent processing. The competencies of the proposed model have been shown for a variety of artificial and natural images. One core property of the proposed circuit is to enhance initially weak and noisy responses along contours. This property has been quantitatively evaluated using a measure of contour saliency. The results show that the model successfully enhances the saliency of contours in noisy images. The measure of contour saliency compares responses at different locations (i.e, at the contour and at other locations). A robust contour processing scheme however should also suppress conflicting responses at the same location, such that only the valid response in the contour directions remains. To quantitatively address this property, a complementary measure of orientation significance is introduced. We have evaluated the temporal evolution of orientation significance for a variety of different contrast and noise values for a number of different realizations of the noise process. In all cases the orientation significance at the contour is considerable increased. At the background, on the contrary, the orientation significance remains almost the same. We have then analyzed the orientation significance for a synthetic orientation distribution to develop some insight into the range of significance values assumed. In particular we have shown that for a prototypic synthetic orientation distribution the significance values at the background cannot decrease below a certain limit.

The model uses contour template tuned for parallel, near colinear orientation. Using input stimuli of noisy circles of varying radii, we have shown that the model can also enhance curved contours to a certain degree. As expected, the amount of enhancement directly relates to the curvature of the contour. We also evaluated the model response to contour fragments as used in an empirical

study by Kapadia et al. (1995). The model results are in good agreement with the empirical data: embedding a single line in a texture of randomly oriented lines results in a decrease of response due to surround suppression. Adding colinear flankers results in a response increase. Finally, we have suggested a model variant using early feedback, which exhibits improved performance on contour saliency.

A number of different model have been proposed for contour grouping, as extensively reviewed in Sec. 4.3. Among the first approaches that utilize recurrent processing for contour extraction is the Boundary Contour System (Grossberg and Mingolla, 1985a,b).A slightly revised version of the original BCS serves as the basic building block for a model of recurrent *intracortical* contour processing at V1 and V2 (Ross et al., 2000). Our model focus on the *intralaminar* processing within V1. Grossberg an coworkers suggest that V1 and V2 circuits are homologous and differ only in the size of the receptive fields, proposing that V2 is basically V1 at larger scale. In contrast, we propose that V1 and V2 have different functional roles, such that, e.g., cells responding to illusory contours occur in V2 and corner selective cells occur in V1. Our model uses *modulating* feedback, i.e., initial bottom-up activity is necessary to generate activity. Consequently, in contrast to the model of Grossberg and coworkers, the proposed model of V1 thus does not allow for the creation of illusory contours. Illusory contours evoke cell responses in V2 (von der Heydt et al., 1984) and have been investigated in a model of V1–V2 interactions (Neumann and Sepp, 1999). For natural stimuli, small responses due to sensor noise occur also between fragmented contours. We have shown that those small responses can be enhanced by the proposed model of V1, making it suitable for contour formation.

Other models selectively integrate activity from end-stop responses (Finkel and Edelman, 1989; Finkel and Sajda, 1992; Heitger et al., 1998; von der Heydt and Peterhans, 1989), while we use activity from initial contrast measurement which is sharpened by feedback modulation.

A alternative approach to model V1 recurrent interaction has been suggested by (Li, 1998, 1999a,b,c). The model of Li differs from our model in a number of features. Li uses a two-layer recurrent model of excitatory and inhibitory neurons, which interact by linear dynamics. In our model, nonlinear shunting equations are used to combine inhibitory and excitatory signals. Further, Li uses two types of highly tuned anisotropic filters for excitatory and inhibitory interactions, which gather input from various orientations. Instead, we employ an excitatory long-range filter which collects only input from the same orientation only, and use a more short-range inhibitory isotropic Gaussian filter which is not tuned for orientation at all. These more basic interaction structures are in agreement with empirical findings. On the other hand, long-range integration from different orientations based on a circularity constraint such as proposed by a large variety of models (Li, 1998; Parent and Zucker, 1989; Yen and Finkel, 1998) are highly speculative and have not been found in vivo.

To conclude, we have shown that basic tasks in early vision processing, such as contour enhancement, noise suppression and preservation of multiple activities at junctions can be realized within an integrated architecture derived from known principles of V1 contour processing.

Chapter 5

# Corner and Junction Detection

## 5.1 Introduction and Motivation

Corners and junctions are points in the image where two or more edges join or intersect. Whereas edges lead to variations of the image intensity along a single direction, corners and junctions are characterized by variations in at least two directions. In other words, edges are intrinsically one-dimensional signals, whereas corners and junctions are intrinsically two-dimensional signals. Compared to regions of homogeneous intensity, edges are rare events. Likewise, compared to edges, corners and junctions are rare events of high information content. Moreover, corners and junctions are invariant under different viewing angles and viewing distances. Both the sparseness of the signal and the invariance under affine transformations and scale variations establish corners and junctions as important image features. Points of intrinsically 2D signal variations such as corners and junctions have also been termed *keypoints* (Heitger et al., 1992; Michaelis, 1997) or *interest points* (Schmid et al., 2000).

Corners and junctions are useful for various higher level vision tasks such as the determination of occlusion relationships, matching of stereo images, object recognition and scene analysis. The importance of corner and junction points for human object recognition has been demonstrated in a number of psychophysical experiments (Attneave, 1954; Biederman, 1985, 1987). Junctions also seem to play an important role in the perception of brightness and transparency (Adelson, 1993, 2000; Anderson, 1997; Metelli, 1974; Todorović, 1997; Watanabe and Cavanagh, 1993; Zaidi et al., 1997). Recently Rubin (2001) proposed that local occlusion cues as signaled by junctions are necessary to trigger modal and amodal surface completion. Rubin showed that other cues such as surface relatability and surface similarity did not lead to the perception of illusory contours or amodal completion when junction cues are removed from otherwise unchanged stimuli.

In physiological studies, cells responding selectively to corners have been reported in monkey visual cortex (Hubel and Wiesel, 1968; Peterhans and von der Heydt, 1990). More recently, Das and Gilbert (1999) showed that correlated activities of V1 cells can signal the presence of smooth outline patterns as well as patterns of orientation discontinuity as occurring at corners and junctions.

In this chapter we propose a new method for corner and junction detection based on circular variance of contour responses within a model hypercolumn. Unlike other explicit approaches as proposed in computer vision (e.g., Harris, 1987; Mokhtarian and Suomela, 1998; Parida and Geiger, 1998), the new method can be regarded as an implicit detection scheme. Explicit schemes are characterized by the formalization and implementation of a special purpose corner detector, related to the idea of a corner detection cell. In contrast to these explicit schemes we propose a model where corners and junctions are implicitly characterized by distributed activity within a hypercolumn, signaling the presence of more than one orientation at a particular position. The proposed model thus follows a proposition of Zucker et al. (1989, p. 72):

> Crossings, corners, and bifurcations are represented at the early processing stages by multiple neurons firing within a "hypercolumn".

The chapter is organized as follows. We shall first give a brief overview of corner detection schemes proposed in the literature (Sec. 5.2) and point to different strategies for the evaluation of corner detectors (Sec. 5.3). Next, we shall detail the new corner detection scheme where corners and junctions are defined implicitly by distributed activity within a hypercolumn (Sec. 5.4). This activity can be robustly computed by recurrent long-range interactions (Chap. 4). Simulations results for a number of artificial and natural images are presented in Sec. 5.5. We shall then apply receiver operator characteristics (ROC) to assess the capabilities of the newly proposed detector in comparison to two other widely used corner detection schemes based on the structure tensor or on Gaussian curvature. Section 5.7 concludes the chapter.

## 5.2 Overview of Corner Detection Schemes

Given the importance of corner and junctions for various visual task, one may not be surprised to find a considerably large number of different computational approaches to corner and junction detection. Following Schmid et al. (2000), the approaches can be classified into three groups, namely *intensity-based methods*, *contour-based methods*, and *parametric model approaches*.

*Intensity-based methods* or *direct approaches* (Rohr, 1994a) detect 2D contrast variations based on the differential geometry of the raw input image (Koenderink and van Doorn, 1987). Intensity-based methods can be further subdivided into approaches based on the structure tensor (Förstner, 1986; Förstner and Gülch, 1987; Harris, 1987; Harris and Stephens, 1988; Nitzberg et al., 1993; Nitzberg and Shiota, 1992; Schmid et al., 2000) and approaches related to Gaussian curvature estimation by computing the Hessian determinant (Beaudet, 1978; Dreschler and Nagel, 1982; Kitchen and Rosenfeld, 1982; Nagel, 1983; Zetzsche and Barth, 1990). Another subclass of intensity-based methods are inspired by the visual system. Corner and junction keypoints are signaled by end-stopped cells which are computed by differentiation of combined even and odd Gabor filters (Heitger et al., 1992). Recently this method has been extended to multiple scales and color images (Würtz and Lourens, 2000).

The second class is defined by *contour-based methods*. These methods first extract the image contours by some edge detection scheme and then detect points of high curvature or inflection points along the image curve (e.g., Medioni and Yasumoto, 1987; Mokhtarian and Suomela, 1998; Pikaz and Dinstein, 1994).

*Parametric model approaches* fit a junction model to the image values within a small neighborhood (Baker et al., 1998; Parida and Geiger, 1998; Rohr, 1994b). These methods often involve a computationally expensive minimization of an energy function measuring the distance of the junction model from the image intensity values.

A more detailed discussion of different methods can be found in Lourens (1998), Mokhtarian and Suomela (1998) and Schmid et al. (2000).

## 5.3 Overview of Evaluation Approaches

A number of different methods have been proposed to evaluate the various approaches to corner detection. The different methods can be classified into methods based on *visual inspection*, *localization accuracy*, and *theoretical analysis* (Schmid et al., 2000).

A simple and popular method relies on *visual inspection* of the detection results: A number of different detectors is applied to a set of images and the results are presented. This methods suffers from a number of drawbacks. First, there is no commonly agreed data base of test images, such that in each study different sets of images are employed. However, a particular detector may perform well on one but worse on another image. Second, the visual inspection is highly

subjective since the number of false positive or missing response as well as the precise localization may not be judged correctly. Third, the results are shown only for a particular choice of the threshold separating corner from non-corner points. Such a threshold is involved in virtually every corner detection schemes, and detection results crucially depend on the proper choice of the threshold. Despite these drawbacks, visual inspection is a common method, useful to provide a first impression of the capabilities of different detection schemes.

*Localization accuracy* is another evaluation method and can be measured based on the correct projection of 3D scene points to 2D image points (Coelho et al., 1991; Heyden and Rohr, 1996). Since this method requires the precise knowledge of 3D points, the evaluation is restricted to simple scenes of, e.g., polyhedral objects.

The performance of various corner detectors can also be assessed by *theoretical analysis* (Deriche and Giraudon, 1990; Rohr, 1994a). Analytical studies are limited to particular configurations such as L-corners.

Recently, Schmid et al. (2000) have proposed an evaluation approach using two new criteria of repeatability and information content. Repeatability refers to the robust detection under different viewing conditions of the same scene, such as variation in viewing angle, distance or illumination, whereas information content measures the distinctiveness of the interest points. These criteria are important in tasks like image matching and tracking.

## 5.4 A Model for Corner and Junction Detection

Corner and junction configurations can be characterized by high responses for two or more orientations at a particular point in the visual space. The neural representation for multiple orientations at a particular point is given by cortical hypercolumn. Thus, intrinsically 2D signal variations can be characterized by significant activity of multiple neurons within a hypercolumn (Zucker et al., 1989). In the previous Chap. 4 we have proposed a model of recurrent long-range interactions in the primary visual cortex for contour enhancement. We have also pointed in Sec. 4.5 to a particular property of the proposed model, namely to preserve multiple activities at corners and junctions. Multiple oriented activities as measured by a simple feedforward mechanism is sensitive to noisy signal variations. By means of the recurrent long-range interactions, these initially noisy activities are evaluated within a larger context. In this recurrent process, only coherent orientation responses are preserved, i.e., responses which are supported by other responses in the spatial neighborhood, while incoherent responses are suppressed. Corners and junctions are thus robustly represented using a distributed representation of high multiple activity within a hypercolumn.

Such a distributed representation may suffice for subsequent neural computations, thus questioning the need for explicit corner detectors. For the purpose of visualization and comparison to other schemes for corner and junction detection however, an explicit representation is utilized. Following the above considerations, corners and junctions can be marked if multiple orientations are active and high overall activity exists within a hypercolumn, as detailed in the following.

For a the resulting activity of the long-range stage $W_\theta$ as defined in Eq. 4.6, the "junctioness" $J_{\mathrm{LR}}$ is given by

$$J_{\mathrm{LR}} = \mathrm{circvar}(W)^2 \sum_\theta W_\theta \ . \tag{5.1}$$

The function "circvar" is a measure of the circular variance within a hypercolumn. The squaring operation enhances the response if the circular variance assumes high values. Circular variance takes values in the range $[0;1]$. A circular variance of 0 denotes a single response, whereas a value of 1 occurs if all orientations have the same activity. Circular variance is thus complementary to

orientation significance as defined in Eq. 4.7:

$$\text{circvar}(W) = 1 - \text{osgnf}(W) = 1 - \frac{|\sum_\theta W_\theta \exp(2\text{i}\theta)|}{\sum_\theta W_\theta} \quad . \tag{5.2}$$

Circular variance has been used in a number of physiological studies to characterize the response properties of cells in V1 (McLaughlin et al., 2000; Pugh et al., 2000; Ringach et al., 1997).

To precisely localize the corner points, the junction map $J_{\text{LR}}$ is first smoothed with a Gaussian ($\sigma = 3$). Corner points are then marked as local maxima whose strength must exceed a fraction $\kappa = 0.25$ of the maximum response in the smoothed junction map. Local maxima are computed within a $3 \times 3$ neighborhood.

The corner detection scheme detailed in this section is defined for the resulting activity of the recurrent long-range interaction $W_\theta$. However, the proposed scheme does not rely on any specific properties of the long-range stage. In fact it is a general scheme which can principally be applied to any input where multiple orientation are locally represented as an (artificial) hypercolumn. Clearly, the results would vary depending on the chosen input. To demonstrate the advantages of the recurrent long-range interaction as opposed to a pure feedforward approach, it is therefore instructive to compare the detection results obtained for two different kinds of input, namely the long-range activity and the complex cell activity. This approach shall be pursued in the next section.

## 5.5 Simulations

In this section we show the competencies of the proposed corner detection scheme for a variety of synthetic and natural images. In particular, the robustness to noise and the localization properties of the new scheme are evaluated. In order to focus on the relative merits of the recurrent long-range interactions for the task of corner and junction detection, the proposed scheme is evaluated using two different kinds of input, namely the activity $W_\theta$ of the long-range stage and the purely feedforward activity $C_\theta$ of the complex cell stage. The model parameters are the same as specified in Sec. 4.4. In particular, the parameterization of the filters employed at the complex cell stage and the long-range stage are the same, resulting in relative RF sizes of complex cells : long-range interaction of about 1 : 4. Similar to Sec. 4.5, a number of 12 recurrent cycles is employed to generate the long-range responses.

A comparative evaluation of the new scheme with respect to other corner detectors can be found in Sec. 5.6.

### 5.5.1 Localization of Generic Junction Configurations

From the outset of corner and junction detection in computer vision, the variety of junction types have been partitioned into distinct classes like T-, L-, and W-junctions (Huffman, 1971). This catalog of junction configurations has been extended by other junction types such as Ψ-junctions, which seem to provide strong cues for inferring surface shading and reflectance (Adelson, 2000; Sinha and Adelson, 1993).

In this first simulation study we probe the model with a large variety of artificial images each depicting a generic junction configuration in the center of the image. We use input images of six different type of junctions, namely L-, T-, X-, Y-, W- and Ψ-junctions. For these images, the localization of the junction points based on the recurrent long-range representation is compared to the results obtained for the purely feedforward complex cell responses. The results are depicted in Fig. 5.1. The localization accuracy is computed as the Euclidean distances between the ground

**Table 5.1.** Localization accuracy of junction points in generic configurations based on complex cell and long-range response. The table shows the Euclidean distance in pixels between the detected location and the ground truth location. For all junctions, the method based on the recurrent long-range interaction can localize the junctions with higher precision (or equal precision for the X-junction).

| junction | complex | long-range |
|---|---|---|
| L | 3.54 | 0.70 |
| T | 2.92 | 0.5 |
| X | 0 | 0 |
| Y | 2.96 | 1.58 |
| W | 2.55 | 0.70 |
| Ψ | 6.5 | 3.5 |

truth location and the location as detected by either method. The localization results are summarized in Tab. 5.1. For all junction types, the localization is considerably better for the method based on the recurrent long-range interaction. The only exception occurs for the X-junction which is perfectly detected by both methods (not shown in Fig. 5.1). Furthermore, the T-junction gives a double response based on the complex cell responses, but accurately yields a single response based on the long-range representation.

To sum up, the evaluation of the local orientation responses within a more global context by means of the recurrent long-range interaction considerably improves the localization performance for a variety of generic junction configurations. Also the double response for the T-junction is eliminated.

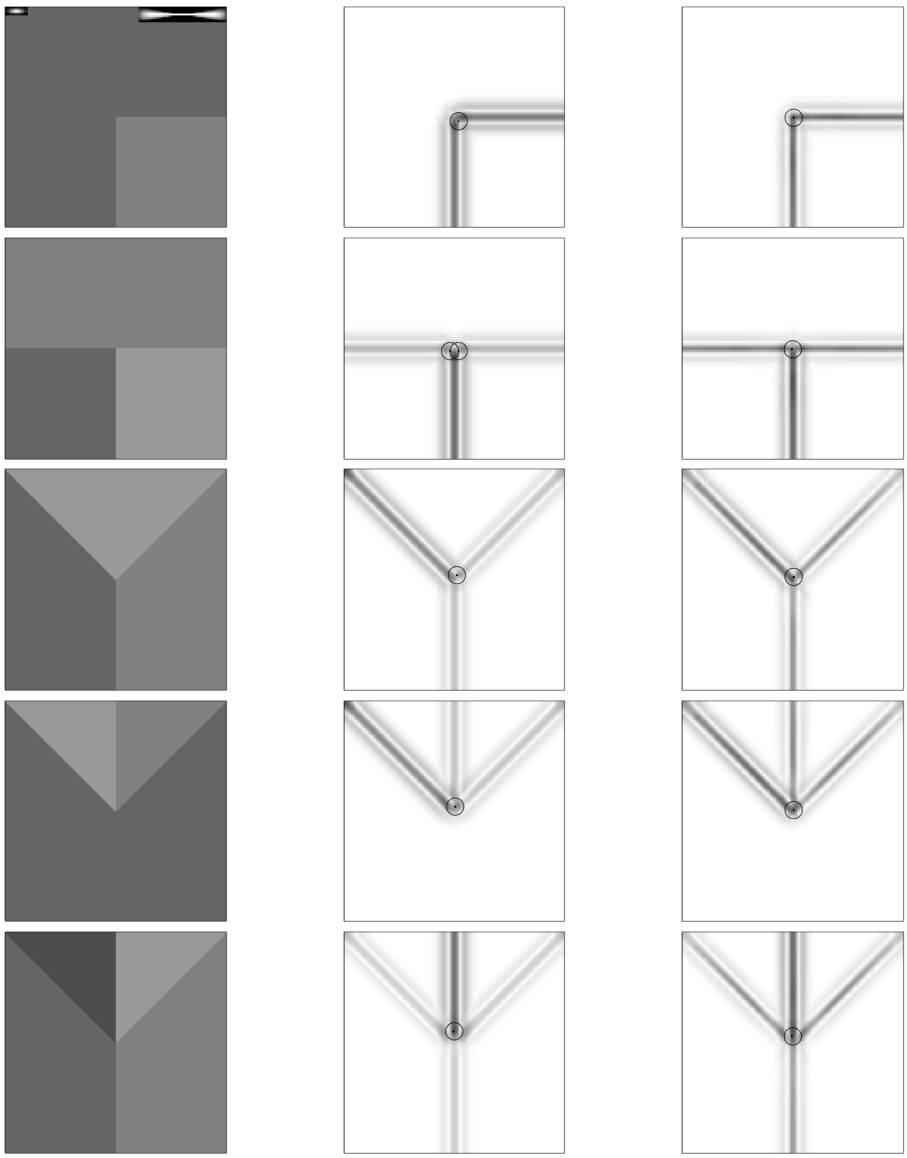

**Fig. 5.1:** Processing of generic junction configurations. The size of the images is $128 \times 128$ pixels. Five different junction configurations *(left row)* are processed, namely L-, T-, Y-, W- and $\Psi$-junctions. The detected corners based on the complex cell responses *(middle)* and on the long-range responses *(right)* are shown. The small insets at the top of the L-junction image show the relative sizes of the filter employed at the complex cell stage *(left inset)* and the long-range stage *(right inset)*. *Black circles* mark the positions of detected corners or junctions. For all junction configurations, the junctions points are localized with higher precision based on the recurrent long-range result. Also the double response for the T-junction is eliminated.

### 5.5.2 Processing of Attneave's Cat

Attneave (1954) has studied the processing and encoding of visual shapes in the context of information theory. He proposed that the image content can be compactly coded as a function of edge transitions instead of a dense field of luminance values. He further proposed that the edge map need not follow precisely the image gradients. Instead, only points of high curvature have to be represented, while edges of low curvature can be abstracted as straight lines. This idea is illustrated in a famous line drawing known as "Attneave's cat" (Fig. 5.2, left). In this drawing, the cat can be easily recognized though all contours of low curvature have been replaced by straight lines. Consequently, points of high curvature have been hypothesized to play an important role in shape representation and object recognition. Further evidence supporting the role of high curvature points comes from studies of Biederman (1985, 1987). Biederman showed that object perception of line drawings is severely impaired when corners, i.e., contours of high curvature, are removed, but largely preserved when contours of low curvature are deleted.

If corners and junctions are such important cues for object recognition they should be detected with high reliability. In the following study we probed the model circuit with Attneave's cat. The results show that the majority of corner and junctions points can be detected by the model using recurrent long-range interaction (Fig. 5.2, right). Compared to the results obtained by a purely feedforward scheme (Fig. 5.2, middle), the number of false responses is considerably reduced.

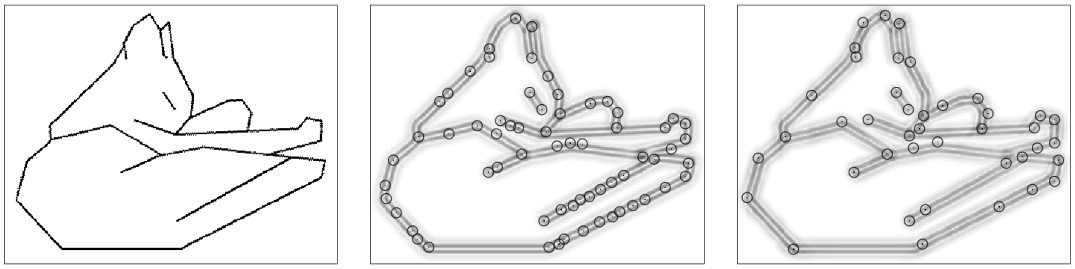

**Fig. 5.2:** Simulation of the corner detection scheme for Attneave's cat *(left)*. The size of the images is $256 \times 330$ pixels. The stimulus is an approximation of the cat's shape using a polygonal approximation based on curvature extrema or corners. The detected corners based on the complex cell responses *(middle)* and on the long-range responses *(right)* are shown and marked with *black circles*. After the recurrent long-range interaction, the corners are well detected, and a number of false positive responses as obtained based on the complex cell responses is eliminated.

### 5.5.3 Natural Images

In this section we show the corner and junction detection performance of both methods for a number of natural images. We begin with a series of three cube images in a laboratory environment (Fig. 5.3). At the complex cell stage, many false responses are detected due to noisy variations of the initial orientation measurement. These variations have been reduced at the long-range stage by the recurrent interaction, such that only the positions of significant orientation variations remain.

We further employed an image used in a corner detection study by Mokhtarian and Suomela (1998) and an image of a staircase. The input image together with the processing results for both the feedforward based complex cell responses and the recurrent long-range processing are depicted in Fig. 5.4. Again, the results obtained based on the long-range responses are superior to the results based on the purely feedforward complex cells responses.

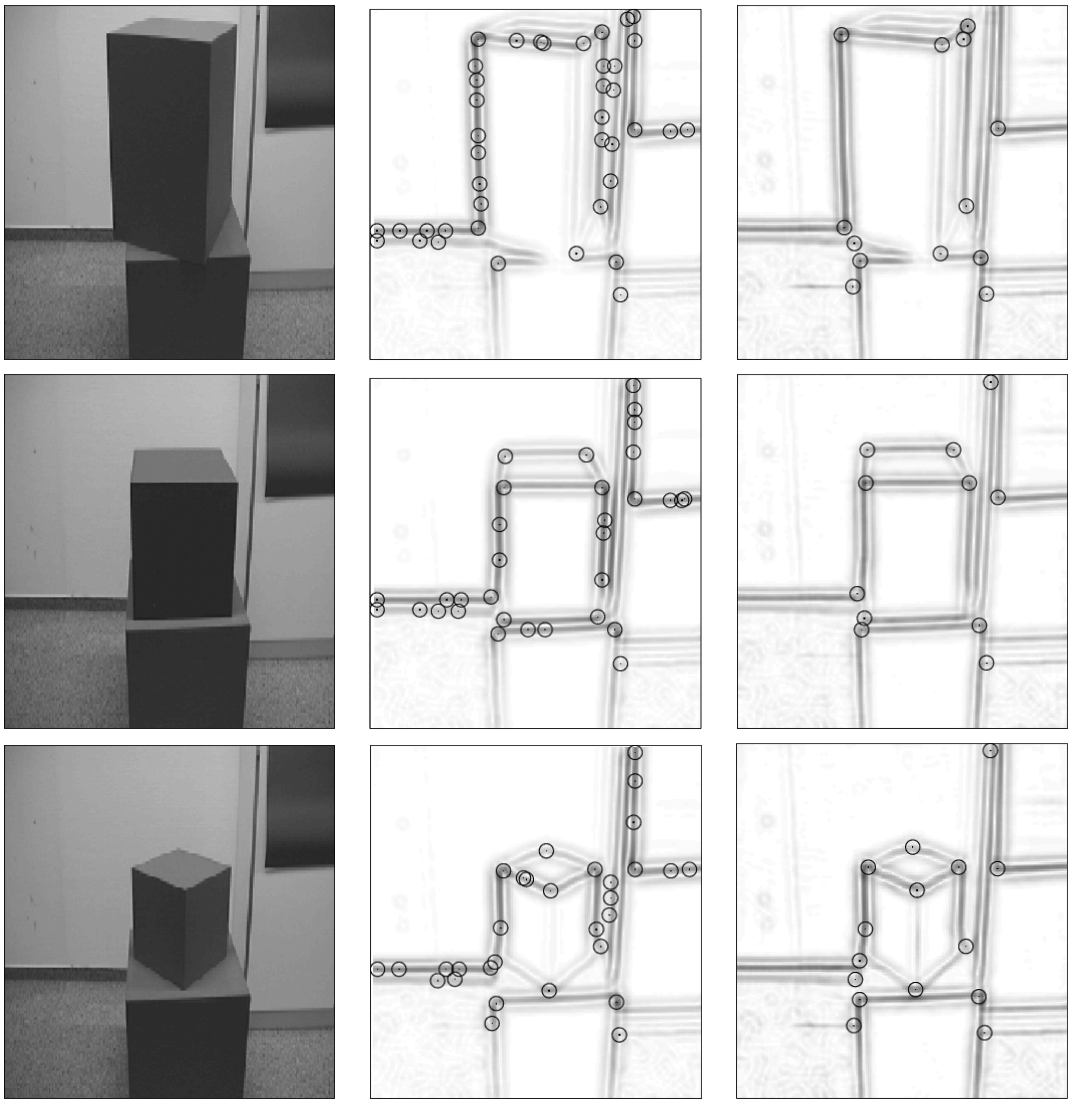

**Fig. 5.3:** Simulation of the corner detection scheme for cube images in a laboratory environment. The size of the images is $230 \times 246$ pixels. For three different images *(left row)* the detected corners based on the complex cell responses *(middle)* and on the long-range responses *(right)* are shown. *Black circles* mark the positions of detected corners or junctions. The recurrent long-range interaction results in a decrease of circular variance along object contours and thus eliminates a number of false positive responses.

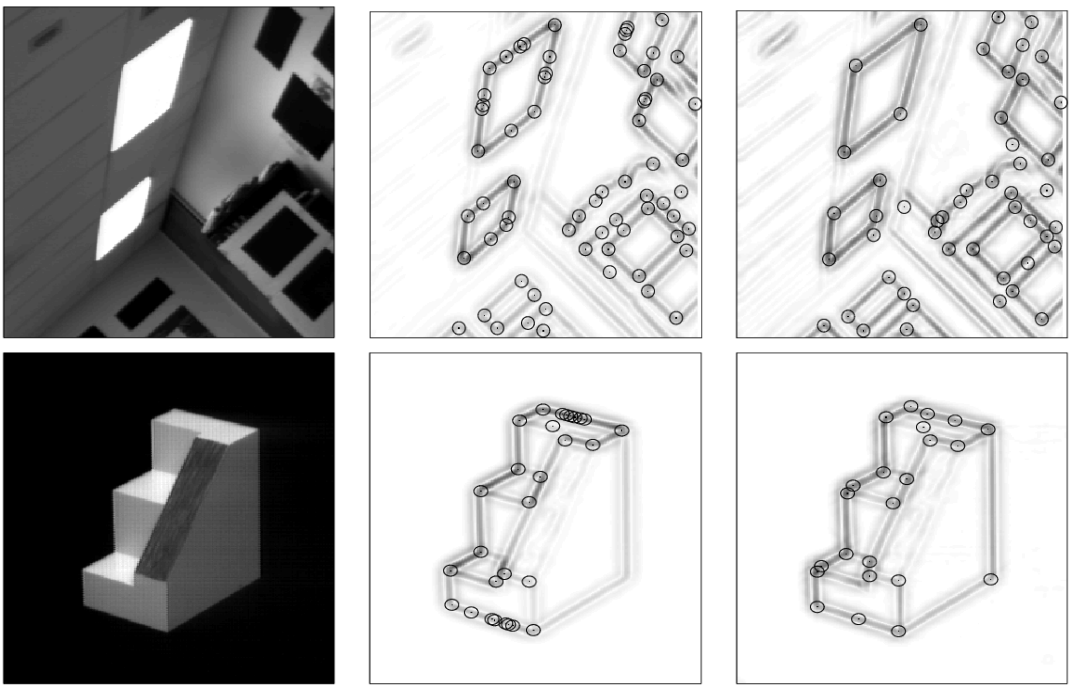

**Fig. 5.4:** Simulation of the corner detection scheme for a laboratory scene from Mokhtarian and Suomela (1998) *(top row)* and a staircase image *(bottom row)*. The size of the laboratory image is $256 \times 256$ pixels, the size of the staircase image is $240 \times 300$ pixels. The detected corners based on the complex cell responses *(middle column)* and on the long-range responses *(right column)* are shown and marked with *black circles*. After the recurrent long-range interaction, the corners are well detected in both images, and the number of false positive responses as obtained based on the complex cell responses is reduced.

## 5.6 Evaluation of Junction Detectors Using Receiver Operating Characteristics

In this section we shall compare the proposed junction detection scheme based on circular variance of long-range responses to other existing methods. We choose two widely used intensity-based methods, namely the method based on the local evaluation of (i) the structure tensor and (ii) the Gaussian curvature. These two methods compute the first- or second-order derivatives of the image intensity values, respectively. Robust computation of the derivatives is thus essential to both methods and can be realized by Gaussian presmoothing of the image or equivalently by using Gaussian derivatives of the proper order (cf. Sec. 3.4.3). In particular, the use of Gaussian derivatives instead of central differences has been shown to improve the Harris corner detector which is based on the structure tensor (Schmid et al., 2000). To allow for a fair comparison of methods, one has to ensure that all junction detectors operate on (at least approximately) the same scale (Lindeberg, 1998; Lindeberg and ter Haar Romeny, 1994b). Motivated by the above considerations we make use of Gaussian derivatives whose standard deviations are parameterized to fit the successive convolution of filter masks used to compute the complex cell responses.

The different detection schemes are evaluated within the framework of receiver operator characteristics (ROC). ROC curves allow to characterize different detectors over the full range of possible biases or thresholds. In virtually all junction detection schemes some kind of thresholding is involved, and the detection performance crucially depends on the determination of the "optimal" threshold value. A threshold-free evaluation of different detectors as provided by ROC analysis allows to separate the sensitivity of the detector from its threshold selection strategy. ROC analysis has its roots in signal detection theory and psychophysics (Coren et al., 1994; Green and Swets, 1974; van Tres, 1968) and is widely used for the evaluation of medical diagnostic techniques (e.g., Swets, 1979). In computer vision, ROC analysis has been used to evaluate edge detection methods (Abdou, 1978; Bowyer et al., 1999), color models (Alexander and Buxton, 1997), image segmentation (Southall et al., 2000), appearance identification (Edwards et al., 1999) and biometric identification using iris recognition (Daugman, 1999).

The section is organized as follows. First, in Sec. 5.6.1 we shall give a description of the two other junction detection schemes included in the comparative study. The detection schemes are based on the structure tensor and on Gaussian curvature, respectively. We shall introduce the basic idea behind each scheme and outline the implementation used. In particular, we motivate the parameterization of the Gaussian derivatives chosen to ensure that each method operates on (approximately) the same scale. A brief overview of ROC analysis is given in Sec. 5.6.2. After this introductory material we describe in Sec. 5.6.3 how ROC analysis can be applied to measure the performance of different junction detectors. Evaluation results obtained for our approach based on circular variance of long-range responses as compared to the methods based on Gaussian curvature or the structure tensor are presented in Sec. 5.6.4.

### 5.6.1 Junction Detectors Used for Comparison

In this section we summarize the basic idea behind the two other junction detection schemes used for comparison, point to different versions as proposed in the literature and outline their present implementation. Before the other methods based on the structure tensor and on Gaussian curvature are introduced, we motivate the use and selection of the same scale for each method to ensure a fair comparison.

### Parameterization of Gaussian Derivatives

The response of any corner detection scheme is scale dependent (Lindeberg, 1998; Lindeberg and ter Haar Romeny, 1994b). Consequently, when different corner detectors are to be compared, each scheme should operate on the same scale. The scale used in the proposed scheme based on long-range responses is determined by the scale of the complex cell responses in the feedforward preprocessing path (Sec. 4.4.2). Therefore we compute the successive convolution of filter masks used in the preprocessing path. The resulting filter $G_{\text{pre}}$ is approximated by an anisotropic Gaussian derivative $G_{\text{fit}}$, whose standard deviations $\sigma_{x,\text{fit}}, \sigma_{y,\text{fit}}$ determine the scale used to compute the derivatives of the other detection schemes used in the comparative study.

The preprocessing essentially consists of difference-of-Gaussians operator $\text{DoG}_{\sigma_\text{c},\sigma_\text{s}}$ at the stage of LGN cells, followed by an anisotropic difference-of-offset-Gaussians $\text{DooG}_{\sigma_x,\sigma_y,\tau}$ at the simple cell stage and a final smoothing with an anisotropic Gaussian $G_{\sigma_x,\sigma_y}$ at the complex cell stage. Note that we omit the orientation $\theta$ of the anisotropic Gaussians assuming a fixed orientation of, e.g., $\theta = 90°$. Further, we merge the computation of simple cell subfields (Eq. 4.2) and their additive combination (Eq. 4.3) in a single DooG operator. The equation defining the resulting preprocessing filter $G_{\text{pre}}$ thus reads

$$G_{\text{pre}} = \text{DoG}_{\sigma_\text{c},\sigma_\text{s}} \star \text{DooG}_{\sigma_x,\sigma_y,\tau} \star G_{\sigma_x,\sigma_y} \ , \tag{5.3}$$

where $\sigma_\text{c} = 1$, $\sigma_\text{s} = 3$, $\sigma_x = 3$, $\sigma_y = 1$ and $\tau = 3$. This mask can be approximated by a Gaussian derivative $G_{\text{fit}}$ with empirically determined standard deviations $\sigma_{x,\text{fit}} = 1.96$ and $\sigma_{y,\text{fit}} = 4.45$. The resulting preprocessing filter mask and its approximation by a Gaussian derivative is depicted in Fig. 5.5.

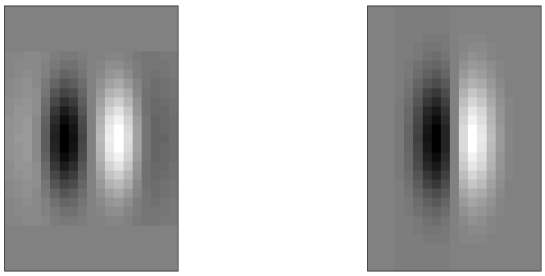

**Fig. 5.5:** *Left:* Weighting function resulting from successive convolution of filter masks used to compute the complex cell responses. *Right:* Fit of a first order Gaussian derivative mask with standard deviations $\sigma_{x,\text{fit}} = 1.96$ and $\sigma_{y,\text{fit}} = 4.45$.

### Structure Tensor Approach

The structure tensor (also called interest operator, scatter matrix or moment tensor (Weickert, 1998)) is an approximation of the local image structure using first-order derivatives. At each image location, the structure tensor measures the local change of the smoothed gradient within a neighborhood. Consider the tensor product $T$ of the image gradients $\nabla I$ defined as

$$T := \nabla I \otimes \nabla I = \nabla I \, \nabla I^\text{T} := \begin{bmatrix} I_x \\ I_y \end{bmatrix} \begin{bmatrix} I_x & I_y \end{bmatrix} = \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \ .$$

The structure tensor $T_\sigma$ then results from local averaging of this tensor product, which is usually implemented by componentwise convolution with a Gaussian $G_\sigma$:

$$T_\sigma := \begin{bmatrix} \overline{I_x^2} & \overline{I_x I_y} \\ \overline{I_x I_y} & \overline{I_y^2} \end{bmatrix} = \begin{bmatrix} G_\sigma * I_x^2 & G_\sigma * I_x I_y \\ G_\sigma * I_x I_y & G_\sigma * I_y^2 \end{bmatrix} \ .$$

The structure tensor $T_\sigma$ is a symmetric, positive semidefinite matrix with orthonormal eigenvectors and real-valued eigenvalues $\lambda_1, \lambda_2$, where $\lambda_1 > \lambda_2$ (Weickert, 1998). Using the notation

$$T_\sigma := \begin{bmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{bmatrix} \ ,$$

the eigenvalues are given by

$$\lambda_1 = \frac{1}{2}\left( t_{11} + t_{22} + \sqrt{(t_{11} - t_{22})^2 + 4t_{12}^2} \right)$$
$$\lambda_2 = \frac{1}{2}\left( t_{11} + t_{22} - \sqrt{(t_{11} - t_{22})^2 + 4t_{12}^2} \right) \ .$$

These eigenvalues and the corresponding rank of $T_\sigma$ characterize the local image structure, as summarized in Tab. 5.2. In particular, high values of the second eigenvalue $\lambda_2$ occur only for intrinsically 2D image structures. Consequently, the second eigenvalue of the structure tensor or its approximation has been proposed as a corner and junction detector.

**Table 5.2.** Description of the local image structure using the eigenvalues of the structure tensor.

| rank $T_\sigma$ | eigenvalues | local image structure |
|---|---|---|
| 0 | $\lambda_1 \approx \lambda_2 \approx 0$ | constant area |
| 1 | $\lambda_1 > 0$ and $\lambda_2 \approx 0$ | intrinsically 1D (edge) |
| 2 | $\lambda_1 > 0$ and $\lambda_2 > 0$ | intrinsically 2D (corner, junction) |

A number of different corner detectors based on the structure tensor have been suggested (Förstner, 1986; Förstner and Gülch, 1987; Harris, 1987; Harris and Stephens, 1988; Nitzberg et al., 1993; Nitzberg and Shiota, 1992; Schmid et al., 2000). The various methods differ by the computation of the first order image derivatives $I_x$ and $I_y$, the averaging function and the particular choice of an approximation of the second eigenvalue.

For example, the corner finder by Harris and Stephens (1988), also known as Plessey feature point detector, computes the derivatives of the image intensity surface by convolution with the mask $\begin{bmatrix} -2 & -1 & 0 & 1 & 2 \end{bmatrix}$ and uses a Gaussian of standard deviation $\sigma = 2$ for smoothing. Instead of computing the second eigenvalue of the matrix, the measure $\det T_\sigma - \alpha \operatorname{trace}(T_\sigma)^2$ is employed, with $\alpha$ being a weighting parameter.

In the present comparative, junction responses based on the structure tensor $J_{\mathrm{ST}}$ are signaled by the second eigenvalue of the structure tensor:

$$J_{\mathrm{ST}} = \frac{1}{2}\left( \overline{I_x^2} + \overline{I_y^2} - \sqrt{(\overline{I_x^2} - \overline{I_y^2})^2 + 4\overline{I_x I_y}^2} \right) \ . \tag{5.4}$$

The first order derivatives of the image $I_x$, and $I_y$ are computed using Gaussian derivatives with standard deviations as determined in Sec. 5.6.1 to fit the scale of the corner detection scheme based on long-range responses. In particular, the Gaussian derivatives are constructed by applying finite difference masks as follows:

$$G_x = G_{\sigma_{x,\mathrm{fit}}, \sigma_{y,\mathrm{fit}}} \star \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$$
$$G_y = G_x^{\mathrm{T}} \ .$$

A Gaussian with $G_\sigma$ with $\sigma = \max\{\sigma_{x,\mathrm{fit}}, \sigma_{y,\mathrm{fit}}\}$ is used for local averaging.

**Gaussian Curvature Approach**

Corner detection schemes using Gaussian curvature are based on second order derivatives of image intensity values. For an image $I$, the Gaussian curvature $K$ is defined as the product of the principal curvatures $\kappa_1$ and $\kappa_2$, i.e, the amount of greatest and least curvature:

$$\mathcal{K} = \kappa_{\min}\kappa_{\max} = \nu \det \underbrace{\begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix}}_{\text{Hessian } H} = \nu \left(I_{xx}I_{yy} - I_{xy}^2\right) \ ,$$

where $\nu = 1/(1+(\operatorname{grad} I)^2)^2$ is a scalar-valued normalization coefficient. The Hessian determinant is invariant against rotation and nonzero for 2D signal variations such as corners. Corners lead to two adjacent local extrema with opposite signs signaling hyperbolic ($\mathcal{K} < 0$) or elliptic points ($\mathcal{K} > 0$) of the image intensity surface.

Beaudet (1978) was the first to employ the concept of Gaussian curvature for the detection of corners and junctions. In this approach, interest points are marked by local maxima of the Hessian determinant $\det H$ (corresponding to saddle points of the image intensity surface). A number of related curvature based approaches have been suggested (Deriche and Giraudon, 1990; Dreschler and Nagel, 1982; Kitchen and Rosenfeld, 1982; Nagel, 1983; Zuniga and Haralick, 1983). Zetzsche and Barth (1990) use the Hessian determinant as a paradigm for the detection of multi-oriented or intrinsically 2D signal variations, since it combines two features regarded as essential, namely a nonlinear combination of orthogonally oriented filters $I_{xx}I_{yy}$ and a compensatory inhibitory interaction $-I_{xy}^2$.

It has been shown that these approaches are based on the same measurement, namely the gradient magnitude and the contour curvature (Noble, 1988). Due to the use of second order derivatives, curvature based approaches are relatively sensitive to noise (Haralick and Shapiro, 1992).

The implementation used in the present comparative study first computes the Hessian determinant:

$$D = I_{xx}I_{yy} - I_{xy}^2 \ .$$

The three derivatives of the image $I_{xx}, I_{yy}$, and $I_{xy}$ are computed using Gaussian derivatives which can be constructed by applying finite difference masks:

$$\begin{aligned} G_{xx} &= G_{\sigma_{x,\text{fit}},\sigma_{y,\text{fit}}} \star \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \\ G_{yy} &= G_{xx}^{\mathrm{T}} \\ G_{xy} &= G_{\sigma_{x,\text{fit}},\sigma_{x,\text{fit}}} \star \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \star \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^{\mathrm{T}} \ . \end{aligned}$$

Since 2D signal variations are characterized by two adjacent extrema of opposite sign, junctions responses are computed as smoothed absolute values of the determinant:

$$J_{\mathrm{GC}} = |D| \star G_\sigma \ , \tag{5.5}$$

where the standard deviation of the Gaussian is set to $\sigma = 3$.

## 5.6.2 Receiver Operating Characteristics (ROC)

Receiver operating characteristic (ROC) is a method to analyze detection rates across all values of a detection threshold (Green and Swets, 1974). ROC analysis has its origin in signal detection theory and is now applied in a number of diverse fields, in particular for the evaluation of medical diagnostic performance.

Consider a general signal detection experiment where the signal is either present or absent and the detector can either respond or not. The relative rates of the four possible outcomes, i.e., true

positive $t_p$, false positive $f_p$, false negative $f_n$ and true negative $t_n$ are conventionally represented in a fourfold table (Tab. 5.3). Dating back to Neyman and Pearson (1933), false negative and false positive responses are sometimes referred to as type I and type II errors, respectively. In the context of medical imaging the true positive $t_p$ rate is called "sensitivity", and the true negative rate $t_n$ is called "specificity".

**Table 5.3.** Fourfold table representing the rates of the four possible outcomes of a general signal detection experiment.

| | | **Signal:** present | absent |
|---|---|---|---|
| **Detection:** | response | true positive $t_p$ (hit) | false positive $f_p$ (false alarm) |
| | no response | false negative $f_n$ (miss) | true negative $t_n$ |

By definition, the values in each column of the outcome matrix sum to unity, i.e.,

$$t_p + f_n = 1 \qquad \text{and} \qquad f_p + t_n = 1 \ .$$

Further, since the detector by definition responds more strongly to the presence of the signal than to its absence, the following inequalities hold:

$$t_p > f_p \qquad \text{and} \qquad t_n > f_n \ .$$

The values of the outcome matrix show the responses for a single bias of the observer, or technically spoken, for a single threshold of the detector. Varying the threshold but not the signal results in different entries in the outcome matrix obtained for the same observer. For example, as the observer's decision criterion varies from a conservative to a more liberal one (technically corresponding to a decrease of the threshold), the rates of both true and false positive responses increase. ROC analysis provides a means to compensate for this variability by considering the outcome matrices over the full range of possible thresholds. A ROC curve is generated by plotting the proportion of true positive responses $t_p$ against the proportion of false positive responses $f_p$ as obtained for all possible values of the threshold.

The theoretical and methodological bases of ROC analysis is provided by signal detection theory. According to signal detection theory, the two cases of signal absent (or noise N) and signal present (or signal-plus-noise SN) generate corresponding sensory responses. Both responses can be described by normal distributions $P_N$ and $P_{SN}$ with different means $\mu_N$ and $\mu_{SN}$ and same standard deviation $\sigma$ (Fig. 5.6). These responses can be thought of, e.g., as the mean firing rates of neurons responding to a particular stimulus. Since the signal evokes a stronger response than the noise, the means of both distributions satisfy $\mu_N < \mu_{SN}$. The distance $d'$ between the signal-plus-noise and the noise distribution (in units of the standard deviation $\sigma$) can be used as a measure of the strength of the signal or, in a complementary view, of the sensitivity of the observer:

$$d' = (\mu_{SN} - \mu_N)/\sigma \ . \tag{5.6}$$

The distance $d'$ can be estimated from true positive rates $t_p$ and false positive rates $f_p$ using the $z$-transformation. The $z$-transformation is the inverted normal distribution (also called inverted error function) and measures the distance of a score from the mean of a distribution in standard deviation units (Ehrenstein and Ehrenstein, 1999). Formaly, the distance $d'$ is given as the difference between $z(t_p)$ and $z(f_p)$:

$$d' = z(t_p) - z(f_p) \ , \tag{5.7}$$

where the $z$-transformation is the inverted error function defined in terms of the error function erf as

$$z(p) := \text{erfinv}(p) = \{x : \text{erf}(x) = p\} \ ,$$

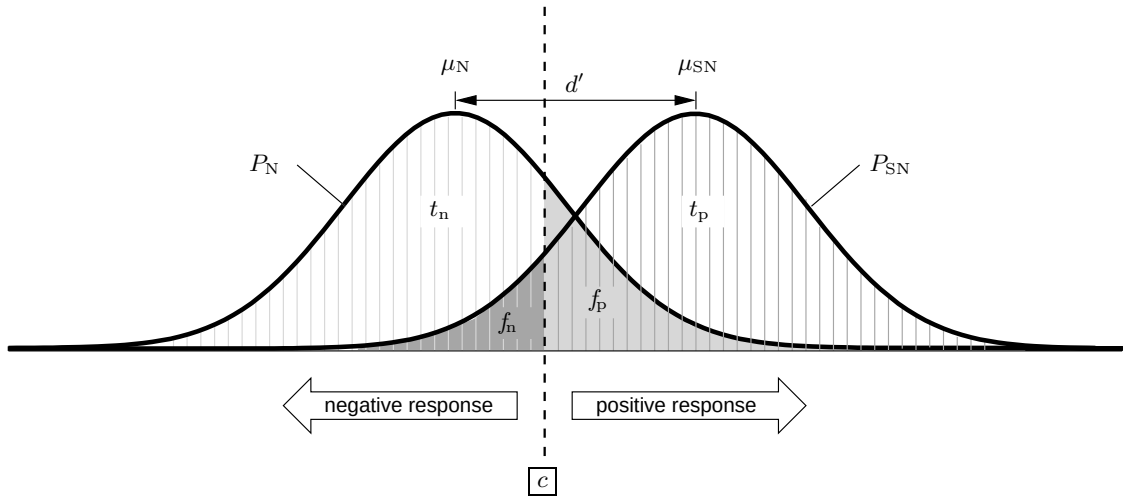


**Fig. 5.6:** Distribution of responses to noise $P_N$ and signal-plus-noise $P_{SN}$ with means $\mu_N$ and $\mu_{SN}$ in a general signal detection experiment. The abscissa denotes the magnitude of sensory response, the ordinate denotes the likelihood of occurrence. The choice of a particular decision criterion $c$ divides the continuum of sensory activities into two areas causing a negative or positive response. Depending on the distance $d'$ between both distributions the criterion $c$ determines the four possible outcomes of the signal detection experiment, i.e., true positive $t_p$, false positive $f_p$, false negative $f_n$ and true negative $t_n$ rate.

where

$$\mathrm{erf}(x) := \int_{-\infty}^{x} g_{\sigma=1}(\xi)\ \mathrm{d}\,\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\xi^2/2)\ \mathrm{d}\,\xi\ .$$

The equivalence between both expressions for $d'$ in Eq. 5.6 and Eq. 5.7 can be shown be rewriting Eq. 5.6 as follows:

$$d' = (\mu_{SN} - \mu_N)/\sigma = (\mu_{SN} - c)/\sigma + (c - \mu_N)/\sigma = z(t_p) - z(f_p)\ ,$$

where $c$ is the decision criterion separating negative and positive responses (see Fig. 5.6).

A value of $d' = 0$ results from chance level performance. Values of 1.0 and 4.65 (where the hit rate reaches 0.99 at a false alarm rate of 0.01) are conventionally considered as moderate respectively near optimal performance (Ehrenstein and Ehrenstein, 1999). For illustration, ROC curves for different values of $d'$ are given in Fig. 5.7.

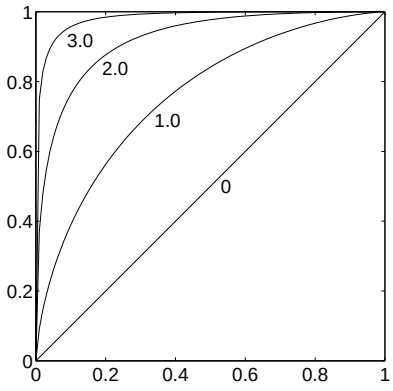


**Fig. 5.7:** ROC curves for eight different values of $d' = 0, 1, 2,$ and 3. The abscissa denotes the false alarm rate $f_p$, the ordinate denotes the hit rate $t_p$.

### 5.6.3 Applying ROC for the Evaluation of Different Junction Detectors

In this section we shall detail how ROC analysis is applied to evaluate different junction detection schemes.

ROC analysis in general is based on ground-truth verification, i.e., the comparison of a detection result with ground truth. Thus, the first step to apply ROC analysis for junction detection is the specification of ground truth junction points for each test image. For artificial images, the ground truth position of junction points are known from the definition of the image or can be rather easily inferred from the gray level variations. For natural images no such ground truth exists. In this case, junction points are marked by visible inspection of an enlarged version of the image. The list of these junction points serves as an approximation of an objective ground truth image.

The next step is the estimation of junction points using a particular junction detection scheme. Three different methods are included in the comparative evaluation which are based on long-range responses (LR), the structure tensor (ST) and on Gaussian curvature (GC). The junction responses $J_{\mathrm{LR}}$, $J_{\mathrm{ST}}$ and $J_{\mathrm{GC}}$ of the different methods are defined in Eq. 5.1, Eq. 5.4 and Eq. 5.5, respectively. The resulting junction responses are normalized to the range $[0; 1]$ to compensate for variations of the response amplitude across different methods.

ROC curves are then computed based on the ground truth image and the normalized junction response as follows. A threshold is varied in $N$ steps over the full range $[0; 1]$ of junction responses, and for each value of the threshold the proportion of true-positive and false-positive responses is computed. To obtain true-positive responses despite localization errors of the methods, responses are accepted within a certain error radius $r_{\mathrm{err}}$ around each ground-truth location.

Finally, the ROC curve characterizing the detection performance of the particular methods is obtained by plotting the true-positive rates against the false positive rates.

To sum up, ROC analysis of the performance of junction detection schemes involves the following five steps:

1. Selection of an input image and determination of the ground truth position of junction points.
2. Application of a particular junction detection scheme to the image.
3. Normalization of the junction responses to the range $[0; 1]$.
4. Variation of a threshold in $N$ steps from 1 to 0 and computation of the respective true-positive $t_{\mathrm{p}}$ and false-positive $f_{\mathrm{p}}$ rate.
5. Plot of the ROC curve, i.e, plotting $t_{\mathrm{p}}$ against $f_{\mathrm{p}}$.

The free parameters of the approach are the number of thresholds $N$ and the error radius $r_{\mathrm{err}}$. We use $N = 40$ which allows for a sufficiently fine resolution of the ROC curves. The error radius is set to $r_{\mathrm{err}} = 3$ pixels.

### 5.6.4 Evaluation Results

In this section we show the results of the ROC analysis applied to a number of artificial and natural images. For each image, ROC curves as obtained for the three different corner detection schemes are displayed. In all plots, solid lines indicate results obtained for the long-range responses (LR), dotted lines indicate results obtained for the structure tensor (ST), and dashed lines indicate results as obtained for the method based on Gaussian curvature (GC).

### Artificial Images

In the first simulation we employ an artificial corner test image from Smith and Brady (1997). The ROC curve for the proposed LR corner detector is well above the ROC curves for the other schemes, indicating a higher sensitivity of the new method (Fig. 5.8).

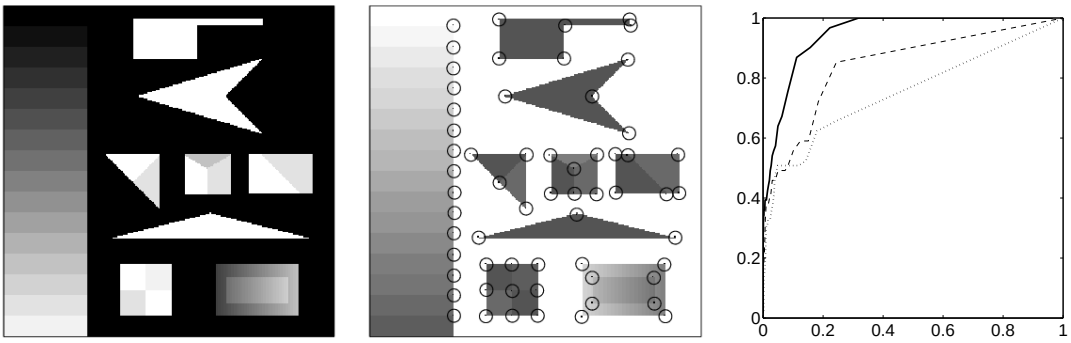


**Fig. 5.8:** *Left:* Artificial test image from Smith and Brady (1997) used for the evaluation of corner detection schemes. *Middle:* Overlay of ground truth junction points. *Right:* ROC curves. The abscissa denotes false positive rate, the ordinate denotes the true positives rate. Curves for the three different corner detection schemes are denoted by *solid* (LR), *dashed* (GC), or *dotted* (ST) *lines*.

### Natural Images

In this section ROC analysis is applied to a number of natural images. The same images as in Sec. 5.5.3 are used.

In a first study we evaluate detection results for three cube images of a laboratory scene. Again, ROC curves as obtained for the LR detector show a higher sensitivity of the new scheme over the full range of thresholds (Fig. 5.9). Note that the ROC curves depicted in Fig. 5.9 show only the first $\approx 10\%$ of the threshold variations for better visualization.

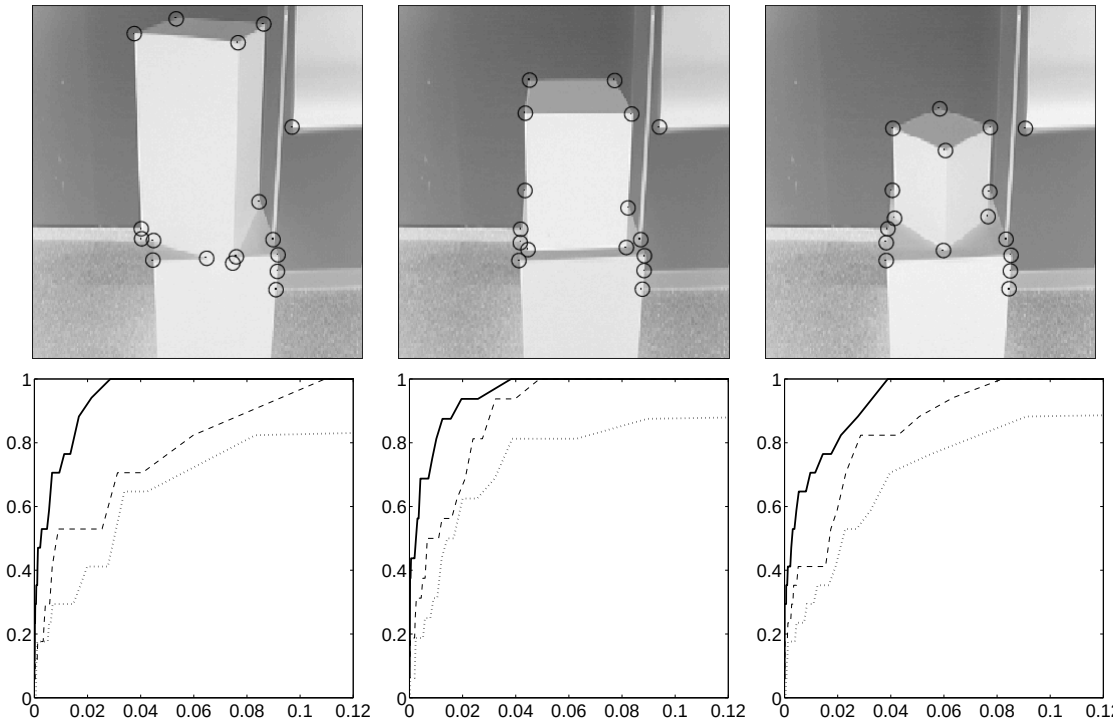


**Fig. 5.9:** *Top row:* Cube images in a laboratory scene with an overlay of manually marked "ground truth" junction points. *Bottom row:* Cutout of the left part of the corresponding ROC curves at $f_{\mathrm{p}} = 0.12$ for better visualization. For larger values of $f_{\mathrm{p}}$, all curves (despite ST) reach a true positive detection rate of 100%. The curves for the three different corner detection schemes are denoted by *solid* (LR), *dashed* (GC) respectively *dotted* (ST) *lines*.

In a final study we employed an image used in a corner detection study by Mokhtarian and Suomela (1998) and an image of a staircase. The simulation results are depicted in Fig. 5.10.

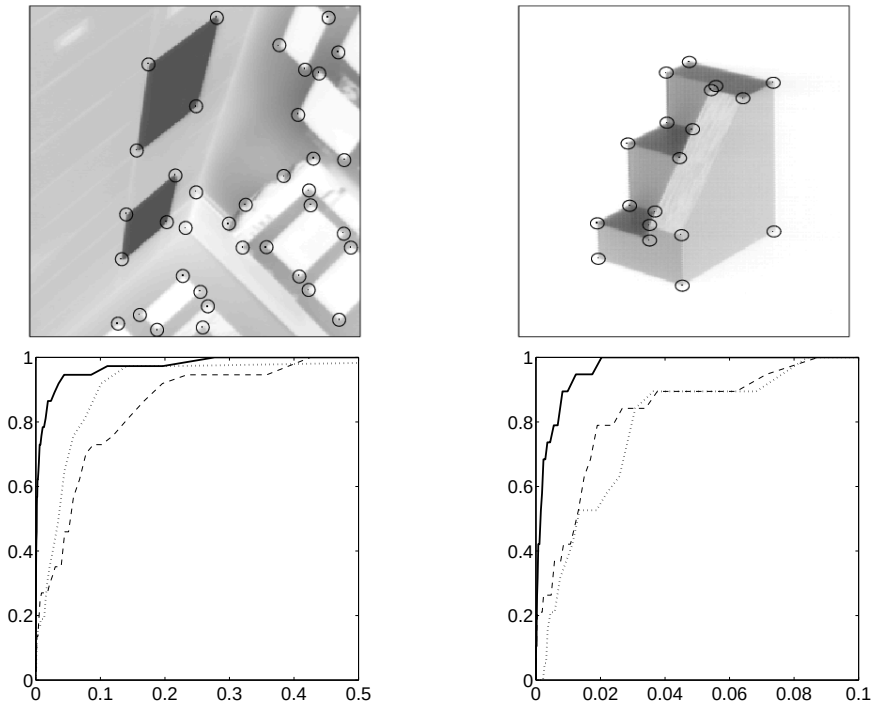


**Fig. 5.10:** *Top row:* A natural corner test image of a laboratory scene from Mokhtarian and Suomela (1998) and a staircase image with an overlay of manually marked "ground truth" junction points. *Bottom row:* Cutout of the left part of the corresponding ROC curves at $f_p = 0.5$ respectively 0.1 for better visualization. For larger values of $f_p$, all curves reach a true positive detection rate of approximately 100%. The curves for the three different corner detection schemes are denoted by *solid* (LR), *dashed* (GC) respectively *dotted* (ST) *lines*.

Compared to the results obtained for natural images, all junction detectors, especially the detectors based on Gaussian curvature and on the structure tensor, exhibit a relatively weak performance on the artificial image (Fig. 5.8). This weaker performance is reflected in the full range $[0; 1]$ of the false positive rate in the plot of the ROC curves necessary to generate a 100% hit rate. The weaker performance results from the 15 T-junctions at the left of the image which are defined by a small contrast at the stems of only $1/16 \approx 6\%$ of the maximally possible luminance variation. The long-range interactions partly compensates for the contrast differences, resulting in the relatively good performance of the newly proposed scheme.

### 5.6.5 Summary of ROC Evaluation

The ROC based evaluation presented in the previous section show that the junction detection method based on long-range responses has an improved performance compared to the results obtained for the structure tensor approach or the Gaussian curvature approach. However, we are not claiming that our approach is generally superior the other approaches. We employed two rather simple implementations of the other approaches, which probably could be tuned to yield a better performance. On the other hand, we have also used a rather straightforward and simple approach to read out the implicit junction information from the distributed representation within an orientation hypercolumn (Eq. 5.1). The main contribution of the present ROC analysis is to demonstrate that the performance of the new method is of the same order of magnitude as standard approaches in computer vision. We conclude that precise junction detection can be

realized by biological plausible circuits. Further, we have shown that the well known method of ROC analysis can be fruitfully applied to evaluate different junction detection schemes.

## 5.7 Discussion and Conclusion

We have proposed a novel method for corner and junction detection based on a distributed representation of orientation information within a hypercolumn, such as hypothesized by Zucker et al. (1989). The explicit representation of a number of orientations in a cortical hypercolumn is shown to constitute a powerful and flexible, multipurpose scheme which can be used to code intrinsically 1D signal variations (Chap. 4) as well as 2D variations like corners and junctions (Chap. 5). A hypercolumnar representation is more powerful than tensor based representation as advocated by Medioni and coworkers (Guy and Medioni, 1996; Medioni et al., 2000). A tensor represents the strength of the two cardinal orientations but cannot adequately capture image structures where more than two orientations met at a particular location or two lines intersect at non-orthogonal angles.

Orientation responses within a hypercolumn can be robustly and reliably computed by using contextual information. We have proposed a model of recurrent long-range interactions to compute coherent orientation responses (Chap. 4). In the context of corner and junction detection we have shown the benefits of using contextual information and recurrent interactions, leading to a considerable increase in localization accuracy and detection performance compared to a simple feedforward scheme.

The proposed junction detection mechanism uses a measure of circular variance for the explicit representation of distributed orientation responses within a hypercolumn. While this measure is motivated mainly computationally, circular variance is essentially a weighted sum of information locally available within a hypercolumn. We hypothesize that a related measure of junction responses can be realized by neurons in the visual cortex. First evidence comes from a study by Das and Gilbert (1999), indicating a graded specialization of neurons for the processing of corners and T-junctions. The degree of selectivity for processing corners was shown to increase with the overlap of the neuron's dendritic arborization with neighboring orientation columns.

A number of studies have stressed the importance of multiple scales for the proper extraction of corner and junction information (Lindeberg, 1998; Mokhtarian and Suomela, 1998; Würtz and Lourens, 2000). The present model operates only on a single scale. However, in the recurrent interaction, the model neurons integrate information over successively increasing ranges of the visual space. Thus, information of different scales is available during the temporal evolution of orientation responses. We have shown for a number of generic junction configurations that accurate, precise localization can be achieved by the proposed model without the need of tracking responses along multiple scales from coarse to fine, as suggested by Mokhtarian and Suomela (1998).

To compare the novel junction detection scheme to other existing approaches in computer vision, we have evaluated two other approaches based on Gaussian curvature and on the structure tensor. We have utilized ROC analysis for the evaluation of junction responses. ROC analysis allows to assess the capabilities of the detectors over the full range of possible thresholds for every test image. Consequently, ROC based evaluation results are not flawed by choice of a particular threshold which can strongly bias the obtained results. The results of ROC analysis for both artificial and natural images show that detection performance of the new scheme is similar, often better compared to the other schemes. Since more sophisticated implementations of the other schemes can be proposed, we do not claim that the novel scheme is generally superior to the other methods. Rather we have shown that robust and accurate junction detection can be realized based on biologically plausible mechanisms.

Chapter 6

# Surface Representation Using Confidence-based Filling-in

## 6.1 Introduction

Experimental studies indicate the existence of distinct perceptual subsystems in human vision, one that is concerned with *contour extraction* and another that assigns *surface properties* to bounded regions. The emerging picture from the experimental investigations is one in which shape outlines are initially extracted, followed by the assignment of attributes such as texture, color, lightness, brightness or transparency to regions (Bressan et al., 1997; Elder and Zucker, 1998; Grossberg and Mingolla, 1985a; Lamme et al., 1999; Rogers-Ramachandran and Ramachandran, 1998). Several perceptual completion phenomena (Pessoa et al., 1998) suggest that, on a functional level, regions inherit local border contrast information by means of "spreading mechanisms" or "filling-in" (Caputo, 1998; Paradiso and Nakayama, 1991). The assignment of surface properties would then be dependent on the determination of stimulus contrast in various feature dimensions, such as luminance, motion direction and velocity, and depth, that would be used to fill-in bounded regions.

The problem of deriving a dense representation of surface quality, such as brightness or color, from local estimates, such as luminance or chromatic border contrast, is inherently ill-posed: there exists no unique solution nor is the solution guaranteed to be stable. Such an inverse problem needs to be regularized in the sense that certain constraints have to be imposed on the space of possible solutions. The constraint of generating a smooth surface, as formalized by minimizing the first order derivatives, leads to a linear diffusion process with a simple reaction term (Neumann et al., 2001).

In filling-in theory, feature signals which provide the source term of the filling-in process are modeled as cells with circular receptive fields (RFs) such as retinal ganglion cells or LGN cells. In previous filling-in models, such cells are modeled to exhibit strong responses even to homogeneous regions (Cohen and Grossberg, 1984; Grossberg and Todorović, 1988). Physiological studies however show that retinal ganglion cells respond strongly only at positions of luminance differences or contrasts (Enroth-Cugell and Robson, 1984). Motivated by these results we use sparse contrast signals with no response to homogeneous regions. The sparse nature of signals necessitates additional confidence signals for the filling-in process. Confidence signals indicate the positions of valid contrast response to be taken as source for the filling-in process. Having established the link between models of perceptual data for biological vision and the mathematical frameworks of regularization theory this leads to the proposal of *confidence-based filling-in* (Neumann et al., 2001).

This chapter is organized as follows: First, we shall review a number of empirical findings supporting the notion of filling-in processes involved in visual perception (Sec. 6.2). In particular, we focus on recent studies which have examined the temporal properties of filling-in. Next, in Sec. 6.3, we briefly review different approaches that have been proposed to model brightness perception.

We shall then examine previous work on the mathematical modeling of filling-in and propose a new scheme of confidence-based filling-in (Sec. 6.4). The relations between filling-in, diffusion processes and regularization theory are briefly reviewed in Sec. 6.5. In Sec. 6.6, the equations defining a fully-fledge model of brightness perception using confidence-based filling-in are given. Simulations showing the competencies of the new model are presented in Sec. 6.7. We show that confidence-based filling allows for a brightness reconstruction which is invariant against shape and size variations of the area to be filled-in. Further, we show that the new approach can account for a number of visual illusions such as simultaneous contrast or the Craik-O'Brien-Cornsweet (COC) effect, and can be successfully employed for the processing of natural images. The proper restoration of reference levels, necessary to represent stimuli such as a luminance staircase is an inherent problem for filling-in models. In Sec. 6.8 we show how this deficit can be overcome by using luminance-modulated contrast signals. Section 6.9 concludes the chapter.

## 6.2 Empirical Evidence for Neural Filling-in

There is a large number of empirical evidence for the existence of neural filling-in mechanisms. In the following we shall briefly review a number of important empirical findings stimulating and supporting filling-in theories of brightness perception. More detailed reviews can be found in Pessoa and Neumann (1998) and Pessoa et al. (1998).

Historically, the idea of filling-in has emerged from observations based on perceptual completion across the blind spot and across pathological or artificial scotomas (Bender and Teuber, 1946; de Weerd et al., 1995; Gerrits and Timmerman, 1969; Ramachandran, 1992; Ramachandran and Gregory, 1991) and from experimental work using stabilized images (Gerrits et al., 1966; Krauskopf, 1963; Riggs et al., 1953; Yarbus, 1967). In stabilized image studies, stimuli consisting of a small disk and an enclosing annulus of different color are used. When the boundary of the interior disk is stabilized on the retina, a spreading of the color of the surrounding annulus into the interior disk is perceived, leading to a final percept of an uniformly colored stimulus. The observations of these demonstrations point toward an important role of color and brightness signal near object boundaries in determining the perception of surface properties of the enclosed region. It has been proposed that the filling-in phenomena found under artificial experimental or pathological conditions reveal a general mechanism of normal vision (Gerrits and Vendrik, 1970; Walls, 1954).

Besides brightness and color, filling-in processes have also been reported for other modalities such as textures (Caputo, 1998; de Weerd et al., 1998; Gyoba, 1997; Ramachandran and Gregory, 1991; Ramachandran et al., 1992; Watanabe and Cavanagh, 1991, 1993), depth (Gillam and Borsting, 1988; Julesz, 1971; Nakayama and Shimojo, 1990a,b) or motion (Watanabe and Cavanagh, 1993).

In the following we shall review a number of recent empirical studies investigating the *temporal dynamics* of filling-in processes for brightness perception.

Evidence for brightness filling-in as a temporal process comes from a psychophysical study by Paradiso and Nakayama (1991). In this study, a visual masking paradigm is used to investigate two issues: First, the role of edge information in determining the brightness of homogeneous regions, and second the temporal dynamics of brightness perception. If brightness perception relies on some form of activity spreading, it should be possible to interrupt this spreading process. In the experiment, a target of a bright disk is followed by a mask (e.g., a smaller circle or a C-shape), which is presented at variable time intervals. For an interstimulus interval of about 50–100 ms, the brightness of the central area is highly dependent on the shape of the mask. For example, for a C-shaped mask, a darkening of the middle region is observed, with the bright region "protruding" inside the C. For a circular-shaped mask, an inner dark disk is perceived. Both these results are consistent with the hypothesis that brightness signals are generated at the borders of their target stimuli and propagate inward. Furthermore, it has been demonstrated that for larger

stimuli maximal suppression occurs later. This finding supports the view that filling-in is an active spreading of neural activity, i.e., a process which takes time.

Recently, a similar masking paradigm has been used to investigate brightness filling-in within texture patterns (Caputo, 1998). Again the spreading could be blocked by the mask if the interstimulus interval is in accordance with the propagation rate required to travel the distance between boundary and mask position. Results of a study employing Craik-O'Brien-Cornsweet (COC) gratings point in the same direction: For higher spatial frequencies of the grating (i.e., for smaller distances) the effect was stronger and persisted to higher temporal frequencies of COC contrast reversal (Davey et al., 1998).

The findings of these psychophysical studies are paralleled by related physiological results regarding the temporal limits of simultaneous contrast or brightness induction (De Valois et al., 1986; Rossi et al., 1996). It was shown that a significant percentage of cells in the visual cortex respond to luminance modulations outside their classical receptive field in correlation with perceived brightness. Thus, correlates of perceived brightness can be found at the very first stage of visual cortical processing, namely in V1 (see also Rossi and Paradiso, 1999). Moreover, the dynamic version of brightness induction occurred only at low temporal frequencies of the sinusoidally modulated surround luminance. This finding indicate that a spreading of neural activity is involved in brightness perception.

In summary, findings from a large number of studies are suggestive of active neural filling-in processes that are initiated at region edges. Using brightness filling-in, the brain generates a spatially organized representation through a continuous propagation of signals, a process that takes time (Neumann et al., 2001; Pessoa et al., 1998).

## 6.3 Review of Models for Brightness Perception

Models of brightness perception can be partitioned into four different classes, namely *filter-based models*, *rule-based models*, *integration models* and *filling-in models*. In the following we shall provide a brief review of different approaches. A more detailed review is given by Kingdom and

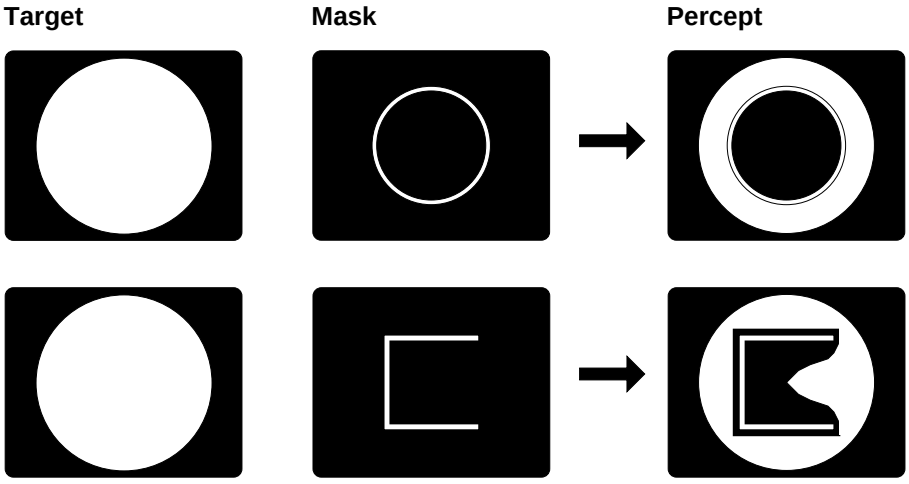


**Fig. 6.1:** Masking paradigm used by Paradiso and Nakayama (1991) to investigate the temporal properties of brightness filling-in. Perceived brightness *(right)* of a homogeneous target *(white disk, left)* is highly dependent on the arrangement of contours in the mask *(middle)*. Target and mask are each displayed for 16 ms, and the temporal delay between target and mask is optimized to yield the strongest suppression effect. (Adapted from Paradiso and Nakayama, 1991.)

Moulden (1989) or Pessoa et al. (1995). A discussion focusing on the prediction of Mach bands by various models can be found in Pessoa (1996).

*Filter-based models* or *feature-based models* (Pessoa, 1996) process the input image by a bank of filters, often at multiple scales. These filter responses are then combined by a measure originating in signal processing theory, such as the local energy of even- and odd-symmetric filters (e.g., Morrone and Burr, 1988) or a weighted sum of rectified responses (e.g., Blakeslee and McCourt, 1999) to account for human brightness perception. A number of filter-based models have been proposed which differ by the type of filter used and the suggested combination of the filter responses (Blakeslee and McCourt, 1997, 1999; du Buf, 1994; du Buf and Fischer, 1995; Fiorentini et al., 1990; Morrone and Burr, 1988; Tolhurst, 1972).

*Rule-based models*, similar to filter-based models, first process the input image with a set of filters, usually for different orientations and at multiple scales, and synthesize the human brightness percept by combining the filter outputs, e.g., by a simple linear combination across orientation and scale (McArthur and Moulden, 1999). The distinctive feature of rule-based models is the use of symbolic interpretation rules at the combination stage. The interpretation rules specify distinct classes of brightness results by analyzing the pattern of zero-crossings in the filter responses. Prominent representatives of rule-based models are the MIRAGE model of Watt and Morgan (1985), the MIDAAS model by Kingdom and Moulden (1992) and a recent extension to 2D (McArthur and Moulden, 1999). As pointed out by Pessoa (1996, p. 3214), "the power of a specific approach stems from its ability to show how related phenomena originate from a common set of mechanisms or processes". In other words, a successful theory should be able to explain and unify a number of seemingly different phenomena by the same underlying principle, and ideally making testable predictions regarding new phenomena. Rule-based models, on the other hand, tend to explain different phenomena by using different rules, and fail to predict the perception of new stimuli based on their fixed set of rules (Pessoa et al., 1995).

*Integration models* compute brightness by integration of local luminance ratios across space (Arend, 1985, 1994; Blake, 1985; Horn, 1974; Hurlbert, 1986). An important class of integration models are models based on retinex theory (Land, 1977, 1983, 1986; Land and McCann, 1971). Retinex models have been proposed in particular to account for color and lightness constancy in so called Mondrian images, i.e., randomly arranged flat 2D patches of different color. In retinex models, color and brightness is recovered by integrating log luminance ratios at edge locations. Recently, a variant of retinex models has been proposed based on center-surround processing (Jobson et al., 1997). Brainard and Wandell (1986) have pointed out a number of drawbacks of retinex theory in modeling human brightness perception, e.g., the overestimation of simultaneous contrast.

*Filling-in models* generate brightness predictions by a diffusive spreading of contrast-sensitive activity at edges. The spreading or filling-in of feature contrast signals is locally controlled by contour signals to block the diffusion at object boundaries. Based on the filling-in theory of Gerrits and Vendrik (1970), filling-in models have been proposed by Grossberg and coworkers (Cohen and Grossberg, 1984; Grossberg, 1983; Grossberg and Mingolla, 1985b; Grossberg and Todorović, 1988). Filling-in models qualitatively account for a wide variety of brightness phenomena, including, e.g., simultaneous contrast, brightness assimilation and the COC effect (Grossberg and Todorović, 1988). More recently, filling-in models have been used to model the perception of Mach bands, low- and high-contrast missing fundamental stimuli and sinusoidal waves (Pessoa et al., 1995) as well as challenging stimuli including Kanizsa stratification, Munker-White illusion, Benary cross and checkerboards which are thought to involve 3D perception (Kelly and Grossberg, 2000). Filling-in models can also account for the temporal dynamics of brightness perception under visual masking conditions (Arrington, 1994). In the domain of image processing, filling-in models have also been employed for the enhancement of synthetic aperture radar images (Grossberg et al., 1995; Mingolla et al., 1999).

Despite the competencies of filling-in models, a drawback has been identified from the onset, namely the failure to restore reference levels such as needed, e.g., to process a luminance staircase.

Various approaches have been proposed to overcome this deficit (Arrington, 1996; Pessoa et al., 1995), but lack empirical evidence or fail for other stimuli such as the COC staircase.

## 6.4 Confidence-based Filling-in

To introduce concepts, we consider the task of generating a continuous representation of surface layout as one of painting or coloring an empty region (Mumford, 1994b). The task thus consists of generating an internal representation of surface properties from given data. Individual surfaces occur at different sizes and with various shapes. Therefore, any such mechanism has to be insensitive to such size and shape variations.

### 6.4.1 BCS/FCS and the Standard Filling-in Equation

Models of brightness perception were among the first to explore the dichotomy of boundary and surface subsystems (e.g., Lamme et al., 1999; Rogers-Ramachandran and Ramachandran, 1998). Based on stabilized image studies it has been proposed that the perception of brightness can be modeled by filling-in processes. Filling-in models suggest that local feature measures undergo a process of lateral spreading, or diffusion, to determine surface appearance (Gerrits and Vendrik, 1970). The basic ideas have been formalized in a model of complementary boundary and surface systems (Boundary Contour System/Feature Contour System, BCS/FCS) by Grossberg and coworkers (Cohen and Grossberg, 1984; Grossberg and Mingolla, 1985a).

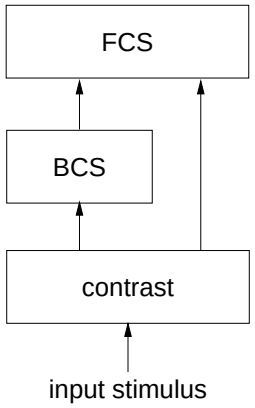


**Fig. 6.2:** Sketch of the BCS/FCS architecture.

In a nutshell, BCS/FCS processing occurs as follows. First, contrast-sensitive signals are extracted from the raw image data. Within the BCS, these contrast signals are processed in a hierarchy of levels resulting in sharp, localized boundary signals, defining a segmentation of the initial input image into compartmental areas. Within the FCS, these boundaries control the lateral spreading or diffusion of contrast-sensitive input signals. A particular property of this architecture is that contrast-sensitive signals are used as input to both the BCS and the FCS. Model simulations based on this architecture qualitatively account for a wide variety of brightness phenomena (e.g., Grossberg and Todorović, 1988; Kelly and Grossberg, 2000; Pessoa et al., 1995). A sketch of the BCS/FCS architecture is depicted in Fig. 6.2.

In the past decades, several models of early vision based on the BCS/FCS architecture have been proposed, with considerable extensions and modifications (Cohen and Grossberg, 1984; Gove et al., 1995; Grossberg and McLoughlin, 1997; Grossberg and Todorović, 1988; Kelly and Grossberg, 2000; Pessoa et al., 1995). In all these models, the same filling-in equation is used, which has been shown

to be equivalent to a linear inhomogeneous diffusion with reaction term (Neumann et al., 2001). In this equation, the reaction term consists of contrast-sensitive input signals $K$ and a passive decay of activity (with rate $\alpha$). The diffusion term describes the nearest-neighbor coupling $\mathcal{N}_i$ of filling-in activities which is locally controlled by permeability signals $P$ (inhomogeneous diffusion). Permeability signals are a monotonically decreasing function of boundary or contour signals, i.e., high contour signals imply low permeability and vice versa. A sketch of the discretized network is depicted in Fig. 6.3.

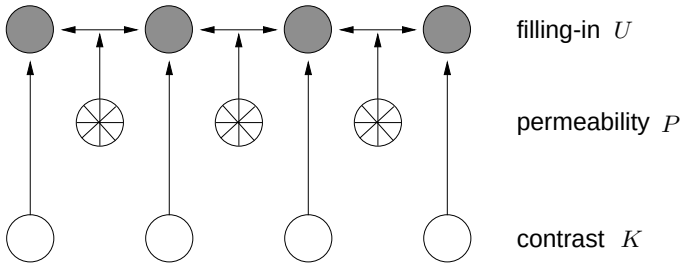


**Fig. 6.3:** Sketch of the discretized filling-in network in the 1D case. At the filling-in layer $U$ *(gray circles)*, each cell interacts with its nearest neighbors in a diffusive process. The lateral coupling between nearest neighbors is locally controlled by space-variant permeability signals $P$ *(wheel icons)*. Input to the filling-in layer is provided by contrast-sensitive signals $K$ *(open circles)*.

In all, the discretized equation for filling-in activity $U$ reads

$$\partial_t U_i = \underbrace{K_i - \alpha U_i}_{\text{reaction term}} + \underbrace{\sum_{j \in \mathcal{N}_i} (U_j - U_i)\, P_{ij}}_{\text{diffusion term}} \ , \tag{6.1}$$

where $\partial_t$ denotes partial differentiation with respect to $t$. Discrete spatial locations are denoted by $i$ and $j$. The nearest neighbor coupling is given by $\mathcal{N}_i = \{i-1, i+1\}$ for the 1D case and $\mathcal{N}_{ij} = \{(i-1,j), (i+1,j), (j-1,i), (j+1,i)\}$ for the 2D case. We will refer to Eq. 6.1 as *standard filling-in* in contrast to the new scheme of confidence-based filling-in which shall be detailed in the following.

### 6.4.2 Confidence-based Filling-in Equation

Previous models of filling-in use a dense representation of contrast-sensitive feature signals as source for the filling-in process. Cells at early stages of the visual system, such as retinal ganglion cells, show strong responses only at luminance discontinuities. Given the sparseness of contrast signals which are zero within homogeneous regions, the visual system has to compute a dense brightness surface from local contrast estimates. Such inverse problems are generally ill-posed in the sense of Hadamard (Bertero et al., 1988; Poggio et al., 1985; Tikhonov and Arsenin, 1977). This means that the existence and uniqueness of a solution and its continuous dependence on the data cannot be guaranteed since the measurements are sparse and may be noisy. The solution to the problem has to be regularized such that proper constraints are imposed on the function space of solutions. Such a constraint is the *smoothness* of the solution, for example. Smoothness can be characterized by minimizing the first order derivatives of the desired solution. The goal is to minimize both the local differences between the measured data and the reconstructed function values (data term) and the stabilizing functional imposed on the function (smoothness term). Minimizing a quadratic functional finally leads to the discretized version of a new filling-in equation, where an additional *confidence signal* $Z$ steers the contribution of the data term (Neumann et al., 2001):

$$\partial_t U_i = Z_i \,(K_i - \alpha U_i) + \sum_{j \in \mathcal{N}_i} (U_j - U_i)\, P_{ij} \ . \tag{6.2}$$

A similar confidence term has been used in the context of visual surface representations (Szeliski, 1990; Terzopoulos, 1986), where the confidence term it assumed to be inversely related to the variance of the measurement of the input data.

Note that confidence-based filling-in is a generalization of standard filling-in: For constant unit-valued confidence signals $Z = 1$ confidence-based filling-in (Eq. 6.2) is equivalent to standard filling-in (Eq. 6.1).

Confidence signals are in the range $[0; 1]$. Zero confidence signals indicate positions where no data are available, while unit-valued confidence signals occur at region boundaries and signal positions of reliable contrast measurements. Consequently, we suggest that an intermediate representation in the processing of contour signals, namely complex cell responses, are involved in the computation of confidence signals. A detailed description of the computation of confidence signals along with the other model equations and parameters used can be found in Sec. 6.6.

## 6.5 Filling-in, Diffusion, and Regularization

It has been suggested by Paradiso and Nakayama (1991) that filling-in processes "may be an important component of the visual system's solution to ill-posed problems". Such ill-posed problems can be solved within the framework of regularization theory (Poggio et al., 1985). On the other hand, membrane regularization is known to be equivalent to biased diffusion (Jähne, 1997; Weickert, 1998), which is essentially the same type of diffusion suggested to model filling-in (Cohen and Grossberg, 1984). These close connections between filling-in, diffusion and regularization have been made rigorous by Neumann et al. (2001). In the following, we shall briefly review the basic ideas behind diffusion processes and regularization, examine the domain of their application, establish their close connections to filling-in, and finally point toward particular extensions that go beyond the scope of the basic filling-in equation.

### 6.5.1 Filling-in

The standard filling-in equation is usually written as

$$\partial_t U_i = -\alpha U_i + K_i + \sum_{j \in \mathcal{N}_i} (U_j - U_i)\, P_{ij} \quad . \tag{6.3}$$

The first term $-\alpha U_i$ is a decay or sink term, the second term $K_i$ are the contrast signal which provide the source term, and the third term describes the nearest neighbor coupling on a discrete grid, which is locally controlled by permeability signals $P$. The equation is linear in $U$ but often referred to as "nonlinear" (e.g., Cohen and Grossberg, 1984; Grossberg and Todorović, 1988).

### 6.5.2 Diffusion

Diffusion is a process that equilibrates concentration differences without creating or destroying mass. The standard diffusion equation has a constant permeability $P$ (also called "diffusivity") and is given by the Laplacian equation

$$\partial_t\, U = P \Delta U = P\,(U_{xx} + U_{yy}) \quad .$$

This standard diffusion equation describes a linear homogeneous diffusion process which is equivalent to Gaussian smoothing (Weickert, 1998).

Using an inhomogeneous or space-variant permeability $P = P(x, y)$ and adding a bias or reaction term which keeps the steady-state solution close to some input $D$ leads to biased inhomogeneous diffusion:

$$\partial_t U = \nabla \cdot (P \nabla U) + \alpha (D - U) \ . \tag{6.4}$$

Setting $K = \alpha D$, the standard filling-in equation (Eq. 6.3) results from a finite difference discretization of Eq. 6.4 (Neumann et al., 2001).

Inhomogeneous diffusion processes have been suggested for image enhancement (e.g., Charbonnier et al., 1994; Fritsch, 1992). Though these formulations employ essentially the same equation as used to model filling-in, a number of differences exists:

(a) The permeability is not computed in a hierarchy of biologically motivated processing levels, but more straight forward based on the image gradients.

(b) The diffusion does not use contrast-sensitive signals resulting from shunting equations as input, rather than the raw input image.

(c) An additional reaction term is usually not used. Instead, the diffusion is stopped at some time to obtain a non-trivial steady state solution. The selection of a stopping time is equivalent to the selection of the decay parameter $\alpha$ (Weickert, 1997b).

Diffusion processes have gained increasing interest in the computer vision community, and a number of extension have been proposed. In particular, the permeability can be computed based on the evolving diffusion activity $U$, i.e., $P = P(U(x, y))$, leading to nonlinear inhomogeneous diffusion (Perona and Malik, 1990b). Further, the scalar-valued permeability can be replaced by a tensor which steers the direction of diffusion away from its usual direction opposed to the direction of the gradient. This type of anisotropic diffusion has been successfully employed for edge enhancement and smoothing along coherent structures (Weickert, 1995, 1996a).

The functionality of these different types of diffusion processes can be understood by analogy to Gaussian smoothing. Standard or Laplacian diffusion is equivalent to smoothing with an isotropic Gaussian of constant standard deviation. Inhomogeneous diffusion is equivalent to smoothing with an (isotropic) Gaussian, but of different or space-variant standard deviations depending on the image location. Anisotropic diffusion—as the name suggests—is equivalent to smoothing with an anisotropic Gaussian, whose main axis and two standard deviations can vary at each image location.

The different types of diffusion equations point toward future directions of research for modeling filling-in processes. In particular, the nonlinear or recurrent computation of permeability signals based on the diffused image may lead to more robust segmentation results and consistent contour representations. Recently, Kelly and Grossberg (2000) have used recurrent interaction between the BCS and FCS (so called "boundary pruning") to achieve mutually consistent representations of contours. Using boundary pruning, a number of empirical data regarding the perception of occluding and occluded objects could be explained.

### 6.5.3 Regularization

Based on the notion of computer vision as inverse optics, a number of problems in computer vision have been identified as inherently ill-posed, such as stereo matching, optical flow computation, shape from shading, or surface reconstruction, among others (Poggio et al., 1985). A problem is considered ill-posed if there exists no unique solution or the solution is not guaranteed to be stable (Hadamard, 1923). Consider the task of surface reconstruction (Szeliski, 1990; Terzopoulos, 1983): given a set of sparse and noise measurements at contour locations, the problem of finding a corresponding brightness or texture representation has no unique solution nor is it guaranteed to be stable. Additional constraints on the proper solution have to be imposed, such as smoothness, to restrict the class of admissible solutions and restore well-posedness.

The mathematical framework for solving ill-posed problems is provided by regularization theory (Tikhonov and Arsenin, 1977). Given a set of data points $D$, a regularized solution $U$ can be obtained by minimizing a suitable energy functional $E[U, D]$. The energy functional is defined as an integral over a so-called Lagrange function F[U,D] which consists of two parts, a data function $F_{\text{data}}$ that ensures that the solution is close to the data points, and a stabilizing or model function $F_{\text{model}}$ that incorporates a priori knowledge about the solution or, equivalently, incorporates the desired smoothness constraints:

$$\begin{aligned} E[U, D] &= \iint_{\Omega} F[U, D] \; \mathrm{d}x \, \mathrm{d}y \quad \rightarrow \min \\ &= \iint_{\Omega} F_{\text{data}}[U, D] \; + \; \lambda \, F_{\text{model}}[U] \; \mathrm{d}x \, \mathrm{d}y \; . \end{aligned}$$

The relative contribution of the smoothness constraint is controlled by a regularization parameter $\lambda$.

The data function is usually defined as

$$F_{\text{data}} = \frac{1}{2}(U - D)^2 \; .$$

A suitable model function that minimizes the variation of the first order derivatives, the so-called membrane function, is given by

$$F_{\text{model}} = F_{\text{membrane}} = \frac{1}{2} \, \widetilde{P}(x, y)(U_x^2 + U_y^2) \; ,$$

where $\widetilde{P}(x, y)$ is a space-variant permeability which allows surface discontinuities at certain positions (Terzopoulos, 1986). Having specified appropriate data and model functions, the solution to the inverse problem is obtained within the calculus of variations (Gelfand and Fomin, 1963). The necessary condition for the existence of a solution that minimizes the energy functional is given by solving the corresponding Euler-Lagrange equation:

$$\frac{\delta F}{\delta U} - \frac{\partial}{\partial x} \frac{\partial F}{\partial U_x} - \frac{\partial}{\partial y} \frac{\partial F}{\partial U_y} = 0 \; .$$

The solution for the Lagrange function $F[U, D]$ with model and data function as specified above reads:

$$0 = (U - D) - \lambda \nabla \cdot (\widetilde{P} \, \nabla U) \; . \tag{6.5}$$

Multiplication of Eq. 6.5 by $-\alpha$ results in

$$0 = \alpha D - \alpha U + \alpha \lambda \nabla \cdot (\widetilde{P} \, \nabla U) \; . \tag{6.6}$$

Setting $K = \alpha D$ and $P = \alpha \lambda \widetilde{P}$, this equation is equivalent to the steady state solution of the biased inhomogeneous diffusion equation (Eq. 6.4), or, alternatively, equivalent to the standard filling-in equation in continuous form (Neumann et al., 2001).

A number of other definitions of the model and data function have been suggested. For example, a confidence measure $Z(x, y)$ as part of the data function allows for a space-variant weighting of the data term depending on the reliability of the measurement.

$$F_{\text{data, conf}} = \frac{1}{2} \, Z(x, y)(U - D)^2 \; . \tag{6.7}$$

In particular, zero-valued confidence allow to locally switch off the data term. Membrane regularization with confidence valued data function motivated the new scheme of confidence-based filling-in (Neumann et al., 2001).

Instead of the membrane stabilizer, a model function that minimizes the variation of the second order derivatives, the so-called thin-plate function, is commonly used

$$F_{\text{model}} = F_{\text{thinplate}} = \frac{1}{2}\,\widetilde{P}(x,y)(U_{xx}^2 + 2U_{xy}^2 + U_{yy}^2) \ .$$

Further, both membrane and thin-plate function can be combined (Szeliski, 1990; Terzopoulos, 1986):

$$F_{\text{model}} = (1-\tau(x,y))\,F_{\text{membrane}} + \tau(x,y)\,F_{\text{thinplate}} \ .$$

Both thin-plate and membrane stabilizers are linear combinations of the first $n$ derivatives of the desired solution $U$, as suggested by Tikhonov and Arsenin (1977).

### 6.5.4 Summary of the Relations of Filling-in to Diffusion and Regularization

In this section we have pointed out the close relationship of filling-in, diffusion processes and regularization (Neumann et al., 2001). Regarding diffusion processes, standard filling-in is equivalent to inhomogeneous biased diffusion. Regarding regularization theory, the steady state solution of the filling-in equation is equivalent to membrane regularization with a space-variant weighting of the smoothing term. These equivalences provides deeper insights into the computational competencies of filling-in and suggests variations and extensions of the filling-in equation commonly used. For example, in regularization theory a confidence measure is often used which allows for a local weighting of the contributions of the data term relative to the model term. This confidence term motivate a new scheme of confidence-based filling-in (Neumann et al., 2001). Regularization theory suggest further extensions, e.g., using a thin-plate stabilizer instead of the membrane model, or a combination of both. Diffusion processes, in a similar fashion, point toward recurrent computations of the permeability function or a tensor-valued permeability implementing an anisotropic diffusion process. Future models of filling-in may benefit from such extensions, leading to refined models of brightness perception.

## 6.6 The Model

In this section a detailed description of the proposed model for brightness perception using confidence-based filling-in is given. Before the equations are presented which define the processing at the various stages, an overview of the model is provided.

### 6.6.1 Model Overview

The model proposed in this chapter stands in the tradition of BCS/FCS models (e.g., Cohen and Grossberg, 1984; Grossberg and Todorović, 1988). A characterizing feature of the BCS/FCS architecture is the use of contrast signals feeding into two different systems, namely the BCS and the FCS. In the BCS, contrast signals are sharpened to generate localized boundary signals; in the FCS, contrast signals are spread in a diffusive process. The use of contrast signals as input for both systems ensures a close spatial relationship of both signals. Unlike previous approaches (Cohen and Grossberg, 1984; Grossberg and Todorović, 1988), the proposed model uses sparse contrast signals and introduces an additional confidence measure in the filling-in equation, defining the new scheme of confidence-based filling-in. The confidence measure is generated from contrast signals and an intermediate stage in the formation of boundary signals, namely complex cell responses. Again, the definition of the confidence measurement based on contrast signals ensures a close topographical relationship between the signals. The topographical organization of the signals simplifies the computation of the filling-in process, since the mapping between the confidence

signals and the corresponding gain-controlled contrast signals is implicitly and precisely provided by the identical positions within the respective topographic maps.

An overview of the model architecture is given in Fig. 6.4.

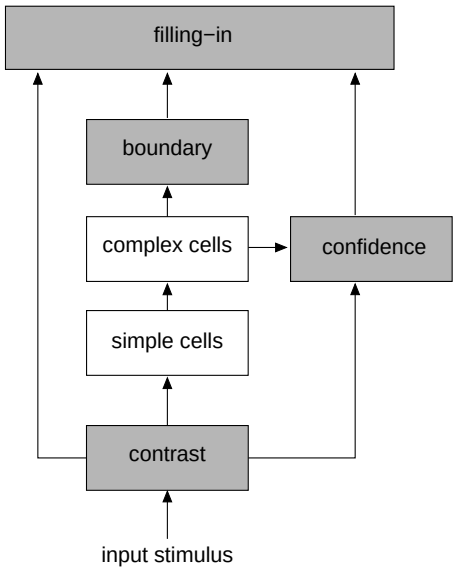


**Fig. 6.4:** Overview of the model architecture. The filling-in stage and the stages which provide the input to the filling-in stage are depicted with *filled gray boxes*.

## 6.6.2 Model Equations

In this section a formal description of the model is given. The model consists of a hierarchy of processing levels, including contrast, confidence and boundary signals, which provide the input to the filling-in stage. In all equations, capital Roman letters denote the 2D maps of activity distributions at the various stages; Greek letters denote positively-valued model parameters.

### Contrast

At the first processing stage, contrast signals are generated from the initial luminance distribution of the input stimulus. Contrast signals occur at luminance differences and provide the source term for the filling-in process, i.e., during filling-in, contrast signals spread or diffuse. Contrast signals are intended to model integrated responses of retinal ganglion cells or LGN cells. Contrast signals $K_{\text{on}}$ and $K_{\text{off}}$ are modeled for two domains, signaling light increments and decrements. The equations defining the contrast signals are detailed in Sec. 3.5.

### Boundary Signals

Boundary signals are modeled in a hierarchy of processing stages, including simple and complex cells, and control the spreading of the contrast signals at the filling-in stage. The equations describing the simple cell circuit and the complex cells are detailed in Sec. 3.5.

Boundary signals result from sharpening of the pooled complex cell signals, yielding localized contour responses at region boundaries. Boundary signals can be computed by a recurrent competitive-cooperative loop (Pessoa et al., 1995) or recurrent long-range interactions motivated by the neural mechanisms in the primary visual cortex (e.g., Grossberg and Raizada, 2000; Hansen

and Neumann, 1999d, 2001b; Li, 1998). For the purpose of the present model, a coarse approximation of this process suffices.

To generate boundary signals, pooled complex cell signals $C_{\text{pool}}$ are first nonlinearly sharpened by multiplying a cell at each location with the activities of its four nearest neighbors:

$$\widetilde{B} = \prod_{\mathcal{N}_{ij}} C_{\text{pool}}(i,j) \ ,$$

where the nearest neighbor coupling is given by $\mathcal{N}_{ij} = \{(i,j),(i,j-1),(i,j+1),(i-1,j),(i+1,j)\}$. Spurious signals below $\kappa_B = 20\%$ of the maximum activity of the sharpened responses $\widetilde{B}$ are discarded by thresholding, resulting in the final boundary representation

$$B = \left[\widetilde{B} - \kappa_B \max\{\widetilde{B}\}\right]^+ \ .$$

### Confidence Measure

Confidence signals are the key component in the proposed scheme of confidence-based filling-in (Neumann et al., 2001). The use of a confidence measure makes the model robust against spurious noisy contrast activities and allows to create a dense brightness surface from sparse contrast data.

Confidence signals are used to locally control the contributions of the data term in the filling-in equation (Eq. 6.9) depending on the validity of the contrast signals: zero-valued confidence signals switch off the data term at positions where no contrast data is available. Nonzero confidence value select valid contrast signals near region boundaries to be passed as source term to the filling-in process. Valid contrast signal are those nonzero contrast signals which give rise to contour signals at region boundaries. Therefore, pooled complex cells $C_{\text{pool}}$ are used to select valid contrast signals. Complex cell signals are computed—via simple cells—from the same contrast signals $K$ which provide the input to the filling-in process (Fig. 6.4). Thus, the use of complex cell signals to define the confidence measure also ensures a close spatial relation between the source term $K$ and the proposed local gain control of the source term by confidence values $Z$ in the filling-in equation (Eq. 6.9).

Confidence signals are computed from pooled complex cell responses $C$ using a function $Z$ with domain-specific contrast signals $K$ as argument. To ensure bounded confidence signals within the range $[0;1]$, a divisive self inhibition is employed. The equation defining the confidence measure reads

$$Z(K) = \frac{KC_{\text{pool}}}{\alpha_Z + KC_{\text{pool}}} + \varepsilon_Z \ . \tag{6.8}$$

The small tonic or base level activity $\varepsilon_Z = 10^{-6}$ of the confidence signals ensures positive-definite filling-in matrices and thus guarantees the existence of a unique solution for the filling-in process. The decay parameter is set to $\alpha_Z = 0.1$. Using the confidence function $Z$, confidence signals for the on- and off-domain are defined as

$$\begin{aligned} Z_{\text{on}} &= Z(K_{\text{on}}) \\ Z_{\text{off}} &= Z(K_{\text{off}}) \ . \end{aligned}$$

### Filling-in

**Confidence-based Filling-in** At the filling-in stage, sparse contrast signals $K$ are subject to an inhomogeneous diffusion process, generating a dense brightness representation. Permeability signals $P$ block the diffusion across boundaries. A confidence measure $Z$ ensures that only valid

contrast are used in the filling-in process. The discretized equation for confidence-based filling-in reads

$$\partial_t U_{ij} = Z_{ij} \left(K_{ij} - \alpha_U U_{ij}\right) + \sum_{n,m \in \mathcal{N}_{ij}} (U_{nm} - U_{ij}) P_{ijnm} \ . \tag{6.9}$$

The passive decay is set to $\alpha_U = 0.5$ and the set $\mathcal{N}_{ij}$ defines the lattice of nearest neighbors for each location $(i, j)$ such as $\mathcal{N}_{ij} = \{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\}$.

Contrast signals feeding the filling-in process occur for two domains, namely on- and off, taking into account the non-negativity of cell response. For the proper function of the filling-in process, on and off contrast signals feed into segregated filling-in networks (Arrington, 1996; Gove et al., 1995; Pessoa et al., 1995) Filling-in activities for the on and off domain $U_{\text{on}}$ and $U_{\text{off}}$ follow the dynamics

$$\begin{aligned} \partial_t U_{\text{on},ij} &= Z_{\text{on},ij} \left(K_{\text{on},ij} - \alpha_U U_{\text{on},ij}\right) + \sum_{n,m \in \mathcal{N}_{ij}} (U_{\text{on},nm} - U_{\text{on},ij}) P_{ijnm} \\ \partial_t U_{\text{off},ij} &= Z_{\text{off},ij} \left(K_{\text{off},ij} - \alpha_U U_{\text{off},ij}\right) + \sum_{n,m \in \mathcal{N}_{ij}} (U_{\text{off},nm} - U_{\text{off},ij}) P_{ijnm} \ . \end{aligned} \tag{6.10}$$

**Permeability Signals** Permeability signals are a complementary representation of boundary signals. At positions of zero-valued boundary signals permeability takes its maximum value. Likewise, at regions contours with high boundary signals permeability is virtually zero. Consequently, permeability is modeled as a decreasing function of boundary signals:

$$\widetilde{P} = \rho \exp(-\lambda B) \ .$$

The permeability coefficient is set to $\rho = 45$, the scale of the boundary signal is set to $\lambda = 500$. The exponential decay ensures a more rapid decline of permeability signals near boundaries than the divisive interaction used in previous filling-in models.

In a discrete implementation, permeability signals occur at the same grid points as filling-in signals. In the above equation however, permeability signals regulate the diffusive coupling *between* adjacent filling-in activities $u_{ij}$ and $u_{mn}$. Permeability signals between positions $(i, j)$ and $(n, m)$ are modeled as the mean of the respective permeability signals at these positions:

$$P_{ijnm} = \frac{\widetilde{P}(i, j) + \widetilde{P}(n, m)}{2} \ .$$

This ensures a symmetric relation of permeabilities between adjacent cells, i.e., $P_{ijnm} = P_{nmij}$.

**Numerical Solution of the Filling-in Equation** As first noted by Arrington (1993), the steady state responses $\partial_t U_{\text{on/off}} = 0$ of the filling-in dynamics as specified in Eq. 6.10 are given by the linear system $\mathbf{M}\mathbf{u}_{\text{on/off}} = \mathbf{k}_{\text{on/off}}$. The matrix $\mathbf{M}$ denotes the lateral interactions, the vector $\mathbf{u}$ denotes the domain specific filling-in activity, and $\mathbf{k}$ denotes the domain specific contrast input which is locally gain-controlled by confidence signals. The 1D vectors $\mathbf{u}$ and $\mathbf{k}$ are obtained from the 2D representations $U$ and $K$ by linewise scanning from upper left to lower right.

For a decay parameter $\alpha_U > 0$ and confidence-values as specified in Eq. 6.8, the matrix $\mathbf{M}$ is guaranteed to be invertible. The steady state solution of the filling-in equation for each domain thus reads

$$\begin{aligned} \mathbf{u}_{\text{on}} &= \mathbf{M}^{-1}\mathbf{k}_{\text{on}} \\ \mathbf{u}_{\text{off}} &= \mathbf{M}^{-1}\mathbf{k}_{\text{off}} \ . \end{aligned}$$

The equations are solved efficiently using the method of preconditioned conjugate gradients (PCG) (Schwarz, 1993).

**Brightness**

Final brightness predictions $O$ are obtained by taking the difference between filling-in activities from the on and off domain:

$$O = U_{\text{on}} - U_{\text{off}} \ . \tag{6.11}$$

## 6.7 Simulations

In this section we present simulation result demonstrating the competencies of the proposed scheme of confidence-based filling-in. The results show that confidence-based filling-in exhibits basic properties which makes it suitable for the computation of surface properties in early vision. For some simulations it is instructive to compare the new scheme of confidence-based filling-in with results generated by the corresponding standard filling-in equation. Keeping all other parameters the same, standard filling-in is simulated by employing a constant unit-valued confidence measure $Z = 1$ in Eq. 6.10. In all simulations, final brightness predictions as specified in Eq. 6.11 are shown.

In a first set of simulations, we demonstrate that confidence-based filling-in—unlike standard filling-in—can generate brightness predictions which are invariant against changes of the shape and size of the regions to be filled in (Sec. 6.7.1). Next we show that brightness reconstruction by confidence-based filling-in is more robust against stimulus noise compared to standard filling-in (Sec. 6.7.2). The mechanism of confidence-based filling-in is then applied to simulate psychophysical data on brightness perception (Sec. 6.7.3). In order to demonstrate the model's capacity to deal with real world data, we finally show results for the processing of real camera images (Sec. 6.7.4). The simulation results shown in this section have also been presented in Neumann et al. (2001), despite Fig. 6.7 and 6.8.

### 6.7.1 Invariance Properties

The first investigation focuses on the properties of the filling-in mechanisms regarding their dependency on the parameter settings and the size of the region to be filled-in. We start with simple luminance patterns showing a central light square of different sizes on a dark background (Fig. 6.5). For the standard filling-in mechanism, the generated brightness signals tend to bow depending on the strength of the permeability coefficient. An increase in the permeability helps to generate flat signals, but the mean brightness is still lower for the large square compared to the small square (Fig. 6.5, middle). On the other hand, the corresponding brightness surfaces generated by confidence-based filling-in remain invariant against these parameter changes: they are always flat, and remain stable for different sizes of the central square (Fig. 6.5, right).

Next, the same mechanisms are applied to another test image that contains shapes of different form and size but of the same luminance level (Fig. 6.6, left). Again, the results reveal a potential weaknesses of standard filling-in: depending on the size or diameter of a pattern, which is unknown a priori, the brightness signals appear at a different amplitude and show different amounts of bowing (Fig. 6.6, middle). With confidence-based filling-in, the brightness patterns appear homogeneous and of almost the same brightness (Fig. 6.6, right). The small variations that occur in the brightness amplitude for confidence-based filling-in are caused by different contrast input measured at the shape contours. Since the contrast measurement is based on an isotropic center-surround mechanism, the responses vary depending on the curvature of the local contour and are slightly higher for higher curvature.

In a final set of simulations we compare the brightness predictions of standard and confidence-based filling-in for circles of different radii and luminance values. For small radii, brightness predictions by both filling-in mechanisms decrease monotonically with increasing radius. This

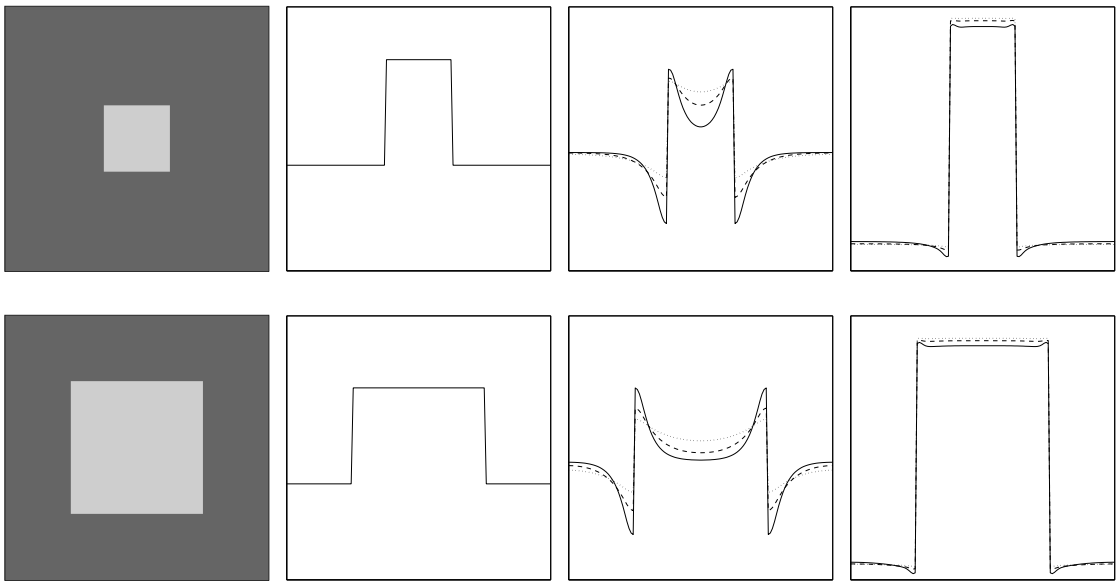

**Fig. 6.5:** Generation of brightness appearance for a rectangular test pattern utilizing mechanisms of standard and confidence-based filling-in. *Left to right:* Stimulus of a central white square on a dark background, corresponding luminance profile, and simulation results for standard filling-in and confidence-based filling-in under variations of the permeability coefficient. *Top row:* Small central square of $32 \times 32$ pixels. *Bottom row:* Large central square of $64 \times 64$ pixels. The total size of both input images is $128 \times 128$ pixels. Brightness predictions for different values of the permeability coefficient $\rho \in \{15, 45, 135\}$ are depicted with *solid*, *dashed*, and *dotted lines*, respectively. For standard filling-in *(third row)*, brightness predictions depend on the value of the permeability coefficient $\rho$ and the size of the central square. For confidence-based filling-in *(rightmost row)*, on the other hand, brightness predictions remains largely unaffected by both kinds of parameter variations.

finding is consistent with an early psychophysical investigation by Hanes (1951). For larger radii, however, the brightness predictions obtained for standard and confidence-based filling-in differ considerably. For standard filling-in, the predicted brightness asymptotically approaches zero, irrespectively of the different luminance values of the input stimuli. For confidence-based filling-in, on the other hand, the brightness predictions saturate at different levels depending on the different luminance values.

The observed property of confidence-based filling-in to generate flat brightness profiles holds true for virtually arbitrarily large homogeneous regions, since the data term in the interior of such regions is switched off by zero confidence values. We conclude that confidence-based filling-in helps to generate a brightness representation that is largely invariant against shape and size transformations, thus improving the robustness of filling-in mechanisms.

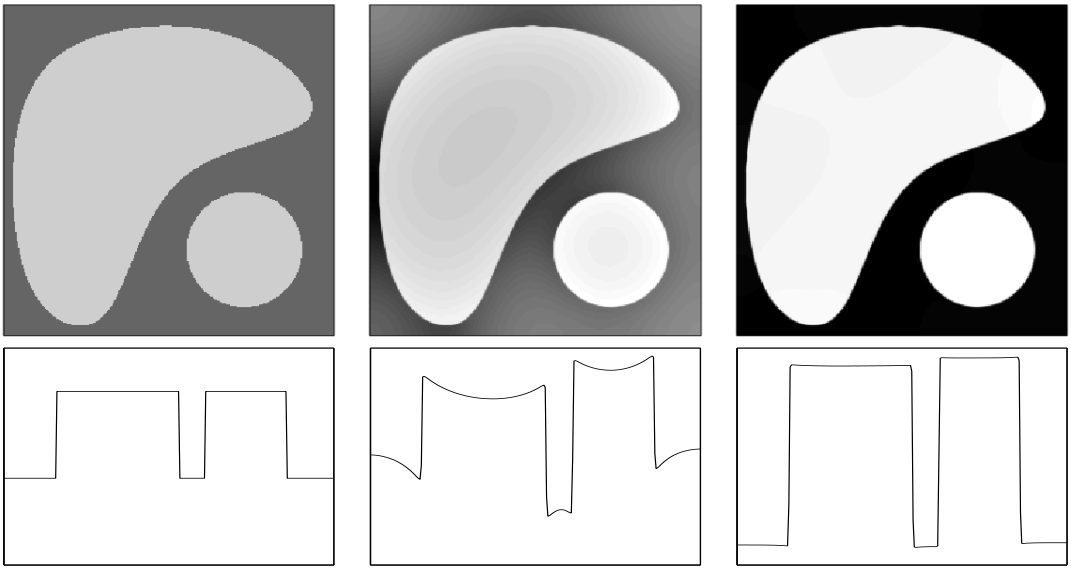

**Fig. 6.6:** Filled-in brightness signals for a test stimulus containing shapes of different size but of the same luminance level. The size of the input stimulus is $256 \times 256$ pixels. *Top row, left to right:* Input luminance pattern, brightness signal generated by standard filling-in and by confidence-based filling-in. *Bottom row:* Corresponding profiles of the luminance function and the brightness predictions taken along the 2D picture diagonals (from upper left to lower right corner).

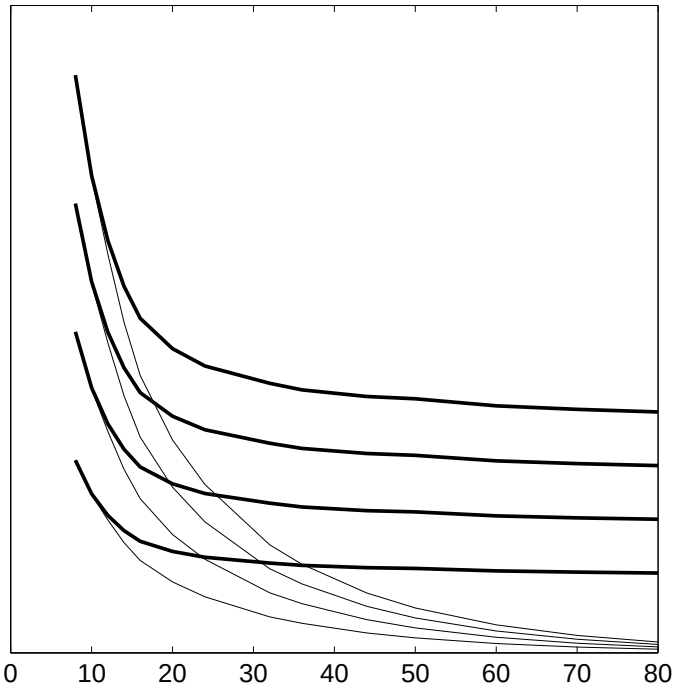


**Fig. 6.7:** Brightness prediction of standard *(solid lines)* and confidence-based filling-in *(bold lines)* for circles of varying diameter and luminance. The ordinate denotes the brightness prediction measured in the center of the circles, the abscissa denotes the diameter of the circle. Circles of four different luminance values are employed, and for each luminance value the diameter of the circle is varied from 8 to 80 pixels. For small radii, brightness predictions by both filling-in mechanisms decrease with increasing stimulus size. For larger radii, however, only confidence-based filling-in leads to a brightness prediction independent of the stimulus size and allows to distinguish the different luminance values in the input stimulus.

### 6.7.2 Noise Robustness

In this section we address the noise suppression properties of filling-in mechanisms. We employ a stimulus of a dark ellipse on light background which is heavily corrupted by high amplitude Gaussian noise (Fig. 6.8, left). The brightness predictions obtained for standard and confidence-based filling-in differ considerably. While confidence-based filling-in generates a smooth, almost flat surface (Fig. 6.8, right), the brightness prediction for standard filling-in is ragged and has high amplitude peaks at the boundaries of the ellipse (Fig. 6.8, middle). The different results are caused by the different treatment of the data term by both filling-in mechanism: For standard filling-in, contrast data at every position are used as input, in particular noisy spurious responses in the interior of the ellipse. For confidence-based filling-in, only the valid contrast signals at the region boundaries contribute to the data term. Contributions of invalid contrast signals in the interior of the ellipse are switched off by zero-valued confidence signals, resulting in a smooth filling-in of the valid contrast data at the contours. We conclude that confidence-based filling-in can generate smooth brightness predictions even in the presence of noise.

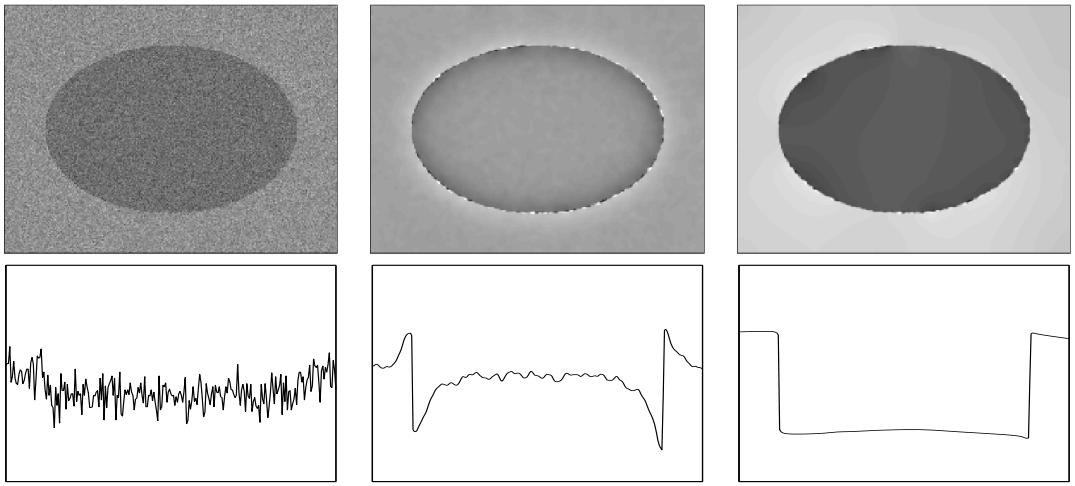

**Fig. 6.8:** Generation of brightness surface for a stimulus of a noisy ellipse utilizing mechanisms of standard and confidence-based filling-in. The size of the input stimulus is $253 \times 189$ pixels. *Top row, left to right:* Input luminance pattern, brightness signal generated by standard filling-in and by confidence-based filling-in. *Bottom row:* Corresponding profiles of the luminance function and the brightness predictions. While the brightness prediction of standard filling-in appears ragged and noisy, confidence-based filling-in generates a smooth and flat brightness surface.

### 6.7.3 Psychophysical Data on Brightness Perception

For a number of luminance pattern, the perceived brightness differs considerably from the physical intensity distributions. Such visual illusions provide important cues to the mechanisms underlying human brightness perception. Filling-in models have been successfully employed to simulate a number of such brightness phenomena. Here we show that the new scheme of confidence-based filling-in can also account for psychophysical data on brightness perception.

#### Simultaneous Contrast

In the first set of simulations, the processing of classical luminance pattern generating simultaneous contrast is examined (Fiorentini et al., 1990; Heinemann, 1972). In simultaneous contrast, the apparent brightness of a central square changes depending on the surround: the same square

looks darker on light background and lighter on dark background (Fig. 6.9, top row). Confidence-based filling-in can successfully account for simultaneous contrast (Fig. 6.9, bottom row), as have previous filling-in models (Grossberg and Todorović, 1988; Neumann, 1996; Pessoa et al., 1995).

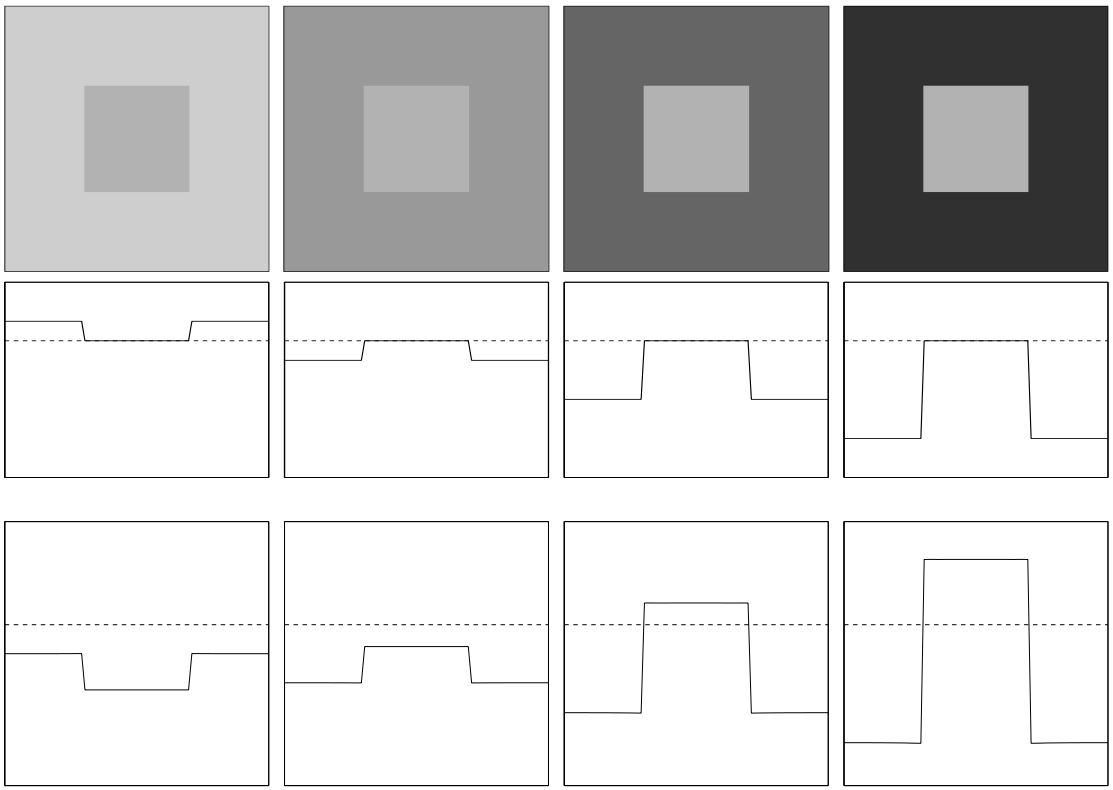

**Fig. 6.9:** Simultaneous contrast stimuli. *Top row:* Input stimuli consisting of the same gray square embedded in different surrounds of decreasing luminance. The size of the images is $80 \times 80$ pixels. *Second row:* Corresponding luminance profiles taken at the center of the input stimuli. The constant luminance level of the central gray square is highlighted by a *dashed horizontal line*. *Bottom row:* Profiles of brightness predictions generated by confidence-based filling. The *dashed horizontal lines* serves as a reference level indicating the mean brightness activity measured at the central regions of all four brightness predictions. The simulations show that confidence-based filling-in successfully accounts for simultaneous contrast, i.e., the increase in perceived brightness of the central square with decreasing luminance of the surrounding background.

### Craik-O'Brien-Cornsweet-Effect

In the next set of simulations we focus on remote border contrast effects and their creation of brightness differences. These cases provide examples of the crucial role of edges in determining the brightness appearance. For example, two regions of equal uniform luminance separated by a cusp edge appear of different brightness, the so-called Craik-O'Brien-Cornsweet (COC) effect (Cornsweet, 1970; Craik, 1940; O'Brien, 1958; see Todorović, 1987, for a review of the COC effect and its theoretical implications). A standard COC stimulus and a COC grating made of adjacent, flipped standard COC stimuli are depicted in Fig. 6.10, left column. These types of stimuli have been identified as the most challenging ones for alternative theories of brightness perception, such as, e.g., filter theories. In fact, yet only filling-in models can properly predict the brightness appearance for COC stimuli and their variants (Blakeslee and McCourt, 1997, 1999).

The processing of a standard COC stimulus is shown in Fig. 6.10, top row. A COC stimulus consist of a cusp edge, separating two regions of equal luminance. Both regions seem to be of different

brightness: the region associated with the negative lobe of the cusp is perceived as uniformly darker compared to the region associated with the positive lobe of the cusp. Confidence-based filling-in correctly predicts this effect (Fig. 6.10, top right), as do previous filling-in models (e.g., Grossberg and Todorović, 1988).

A COC grating (Fig. 6.10, bottom row) consists of a sequence of cusp edges having pairwise opposite contrast polarities. This stimulus is perceived as a series of alternating dark and bright stripes similar to a square wave. As reviewed in Sec. 6.2, the temporal dynamics of brightness perception for such a COC grating are consistent with a neural filling-in mechanism (Davey et al., 1998). Confidence-based filling-in, at equilibrium, correctly predicts the final brightness percept of a square-wave pattern (Fig. 6.10, bottom right).

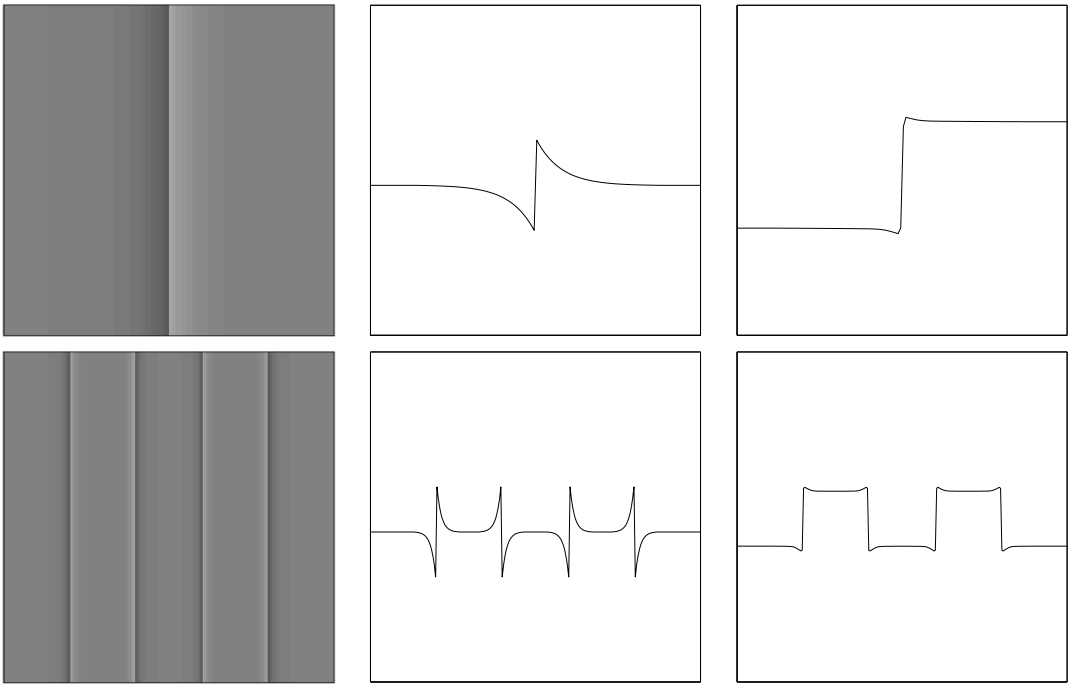

**Fig. 6.10:** Filled-in brightness signals for a standard COC stimulus *(top row)* and a COC grating *(bottom row)* made of cusps of opposite contrast polarity. *Left to right:* Input luminance pattern, corresponding profile, and the profile of the brightness pattern generated by confidence-based filling-in. Confidence-based filling-in correctly predicts the perceived brightness for both kinds of COC stimuli.

### 6.7.4 Real World Application

In order to demonstrate the functional significance of the proposed mechanism we show the processing results for a camera image of a real object. In order to exclude any possible influences from 3D effects, e.g., by shadowing or variations in surface orientations, we use a matte black card-board that is attached to a homogeneous background surface. This intrinsically flat scene is directly illuminated by a point-like light source at a distance of approximately 2 m. The illumination by a single light source generates a significant intensity gradient in the original intensity image. The target surface is imaged from two different distances at about 2 and 1 m, respectively (Fig. 6.11, left column).

Simulation results show that the mechanism of confidence-based filling-in is capable of generating a representation of homogeneous surface properties (Fig. 6.11, third and fourth column). The result is independent of the projected region size, thus showing the property of size invariance. Further, the illumination gradient is discounted and the noise is successfully suppressed.

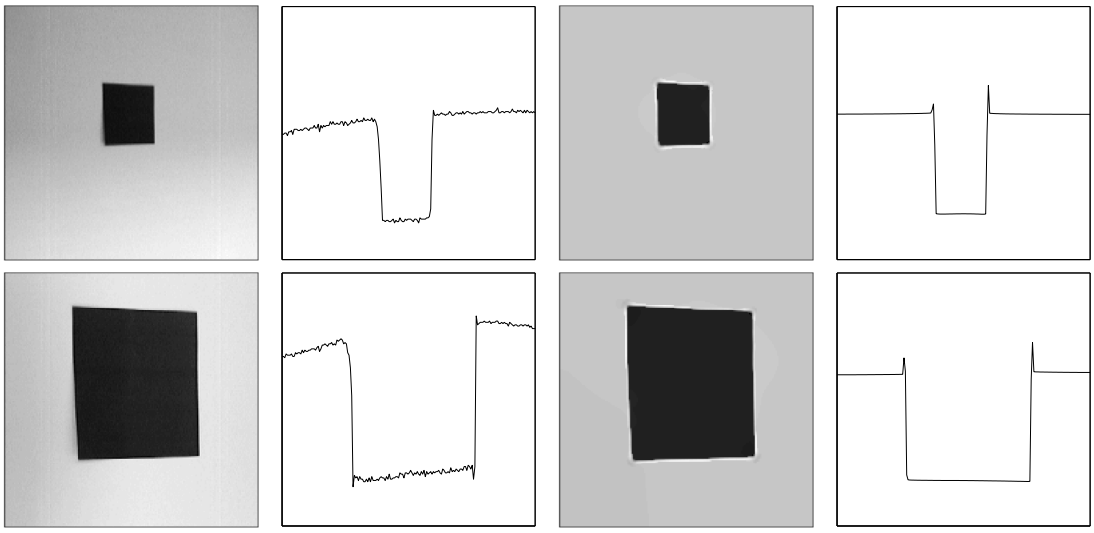

**Fig. 6.11:** Processing the camera images of a flat 3D object acquired from different distances. The target object is a matte card-board of low reflectance which is attached on a lighter background surface and illuminated by a primary light source generating a visible illumination gradient. *Top row:* Input intensity image of the object at a larger viewing distance and a profile section *(left pair)* together with the corresponding filling-in result and a profile section *(right pair). Bottom row:* Corresponding input representations and processing results for the object at a closer viewing distance. The results show three properties of confidence-based filling-in: i) generation of a homogeneous brightness surface for regions of different sizes, ii) reduction of noise, and iii) discount of the illumination gradient.

## 6.8 Restoration of Reference Levels

A main deficit of filling-in mechanisms is the proper restoration of reference levels. Input to the filling-in stage is generated by contrast signals. These contrast signals result from an initial center-surround processing which does not respond to homogeneous regions, but discounts the illuminant by suppressing the DC component of the input signal. As a desired result, the responses are invariant against changing illumination (Land, 1983). At the same time, the use of DC-free contrast signals destroys all information about the reference levels of the input stimulus. For a number of visual stimuli however, e.g., a luminance staircase or a luminance pyramid, the proper restoration of reference levels is essential for a valid brightness reconstruction.

To illustrate the problem, consider the case of a luminance staircase which consists of adjacent homogeneous regions of increasing luminance (Fig. 6.12, top). At each dark-light transition, juxtaposed on and off contrast signals are generated which feed into segregated on- and off-filling-in systems, respectively (Fig. 6.12, middle). Since the relative contrast is the same at each dark-light transition, and filling-in of the contrast signals is confined to a single region by localized boundary signals, the filled-in brightness representations for each domain occur only at a single non-zero amplitude (Fig. 6.12, middle). Combination of both on and off domains results in a final brightness prediction where the different luminance values of the three innermost steps are represented by only a single level (Fig. 6.12, bottom). In general, brightness filling-in of contrast signals for a luminance staircase can generate only three different levels, irrespectively of the number of luminance steps in the input stimulus.

Several different solutions have been proposed for the reconstruction of reference levels, such as directional filling-in (Arrington, 1996) or the use of a dense representation of luminance information which is incorporated either implicitly by using contrast-sensitive signals with nonzero DC level (Grossberg and Todorović, 1988) or explicitely by proposing a separate luminance channel (Neumann, 1996; Pessoa et al., 1995). Both types of solutions can only partly solve the prob-

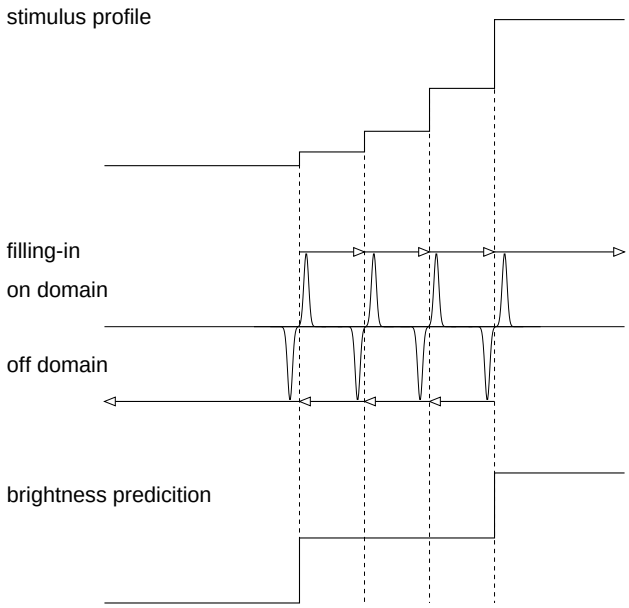


**Fig. 6.12:** Filling-in of contrast signals for a staircase stimulus. *Top:* Profile of a luminance staircase. *Middle:* Juxtaposed on and off contrast at luminance discontinuities are filled in for two separated domains *(arrows)*. For the purpose of illustration, the off domain signals are mirrored at the $x$-axis. *Bottom:* Final brightness prediction generated by subtracting off from on contrast signals. The sketch illustrates a weakness of filling-in processes: the luminance steps give rise to contrast signals of the same amplitude, which cannot be used to restore the different luminance values in the input stimulus.

lem. Directional filling proposes semi-permeable boundaries with allow on and off contrast signals to flow in the direction of increasing respectively decreasing brightness. This approach however fails to discriminate a sequence of COC flanks from a luminance staircase. Further, directional filling-in cannot account for luminance staircases with added pedestals or other line-like structures of opposite polarity, which cause the directional filling-in to stop. Approaches based on a dense luminance information, on the other hand, are difficult to reconcile with ample empirical evidence suggesting that retinal ganglion cells do not show significant responses to homogeneous regions.

Motivated by these considerations we suggest an alternative approach to the restoration of reference levels, based on a localized coding of luminance information at contrast locations. In this approach, sparse contrast signals are modulated to carry an additional luminance component.

We suggest that luminance-modulated contrast signals can be generated from an initial representation of contrast-sensitive signals $X_{\text{on}}$ and $X_{\text{off}}$ which contain a DC component. Such contrast-sensitive signals result from shunting interactions as specified in Eq. 3.3 and Eq. 3.2. These responses can be viewed as a multiplexed representation of a pure contrast signals $K_{\text{on/off}}$ and and a pure luminance signal $L$ (Neumann, 1996). On and off contrast signals for each domain $K_{\text{on}}$ and $K_{\text{off}}$ and the luminance signal can be generated from the contrast-sensitive responses $X_{\text{on}}$ and $X_{\text{off}}$ as follows:

$$\begin{aligned} K_{\text{on}} &= \eta \left[ X_{\text{on}} - X_{\text{off}} - \kappa \right]^+ \\ K_{\text{off}} &= \eta \left[ X_{\text{off}} - X_{\text{on}} - \kappa \right]^+ \\ L &= X_{\text{on}} + X_{\text{off}} \ , \end{aligned}$$

where $\eta = 100$ is scaling parameter and $\kappa = 0.005$ is a small threshold to suppress spurious contrast responses. Luminance-modulated contrast signals $\widetilde{K}_{\text{on}}$ and $\widetilde{K}_{\text{off}}$ for the two domains are

then modeled as

$$\widetilde{K}_{\text{on}} = L\left(\frac{K_{\text{on}}}{\alpha_{\widetilde{K}} + K_{\text{on}}}\right) + \zeta_{\widetilde{K}} K_{\text{on}}$$
$$\widetilde{K}_{\text{off}} = L\left(\frac{K_{\text{off}}}{\alpha_{\widetilde{K}} + K_{\text{off}}}\right) + \zeta_{\widetilde{K}} K_{\text{off}} \ , \tag{6.12}$$

where $\alpha_{\widetilde{K}} = 0.01$ and $\zeta_{\widetilde{K}} = 0.005$. In these equations, luminance signals $L$ are selectively gated by compressed contrast signals $K/(\alpha_{\widetilde{K}} + K)$. The compression ensures that the gating of luminance signals is largely unaffected by the particular amplitude of the contrast signal. Brightness contrast effects motivate a small additional contrast-dependent term $\zeta_{\widetilde{K}} K$.

Luminance-modulated contrast responses as described in Eq. 6.12 result from the steady-state responses of a disinhibition circuit which follows the dynamics

$$\partial_t Y_{\text{on}} = -\alpha_{\widetilde{K}} Y_{\text{on}} + \alpha_{\widetilde{K}} L - Y_{\text{on}} K_{\text{on}}$$
$$\partial_t \widetilde{K}_{\text{on}} = -\widetilde{K}_{\text{on}} + L + \zeta_{\widetilde{K}} K_{\text{on}} - Y_{\text{on}} \ .$$

The respective equations for the off domain are obtained analogously.

To further illustrate the functionality of the proposed mechanism, consider a sketch of the underlying circuit depicted in Fig. 6.13. In this circuit the luminance signal $L$ provide an excitatory

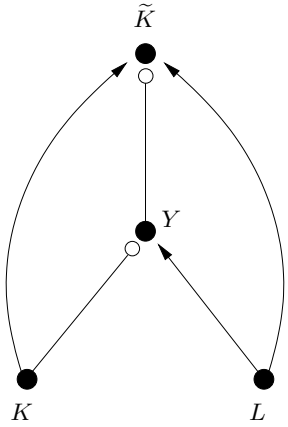


**Fig. 6.13:** Sketch of the proposed circuit generating luminance-modulated contrast signals $\widetilde{K}$ from contrast signals $K$ and luminance signals $L$. Arrows denote excitatory input, *circles* at the end of lines denote inhibitory input.

input to the final stage of the luminance-modulated contrast signal $\widetilde{K}$. Self-inhibition of the luminance signal by means of an an intermediate stage $Y$ cancel any response from the luminance signal alone. The self-inhibition in turn is subject to shunting inhibition from contrast signals $K$. The disinhibition ensures that luminance signals are only present at locations of non-zero contrast signals.

Simulation results for a battery of test stimuli using confidence-based filling-in of luminance-modulated contrast signals are depicted in Fig. 6.14. The final brightness predictions are generated by

$$O = U_{\text{on}} - 0.4\, U_{\text{off}} \ .$$

The imbalance between the on and off domain ensures that a luminance-modulated signal of one domain is not completely extinguished by a luminance-modulated signal of the same strength from the complementary domain. In the simulations, a small permeability of $\rho = 1$ is used.

In a first set of simulations we show that the newly proposed scheme successfully generates reference levels in the case of a luminance staircase and a luminance pyramid (Fig. 6.14, upper two rows). Next we employ the circuit for the processing of a standard COC stimulus and a sequence of COC flanks of same polarity. Simulation results account for the generation of two reference brightness levels by a standard COC stimulus, and the different perception of a COC sequence compared to a luminance staircase. Finally, we show that filling-in of luminance-modulated contrast signals can also account for brightness contrast phenomena such as occurring in the classical simultaneous contrast stimulus (Fig. 6.14, bottom row). Modeling both brightness contrast effects and the generation of different reference levels as occurring, e.g., in a luminance staircase, impose a particular challenge to any brightness reconstruction scheme, since the reconstruction has to be based on the sensitive trade-off between both contrast and luminance signals.

In all, the results show that the newly proposed scheme can account for a variety of brightness phenomena. In particular, it has been demonstrated that filling-in of luminance-modulated contrast signals can be successfully employed for the proper restoration of reference levels.

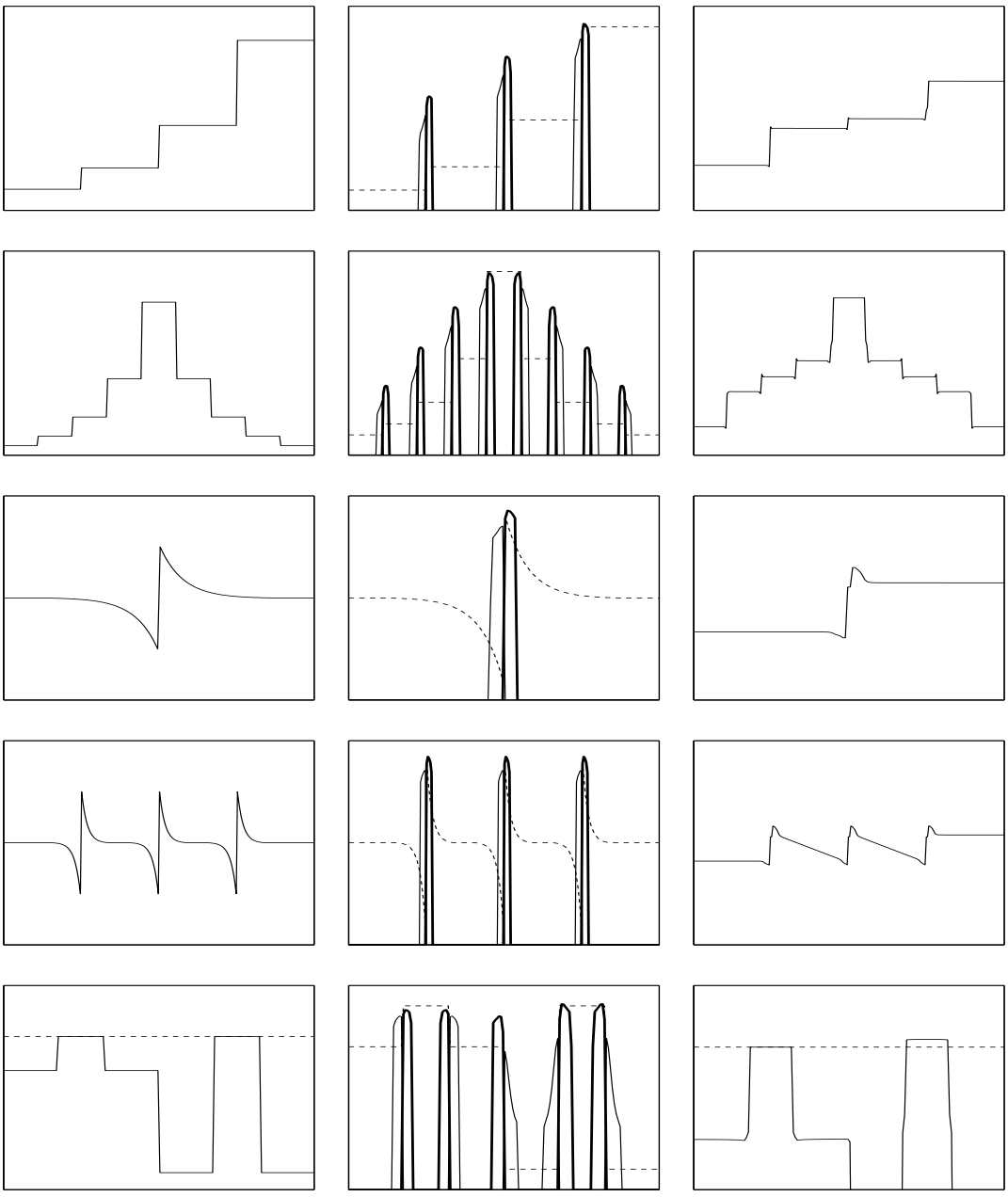

**Fig. 6.14:** Brightness reconstruction using luminance-modulated contrast signals for a number of test stimuli, including a luminance staircase, luminance pyramid, COC, COC sequences of same contrast polarity, and simultaneous contrast *(from top to bottom). Each row, from left to right*, shows a profile of the input stimulus, the corresponding luminance-modulated on contrast *(bold)* and off contrast signals *(solid)* together with an overlay of the stimulus profile *(dashed)*, and the final brightness prediction by confidence-based filling-in.

## 6.9 Discussion and Conclusions

We have suggested a model for human brightness perception based on a newly proposed mechanism of confidence-based filling-in. Confidence signals are used to discriminate positions of reliable measurements of contrast data from positions where no data is available. Incorporating a confidence-measure in the filling-in equation allows to generate a continuous, dense brightness surface from sparse contrast data as signaled by retinal ganglion cells. Unlike standard filling-in, the suggested mechanism allows to generate a size invariant brightness representation even on the basis of sparse input data. The selective choice of valid contrast signals as input to the filling-in stage by means of the proposed confidence measure renders the brightness reconstruction robust against noisy input data. Furthermore, a number of perceptual phenomena such as brightness contrast effects or the COC effect can be successfully modeled. An application to real world data further demonstrates the functional significance of the proposed scheme.

A number of filling-in models have been proposed which successfully account for different brightness phenomena (Cohen and Grossberg, 1984; Grossberg and Todorović, 1988; Pessoa et al., 1995). However, these schemes rely on the existence of a dense luminance signal, which is incorporated either implicitly (Grossberg and Todorović, 1988) or explicitly (Neumann, 1996; Pessoa et al., 1995) into the filling-in process. The dense luminance signal assumed in these models is non-zero in homogeneous regions and of the same order of magnitude as the contrast response to luminance discontinuities. Empirical evidence, however, does not support a transmission of such a dense luminance signals along the primary visual pathway (Enroth-Cugell and Robson, 1966). Instead, retinal ganglion cells respond vigorously only at luminance discontinuities but not inside homogeneous regions. The present model, for the first time, suggest how the seemingly contradicting findings of sparse contrast signals and the perception of dense brightness surfaces can be reconciled using confidence-based filling-in.

The proper restoration of reference levels has been problematic for filling-in models. We suggest a possible mechanism based on sparse luminance-modulated contrast signals. The proposed scheme can account for a number of stimuli, including luminance staircases, COC flanks, and simultaneous contrast. We suggest that the use of multiple scales is another important component to be incorporated into filling-in models for brightness perception (e.g., Sepp and Neumann, 1999).

Filling-in models are based on the assumption of distributed, topographically organized maps of boundaries and regions (Komatsu et al., 1996). The need for an explicit and intrinsically redundant representation of extended brightness regions which is crucial to all kinds of filling-in models has been subject to intense debate. Instead of such an explicit representation, alternative approaches have suggested a non-topographic or sparse coding of contrasts and boundaries using a compact symbolic code (Marr and Hildreth, 1980), or a sparse localized code (Pollen and Ronner, 1983). Regarding filling-in, both visual scientists (e.g., Kingdom and Moulden, 1992), and philosophers (Dennet, 1992) have argued against the logical need for a neural spreading of activity and against a continuous representation of the brightness profiles (for a review see Pessoa et al., 1998). It has been proposed that instead of filling-in, the brain could simply assign a symbolic brightness label to a bounded region or could ignore the absence of direct neural support (Dennet, 1992). This kind of logical reasoning has to face the fact that there is now ample empirical evidence, mostly from psychophysics, that brightness perception indeed depends on an spreading, or filling-in, of neural activity. Consequently, we conclude that the question no longer is *whether*, rather than "*why* does the brain fill in?" (Pessoa and Neumann, 1998).

It has been suggested that filling-in processes are involved in behavioral tasks such as grasping or are an important component for the recognition of occluded objects (Grossberg, 1994). The validation of these suggestions is a challenging task for future research. Models of surface completion are helpful by integrating empirical results into a precise computational and algorithmic description.

Chapter 7

# Conclusion and Outlook

In this concluding chapter we shall summarize the main findings of this work and point toward future investigations.

## 7.1 Conclusion

In the present work we have suggested, examined and evaluated mechanisms of early and midlevel vision for the robust and reliable processing and representation of various visual primitives. These visual primitives can be coarsely classed within two parallel systems, one system for the processing of discontinuities such as contrasts, contours and corners, and another system for the representation of homogeneous surface qualities such as brightness. The two systems and their subsystems are not defined by strictly segregated modules which solve the particular task by highly specialized mechanisms. Instead, building upon earlier suggestions by Grossberg and coworkers, we have devised a unified architecture of interacting systems for the processing of contrasts, contours, junctions and brightness surfaces. We have shown how various tasks in low and midlevel vision such as contrast enhancement and noise suppression, contour grouping, junction detection and robust surface representation can be solved efficiently based on common mechanisms and representations of interacting (sub-)systems within a single architectural framework. We hypothesize that such a coherent representation of the external world provides the necessary base for successful interaction within the environment.

In the following we shall summarize the proposed mechanisms and findings of the four main systems we have studied in detail, namely contrast processing, contour grouping, corner and junction detection and surface representation.

### 7.1.1 Contrast Processing

The extraction of contrast information at luminance discontinuities is the first main processing step in natural and the most artificial vision systems. Based on physiological findings we have suggested a mechanism of dominating opponent inhibition (DOI) for robust contrast processing. We have integrated the DOI mechanism into an existing nonlinear simple cell circuit and have shown that the model with DOI can account for empirical data on luminance gradient reversal. Next we have applied the same model with identical parameter settings for the processing of both natural and artificial images. We have shown in a number of simulations that the new mechanism of DOI results in a significant suppression of responses to noisy regions. The suppression is largely invariant against the amount of the noise level. This adaptive processing has been further examined in a stochastic analysis, indicating that the amount of suppression by DOI is proportional to the noise level. Further we have demonstrated that the proposed circuit remains sensitive even to small contrast changes. The results lead to the proposal of a new functional role of the dominant inhibition as observed empirically, namely to allow for robust contrast processing under suboptimal, noisy viewing conditions such as occurring, e.g., under weak illumination.

We have also addressed the intensively studied issue of contrast-invariant orientation tuning. We have shown that contrast-invariant orientation tuning can be achieved in a purely feedforward model based on inhibition. Inhibition can be either introduced at the LGN level or between simple cells of opposite contrast polarity. In either case, inhibition suffices to render the tuning of the simple cells invariant against contrast changes. Dominating opponent inhibition serves to sharpen the tuning of the simple cells. We have shown that the resulting tuning width of model simple cells is in accordance with empirical findings. More important, the results demonstrate that recurrent intracortical interactions are not necessary for the generation of contrast-invariant orientation tuning.

Overall, we have devised and evaluated a biologically plausible mechanism of dominating opponent inhibition which can account for a number of empirical findings. The application to image processing motivates a functional role of the mechanism, namely the adaptive suppression of noise.

### 7.1.2 Contour Grouping

Local contrast measurements are often fragmented and noisy. The evaluation of such initial contrast estimates within a more larger context by grouping of neighboring responses results in a coherent representation of contour information. Contrast sensitive neurons in the primary visual cortex have rather small classical receptive fields, but interact by means of horizontal long-range connections with cells of similar orientation preferences. It has be hypothesized that these long-range interactions play an important role in the grouping of coherent activity and instantiate a neural implementation of the Gestalt rule of good continuation. Based on these considerations we have developed a model of recurrent long-range interactions in V1 for contour processing. Unlike other models, where long-range connections are described by complex cocircularity patterns, the long-range connections in our model are confined to cells with parallel, near colinear receptive fields in accordance with empirical findings. We have shown for several artificial and natural images that the recurrent long-range processing results in a selective enhancement of coherent activity at contour locations. The competencies of the model have been quantitatively evaluated using two measures of contour saliency and orientation significance. The evaluation results show for a variety of stimuli of different contrast and noise values that both contour saliency and orientation significance is enhanced during recurrent long-range interaction. We have further examined the properties of the colinear interaction scheme when applied to curved patterns, showing that the model also successfully enhances curved pattern to a certain degree, depending on curvature.

To address the biological relevance of the model we have simulated model responses to stimuli of contour fragments as used in an empirical study by Kapadia et al. (1995). In this study, the response to a central bar element decreases when the bar is embedded in a texture of randomly oriented bars, but increases if colinear flankers are added on either side of the central bar. Our model qualitatively reproduces both effects of surround suppression and facilitation depending on the layout of the surrounding bars.

We have also suggested and examined a model variant using early feedback of grouped responses. This model variant shows an even stronger enhancement of contour saliency compared to the standard model of recurrent long-range interactions. These results may suggest a functional role for the layout of different feedback projections in V1.

The proposal and evaluation of a novel model of colinear recurrent long-range interactions has been complemented by an extensive literature survey regarding both empirical and computational work on long-range interactions and contour grouping. We have reviewed a large body of empirical evidence for long-range interactions from different disciplines, including anatomy, physiology, psychophysics and statistics. With regard to other computational and modeling approaches, we have first discussed the basic computational elements underlying most computational schemes and have suggested an overall framework which allows to classify different approaches. Having established

this framework, we have discussed several important approaches to contour grouping, including both biologically motivated models and more computationally inspired algorithms.

To sum up, we have shown that basic tasks of early and midlevel vision such as noise suppression and contour enhancement can be be realized within a biological plausible architecture based on colinear recurrent long-range interactions. Besides the competencies discussed so far, the models further realizes a neural mechanism for the robust representation of corners and junctions, as shall be detailed in the following.

### 7.1.3 Corner and Junction Detection

Corners and junctions are relatively rare events in an image, carrying high information content, and are invariant against moderate changes in viewing direction and viewing distance. Junctions provide important cues in various visual tasks, such as the determination of occlusion relationship for figure-ground separation, transparency perception, and object recognition, among others. Thus, the proper detection of corner and junction points is important for both biological and artificial vision systems.

We have proposed a novel scheme for the robust and reliable representation and detection of junction points based on a distributed representation within an orientation hypercolumn. In this scheme, corners and junctions are implicitly characterized by high activity for more than two orientations within a hypercolumn. The recurrent colinear long-range processing as summarized above is employed to generate a robust, coherent hypercolumnar representation of orientations at each location. A measure of circular variance is used to extract corner and junction points from this distributed representation.

In a first set of simulations we compare the detected junctions based on recurrent long-range responses to junction responses as obtained for a purely feedforward model of complex cells. We show for a number of artificial and natural images that localization accuracy and positive correctness is improved by recurrent long-range interaction in comparison with initial feedforward complex cell processing. In a second set of simulations we compare the new scheme with two widely used junction detector schemes in computer vision, based on Gaussian curvature and the structure tensor. Receiver operator characteristic (ROC) analysis is used for a threshold-free evaluation of the different approaches. We show for both artificial and natural images that the new approach performs superior to the standard schemes.

To conclude, we have shown that junctions can be robustly and reliably represented by a suggested biological mechanism based on a distributed hypercolumnar representation and recurrent colinear long-range interactions. Further, we have suggested how the well-established method of ROC analysis can used for a threshold-free evaluation of different junction detection schemes.

### 7.1.4 Surface Representation

The systems summarized so far deal with the processing of discontinuities, i.e., contrast, contours, and corners. In this section we describe a complementary system for the representation of homogeneous brightness surfaces.

Responses of retinal ganglion cells transmitted along the primary visual pathway are strong only at locations of luminance discontinuities but not inside homogeneous regions. How does the perception of such homogeneous regions occurs in the absence of direct neural support? Motivated by regularization theory we suggest a new scheme of confidence-based filling-in for the reconstruction of a dense brightness surface from sparse contrast data. The confidence measure allows to distinguish locations were no contrast input is available from locations of valid contrast signals. The competencies of the new scheme are demonstrated in a number of simulations. In a first set

of simulations we compare the proposed scheme with a standard filling-in mechanism having no confidence measure. We show that only confidence-based filling-in can generate brightness predictions from sparse contrast data which are invariant against the size and shape of the region to be filled in. Further, we show that confidence-based filling-in can generate smooth brightness surfaces even from noisy contrast data, in contrast to standard filling-in. Next we have demonstrated that confidence-based filling-in can account for a number psychophysical data on human brightness perception, such as simultaneous contrast and COC effect. In a final set of simulations we apply confidence-based filling-in for the processing of natural images. The simulation results on natural images further demonstrate three properties of the proposed scheme, namely the generation of a size-invariant brightness representations, the suppression of noise, and the discount of the illuminant.

The reconstruction of reference levels is of particular importance for the proper prediction of the brightness perception of stimuli such as a luminance staircase or a luminance pyramid. Filling-in schemes based on contrast data alone fail to account for the perception of these stimuli. Previous approaches have tried to overcome this failure by using a dense representation of luminance information, which is difficult to reconcile with empirical findings. We have suggested a new approach for the reconstruction of reference levels, where local, sparse contrast signals are modulated to carry an additional luminance component. We have shown for a number of test stimuli, including luminance staircase, a standard COC flank, sequences of COC flanks, and simultaneous contrast, that the newly proposed scheme can successfully predict human brightness perception.

To sum up, we have suggested a novel mechanisms of confidence-based filling-in which allows to reconstruct a dense brightness surface from sparse contrast data. The proposed mechanisms exhibits desired invariance properties and successfully accounts for a number of brightness phenomena.

## 7.2 Outlook

The computational mechanisms and the functional architecture proposed and studied in this work suggests several extensions and directions of future research. These extensions can be grouped into three main categories, namely the study of further interactions between subsystems, the application of mechanisms to model the processing of other visual modalities, and technical applications. In the following we shall discuss sample extensions for each category.

### 7.2.1 Interaction between Subsystems

The interaction between the different subsystems is a guiding idea of the architecture proposed in this work. A number of additional interactions can be introduced to further enhance the robustness and the computational competencies of the system.

First, there is a close and mutual interaction between form and surface perception: form information defines the surface boundaries (as realized in the present model), but likewise, the surface layout determines the contour formation. Filling-in results in a regularized layout of surfaces, and contour processing based such a regularized solution may be more robust. For example, in a technical context it has been shown that edge detection based on diffused images outperforms classical edge detection methods (Perona and Malik, 1990b). Consequently, we suggest that contour information can be determined based on the filling-in results, and is feed back to enhance contour locations as computed in the feedforward stream. This architecture of two interacting systems would be similar to the architecture used in the V1 model developed in this work.

Second, besides contour information, junctions are known to influence the perception of surface color. Thus, junction information could be used as an additional stream for steering the filling-in mechanism. Having suggested the computation of contour information based on a regularized

surface representation naturally leads to the idea of also computing corners and junctions from the resulting distributed contour representation, and feed back and compare the information to the feedforward computations of junctions.

Third, the computational theory underlying the recurrent processing of two mutually connected cortical areas or regions which guided the development of the present V1 model has also been realized in a model of V1–V2 cortico-cortical recurrent processing (Neumann and Sepp, 1999). Unlike the present V1 model, the V1–V2 model accounts for illusory contour formation and uses more complex bipole patterns along with nonlinear interactions between subfields, but does not include an intracortical recurrent processing within V1 (see Sec. 4.3.2 for a detailed review). Similar to the cortical architecture, both intracortical and cortico-cortical recurrent processing streams as realized in the different models could be integrated. It is suggested that V1 responses are modulated by feedback connections from the higher cortical area V2 which integrates information from a broader context.

### 7.2.2 Application to other Modalities

In the present work we have studied mechanisms for the processing of contrast, contours, corners and surface brightness. The application and extension of the present computational mechanisms to the processing of other visual modalities such as texture, motion, depth and color and the integration into a common architecture is a promising approach. Such an extension would probably lead to a deeper understanding of the functional competencies of the proposed mechanisms, and their integration into a single architecture opens various directions of their mutual interactions to further increase the robustness of each processing stream.

### 7.2.3 Technical Applications

In Sec. 3.9 we have described a pioneering application of the contrast detection scheme based on dominating opponent inhibition as a feature detector for object recognition. In a similar fashion, other subsystems within the present architecture could provide features for higher level tasks. In particular, the known importance of corners and junctions for object recognition suggests the use of the proposed junction detection scheme as a feature detector for an artificial object recognition system. The comparison of classification results obtained with the new junction detector to results as obtained with other, classical methods would also provide a high-level, task-oriented evaluation of different junction detection schemes.

The computational methods developed in this work may also stimulate novel paradigms for interactive image editing and transformation. Recently it has been suggested how methods of contour grouping and surface color reconstructions can be employed for interactive "image editing in the contour domain" (Elder and Goldberg, 2001). In this approach, the basic element of an image editing system is a contour rather than a pixel. Frequently occurring image manipulation tasks such as the deletion of an object can be more efficiently realized based on a contour representation. Surface interpolation similar to filling-in processes then converts the modified contour representation into a dense gray-scale image. As pointed out by Elder and Goldberg (2001), such an interactive image editing system also provides a testbed for different computational approaches to contour grouping and surface interpolation.

### 7.2.4 Summary of the Outlook

The present architecture benefits from close interactions of different subsystems, pointing toward the importance different streams and complementary representations for a robust extraction of visual properties. In the long term view, a particularly promising approach might be the integration

of different visual modalities from a successively increasing subset of brightness, texture, color, motion and depth processing streams. We hypothesize that particular tasks such as figure-ground segregation cannot be solved by a single, highly specialized, ever so computationally advanced module which is restricted to the processing of one modality. Instead, robust visual processing may necessarily depend on the interaction between several streams. In this spirit, the rather limited performance of technical systems when compared to natural vision systems might be overcome by exploiting the full set of visual information available from the environment.

The rather long list of possible extension reflects the still large gap between biological and technical vision systems. On the other hand, the overall processing system can be decomposed into a number of—at least partially—known functional mechanisms such as recurrent interaction between regions, distributed representation in complementary systems, competitive-cooperative center-surround processing, long-range integration of context information, and filling-in, among others. The present work contributes to a deeper understanding of the functional significance and the computational competencies of these mechanisms.

Appendix A

# Mathematical Supplement

## A.1 Gaussian and DoG Filter Functions

### A.1.1 Function Definitions

**1D Gaussian**

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \ .$$

**Isotropic 2D Gaussian**

$$G_\sigma = g_\sigma(x)\, g_\sigma(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \ .$$

**Anisotropic offset rotated 2D Gaussian**

$$G_{\sigma_x,\sigma_y,\mu_x,\mu_y,\theta} = \frac{1}{2\pi\,\sigma_x\sigma_y} \exp\left(-\frac{(x_{\text{rot}}-\mu_x)^2}{2\sigma_x^2} - \frac{(y_{\text{rot}}-\mu_y)^2}{2\sigma_y^2}\right) \ ,$$

where

$$\begin{aligned}\begin{pmatrix} x_{\text{rot}} \\ y_{\text{rot}} \end{pmatrix} &= \text{Rot}_{-\theta} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} x\,\cos\theta + y\,\sin\theta \\ -x\,\sin\theta\ + y\,\cos\theta \end{pmatrix} \ .\end{aligned}$$

**Difference of Gaussians (DoG)**

The DoG is defined by the difference of a center Gaussian with small standard deviation $\sigma_\text{c}$ and a surround Gaussian with larger standard deviation $\sigma_\text{s}$:

$$\text{DoG}_{\sigma_\text{c},\sigma_\text{s}} = G_{\sigma_\text{c}} - G_{\sigma_\text{s}} \ .$$

### A.1.2 Gaussian Derivatives

The derivatives of the 1D Gaussian up to fourth order are given by

$$\begin{aligned} g_x(x) &= -\frac{x}{\sigma^2}\, g(x) \\ g_{xx}(x) &= \frac{1}{\sigma^2}\Big(\frac{x^2}{\sigma^2} - 1\Big)\, g(x) \\ g_{xxx}(x) &= -\frac{1}{\sigma^3}\Big(\frac{x^3}{\sigma^3} - 3\frac{x}{\sigma}\Big)\, g(x) \\ g_{xxxx}(x) &= \frac{1}{\sigma^4}\Big(\frac{x^4}{\sigma^4} - 6\frac{x^2}{\sigma^2} + 3\Big)\, g(x) \ . \end{aligned}$$

The general for can be expressed using Hermite polynomials $H_n$:

$$g_{x^n} = (-1)^n\, \frac{1}{\sigma^n}\, H_n\Big(\frac{x}{\sigma}\Big)\, g(x) \ .$$

The first eight Hermite polynomials $H_n(x), n = 0, \ldots, 7, x \in \mathbb{R}$ are defined as (cf. Lindeberg and ter Haar Romeny, 1994a):

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= x \\ H_2(x) &= x^2 - 1 \\ H_3(x) &= x^3 - 3x \\ H_4(x) &= x^4 - 6x^2 + 3 \\ H_5(x) &= x^5 - 10x^3 + 15x \\ H_6(x) &= x^6 - 15x^4 + 45x^2 - 15 \\ H_7(x) &= x^7 - 21x^5 + 105x^3 - 105x \ . \end{aligned}$$

The Hermite polynomial of higher order can be computed using the recurrence relation

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x) \ ,$$

or the closed formula (cf. Michaelis, 1997):

$$H_n = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^i}{2^i} \frac{n!}{i!(n-2i)!}\, x^{n-2i} \ .$$

The corresponding equations for the two-dimensional Gaussian can be obtained using the separability of the Gaussian

$$G(x, y) = g(x)g(y) \ ,$$

which allows to express the derivatives of the 2D Gaussian in terms of the 1D derivatives. In general, the 2D derivatives obey

$$G_{x^n y^m}(x, y) = g_{x^n}(x)\, g_{y^m}(y) \ .$$

Given the partial derivative operators up to order $n$ in $x$-, $y$- and mixed $(x, y)$- directions allows to construct the $n$-th order directional derivative in any direction $\theta$:

$$\partial_{\theta^n} G = (\cos\theta G_x + \sin\theta G_y)^n \ .$$

### A.1.3 Maximal Response of a DoG-Operator to a Step Edge

Without loss of generality, we assume a gauge coordinate system $(x, y)$ so that the step edge varies only along the $x$-axis. Since the DoG is defined by Gaussians, which are separable along the $x$- and $y$-axis, the problem reduces to the 1D case. The step edge is formalized as a Heaviside function $H$

$$H(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} .$$

The maximal response is determined as maximum of the convolution of a 1D-DoG with the Heaviside function. From elementary analysis, the necessary condition for extrema is given by the zero points of the first derivative. Since convolution with the Heaviside function is equivalent to integration (Bracewell, 1965), the zero points of the 1D-DoG are to be computed:

$$\begin{aligned} & g_{\sigma_\mathrm{c}} - g_{\sigma_\mathrm{s}} = 0 \\ \Rightarrow \quad & \frac{1}{\sqrt{2\pi}\sigma_\mathrm{c}} \exp\left(-\frac{x^2}{2\sigma_\mathrm{c}^2}\right) - \frac{1}{\sqrt{2\pi}\sigma_\mathrm{s}} \exp\left(-\frac{x^2}{2\sigma_\mathrm{s}^2}\right) = 0 \\ \Rightarrow \quad & \exp\left(-\frac{x^2}{2\sigma_\mathrm{c}^2} + \frac{x^2}{2\sigma_\mathrm{s}^2}\right) = \exp\left(\frac{x^2(\sigma_\mathrm{c}^2 - \sigma_\mathrm{s}^2)}{2\sigma_\mathrm{c}^2\sigma_\mathrm{s}^2}\right) = \frac{\sigma_\mathrm{c}}{\sigma_\mathrm{s}} \\ \Rightarrow \quad & x = \pm\sqrt{\ln\left(\frac{\sigma_\mathrm{c}}{\sigma_\mathrm{s}}\right)\frac{2\sigma_\mathrm{c}^2\sigma_\mathrm{s}^2}{\sigma_\mathrm{c}^2 - \sigma_\mathrm{s}^2}} \\ \Rightarrow \quad & x = \pm\sigma_\mathrm{c}\sigma_\mathrm{s}\sqrt{\frac{2(\ln\sigma_\mathrm{c} - \ln\sigma_\mathrm{s})}{\sigma_\mathrm{c}^2 - \sigma_\mathrm{s}^2}} . \end{aligned}$$

If the $\sigma_\mathrm{s}$ is defined as $s\sigma_c$ for some $s > 1$, the above equation writes

$$x = \pm\sigma_\mathrm{c}\sqrt{\frac{s^2 \ln s^2}{s^2 - 1}} .$$

## A.2 Discrete Approximations of the Laplacian

The Laplacian of a 2D image $I$ is defined as

$$\Delta I \equiv \nabla^2 I = I_{xx} + I_{yy} .$$

A straight forward approximation of the second order derivative combines two discrete first-order derivative filters, namely forward and backward differences:

$$D_{xx} = {}^{-}D_x * {}^{+}D_x = \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} .$$

The Laplacian filter $L$ is then given by the mask

$$L = D_{xx} + D_{yy} = \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} .$$

The corresponding approximation using crossed derivatives is given by (Lindeberg, 1990)

$${}^{\times}L = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 1 \end{bmatrix} .$$

The combination of both approximations in the way $L+2\,{}^{\times}L$ leads to the definition of the Laplacian in a 8-neighborhood:

$$ {}^{8}L = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad . $$

The combination of both approximation in the way $\frac{2}{3}L + \frac{1}{3}{}^{\times}L$ leads to the approximation of the Laplacian with the least spatial anisotropy (Dahlquist et al., 1974; Lindeberg, 1993):

$$ {}^{9}L = \frac{1}{6} \begin{bmatrix} 1 & 1 & 4 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \quad . $$

Another important approximation of the Laplacian can be formulated in terms of 2D binomial masks $B^{p,p}$ of order $p$ (cf. Jähne, 1997)

$$ L \approx B^{p,p} - 1 \ , $$

where the 2D binomial masks $B^{p,p}$ are given by convolution of 1D binomial masks $B^p$

$$ B^{p,p} = B^p * (B^p)^{\mathrm{T}} \ . $$

The smallest symmetric mask is given for $p = 2$

$$ {}^{B^{2,2}}L = B^{(2,2)} - 1 = \frac{1}{16} \left( \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad . $$

The discrete approximations of the Laplacian are isotropy only for small frequencies. The approximation using binomial mask however has significantly reduced anisotropy for smaller frequencies than the discretization based on forward and backward differences (Jähne, 1997).

## A.3 Simple Cell Model

### A.3.1 Response of a Linear Simple Cell Model

Complex cell responses are insensitive to the polarity of a dark-light transition and are modeled as the sum of simple cell responses of opposite contrast polarity. Here we show that the complex cell responses $C_\theta$ for a *linear* simple cell model are given by the absolute value of a linear filter operation applied to the input intensity $I$. The linear filtering is realized as the convolution of the input intensity $I$ with a DoG filter followed by a convolution of a difference of offset elongated Gaussians or subfield masks $G_{\theta,\mathrm{right}} - G_{\theta,\mathrm{left}}$:

$$ C_\theta = \left| I * \mathrm{DoG} * (G_{\theta,\mathrm{right}} - G_{\theta,\mathrm{left}}) \right| \ . $$

It is shown that the nonlinearities introduced by half-wave rectification at various stages finally lead to the above expression:

$$ \begin{aligned} C_\theta &= S_{\mathrm{ld},\theta} + S_{\mathrm{dl},\theta} \\ &= \left[ \widetilde{S}_{\mathrm{ld},\theta} - \widetilde{S}_{\mathrm{dl},\theta} \right]^+ + \left[ \widetilde{S}_{\mathrm{dl},\theta} - \widetilde{S}_{\mathrm{ld},\theta} \right]^+ \end{aligned} $$

Using the equality $[a]^+ + [-a]^+ = |a|$:

$$ \begin{aligned} &= \left| \widetilde{S}_{\mathrm{ld},\theta} - \widetilde{S}_{\mathrm{dl},\theta} \right| \\ &= \left| R_{\mathrm{on},\theta,\mathrm{left}} + R_{\mathrm{off},\theta,\mathrm{right}} - R_{\mathrm{off},\theta,\mathrm{left}} - R_{\mathrm{on},\theta,\mathrm{right}} \right| \end{aligned} $$

Using the linearity of convolution and the equation $X = X_{\text{on}} - X_{\text{off}}$, the subfield activities $R$ write:

$$= \left| [X * G_{\theta,\text{right}}]^+ + [-X * G_{\theta,\text{left}}]^+ - [-X * G_{\theta,\text{right}}]^+ - [X * G_{\theta,\text{left}}]^+ \right|$$

Using the equality $[a]^+ - [-a]^+ = a$:

$$\begin{aligned} &= \left| X * G_{\theta,\text{right}} - X * G_{\theta,\text{left}} \right| \\ &= \left| X * (G_{\theta,\text{right}} - G_{\theta,\text{left}}) \right| \\ &= \left| I * \text{DoG} * (G_{\theta,\text{right}} - G_{\theta,\text{left}}) \right| \ . \end{aligned}$$

### A.3.2 Third Stage of the Nonlinear Simple Cell Model

A detailed derivation for the equation of the third stage $\widetilde{S}$ of the nonlinear simple cell model is given. In particular, the derivation clarifies that a symmetric relationship between on- and off-channel is achieved by setting $\delta_S = \beta_S \gamma_S$.

For the on-channel, the equations of the first two stages of the nonlinear simple cell read

$$\begin{aligned} S_{\text{on}}^{(1)} &= \frac{R_{\text{on}}}{\alpha_S + \beta_S R_{\text{off}}} \ , \\ S_{\text{on}}^{(2)} &= \frac{R_{\text{on}}}{\gamma_S + \delta_S S_{\text{on}}^{(1)}} \ . \end{aligned}$$

Inserting the equation for $S_{\text{on}}^{(1)}$ into the equation for $S_{\text{on}}^{(2)}$ yields

$$\begin{aligned} S_{\text{on}}^{(2)} &= \frac{R_{\text{on}}}{\gamma_S + \delta_S \frac{R_{\text{on}}}{\alpha_S + \beta_S R_{\text{off}}}} \\ &= \frac{R_{\text{on}}\, c_1}{\gamma_S c_1 + \delta_S R_{\text{on}}} \ , \qquad \text{where } c_1 = \alpha_S + \beta_S R_{\text{off}} \ . \end{aligned}$$

Analogously operations for the off-channel yield

$$S_{\text{off}}^{(2)} = \frac{R_{\text{off}}\, c_2}{\gamma_S c_2 + \delta_S R_{\text{off}}} \ , \qquad \text{where } c_2 = \alpha_S + \beta_S R_{\text{on}} \ .$$

Activation of the third stage can then be written as

$$\begin{aligned} \widetilde{S} &= S_{\text{on}}^{(2)} + S_{\text{off}}^{(2)} \\ &= \frac{R_{\text{on}} c_1}{\gamma_S c_1 + \delta_S R_{\text{on}}} + \frac{R_{\text{off}} c_2}{\gamma_S c_2 + \delta_S R_{\text{off}}} \\ &= \frac{R_{\text{on}}\, c_1}{\alpha_S \gamma_S + \beta_S \gamma_S R_{\text{off}} + \delta_S R_{\text{on}}} + \frac{R_{\text{off}}\, c_2}{\alpha_S \gamma_S + \beta_S \gamma_S R_{\text{on}} + \delta_S R_{\text{off}}} \ . \end{aligned}$$

The denominators for on- and off-channel activities only differ in the scaling parameters $\beta_S \gamma_S$ and $\delta_S$ for the subfield activities $R_{\text{on}}$ and $R_{\text{off}}$. Setting $\delta_S = \beta_S \gamma_S$, i.e., assuming a symmetric relationship between on- and off-channel, yields the final concise equation for the third stage:

$$\widetilde{S} = \frac{R_{\text{on}}\, c_1 + R_{\text{off}}\, c_2}{\alpha_S \gamma_S + \beta_S \gamma_S (R_{\text{on}} + R_{\text{off}})} = \frac{\alpha_S (R_{\text{on}} + R_{\text{off}}) + 2\beta_S (R_{\text{on}} R_{\text{off}})}{\alpha_S \gamma_S + \beta_S \gamma_S (R_{\text{on}} + R_{\text{off}})} \ .$$

For $\alpha_S \ll \beta_S$, as in the case of the parameter values of Chap. 3, where $\alpha_S = 1$ and $\beta_S = 10\,000$, $\widetilde{S}$ can be approximated by

$$\widetilde{S} \approx \frac{2}{\gamma_S} \frac{R_{\text{on}} R_{\text{off}}}{R_{\text{on}} + R_{\text{off}}} \ .$$

Thus, the third stage of the simple cell model is basically a multiplication of the subfield responses normalized by the addition of the subfield responses.

## A.4 Elementary Connection Patterns Derived from Basic Symmetry Relations

In this section we show how elementary connection patterns found in grouping functions, namely parallelism and cocircularity, can be be derived a priori from symmetry constraints. The term "a priori" is used here to contrast the present derivation from psychophysical methods to determine proper grouping schemes (e.g., Field et al., 1993; Kellman and Shipley, 1991). Such psychophysically revealed constraints are useful to further confine the basic patterns to be used in concrete model implementations.

For the following explanation, consider the grouping between two edge elements $\mathbf{e_1}$ and $\mathbf{e_2}$. A fundamental assumption on the properties of the grouping function is *symmetry*, i.e., any of the two elements can serve as the reference element. First consider the edge $\mathbf{e_1}$ as the reference element. Without loss of generality, the relative position of the second edge $\mathbf{e_2}$ is given by the triplet $(r, \varphi, \theta)$. For the reverse arrangement, the relative position $(\widetilde{r}, \widetilde{\varphi}, \widetilde{\theta})$ of the edge $\mathbf{e_1}$ with respect to the reference element edge $\mathbf{e_2}$ is given by

$$\widetilde{r} = r$$
$$\widetilde{\varphi} = \pi - \theta + \varphi \tag{A.1}$$
$$\widetilde{\theta} = \begin{cases} \pi - \theta & \text{if } \theta \neq 0, \\ 0 & \text{if } \theta = 0. \end{cases} \tag{A.2}$$

See Fig. A.1 for an illustration of the geometric relationship.

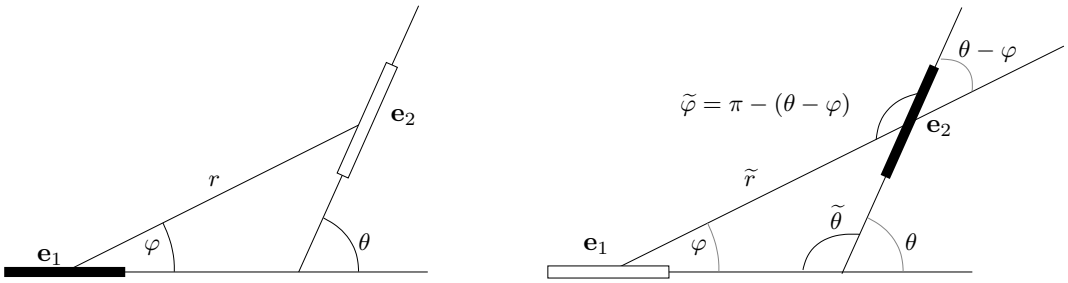


**Fig. A.1:** Geometrical relations between two edge elements $\mathbf{e}_1$ and $\mathbf{e}_2$, each of which serving as a reference element. *Left:* For edge $\mathbf{e}_1$ serving as the reference element, the relative position of edge $\mathbf{e}_1$ is given by the triplet $(r, \varphi, \theta)$ (cf. Fig. 4.14 on p. 106). *Right:* For the reverse relationship with edge $\mathbf{e}_2$ serving as the reference element, the relative position of edge $\mathbf{e}_2$ is given by the triplet $(\widetilde{r}, \widetilde{\varphi}, \widetilde{\theta}) = (r, \pi - \theta + \varphi, \pi - \theta)$. *Solid black rectangles* denote the respective reference edge, *open rectangles* denote the neighboring edge.

The two ensembles as depicted in Fig. A.1 can be aligned at their respective reference elements to define two positions with the same grouping strength. Clearly, for arbitrary values of $\varphi$ and $\theta$, these positions do not occur at symmetric locations (Fig. A.2).

In the following we examine the specific constraints which are imposed on $\varphi$ and $\theta$ if an additional assumption on a symmetric location of the two edges is made.

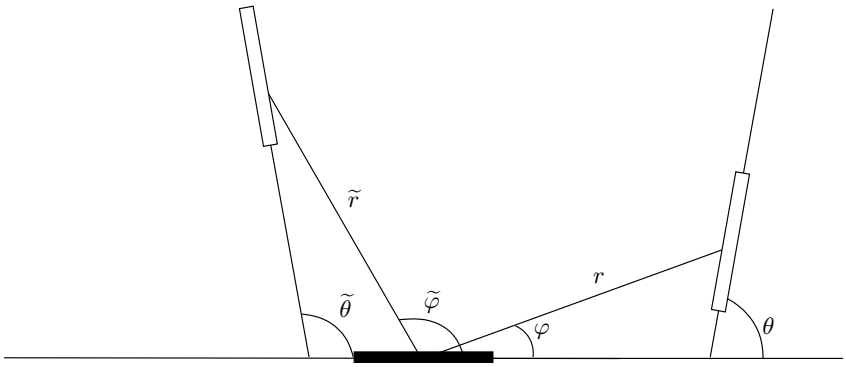


**Fig. A.2:** An ensemble of two edge elements with fixed relative positions is aligned such that each of the two edge elements serves as the reference element *(solid black rectangle)*. For arbitrary values of $\varphi$ and $\theta$, the two positions of the respective neighboring edge *(open rectangles)* do not occur at symmetric locations.

### A.4.1 A Mirror-symmetric Arrangement Implies Cocircularity

If we assume a mirror-symmetric arrangement of the positions, the relative angular position and orientation of the neighboring element suffices

$$\widetilde{\varphi} = \pi - \varphi \tag{A.3}$$

$$\widetilde{\theta} = \pi - \theta \ . \tag{A.4}$$

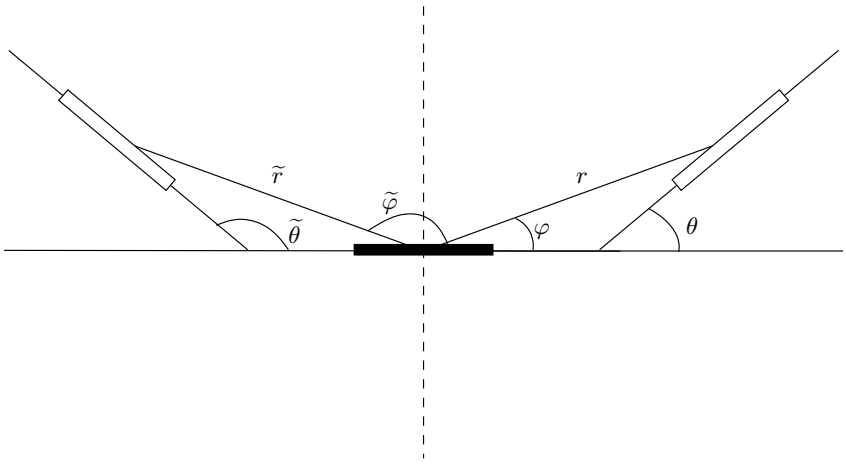


**Fig. A.3:** Alignment of an ensemble of two edge elements under the constraint of mirror-symmetric positions.

Equation A.4 is fulfilled by definition of $\theta$ (Eq. A.2). Equating Eqn. A.1 and A.3

$$\pi - \theta + \varphi = \pi - \varphi$$

and solving for $\theta$ results in the *cocircularity condition*

$$\theta = 2\varphi \ .$$

A sample plot of the pattern of cocircular edge elements on a regular grid is given in Fig. A.4.

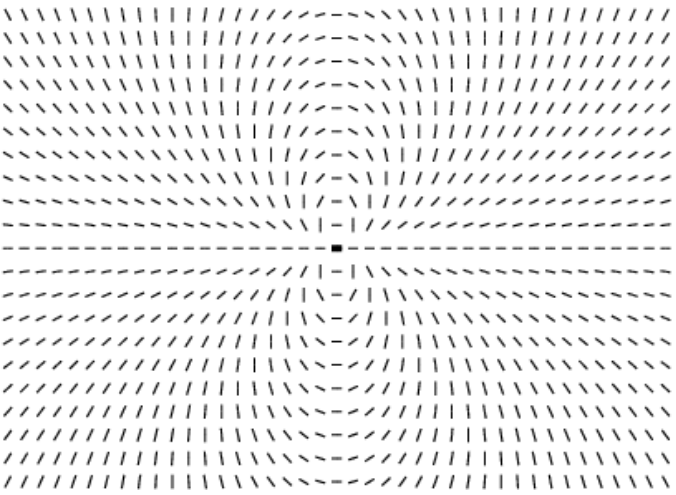

**Fig. A.4:** Cocircularity as an elementary connection pattern. The *bold central line segment* indicates the reference element.

### A.4.2 A Point-symmetric Arrangement Implies Parallelism

Similarly, if we assume a point-symmetric arrangement of the positions, the relative angular position and orientation of the neighboring element suffices

$$\widetilde{\varphi} = \pi + \varphi \tag{A.5}$$

$$\widetilde{\theta} = \pi + (\pi - \theta) = -\theta \ . \tag{A.6}$$

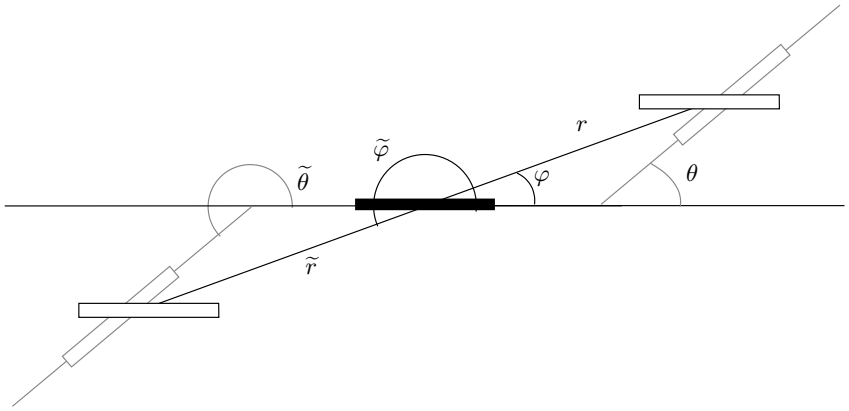


**Fig. A.5:** Alignment of an ensemble of two edge elements under the constraint of point-symmetric positions.

Equating Eqn. A.1 and A.5

$$\pi - \theta + \varphi = \pi + \varphi$$

and solving for $\theta$ results in *parallel orientation*, i.e,

$$\theta = 0 \ .$$

This condition also fulfills Eqn. A.2 and A.6.

A sample plot of the pattern of parallel edge elements on a regular grid is given in Fig. A.6.

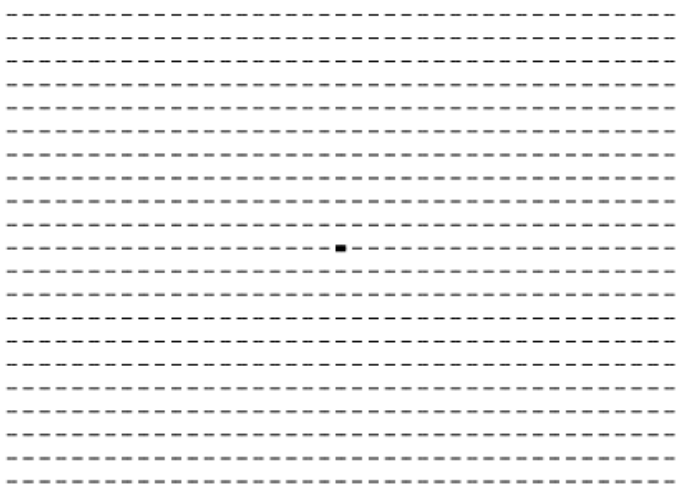

**Fig. A.6:** Parallelism as an elementary connection pattern. The *bold central line segment* indicates the reference element.

Interestingly, the same basic pattern of parallelism and cocircularity have been found in the second order statistics of contour cooccurrences in natural scenes (Geisler et al., 2001; Sigman et al., 2001).

**Appendix B**

# A Review of Diffusion Filtering for Image Processing

In Chap. 6 we have pointed out that filling-in is equivalent to a linear inhomogeneous diffusion equation with a reaction term. In this section we provide an overview over different kinds of diffusion equations that have been proposed for image processing. We suggest a taxonomy for the different types of equations and point out the relativ merits and drawbacks of the particular methods.

The section is organized as follows. We shall first derive the basic diffusion equation from the two characterizing properties of all diffusion processes, namely equilibration and conservation. We shall then introduce the terminology for different kinds of diffusion processes and suggest a tentative taxonomy. Finally, examples of different important kinds of diffusion processes are reviewed.

Further details, especially with regard to nonlinear and anisotropic diffusion can be found in an excellent review of Weickert (1997b) and a book by the same author (1998).

## B.1 The Basic Diffusion Equation

In physics, diffusion equations are widely used to describe different kinds of transport processes which obey two properties. The first property is a local one, stating that diffusion is a process that equilibrates concentration differences. The second property is a temporal one, stating that the process is conservative, i.e., mass is neither created nor destroyed. Both properties can be mathematically formalized, yielding the general diffusion equation.

The equilibration property is expressed in Fick's law: The concentration gradient $\operatorname{grad} u$ creates a flux $\mathbf{j}$ which aims to compensate for the concentration differences. The flux is opposed to the direction of the concentration gradient and proportional to it. The relation between the flux and the gradient is expressed by the diffusivity $D$, which is in general a positive definite symmetric matrix

$$\mathbf{j} = -D \operatorname{grad} u \ . \tag{B.1}$$

The flux is always away from the area of higher concentration, thus the flux is positive when moving down the gradient. The negative sign in Eq. B.1 cancels the negative gradient along the direction of positive flux.

The property of neither creating nor destroying mass is expressed by the continuity equation

$$\partial_t u = -\operatorname{div} \mathbf{j} \ , \tag{B.2}$$

where $\partial_t$ denotes partial derivation with respect to time $t$.

Inserting Fick's law (Eq. B.1) into the continuity equation (B.2) yields the diffusion equation

$$\partial_t u = \operatorname{div}(D \operatorname{grad} u) \ . \tag{B.3}$$

The diffusion equation can be written using the Nabla operator $\nabla$

$$\partial_t u = \nabla \cdot (D \nabla u) \ .$$

The basic diffusion equation can be extended by an additional reaction term which keeps the steady-state solution close to the original input $I$

$$\partial_t u = \nabla \cdot (D \nabla u) + \alpha (I - u) \ , \qquad \alpha > 0 \ .$$

Diffusion equations that incorporate a reaction term are also called biased diffusion.

In physics the transport of particles, e.g., to diffuse dopant atoms into a semiconductor, or the propagation of heat is described by diffusion equations. In this case $u$ is identified which the number of particles or temperature, respectively. To transfer diffusion processes to image processing, the concentration $u$ is identified with the gray value of an image.

## B.2 Terminology

Different diffusion types result from different properties of the diffusivity $D$, which is in general a positive definite symmetric matrix. Diffusion can be characterized with regard to three different properties: isotropy, homogeneity and linearity .

*Isotropy:* If the flux $\mathbf{j}$ is parallel to the the gradient, the diffusion is called isotropic. In this case the diffusivity $D$ is a scalar. In the general case of anisotropic diffusion, the flux is not parallel to the gradient but rotated toward the orientation of interesting features.

*Homogeneity:* If the diffusivity is constant over the whole image domain, the diffusion is homogeneous. A diffusion with a spatially varying diffusivity is called inhomogeneous. Sometimes homogeneous diffusion is called space-invariant and inhomogeneous diffusion is called space-variant.

*Linearity:* If the diffusivity $D$ is a function of the evolving image $u$, the diffusion is nonlinear. If the diffusivity is independent of $u$, the diffusion in linear. In the case of biased diffusion the diffusion is nonlinear if the reaction term depends nonlinearly on $u$.

Some of the different possible characterizations are not independent of each other. Consider the case of homogeneous diffusion. This type of diffusion is always linear and isotropic. It is *linear*, because in nonlinear diffusion, the diffusivity $D$ depends on the evolving image which is a function of the location, making $D$ inhomogeneous. It is *isotropic* by an analogous argument, since in anisotropic diffusion the flux is rotated by $D$ depending on the underlying structure of the image.

Based on this considerations a taxonomy of diffusion equations is suggested (see Fig. B.1). Such a taxonomy is useful because it allows a subsumption of different types of diffusion as proposed in the literature within a general scheme. Nevertheless, the taxonomy is rather tentative. In particular, the inclusion relationships that govern different types of diffusion, as detailed below, cannot be reflected within a hierarchical taxonomy.

## B.3 Different Types of Diffusion

In this section we shall provide an overview over different types of diffusion equations used in image processing. Starting with the simplest and most popular form of diffusion, the linear homogeneous diffusion, we shall proceed toward more complex forms of inhomogeneous diffusion types, where information about the the local structure of the quantity to be diffused is used to regulate the diffusion process. Important cases in inhomogeneous diffusion are the classes of nonlinear diffusion processes, both isotropic and anisotropic. These basic diffusion types are governed by inclusion

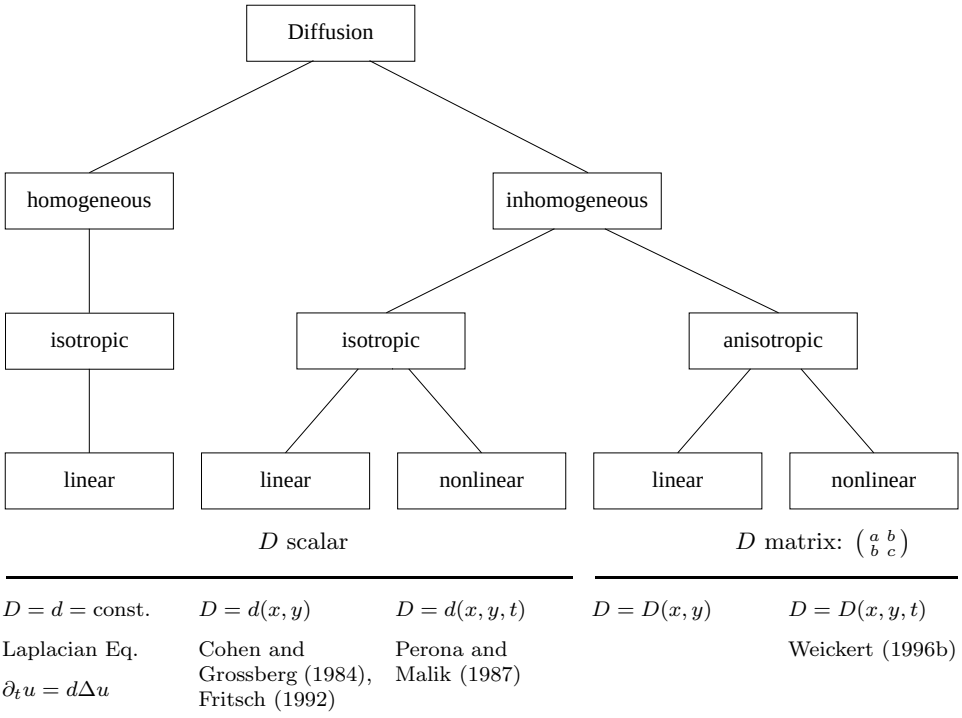


**Fig. B.1:** Diffusion taxonomy together with example references.

relationships: linear diffusion can be regarded as a special case of nonlinear isotropic diffusion, which in turn is a special case of nonlinear anisotropic diffusion (cf. Weickert, 2000).

In the following the relative merits and drawbacks for each type of diffusion are described.

### B.3.1 Linear Homogeneous Diffusion

The simplest and most widely-used class of diffusion equation is the linear homogeneous diffusion. Linear homogeneous diffusion is characterized by a scalar-valued diffusivity $d$, which is constant over the whole image domain. In this case, the general diffusion equation reduces to the Laplacian equation

$$\partial_t \, u = d \, \Delta u \equiv d \, (u_{xx} + u_{yy}) \ .$$

If the diffusion process starts with an image $I$, i.e., the initial condition is given by

$$u(x, y, t = 0) = I(x, y) \ .$$

This equation has the unique solution

$$u(x, y, t) = G_{\sqrt{2t}} \ast I, \qquad t > 0 \ ,$$

where $G_\sigma$ is the 2D-Gaussian with standard deviation $\sigma$. The diffused image at a given time $t$ can thus be computed by convolving the original image $I$ with a Gaussian whose standard deviation depends on $t$ by $\sigma = \sqrt{2t}$. In other words, linear homogeneous diffusion is equivalent to Gaussian smoothing.

**Relation to Scale-Space** Linear diffusion is intimately related to the concept of scale-space. In fact, linear diffusion is equivalent to linear scale-space theories (Weickert, 1998). A scale space is a family of gradually smoother versions of an initial image $I$. This family or set can be represented by applying a general scale space operator $\mathcal{S}$ for different scales $s > 0$ to the image:

$$\{\mathcal{S}_s I \mid s > 0\} \ .$$

The scale space operator is constructed to fulfill certain properties, which are formalized as some initial conditions or axioms.

The most general property is *linearity*, which states that applying the scale space operator $\mathcal{S}_s$ to a linear combination of two images $I_1$ and $I_2$ is equivalent to a linear combination of the operator applied to the individual images:

$$\mathcal{S}_s(aI_1 + bI_2) = a\mathcal{S}_s I_1 + b\mathcal{S}_s I_2 \ .$$

Another basic property of a scale space is that the total amount of information decreases with increasing scale, or, intuively, that local extrema cannot "popping up from nowhere" (Lindeberg, 1990). This can be formalized as the extremum or *maximum-minimum* principle, which states that local extrema are not enhanced:

$$\inf I \leq \mathcal{S}_s I \leq \sup I \ .$$

The requirement of non-enhancement of local extrema is the 2D analogon of the requirement of non-creation of local extrema in 1D (Witkin, 1983). The principle of non-creation of local extrema cannot be directly generalized to 2D, since in dimensions higher than one the number of local extrema may increase for any non-trivial kernel (Lifshitz and Pizer, 1990; Lindeberg, 1990).

Another frequently used axiom is the *semi-group property*. This property states that the subsequent application of two operators with different scale parameters $s_1, s_2$ is equivalent to applying a single operator with a scale parameter $s$ given by some function $s = f(s_1, s_2)$

$$\mathcal{S}_{s_2}(\mathcal{S}_{s_1} I) = \mathcal{S}_s I$$

Different axiomatics have been used in the literature to define a scale space (see Weickert (1998) for an overview). Interestingly, under the axiom of a *linear* scale space, all these different axiomatics leads to the same operator, namely Gaussian convolution. This confirms the evidence that Gaussian scale space is unique within a linear framework.

A detailed treatment of linear scale space can be found in the monographs of Lindeberg (1994) and Florack (1997). See also Weickert (1998) and references therein. Historically, the idea of scale space has been developed presumably independently by Iijimma (1963) and Witkin (1983), as detailed in Weickert et al. (1997a).

**Numerical Aspects** The diffusion equation can be implemented by essentially two different methods: By approximation of the diffusion equation or by approximation of Gaussian convolution.

The diffusion equation can be spatially discretized by finite difference schemes and temporally discretized by explicit schemes such as the Euler scheme.

The Gaussian convolution can be implemented by spatial convolution of the image with a discrete approximation of the Gaussian, a so-called Gaussian kernel. The Gaussian kernel samples the Gaussian function at discrete points within a limited interval. This interval is defined by some multiple of the standard deviation of the Gaussian. For an image of $N$ pixels and a kernel size of $M$ pixels, the computational complexity is $O(MN)$. By separating the Gaussian kernel into two 1D kernels, the complexity is given by $O(\sqrt{M}N)$. Since convolution in the spatial domain is equivalent to multiplication in the frequency domain, an efficient computation of the Fourier transform, the

Fast Fourier Transform (FFT) can be used. In this case, the computational complexity reduces to $O(N \log N)$. The FFT approach is faster if the kernel is large compared to the image. In practice, the efficiency further depends on the kind of algorithm that is used to implement the FFT. A detailed description of algorithms can be found, e.g., in Press et al. (1992) or Blahut (1985). Last but not least, execution time depends by the underlying hardware architecture which determines the efficiency of addition, multiplication and storage operations.

Another very efficient scheme that can be used both for the approximation of the diffusion equation or the convolution makes use of multigrid methods such as the Gaussian pyramid (Burt, 1981; Burt and Adelson, 1983). The Gaussian pyramid is a set of increasingly smoother images of decreasing size. The pyramid is constructed in a recursive process. At each step, the next level of the pyramid is generated by applying a kernel of fixed size to the previous level and storing the result at a coarser grid, i.e., in a smaller image. Since the kernel size is fixed, the computational complexity reduces to $O(N)$.

**Properties and Limitations** Linear diffusion has the desired property to suppress small structures and noisy. Since linear diffusion is equivalent to Gaussian smoothing, all properties of Gaussian smoothing hold true for linear diffusion. In particular, the transfer function of the Gaussian is also a Gaussian, causing monotonically higher dampening at higher frequencies. However, linear diffusion has some drawbacks. Since the process is by definition insensitive to any information of the image structure, not only noise is reduced, but also meaningful structures such as edges and corners are blurred and dislocated.

The remaining part of this section describes how this shortcomings, in particular the blurring and dislocation of edges, can be overcome by using inhomogeneous or nonlinear diffusion equations.

### B.3.2 Linear Inhomogeneous Diffusion

In linear inhomogeneous diffusion, the diffusivity $D$ is no longer a constant scalar, but depends on the structure of the underlying image, i.e., $D = d(x, y)$. The objective is to retain meaningful structures such as edges. A suitable computation of the diffusivity $d(x, y)$ yields low values in the vicinity of meaningful structures to suppress or eventually stop the diffusion at such locations.

Inhomogeneous diffusion for image processing has been suggested by, e.g., Fritsch (1992) or Charbonnier et al. (1994). Charbonnier et al. use the absolute value of the gradient of the image $|\nabla I|$ to reduce the diffusivity $d(x, y)$ for high values by setting

$$d(x, y) = \frac{1}{\sqrt{1 + |\nabla I(x, y)|^2 / \lambda^2}} \ , \qquad \lambda > 0 \ .$$

Although the computation of $d$ is nonlinear, the resulting diffusion equation remains linear

$$\partial_t \, u = \nabla \cdot (d(x, y) \, \nabla u) \ .$$

**Comparison to Filling-In** Filling-in as proposed by Cohen and Grossberg (1984) for the 1D case and extended to 2D by (Grossberg and Todorović, 1988) is essentially a linear inhomogeneous diffusion equation. Some differences to the abovementioned formulations as used for image processing can be observed:

(a) The diffusivity is not computed by rather straight forward computations based on the image gradient. Instead, a hierarchy of biologically motivated processing levels containing simple and complex cells is employed.

(b) The diffusion is not computed on the input image, rather than on a preprocessed contrast image resulting from shunting equations.

(c) An additional reaction term is introduced, resulting in a non-trivial steady state solution.

Also, Grossberg and colleagues deviate from the terminology introduced above and term the diffusion "nonlinear".

**Properties and Limitations** Compared to homogeneous diffusion, the blurring and dislocation of edges is successfully reduced for inhomogeneous diffusion. Some drawbacks, however, remain. For long diffusion times, artefacts reflecting the differential structure of the initial image may occur (Weickert, 1997b). This shortcoming can be overcome by computing the diffusivity not only once on the initial image, but iteratively at each time step on the evolving image. This leads to *nonlinear* inhomogeneous isotropic diffusion. A second drawback of inhomogeneous diffusion is that noise at edges is not smoothed. This is an intrinsic property of inhomogeneous isotropic diffusion and does remain in the nonlinear case. Noise at edges can be smoothed by using anisotropic diffusion.

### B.3.3 Nonlinear Isotropic Diffusion

In nonlinear inhomogeneous diffusion, the diffusivity $D$ is a space-variant function which depends on the structure of the evolving image, i.e., $D = d(x, y, t) = d(x, y, u(t))$. This is in contrast to the linear inhomogeneous diffusion, where the diffusivity is independent of the time $t$ and is computed only once based on the initial input image. Nonlinear isotropic diffusion has first been published by Perona and Malik (1987). They suggested a diffusivity

$$d(x, y, t) = \frac{1}{1 + |\nabla u(x, y, t)|^2/\lambda^2} \quad , \qquad \lambda > 0 \quad .$$

Note that Perona and Malik named the diffusion "anisotropic", though the diffusivity is still a scalar valued function.

**Properties and Limitations** Like linear inhomogeneous diffusion, nonlinear diffusion does not blur structures such as edges, which remain stable over a long time. It has been demonstrated that edge detection based on images which are preprocessed using nonlinear homogeneous diffusion outperform classical methods such as the Canny edge detector (Perona and Malik, 1990b). In contrast to the linear variant the nonlinear diffusion is more adaptive, since the diffusivity is iteratively computed on the evolving image.

However, nonlinear diffusion also has some drawbacks. Like the linear inhomogeneous diffusion, noise at edges cannot be reduced. Further, the original formulation as proposed by Perona and Malik resemble the backward diffusion type if the gradient is higher than the contrast parameter $\lambda$. In this case, edges are sharpened, but the whole equation becomes ill-posed and needs to be regularized (Nitzberg and Shiota, 1992; Weickert, 1998). In practical implementations, this ill-posedness causes the staircasing effect, where a smoothed step edges is segmented into piecewise constant regions. Several schemes have been suggested to regularize the Perona-Malik diffusion, e.g., to compute the diffusivity on a smoothed version of the evolving image (Catté et al., 1992). Nevertheless, even the regularized version cannot smooth small structures which a sufficiently large gradient such as high amplitude noise.

**Numerical Aspects** The nonlinear diffusion is computationally more expensive than the linear case, since the diffusivity has to be computed for each time step. Further, explicit schemes such as the Euler-forward forward discretization are only stable for small time steps. More precisely, the explicit scheme in matrix-vector form is given by

$$\frac{u^{k+1} - u^k}{\tau} = A(u^k)u^k \quad ,$$

where $\tau$ is the time step size, $k \in \mathbb{N}_0$ are the discrete times steps at discrete times $t_k := k\tau$, and $A(u^k) = (a_{ij}(u^k))$ is the system matrix of nearest neighbor interactions. The explicit scheme leads to the iteration scheme

$$u^{k+1} = (I + \tau A(u^k))u^k \ ,$$

where $I$ is the unit matrix. As detailed by Weickert et al. (1998), the explicit scheme is stable for time step sizes

$$\tau < \frac{1}{\max_i \sum_{j \neq i} a_{ij}(u^k)} \ .$$

For example, a unit-valued grid size $h = 1$ and a diffusivity which is bounded from above by 1 results in the stability condition $\tau < 1/(2m)$ for $m$ dimensions, i.e, $\tau < 1/4$ for 2D images.

A semi-implicit descritization of the nonlinear diffusion equation is given by

$$\frac{u^{k+1} - u^k}{\tau} = A(u^k)u^{k+1} \ ,$$

which leads to the iteration scheme

$$u^{k+1} = (I - \tau A(u^k))^{-1}u^k \ .$$

The semi-implicit scheme is unconditionally stable, i.e., the time step size $\tau$ can be chosen arbitrarily large. However, the semi-implicit scheme requires to solve a linear system of $N$ equations for an image of $N$ pixels. The system matrix $A$ is symmetric, diagonally dominant, and positive definite. Typical sizes of the image and the stencil result in an extremely sparse matrix. More precisely, the finite difference discretization on a $n \times n$ stencil results in a system matrix of at most $n^2$ nonvanishing entries per row. For an image of size $256 \times 256$ and a stencil of size $3 \times 3$, the system matrix is of size $256^2 \times 256^2$ with at most 9 nonvanishing entries per row, i.e., only a fraction of $9 \cdot 256^2/(256^2 \cdot 256^2) = 9/65536 = 1.4\,10^{-4}$ entries are non-zero. Standard direct algorithms such as Gaussian elimination are not suitable, since they destroy the zeros and cause an enormous storage and computation effort. Iterative algorithms such as Gauß-Seidel or successive overrelaxation (SOR) can be applied, but their convergence is rather slow (Weickert, 1998). Iterative methods such as preconditioned conjugate gradient algorithms are faster, but require more storage.

For an efficient computation of the semi-implicit scheme, an algorithm based on additive operator splitting (AOS) has been proposed (Weickert, 1997a; Weickert et al., 1998). The AOS scheme is a linear decomposition of the system matrix into multiple tridiagonal matrices. These matrices can be efficiently inverted using the Thomas algorithm, which is a Gaussian elimination algorithm for tridiagonal matrices. The AOS scheme results in an linear system which is not identical to the original one but has the same order of approximation both in space and time. AOS schemes have been employed by Weickert and coworkers to solve the semi-implicit discritization of the regularized version of the Perona-Malik equation as proposed by Catté et al. (1992). The AOS schemes are about 10 times faster than standard explicit schemes. The AOS scheme is also suitable for a parallel implementation which has been shown to results in a speed-up of another order of magnitude (Weickert et al., 1997b).

### B.3.4 Nonlinear Anisotropic Diffusion

In anisotropic diffusion, the diffusivity is no longer a scalar but a tensor. The tensor causes a rotation of the flux so that the flux is not always parallel and opposed to the direction of the gradient. Instead, the flux is rotated depending on the structure of the evolving image, e.g., in a direction orthogonal to the gradient. Anisotropic diffusion can be either linear or nonlinear.

Anisotropic nonlinear diffusion for noise reduction and edge enhancement, the so-called edge-enhancing anisotropic diffusion has been proposed by Weickert (1996a). The diffusion tensor $D$

has two eigenvectors, one eigenvector $\mathbf{v}_1$ being parallel to the smoothed gradient $\nabla u_\sigma$, the other eigenvector $\mathbf{v}_2$ being orthogonal to the smoothed gradient, i.e., the vector points along the edge. The corresponding eigenvalue $\lambda_1$ is given by a decreasing function of the gradient, the eigenvalue $\lambda_2$ has unit value

$$\lambda_1 = \begin{cases} 1 & |\nabla u_\sigma| = 0 \\ 1 - \exp\Big(\dfrac{-C_m}{(|\nabla u_\sigma|/\lambda)^m}\Big) & |\nabla u_\sigma| > 0 \end{cases}$$
$$\lambda_2 = 1 \ .$$

with parameters $\lambda > 0$, $m \in \mathbb{N}$ and $C_m \in \mathbb{R}^+$. The diffusion tensor $D$ can then be constructed using the spectral theorem or principal axes theorem (cf. Strang, 1993, pp. 273)

$$D = Q \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} Q^{\mathrm{T}} \quad \text{where } Q = [\mathbf{v}_1 \ \mathbf{v}_2] \ .$$

This type of edge-enhancing diffusion is an anisotropic regularization of the isotropic Perona-Malik diffusion (Weickert, 1998).

Another example of anisotropic diffusion make use of the structure tensor to compute the diffusion tensor (Weickert, 1995). This coherence-enhancing diffusion is suitable to smooth and close gaps for coherent, intrinsically one-dimensional structures.

**Properties and Limitations** Compared to isotropic diffusion processes, anisotropic diffusion has several advantages. Noise at edges can be smoothed or flowlike structures can be enhanced. However, the discretization of the anisotropic diffusion processed in not trivial. Standard approximations by central differences cannot be applied and more sophisticated schemes are needed (Weickert, 1998).

## B.4 Summary of Diffusion Equations

Diffusion equations have become of increasing interest in the computer vision community. The simplest form of linear diffusion is equivalent to the concept of Gaussian scale-space. Isotropic, inhomogeneous extensions of the basic diffusion equation incorporate information of the image structure to steer the diffusion process, allowing the preservation of edges and corners. The extention to anisotropic diffusion comprise the most general class of diffusion equations. Anisotropic diffusion is successfully employed for the enhancement of edges and coherent structures, especially in the domain of medical image processing (for references, cf. Weickert, 1998, pp. 26–27).

# Glossary

**amodal completion** Perceived completion of an occluded object behind a visible occluder. The term "amodal" refers to the absence of any sensory information such as color, brightness, texture or depth from the region behind the occluder. (after Kellman and Shipley, 1991)

**corpus callosum** The corpus callosum is a large bundle of axons connecting the two cerebral hemispheres. (after Kandel et al., 1991)

**electrical synapse** Electrical synapses couple neurons via direct flow of ionic current and allow for a virtually instantaneous transmission of signals. In contrast to chemical synapses, which are unidirectional, electrical synapses are usually bidirectional. Electrical synapses are not restricted to a pair of cells but can connect groups of neurons, allowing the flow of signals within a syncytial network of cells. (after Kandel et al., 1991)

**GABA** Gamma-aminobutyric acid is the main inhibitory transmitter in the brain. (after Kandel et al., 1991)

**gap junction** Gap junctions and their bridging channels are the ultrastructural components of an electrical synapse. The space between cells at a gap junction is only 35 nm. Gap junctions are found, e.g., in the retina between rods and cones or between horizontal cells. (after Kandel et al., 1991)

**HRP** Horseradish peroxidase is a marker enzyme that is widely used for retrograde labeling. Retrograde labeling is a technique used to label cell bodies belonging to specific terminals: A dye is injected into the cortex, taken up at cell terminals and transported back to the cell bodies (i.e., in the retrograde direction). The reaction product is usually visualized histochemically. (after Kandel et al., 1991)

**LGN** The lateral geniculate nucleus is a thalamic structure in the diencephalon which is the primary target of retinal ganglion cells.

**modal completion** Perceived completion of areas not delimited by physical differences, as occurring across the blindspot or in illusory contour formation. (after Kellman and Shipley, 1991)

**pretectum** The pretectum is a structure in the midbrain which receives input from retinal ganglion cells and controls pupillary reflexes. (after Kandel et al., 1991)

**pulvinar** The pulvinar is the most posterior and largest part of the thalamus and contains numerous subdivisions. The pulvinar makes bidirectional connections with the parietal-temporal-occipital association cortex and receives input from the superior colliculus and the primary visual cortex. These diverse projections suggest that the pulvinar integrates sensory information and is involved in visual attention. (after Kandel et al., 1991)

**RF** The receptive field of a cell is that part of the visual field which, when stimulated, yields a measurable response from the cell. (Zeki, 1993)
The term was originally introduced by Hartline (1938) as "the region of the retina which must be illuminated in order to obtain a response in any given fiber".

**shunting inhibition** In shunting inhibition, the inhibitory contribution depends on the activity of the receiving node, resulting in a nonlinear, divisive inhibition.

**somatic sensory system** The somatic sensory system is concerned with the detection of physical forces acting on the body and can be divided into four major modalities: *touch* such as light touch, vibration, or pressure, *proprioception* sensing static position and movements of limbs and body, *nociception* signaling pain and *temperature sense*. (after Kandel et al., 1991)

**superior colliculus** The superior colliculus is a structure in the midbrain which receives input from retinal ganglion cells and controls saccadic eye movement. (after Kandel et al., 1991)

**visual field** The visual field is the view seen by the two eyes without movement of the head. (Kandel et al., 1991)

# Bibliography


Abdou, I. E. (1978). Quantitative methods for edge detection. Tech. Rep. USCIPI Report 830, Image Processing Institute, Univ. of California.

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262(5142):2042–2044.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In Gazzaniga, M. S. (ed.), *The New Cognitive Neurosciences*, pp. 339–351. MIT Press, Cambridge, MA, 2 edn.

Adorján, P., Levitt, J. B., Lund, J. S., and Obermayer, K. (1999). A model for the intracortical origin of orientation tuning in macaque striate cortex. *Visual Neurosci.*, 16:303–318.

Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A. C., and Nelson, J. C. (1994). Polyneuronal innervation of spiny stellate neurons in cat visual cortex. *J. Comp. Neurol.*, 341(1):39–49.

Albus, K. and Wahle, P. (1994). The topography of tangential inhibitory connections in the postnatally developing and mature striate cortex of the cat. *Europ. J. Neurosci.*, 6(5):779–792.

Albus, K., Wahle, P., Lübke, J., and Matute, C. (1991). The contribution of GABA-ergic neurons to horizontal intrinsic connections in upper layers of the cat's striate cortex. *Exp. Brain Res.*, 85(1):235–239.

Alexander, D. C. and Buxton, B. F. (1997). Modelling of single mode distributions of color data using directional statistics. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '97)*, pp. 319–324.

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Ann. Rev. Neurosci.*, 8(2):407–430.

Aloimonos, Y., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. *Int. J. Comput. Vision*, 2(3):333–356.

Alter, T. D. and Basri, R. (1996). Extracting salient curves from images: an analysis of the saliency network. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVP 1996), San Francisco, CA*, pp. 13–20. A preliminary version has been published as Mass. Inst. Technol. Artif. Intell. Lab. Memo 1550, Cambridge, MA, 1995.

Amir, Y., Harel, M., and Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in the macaque monkey visual cortex. *J. Comp. Neurol.*, 334:19–46.

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

Arend, L. (1985). Spatial gradient illusions and inconsistent integrals. *Invest. Ophthalmol. Visual Sci.*, 26:280.

Arend, L. (1994). Surface colors, illumination, and surface geometry: intrinsic-image models of human color perception. In Gilchrist, A. L. (ed.), *Lightness, Brightness, and Transparency*, pp. 159–213. Erlbaum, Hillsdale, NJ.

Arrington, K. F. (1993). *Neural Network Models for Color and Brightness Perception and Binocular Rivalry*. Ph.D. thesis, Boston Univ., Graduate School, Dept. of Cognitive and Neural Systems.

Arrington, K. F. (1994). The temporal dynamics of brightness filling-in. *Vision Res.*, 34(24):3371–3387.

Arrington, K. F. (1996). Directional filling-in. *Neural Comput.*, 8:300–318.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.*, 61(3):183–193.

August, J., Siddiqi, K., and Zucker, S. W. (1999). Contour fragment grouping and shared, simple occluders. *Computer Vision Image Understanding*, 76(2):146–162.

Baker, S., Nayar, S., and Murase, H. (1998). Parametric feature detection. *Int. J. Comput. Vision*, 27(1):27–50.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *J. Physiol. (Lond.)*, 119:69–88.

Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press, London.

Baylor, D. A., Lamb, T. D., and Yau, K.-W. (1979). Responses of retinal rods to single photons. *J. Physiol. (Lond.)*, 288:613–634.

Beaudet, P. R. (1978). Rotationally invariant image operators. In *4th Int. Joint Conf. on Pattern Recognition*, pp. 578–583. Kyoto, Japan.

Beck, J. (1983). Textural segmentation, second-order statistics and textural elements. *Biol. Cybern.*, 48:125–130.

Beck, J., Rosenfeld, A., and Ivry, R. (1989). Line segregation. *Spat. Vis.*, 4:75–101.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA*, 92:3844–3848.

Bender, M. B. and Teuber, H. L. (1946). Phenomena of fluctuation, extinction and completion in visual perception. *Arch. Neurol. Psychiat.*, 55:627–658.

Bertero, M., Poggio, T. A., and Torre, V. (1988). Ill-posed problems in early vision. *Proc. IEEE*, 76(8):869–889.

Berzins, V. (1984). Accuracy of Laplacian edge detectors. *Computer Vision, Graphics, Image Proc.*, 27:195–210.

Biederman, I. (1985). Human image understanding: recent research and a theory. *Computer Vision, Graphics, Image Proc.*, 32(1):29–73.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.*, 94(2):115–147.

Bishop, P. O., Coombs, J. C., and Henry, G. H. (1973). Receptive fields of simple cells in the cat striate cortex. *J. Physiol. (Lond.)*, 231:31–60.

Blahut, R. (1985). *Fast Algorithms for Digital Signal Processing*. Addison-Wesley, Reading, MA.

Blake, A. (1985). Boundary conditions for lightness computation in Mondrian worlds. *Computer Vision, Graphics, Image Proc.*, 32:314–327.

Blakemore, C., Carpenter, R. H., and Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228(5266):37–39.

Blakeslee, B. and McCourt, M. E. (1997). Similar mechanisms underlie simultaneous brightness contrast and grating induction. *Vision Res.*, 37:2849–2869.

Blakeslee, B. and McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision Res.*, 38:4361–4377.

Blasdel, G. G. and Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. *J. Neurosci.*, 4(3):880–895.

Blum, H. (1973). Biological shape and visual science. *J. Theor. Biol.*, 38:205–287.

Bolz, J. and Gilbert, C. D. (1986). Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature*, 320(6060):362–365.

Bolz, J. and Gilbert, C. D. (1989). The role of horizontal connections in generating long receptive fields in the cat visual cortex. *J. Eur. Neurosci.*, 1:263–268.

Bolz, J., Gilbert, C. D., and Wiesel, T. N. (1989). Pharmacological analysis of cortical circuitry. *Trends Neurosci.*, 12(8):292–296.

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neurosci.*, 2(1):41–55.

Borg-Graham, L. J., Monier, C., and Frégnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393:369–373.

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.*, 17(6):2112–2127.

Bowyer, K. W., Kranenburg, C., and Dougherty, S. (1999). Edge detector evaluation using empirical ROC curves. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '99), Fort Collins, CL*, vol. 1, pp. 354–359.

Bracewell, R. (1965). *The Fourier Transform and Its Applications*. McGraw-Hill, New York.

Brainard, D. H. and Wandell, B. A. (1986). Analysis of the retinex theory of color vision. *J. Opt. Soc. Am. (A)*, 3(10):1651–1661.

Bressan, P., Mingolla, E., Spillmann, L., and Watanabe, T. (1997). Neon color spreading: a review. *Perception*, 26:1353–1366.

Brookes, A. and Stevens, K. A. (1991). Symbolic grouping versus simple cell models. *Biol. Cybern.*, 65:375–380.

Bullier, J., Hupé, J. M., James, A. C., and Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *J. Physiol. (Paris)*, 90(3–4):217–220.

Bullier, J., McCourt, M. E., and Henry, G. H. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Exp. Brain Res.*, 70:90–98.

Burr, D., Morrone, C., and Maffei, L. (1981). Intra-cortical inhibition prevents simple cells from responding to textured visual pattern. *Exp. Brain Res.*, 43:455–548.

Burt, P. J. (1981). Fast filter transforms for image processing. *Comput. Graph. Image Proc.*, 16:20–51.

Burt, P. J. and Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.*, 31:532–540.

Cajal, S. R. (1892). *The Structure of the Retina*. Thomas, Springfield, IL. 1972. Translated by S. A. Thorpe and M. Glickstein.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Ann. Rev. Neurosci.*, 21:47–74.

Callaway, E. M. and Wiser, A. K. (1996). Contributions of individual layer 2–5 spiny neurons to local circuits in macaque primary visual cortex. *Visual Neurosci.*, 13:907–922.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698.

Caputo, G. (1998). Texture brightness filling-in. *Vision Res.*, 38(6):841–851.

Carandini, M. (1996). *Linearity, gain control and spike encoding in the primary visual cortex*. Ph.D. thesis, New York University, New York.

Carandini, M. and Ferster, D. (2000). Orientation tuning of membrane potential and firing rate responses in cat primary visual cortex. *J. Neurosci.*, 20:470–484.

Carandini, M. and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264(5163):1333–1336.

Carandini, M., Heeger, D. J., and Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.*, 17(21):8621–8644.

Carandini, M., Heeger, D. J., and Movshon, J. A. (1999). Linearity and gain control in V1 simple cells. In Ulinski, P. S., Jones, E. G., and Peters, A. (eds.), *Models of Cortical Circuits*, vol. 13 of *Cerebral Cortex*, chap. 7, pp. 401–443. Kluwer Academic/Plenum Publishers, New York.

Carandini, M. and Ringach, D. L. (1997). Predictions of a recurrent model of orientation selectivity. *Vision Res.*, 37(21):3061–3071.

Carpenter, G. A. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21:77–88.

Catté, F., Lions, P.-L., Morel, J.-M., and Coll, T. (1992). Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal of Numerical Analysis*, 29(1):182–193.

Charbonnier, P., Blanc-Féraud, L., Aubert, G., and Barlaud, M. (1994). Two deterministic half quadratic regularization algorithms for computed imaging. In *Proc. IEEE Int. Conf. Image Proc. (ICIP-94, Austin. Nov. 13–16, 1994)*, vol. 2, pp. 168–172. IEEE Computer Society Press, Los Alamitos.

Cheng, H., Chino, Y. M., Smith III, E. L., Hamamoto, J., and Yoshida, K. (1995). Transfer characteristics of lateral geniculate nucleus x neurons in the cat: effects of spatial frequency and contrast. *J. Neurophysiol.*, pp. 2548–2557.

Coelho, C., Heller, A., Mundy, J. L., Forsyth, D., and Zisserman, A. (1991). An experimental evaluation of projective invariants. In *Proc. DARPA-ESPRIT Workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pp. 273–293.

Cohen, M. and Grossberg, S. (1984). Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. *Percept. Psychophys.*, 36:428–456.

Coren, S., Ward, L. M., and Enns, J. T. (1994). *Sensation and Perception*. Hartcourt Brace, Fort Worth.

Cornsweet, T. N. (1970). *Visual Perception*. Harcourt Brace Jovanovich, San Diego.

Cragg, B. G. (1969). The topography of the afferent projections in the circumstriate visual cortex of the monkey studied by the Nauta method. *Vision Res.*, 9:733–747.

Craik, K. J. W. (1940). *Visual Adaption*. Ph.D. thesis, Cambride Univ.

Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hyphothesis. *Proc. Natl Acad. Sci. USA*, 81:4586–4590.

Crick, F. and Koch, C. (1998). Why is the a hierarchy of visual cortical and thalamic areas: the no-strong loops hypothesis. *Nature*, 391:245–250.

Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E., and Crane, E. P. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science*, 236:579–582.

Dahlquist, G., Björk, Å., and Anderson, N. (1974). *Numerical Methods*. Prentice-Hall, Englewood Cliffs.

Dakin, S. C. (1997). The detection of structure in Glass patterns: psychophysical and computational models. *Vision Res.*, 37:2227–2246.

Dakin, S. C. and Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *J. Opt. Soc. Am.*, 15(6):1486–1499.

Das, A. (1996). Orientation in visual cortex: a simple mechanism emerges. *Neuron*, 16(3):447–480.

Das, A. and Gilbert, C. D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*, 399:655–661.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.*, 20:847–856.

Daugman, J. G. (1984). Spatial visual channels in the Fourier plane. *Vision Res.*, 24(9):891–910.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. (A)*, 2(7):1160–1169.

Daugman, J. G. (1999). Biometric decision landscapes. Tech. Rep. TR 482, The Computer Laboratory, Univ. of Cambridge.

Davey, M. P., Maddess, T., and Srinivasan, M. V. (1998). The spatiotemporal properties of the Craik-O'Brien-Cornsweet effect are consistent with 'filling-in'. *Vision Res.*, 38:2037–2046.

Davis, E. R. (1987). Design of optimal Gaussian operators in small neighbourhoods. *Image Vision Comput.*, 5(3):199–205.

De Valois, R. L. and De Valois, K. K. (1988). *Spatial Vision*. Oxford Univ. Press, New York.

De Valois, R. L., Webster, M. A., De Valois, K. K., and Lingelbach, B. (1986). Temporal properties of brightness and color induction. *Vision Res.*, 26:887–897.

De Valois, R. L., Yund, E. W., and Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.*, 22(5):531–544.

de Weerd, P., Desimone, R., and Ungerleider, L. G. (1998). Perceptual filling-in: a parametric study. *Vision Res.*, 38:2721–2734.

de Weerd, P., Gattass, R., Desimone, R., and Ungerleider, L. G. (1995). Response of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature*, 377:731–734.

DeAngelis, G. C., Freeman, R. D., and Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *J. Neurophysiol.*, 71(1):347–374.

Dennet, D. C. (1992). "Filling-in" versus finding-out: a ubiquitous confusion in cognitive science. In Pick Jr., H. L., van den Broek, P., and Knill, D. C. (eds.), *Cognition: Conceptual and Methodological Issues*. American Psychological Association, Washington, D. C.

Deriche, R. and Giraudon, G. (1990). Accurate corner detection: an analytical study. In *Proc. 3rd Int. Conf. Computer Vision*, pp. 66–70. Osaka, Japan.

DeYoe, E. A. and Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.*, 11(5):219–226.

Dobbins, A., Zucker, S. W., and Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438–441.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., and Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269(5226):981–985.

Douglas, R. J., Martin, K. A. C., and Whitteridge, D. (1988). Selective responses of visual cortical cells do not depend on shunting inhibition. *Nature*, 332(6165):642–644.

Dowling, J. E. (1987). *The Retina: An Approachable Part of the Brain*. The Belknap Press, Harvard Univ. Press, Cambridge, MA.

Dreschler, L. and Nagel, H.-H. (1982). Volumetric model and 3D-trajectory of a moving car derived from monocular TV-frame sequences of a street scene. *Computer Vision, Graphics, Image Proc.*, 20(3):199–228.

Dresp, B. (1993). Bright lines and edges facilitate the detection of small light targets. *Spat. Vis.*, 7:213–225.

Dresp, B. and Bonnet, C. (1995). Subthreshold summation with illusory contours. *Vision Res.*, 35:1071–1078.

du Buf, H. J. M. (1994). Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biol. Cybern.*, 69:449–461.

du Buf, H. J. M. and Fischer, S. (1995). Modeling brightness perception and syntactical image coding. *Opt. Engineer.*, 34(7):1900–1911.

Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Rev. Neurosci.*, 2:920–926.

Eckhorn, R. (1994). Oscillatory and non-oscillatory synchronization in the visual cortex and their possible roles in associations of visual features. *Prog. Brain Res.*, 102:405–426.

Edwards, G. J., Taylor, C. J., and Cootes, T. F. (1999). Improving identification performance by integrating evidence from sequences. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '99), Fort Collins, CL*, vol. 1, pp. 486–491.

Ehrenstein, W. H. and Ehrenstein, A. (1999). Psychophysical methods. In Windhorst, U. and Johansson, H. (eds.), *Modern Techniques in Neuroscience Research*, chap. 43. Springer, Berlin Heidelberg.

Elder, J. H. and Goldberg, R. M. (2001). Image editing in the contour domain. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(3):1–6.

Elder, J. H. and Zucker, S. W. (1994). A measure of closure. *Vision Res.*, 34(24):3361–3369.

Elder, J. H. and Zucker, S. W. (1998). Evidence for boundary-specific grouping. *Vision Res.*, 38(1):143–152.

Engel, A. K., König, P., Gray, C. M., and Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: intercolumnar interactions as determined by cross-correlation analysis. *Europ. J. Neurosci.*, 2:558–606.

Engel, A. K., Kreiter, A. K., König, P., and Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc. Natl Acad. Sci. USA*, 88:6048–6052.

Enroth-Cugell, C. and Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol. (Lond.)*, 187:517–552.

Enroth-Cugell, C. and Robson, J. G. (1984). Functional characteristics and diversity of cat retinal ganglion cells. *Invest. Ophthalmol. Visual Sci.*, 25:250–267.

Euler, L. (1744). *Methodus inveniendi lineas curvas maximi minimive proprietat gaudentes*. Lausanne.

Fahle, M. and Poggio, T. A. (1981). Visual hyperacuity: spatiotemporal interpolation in human vision. *Proc. R. Soc. London (B)*, 231:451–477.

Farné, M. (1968). Alcune osservazioni con linee virtuali e margini quasi percettivi. *Bollettino della Societaà Italiana di biologia sperimentale*, 44(19):1613–1616.

Fechner, G. T. (1889). *Elemente der Psychophysik*. Breitkopf & Härtel, Leibzip.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1:1–47.

Ferster, D. (1987). Origin of orientation-selective EPSPs in simple cells of cat visual cortex. *J. Neurosci.*, 7(6):1780–1791.

Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.*, 8(4):1172–1180.

Ferster, D. (1989). The synaptic inputs to simple cells in the cat visual cortex. In Lam, D. and Gilbert, C. D. (eds.), *Neural Mechanisms of Visual Perception*, chap. 3, pp. 63–85. Portfolio Publ. Co., The Woodlands, Texas.

Ferster, D., Chung, S., and Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380(6571):249–252.

Ferster, D. and Koch, C. (1987). Neuronal connections underlying orientation selectivity in cat visual cortex. *Trends Neurosci.*, 10(2):487–492.

Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in visual cortex. *Ann. Rev. Neurosci.*, 23:441–471.

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Res.*, 33(2):173–193.

Finkel, L. H. and Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: a computer simulation of interactive functionally segregated visual areas. *J. Neurosci.*, 9(9):3188–3208.

Finkel, L. H. and Sajda, P. (1992). Object discrimination based on depth-from-occlusion. *Neural Comput.*, 4:901–921.

Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P. H., and Thomas, J. P. (1990). The perception of brightness and darkness: Relations to neuronal receptive fields. In Spillmann, L. and Werner, J. S. (eds.), *Visual Perception: The Neurophysiological Foundations*, pp. 129–161. Academic Press, San Diego.

Fitzpatrick, D. (1996). The functional organization of local circuits in visual cortex: insights from the study of tree shrew striate cortex. *Cereb. Cortex*, 6:329–341.

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Curr. Opin. Neurobiol.*, 10(4):438–443.

Florack, L. M. J. (1997). *Image Structure*. Kluwer, Dordrecht.

Fodor, J. and Phylyshyn, Z. (1981). How direct is visual perception?: Some reflections on Gibson's "ecological approach". *Cognition*, 9:139–196.

Förstner, W. (1986). A feature based correspondence algorithm for image matching. In *Int. Arch. Photogramm. Remote Sensing*, vol. 26, pp. 176–189.

Förstner, W. and Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and centers of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland*, pp. 281–305.

Forsyth, D. and Ponce, J. (2000). Computer vision – a modern approach. Online manuscript http://www.cs.berkeley.edu./~daf/book2chaps.html. Third draft. To be published by Prentice-Hall.

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(9):891–906.

Freeman, W. T. and Roth, M. (1995). Orientation histograms for hand gesture recognition. Tech. Rep. TR-94-03, Mitsubishi Electric Research Labatories, Cambridge Research Center.

Fritsch, D. S. (1992). A medial description of greyscale image structure by gradient-limited diffusion. In Robb, R. A. (ed.), *Visualization in Biomedical Computing*, vol. 1808, pp. 105–117. SPIE.

Furman, G. (1965). Comparison of models for subtractive and shunting lateral-inhibition in receptor-neuron fields. *Kybernetik*, 2:257–274.

Gabor, D. (1946). Theory of communication. *IEEE Part III*, 93(26):429–457.

Garey, L. J. and Powell, T. P. (1971). An experimental study of the termination of the lateral geniculo-cortical pathway in the cat and monkey. *Proc. R. Soc. London (B)*, 179(54):41–63.

Gegenfurtner, K. R., Mayser, H., and Sharpe, L. T. (1999). Seeing movement in the dark. *Nature*, 398:475–476.

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.*, 41(6):711–724.

Gelfand, I. M. and Fomin, S. V. (1963). *Calculus of Variations*. Prentice-Hall, Englewood Cliffs.

Georgopoulos, A. P., Schwarz, A. B., and Kettner, R. E. (1986). Neural population coding of movement direction. *Science*, 233:1416–1419.

Gerrits, H. J. M., de Haan, B., and Vendrik, A. J. H. (1966). Experiments with retinal stabilized images. Relations between the observations and neural data. *Vision Res.*, 6:427–440.

Gerrits, H. J. M. and Timmerman, G. J. M. E. N. (1969). The filling-in process in patients with retinal scotoma. *Vision Res.*, 9:439–442.

Gerrits, H. J. M. and Vendrik, A. J. H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Exp. Brain Res.*, 11:411–430.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA.

Gigus, Z. and Malik, J. (1991). Detecting curvilinear structure in images. Tech. Rep. UCB/CSD 91/619, Univ. of California, Computer Science Division (EECS).

Gilbert, C. D. (1977). Laminar differences in receptive field properties of cells in cat primary visual cortex. *J. Physiol. (Lond.)*, 268:391–421.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9(1):1–13.

Gilbert, C. D. (1993). Circuitry, architecture, and functional dynamics of visual cortex. *Cereb. Cortex*, 3(5):373–386.

Gilbert, C. D. (1998). Adult cortical dynamics. *Physiol. Rev.*, 78(2):467–485.

Gilbert, C. D., Das, A., Ito, M., Kapadia, M. K., and Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc. Natl Acad. Sci. USA*, 93(2):615–622.

Gilbert, C. D. and Wiesel, T. N. (1979). Morphology and intracortical projections of functionally characterised neurons in the cat visual cortex. *Nature*, 280:120–125.

Gilbert, C. D. and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *J. Neurosci.*, 3:1116–1133.

Gilbert, C. D. and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.*, 9(7):2432–2442.

Gilbert, C. D. and Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res.*, 30(11):1689–1701.

Gillam, B. and Borsting, E. (1988). The role of monocular regions in stereoscopic displays. *Perception*, 17:603–608.

Gizzi, M. S., Katz, E., Schumer, R. A., and Movshon, J. A. (1990). Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *J. Neurophysiol.*, 63(6):1529–1543.

Glass, L. (1969). Moire effect from random dots. *Nature*, 223:578–580.

Glezer, V. D., Tscherbach, T. A., Bondarko, V. A., and Gauzelman, V. E. (1980). Linear and nonlinear properties of simple and complex receptive fields in area 17 of the cat visual cortex: a model of the fields. *Biol. Cybern.*, 37(4):195–208.

Gonzalez, R. C. and Woods, R. E. (1993). *Digital Image Processing*. Addison-Wesley, Reading, MA.

Gove, A., Grossberg, S., and Mingolla, E. (1995). Brightness perception, illusory contours and corticogeniculate feedback. *Visual Neurosci.*, 12:1027–1052.

Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337.

Green, D. M. and Swets, J. A. (1974). *Signal Detection Theory and Psychophysics*. Krieger, Huntington, NY.

Gregory, R. L. (1968). Visual illusions. *Sci. Am.*, 219:66–76.

Gregory, R. L. (1972). Cognitive contours. *Nature*, 238:51–52.

Grossberg, S. (1970). Neural pattern discrimination. *J. Theoret. Biol.*, 27:291–337.

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

Grossberg, S. (1983). The quantized geometry of visual space: the coherent computation of depth, form, and lightness. *Behav. Brain. Sci.*, 6:625–692.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Percept. Psychophys.*, 55(1):48–121.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychol. Rev.*, 104(3):618–658.

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spat. Vis.*, 12(2):163–185.

Grossberg, S. (2000). The complementary brain: unifying brain dynamics and modularity. *Trends Cogn. Sci.*, 4(6):233–246.

Grossberg, S. and McLoughlin, N. (1997). Cortical dynamics of three-dimensional surface perception: binocular and half-occluded scenic images. *Neural Networks*, 10:1583–1605.

Grossberg, S. and Mingolla, E. (1985a). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.*, 92:173–211.

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

Grossberg, S. and Mingolla, E. (1987). Neural dynamics of surface perception: boundary webs, illuminants, and shape-from shading. *Computer Vision, Graphics, Image Proc.*, 37(1):116–135.

Grossberg, S., Mingolla, E., and Ross, W. D. (1997). Visual brain and visual perception: how does the cortex do perceptual grouping? *Trends Neurosci.*, 20(3):106–111.

Grossberg, S., Mingolla, E., and Williamson, J. R. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks*, 8(7–8):1005–1028.

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 Res.*, 40(10):1413–1432.

Grossberg, S. and Todorović, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys.*, 43:241–277.

Grossberg, S. and Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cereb. Cortex*, 11(1):37–58.

Guy, G. and Medioni, G. (1996). Inferring global perceptual contours from local features. *Int. J. Comput. Vision*, 20(1/2):113–133.

Gyoba, H. (1997). Loss of a forest: perceptual fading and filling-in of static texture patterns. *Perception*, 26:1317–1320.

Hadamard, J. (1923). *Lectures on Cauchy Problems in Linear Partial Differential Equations*. Yale Univ. Press, New Haven.

Hammond, P. and MacKay, D. (1983). Influence of luminance gradient reversal on simple cells in feline striate cortex. *J. Physiol.*, 337:69–87.

Hanes, R. M. (1951). Suprathreshold area brightness relationships. *J. Opt. Soc. Am.*, 41:28–31.

Hansel, D. and Sompolinsky, H. (1996). Chaos and synchrony in a model of a hypercolumn in visual cortex. *J. Comput. Neurosci.*, 3:7–34.

Hansen, T. (2002). A new theory of the Poggendorff illusion based on stereoscopic vision. In Bülthoff, H. H., Gegenfurtner, K. R., Mallot, H. A., and Ulrich, R. (eds.), *Beiträge zur 5. Tübinger Wahrnehmungskonferenz (TWK 2002), Tübingen, Germany*, p. 70. Knirsch, Kirchentellinsfurt, Germany.

Hansen, T., Baratoff, G., and Neumann, H. (1999a). Dominating opponent inhibition of on and off pathways for robust contrast detection. In Förstner, W., Buhmann, J. M., Faber, A., and Faber, P. (eds.), *Mustererkennung 1999*, pp. 232–239. Springer, Berlin Heidelberg.

Hansen, T., Baratoff, G., and Neumann, H. (2000). A simple cell model with dominating opponent inhibition for robust contrast detection. *Kognitionswissenschaft*, 9(2):93–100.

Hansen, T. and Neumann, H. (1999a). Contour enhancement and corner detection by recurrent V1 long-range interactions. In *Proc. Cognitive Neuroscience*. Bremen, Germany.

Hansen, T. and Neumann, H. (1999b). Contrast processing and contour enhancement: a model of recurrent long-range interactions in V1. In *Proc. 3. Int. Conf. on Cognitive and Neural Systems (ICCNS 1999)*. Boston, MA.

Hansen, T. and Neumann, H. (1999c). A functional model of recurrent V1 long-range interactions. In Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R., and Mallot, H. A. (eds.), *Beiträge zur 2. Tübinger Wahrnehmungskonferenz (TWK 1999), Tübingen, Germany*, p. 62. Knirsch, Kirchentellinsfurt, Germany.

Hansen, T. and Neumann, H. (1999d). A model of V1 visual contrast processing utilizing long-range connections and recurrent interactions. In *Proc. 9. Int. Conf. on Artificial Neural Networks (ICANN99)*, pp. 61–66. Edinburgh, UK.

Hansen, T. and Neumann, H. (1999e). A neural model of dominating opponent inhibition in visual on and off pathways for robust contour extraction. In *Proc. 4. Fachtagung der Gesellschaft für Kognitionswissenschaft (KogWis99)*, pp. 165–170. Bielefeld, Germany.

Hansen, T. and Neumann, H. (2000). Recurrent long-range interaction for corner detection. In Bülthoff, H. H., Fahle, M., Gegenfurtner, K. R., and Mallot, H. A. (eds.), *Beiträge zur 3. Tübinger Wahrnehmungskonferenz (TWK 2000), Tübingen, Germany*, p. 71. Knirsch, Kirchentellinsfurt, Germany.

Hansen, T. and Neumann, H. (2001a). Feedforward models of contrast-invariant orientation tuning. In Bülthoff, H. H., Gegenfurtner, K. R., Mallot, H. A., and Ulrich, R. (eds.), *Beiträge zur 4. Tübinger Wahrnehmungskonferenz (TWK 2001), Tübingen, Germany*, p. 177. Knirsch, Kirchentellinsfurt, Germany.

Hansen, T. and Neumann, H. (2001b). Neural mechanisms for representing surface and contour features. In Wermter, S., Austin, J., and Willshaw, D. (eds.), *Emergent Neural Computational Architectures Based on Neuroscience*, LNCS/LNAI 2036, pp. 139–153. Springer, Berlin Heidelberg.

Hansen, T. and Neumann, H. (2002a). A computational model of recurrent, colinear long-range interaction in V1 for contour enhancement and junction detection. In *Proc. Vision ScienceS Society 2nd Annual Meeting (VSS 2002)*, p. 42. Sarasota, FL.

Hansen, T. and Neumann, H. (2002b). A neural mechanism for robust junction representation in the visual cortex. In *Proc. 6. Int. Conf. on Cognitive and Neural Systems (ICCNS 2002)*. Boston, MA.

Hansen, T., Riedel, K. O., Pessoa, L., and Neumann, H. (1999b). Regularization and 2D brightness filling-in – theoretical analysis and numerical simulations. In *Proc. 3. Int. Conf. on Cognitive and Neural Systems (ICCNS 1999)*. Boston, MA.

Hansen, T., Sepp, W., and Neumann, H. (1999c). Recurrent long-range interactions in early vision. In *Proc. EmerNet: International Workshop on Emergent Neural Computational Architectures Based on Neuroscience*. Edinburgh, UK.

Hansen, T., Sepp, W., and Neumann, H. (2001). Recurrent long-range interactions in early vision. In Wermter, S., Austin, J., and Willshaw, D. (eds.), *Emerging Neural Architectures Based on Neuroscience*, LNCS/LNAI 2036, pp. 127–138. Springer, Berlin Heidelberg.

Haralick, R. M. and Shapiro, L. G. (1992). *Computer and Robot Vision*, vol. 1. Addison-Wesley, Reading, MA.

Harris, C. J. (1987). Determination of ego-motion from matched points. In *Proc. Alvey Vision Conference*, pp. 189–192. Cambridge, UK.

Harris, C. J. and Stephens, M. (1988). A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference (AVC 88)*, pp. 147–151. Manchester, UK.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.*, 121:400–415.

Hashimoto, M. and Sklansky, J. (1987). Multiple-order derivatives for detecting local image characteristics. *Computer Vision, Graphics, Image Proc.*, 39:28–55.

Heath, M. D., Sarkar, S., Sanocki, T., and Bowyer, K. W. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(12):1338–1359.

Hecht, S., Schlaer, S., and Pirenne, M. H. (1942). Energy, quanta and vision. *J. Gen. Physiol.*, 25:819–840.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neurosci.*, 9:181–197.

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiol.*, 70(5):1885–1898.

Heeger, D. J., Simoncelli, E. P., and Movshon, J. A. (1996). Computational models of cortical visual processing. *Proc. Natl Acad. Sci. USA*, 93(2):623–627.

Heggelund, P. (1981). Receptive field organization of simple cells in cat striate cortex. *Exp. Brain Res.*, 42:89–98.

Heggelund, P. (1986a). Quantitative studies of the discharge fields of single cells in cat striate cortex. *J. Physiol.*, 373:277–292.

Heggelund, P. (1986b). Quantitative studies of enhancement and suppression zones in the receptive field of simple cells in cat striate cortex. *J. Physiol.*, 373:293–310.

Heggelund, P., Krekling, S., and Skottun, B. C. (1983). Spatial summation in the receptive field of simple cells in the cat striate cortex. *Exp. Brain Res.*, 52(1):87–98.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing and test-field luminances. *J. Exp. Psychol.*, 50:89–96.

Heinemann, E. G. (1972). Simultaneous brightness induction. In Jameson, D. and Hurvich, L. (eds.), *Handbook of Sensory Physiology*, vol. VII-4, pp. 146–169. Springer, Berlin Heidelberg.

Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., and Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res.*, 32(5):963–981.

Heitger, F. and von der Heydt, R. (1993). A computational model of neural contour processing: Figure-ground segregation and illusory contours. In *Proc. 4th Int. Conf. Computer Vision (ICCV 1993)*, pp. 32–40.

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., and Kübler, O. (1998). Simulation of neural contour mechanisms: Representing anomalous contours. *Image Vision Comput.*, 16:407–421.

Hérault, L. and Horaud, R. (1993). Figure-ground discrimination: a combinatorical optimization approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:899–914.

Hess, R. F. and Dakin, S. C. (1997). Absence of contour linking in peripheral vision. *Nature*, 390:602–604.

Hess, R. F. and Dakin, S. C. (1999). Contour integration in the peripheral field. *Vision Res.*, 39(5):947–959.

Heyden, A. and Rohr, K. (1996). Evaluation of corner extraction schemes using invariance methods. In *Proc. 13th Int. Conf. Pattern Recognition, Vienna, Austria*, vol. 1, pp. 895–899.

Hirsch, J. A., Alonso, J. M., Reid, R. C., and Martinez, L. M. (1998). Synaptic integration in striate cortical simple cells. *J. Neurosci.*, 18(22):9517–9528.

Hirsch, J. A. and Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *J. Neurosci.*, 11(6):1800–1809.

Hodgkin, A. L. (1964). *The conduction of nervous impulses*. Liverpool Univ. Press, Liverpool.

Horn, B. K. P. (1974). Determining lightness from an image. *Computer Graphics Image Proc.*, 3:277–299.

Horn, B. K. P. (1983). The curve of least energy. *ACM Trans. Math. Software*, 9:441–460.

Horton, J. C. and Sherk, H. (1984). Receptive field properties in the cat's lateral geniculate nucleus in the absence of on-center retinal input. *J. Neurosci.*, 4(2):374–380.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195:215–243.

Hubel, D. H. and Wiesel, T. N. (1974). Uniformity of monkey striate cortex: parallel relationship between field size, scatter, and magnification factor. *J. Comp. Neurol.*, 158:295–306.

Hubel, D. H. and Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. London (B)*, 198:1–59.

Huffman, D. A. (1971). Impossible objects as nonsense sentences. In Meltzer, B. and Michic, D. (eds.), *Machine Intelligence 6*, pp. 295–323. Edinburgh University Press, Edinburgh.

Hummel, R. A. and Zucker, S. W. (1983). On the foundation of relaxation labeling processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 5(3):267–287.

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

Hurlbert, A. C. (1986). Formal connections between lightness algorithms. *J. Opt. Soc. Am. (A)*, 3:1684–1693.

Iijimma, T. (1963). Theory of pattern recognition. *Electronics and Communications in Japan*, pp. 123–143.

Ito, M. and Gilbert, C. D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkey. *Neuron*, 22:593–604.

Iverson, L. A. (1994). *Toward Discrete Geometric Models for Early Vision*. Ph.D. thesis, Electrical Engineering, McGill University, Montreal.

Iverson, L. A. and Zucker, S. W. (1995). Logical/linear operators for image curves. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(10):982–996. C source code available from http://www.ai.sri.com/~leei/loglin.html.

Jähne, B. (1997). *Digital Image Processing. Concepts, Algorithms, and Scientific Applications*. Springer, Berlin Heidelberg, 4 edn. A German edition has been published under the title "Digitale Bildverarbeitung", Springer, Berlin Heidelberg, 1997.

Jähne, B., Scharr, H., and Körkel, S. (1999). Principles of filter design. In Jähne, B., Haußecker, H., and Geisler, P. (eds.), *Handbook on Computer Vision and Applications*, vol. 2: Signal Processing and Pattern Recognition, pp. 125–152. Academic Press, San Diego.

Jobson, D. J., Rahman, Z.-U., and Woodell, G. A. (1997). Properties and performance of a center/surround retinex. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(3):451–462.

Johnson, R. R. and Burkhalter, A. (1996). Microcircuitry of forward and feedback connections within rat visual cortex. *J. Comp. Neurol.*, 368:383–398.

Johnson, R. R. and Burkhalter, A. (1997). A polysynaptic feedback circuit in rat visual cortex. *J. Neurosci.*, 17(18):7129–7140.

Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58:1187–1211.

Jones, J. P., Stepnoski, A., and Palmer, L. A. (1987a). The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58:1212–1232.

Jones, J. P., Stepnoski, A., and Palmer, L. A. (1987b). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58:1233–1258.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Univ. of Chicago Press, Chicago.

Kandel, E. R., Schwarz, J. R., and Jessell, T. M. (eds.) (1991). *Principles of Neural Science*. Elsevier, New York, 3 edn. A German edition has been published under the title "Neurowissenschaften", Spektrum Akademischer Verlag, Heidelberg, 1995.

Kanizsa, G. (1976). Subjective contours. *Sci. Am.*, 234(4):48–52.

Kanizsa, G. (1979). *Organization in vision*. Praeger Publishers, 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(4):843–856.

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc. Natl Acad. Sci. USA*, 96(21):12073–12078.

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *J. Neurophysiol.*, 84(4):2048–2062.

Kaplan, E., Lee, B. B., and Shapley, R. M. (1990). New views of primate retinal function. In Osborne, N. and Chader, J. (eds.), *Progress in Retinal Research*, vol. 9, chap. 7, pp. 273–336. Pergamon Press, Oxford.

Kastner, S., Nothdurft, H. C., and Pigarev, I. N. (1997). Neural correlates of pop-out in cat striate cortex. *Vision Res.*, 37:371–376.

Kellman, P. J. and Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cogn. Psychol.*, 23(2):141–221.

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

Kestler, H. A., Simon, S., Baune, A., Hagenbuchner, M., Schwenker, F., and Palm, G. (1999). A hierarchical neural object classifier for subsymbolic-symbolic coupling. In *Proc. DAGM*, pp. 232–239. Bonn, Germany.

Kingdom, F. and Moulden, B. (1989). Border effects on brightness: a review of findings, models and issues. *Spat. Vis.*, 3(4):225–262.

Kingdom, F. and Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Res.*, 32(8):1565–1582.

Kisvarday, Z. F. and Eysel, U. T. (1992). Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17). *Neuroscience*, 46:275–286.

Kisvarday, Z. F. and Eysel, U. T. (1993). Functional and structural topography of horizontal inhibitory connections in cat visual cortex. *Europ. J. Neurosci.*, 5(12):1558–1572.

Kisvarday, Z. F., Kim, D. S., Eysel, U. T., and Bonhoeffer, T. (1994). Relationship between lateral inhibitory connections and the topography of the orientation map in cat visual cortex. *Europ. J. Neurosci.*, 6(10):1619–1632.

Kitchen, L. and Rosenfeld, A. (1982). Gray-level corner detection. *Patt. Recog. Letters*, 1(2):95–102.

Knierim, J. J. and Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.*, 67(4):961–980.

Knight, B. W. (2000). Dynamics of encoding in neuron populations: some general mathematical features. *Neural Comput.*, 12(3):473–518.

Kodak 2001 (2001). Kodak digital images offering. Online available from http://www.kodak.com/digitalImages/samples/imageIntro.shtml.

Koenderink, J. J. and van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biol. Cybern.*, 55(6):367–375.

Koenderink, J. J. and van Doorn, A. J. (1990). Receptive field families. *Biol. Cybern.*, 58:163–177.

Koenderink, J. J. and van Doorn, A. J. (1992). Generic neighborhood operators. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(6):597–605.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Hartcourt, New York.

Köhler, W. (1929). *Gestalt Psychology*. Liveright, New York.

Kolb, H., Fernandez, E., and Nelson, R. (2000). Webvision – the organization of the vertebrate retina. Online manuscript http://webvision.med.utah.edu.

Komatsu, H., Murakami, I., and Kinoshita, M. (1996). Surface representation in the visual system. *Brain. Res. Cogn. Brain. Res.*, 5(1):97–104.

Kovács, I. (1996). Gestalten of today: early processing of visual contours and surfaces. *Behav. Brain Res.*, 82(1):1–11.

Kovács, I. and Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proc. Natl Acad. Sci. USA*, 90:7495–7497.

Kovács, I. and Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370:644–646.

Krauskopf, J. (1963). Effect of retinal image stabilization on the appearance of heterochromatic targets. *J. Opt. Soc. Am.*, 53:741–744.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.*, 16:37–68.

Külpe, O. (1893). *Grundriss der Psychologie*. Wilhelm Engelmann, Leipzig.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.*, 15(2):1605–1615.

Lamme, V. A. F., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex*, 9(4):406–413.

Lamme, V. A. F. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.*, 23(11):571–579.

Lamme, V. A. F. and Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. *Front. Biosci.*, 5:d232–243.

Lamme, V. A. F., Supèr, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.*, 8(4):529–535.

Land, E. H. (1977). The retinex theory of color vision. *Sci. Am.*, 237(6):108–128.

Land, E. H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proc. Natl Acad. Sci. USA*, 80(16):5163–5169.

Land, E. H. (1986). Recent advances in retinex theory. *Vision Res.*, 26(1):7–21.

Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11.

Lee, D. (1990). Coping with discontinuities in computer vision: Their detection, classification, and measurement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(4):321–344.

Lee, S.-H. and Blake, R. (2001). Neural synergy in visual grouping: when good continuation meets common fate. *Vision Res.*, 41:2057–2064.

Lee, T. S., Mumford, D., Romero, R., and Lamme, V. A. F. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.*, 38:2429–2454.

Lehky, S. R. and Sejnowsky, T. J. (1998). Seeing white: qualia in the context of population decoding. *Neural Comput.*, 11:1261–1280.

Leite, J. A. F. and Hancock, E. R. (1997). Iterative curve organization with the EM algorithm. *Patt. Recog. Letters*, 18:143–155.

Lesher, G. W. (1995). Illusory contours: toward a neurally based perceptual theory. *Psychonomic Bulletin and Review*, 2(3):279–321.

LeVay, S. and Gilbert, C. D. (1976). Laminar patterns of geniculocortical projection in the cat. *Brain. Res.*, 113(1):1–19.

Levine, M. D. (1985). *Vision in Man and Machine*. McGraw-Hill, New York.

Li, C. Y. and Li, W. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons—classification and tuning properties. *Vision Res.*, 18:2337–2355.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Comput.*, 10(4):903–940.

Li, Z. (1999a). Visual segmentation by contextual influences via intra-cortical interactions in the primary visual cortex. *Network: Comput. Neural Syst.*, 10(2):187–212.

Li, Z. (1999b). Pre-attentive segmentation in the primary visual cortex. *Spat. Vis.*, 13:25–50.

Li, Z. (1999c). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc. Natl Acad. Sci. USA*, 96(18):10530–10535.

Li, Z. (2000). Can V1 mechanisms account for figure-ground and medial axis effects. In *Advances in Neural Information Processing Systems 12 (NIPS'99)*. MIT Press, Cambride, MA.

Li, Z. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Comput.*, 13(8):1749–1780.

Li, Z. and Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. *Network: Comput. Neural Syst.*, 10:59–77.

Lifshitz, L. M. and Pizer, S. M. (1990). A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(6):529–541.

Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12:234–254.

Lindeberg, T. (1993). Discrete derivative approximations with scale-space properties: a basis for low-level feature extraction. *J. Math. Imag. Vision*, 3(4):349–376. Also available online from http://bion.nada.kth.se/~tony/earlyvision.html.

Lindeberg, T. (1994). *Scale-space theory in computer vision*. Kluwer, Boston.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *Int. J. Comput. Vision*, 30(2):77–116.

Lindeberg, T. and ter Haar Romeny, B. M. (1994a). Linear scale-space I: basic theory. In ter Haar Romeny, B. M. (ed.), *Geometry-Driven Diffusion in Computer Vision*, chap. 1, pp. 1–38. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Lindeberg, T. and ter Haar Romeny, B. M. (1994b). Linear scale-space II: early visual operations. In ter Haar Romeny, B. M. (ed.), *Geometry-Driven Diffusion in Computer Vision*, chap. 2, pp. 39–72. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Livingstone, M. S. (1996). Oscillatory firing and interneuronal correlations in squirrel-monkey striate cortex. *J. Neurophysiol.*, 75:2467–2485.

Livingstone, M. S. and Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *J. Neurosci.*, 4(1):309–356.

Livingstone, M. S. and Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.*, 7:3416–3468.

Lourens, T. (1998). *A biologically plausible model for corner-based object recognition from color images*. Ph.D. thesis, Univ. of Groningen, The Netherlands, Shaker Publishing B.V., Maastricht, The Netherlands.

Lüdtke, N., Wilson, R. C., and Hancock, E. R. (2000). Tangent fields from population coding. In Lee, S.-W., Bülthoff, H. H., and Poggio, T. A. (eds.), *Biologically Motivated Computer Vision (BMCV 2000)*, LNCS 1811, pp. 584–593. Springer, Berlin Heidelberg.

Lund, J. S., Fitzpatrick, D., and Humphrey, A. L. (1985). The striate visual cortex of the tree shrew. In Jones, E. G. and Peters, A. (eds.), *Cerebral Cortex*, pp. 157–205. Plenum Press, New York.

MacLeod, D. I. A. (1972). Rods cancel cones in flicker. *Nature*, 235:173–174.

Maffei, L. and Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Res.*, 13(7):1255–1267.

Maffei, L. and Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Res.*, 16:1131–1139.

Mahamud, S., Thornber, K. K., and Williams, L. R. (1999). Segmentation of salient closed contours from real images. In *Proc. 7th Int. Conf. Computer Vision (ICCV '99), Corfu, Greece*, pp. 577–586. IEEE Computer Society Press, Los Alamitos, CA.

Malach, R., Amir, Y., Harel, M., and Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in the primate striate cortex. *Proc. Natl Acad. Sci. USA*, 90:10469–10473.

Mallot, H. A. (2000). *Computational Vision: Information Processing in Perception and Visual Behavior*. MIT Press, Cambridge, MA. Translated by J. S. Allen. A German edition has been published under the title "Sehen und die Verarbeitung visueller Information", Vieweg, Wiesbaden, 2001.

Marr, D. (1982). *Vision*. W. H. Freeman & Co., San Francisco.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. London (B)*, 207:187–217.

Martin, K. A. C. (1984). Neural circuits in cat striate cortex. In Jones, E. G. and Peters, A. (eds.), *Cerebral Cortex*, vol. 2, pp. 241–284. Plenum Press, New York.

Martin, K. A. C. (1988). The Wellcome Prize lecture. From single cells to simple circuits in the cerebral cortex. *Q. J. Exp. Physiol.*, 73(5):637–702.

Martin, K. A. C. and Whitteridge, D. (1984). Form, function and intracortical porjections of spiny neurons in the visual striate cortex of the cat. *J. Physiol. (Lond.)*, 353:463–504.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.*, 70(11):1297–1300.

Mason, C. and Kandel, E. R. (1991). Central visual pathways. In Kandel, E. R., Schwarz, J. R., and Jessell, T. M. (eds.), *Principles of Neural Science*, chap. 29, pp. 420–439. Elsevier, New York, 3 edn.

Matsubara, J. A. and Boyd, J. D. (1992). Presence of GABA-immunoreactive neurons within intracortical patches in area 18 of the cat. *Brain. Res.*, 583(1–2):161–170.

Maunsell, J. H. R. and Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Ann. Rev. Neurosci.*, 10:363–401.

McArthur, J. A. and Moulden, B. (1999). A two-dimensional model of brightness perception based on spatial filtering consistent with retinal processing. *Vision Res.*, 39:1199–1219.

McGuire, B. A., Gilbert, C. D., Rivlin, P. K., and Wiesel, T. N. (1991). Targets of horizontal connections in macaque primary visual cortex. *J. Comp. Neurol.*, 305(3):370–392.

McGuire, B. A., Hornung, J. P., Gilbert, C. D., and Wiesel, T. N. (1984). Patterns of synaptic input to layer 4 of cat striate cortex. *J. Neurosci.*, 4(12):3021–3033.

McLaughlin, D., Shapley, R., Shelley, M. J., and Wielaard, D. J. (2000). A neural network model of macaque primary visual cortex (V1): orientation selectivity and dynamics in the input layer 4c$\alpha$. *Proc. Natl Acad. Sci. USA*, 97(14):8087–8092.

Medioni, G., Lee, M.-S., and Tang, C.-K. (2000). *A Computational Framework for Segmentation and Grouping*. Amsterdam, The Netherlands. Elsevier.

Medioni, G. and Yasumoto, Y. (1987). Corner detection and curve representation using cubic B-splines. *Computer Vision, Graphics, Image Proc.*, 39:267–278.

Metelli, F. (1974). The perception of transparency. *Sci. Am.*, 230(4):90–98.

Meyer, G. E. and Garges, C. (1979). Subjective contours and the Poggendorf illusion. *Percept. Psychophys.*, 26:302–304.

Michaelis, M. (1997). *Low level image processing using steerable filters*. Ph.D. thesis, Christian-Albrechts-Universität Kiel. Bericht Nr. 9716.

Miikkulainen, R. and Sirosh, J. (1996). Introduction: the emerging understanding of lateral interactions in the cortex. In Sirosh, J., Miikkulainen, R., and Choe, Y. (eds.), *Lateral Interactions in the Cortex: Structure and Function*. The UTCS Neural Networks Research Group, Austin, TX. Electronic book. http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

Mingolla, E., Ross, W. D., and Grossberg, S. (1999). A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. *Neural Networks*, 12:499–511.

Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.*, 6:414–417.

Mitchison, G. and Crick, F. (1982). Long axons within the striate cortex: their distribution, orientation, and pattern of connections. *Proc. Natl Acad. Sci. USA*, 79:3661–3665.

Mokhtarian, F. and Suomela, R. (1998). Robust image corner detection through curvature scale space. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(12):1376–1381.

Morrone, M. and Burr, D. (1988). Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. London (B)*, 253:221–245.

Mumford, D. (1991). On the computational architecture of the neocortex II: the role of cortico-cortical loops. *Biol. Cybern.*, 65:241–251.

Mumford, D. (1994a). Elastica and computer vision. In Bajaj, C. L. (ed.), *Algebraic Geometry and its Applications*, chap. 31, pp. 491–506. Springer, New York Berlin Heidelberg.

Mumford, D. (1994b). Neural architectures for pattern-theoretic problems. In Koch, C. and Davis, J. L. (eds.), *Large-Scale Neuronal Theories of the Brain*. MIT Press, Cambridge, MA.

Nagel, H.-H. (1983). Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics, Image Proc.*, 21:85–117.

Nakayama, K. and Shimojo, S. (1990a). Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Res.*, 30:1811–1825.

Nakayama, K. and Shimojo, S. (1990b). Toward a neural understanding of visual surface representation. *Cold Spring Harb. Symp. Quant. Biol.*, 55:911–924.

Nakayama, K. and Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, 257(5075):1357–1363.

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(1):55–68.

Nalwa, V. S. (1993). *A Guided Tour of Computer Vision*. Addison-Wesley, Reading, MA.

Nelson, J. I. and Frost, B. J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. *Brain. Res.*, 139(2):359–365.

Nelson, J. I. and Frost, B. J. (1985). Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex. *Exp. Brain Res.*, 61(1):54–61.

Nelson, S. B., Toth, L., Sheth, B., and Sur, M. (1994). Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science*, 265(5173):774–777.

Neumann, H. (1996). Mechanisms of neural architecture for visual contrast and brightness perception. *Neural Networks*, 9(6):921–936.

Neumann, H. and Mingolla, E. (2001). Computational neural models of spatial integration and perceptual grouping. In Shipley, T. F. and Kellman, P. J. (eds.), *From Fragments to Objects: Segmentation and Grouping in Vision*, vol. 130 of *Advances in Psychology*, chap. 12, pp. 353–400. Elsevier Science, Amsterdam, The Netherlands. In preparation.

Neumann, H. and Pessoa, L. (1994). A simple cell model with multiple spatial frequency selectivity and linear/non-linear response properties. In *Proc. World Congress on Neural Networks (WCNN '93), Vol. I–IV*, pp. (IV) 290–298. San Diego, CA.

Neumann, H., Pessoa, L., and Hansen, T. (1999). Interaction of ON and OFF pathways for visual contrast measurement. *Biol. Cybern.*, 81(5–6):515–532.

Neumann, H., Pessoa, L., and Hansen, T. (2001). Visual filling-in for computing perceptual surface properties. *Biol. Cybern.*, 85(5):355–369.

Neumann, H., Pessoa, L., and Mingolla, E. (1998). A neural architecture of brightness perception: Non-linear contrast detection and geometry-driven diffusion. *Image Vision Comput.*, 16:423–446.

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

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. Lond. A*, 231:289–337.

Nieuwenhuys, R., Voogd, J., and van Huijzen, C. (1988). *The Human Central Nervous System*. Springer, Berlin Heidelberg.

Nitzberg, M., Mumford, D., and Shiota, T. (1993). *Filtering, segmentation and depth*, vol. 662 of *LNCS*. Springer, Berlin Heidelberg.

Nitzberg, M. and Shiota, T. (1992). Nonlinear image filtering with edge and corner enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14:826–833.

Noble, J. A. (1988). *Descriptions of Image Surfaces*. Ph.D. thesis, Robotics Research Group, Dept. of Engineering Science, Oxford Univ., UK.

Norman, J. (2002). Two visual systems and two theories of perception: an attempt to reconcile the constructivist and ecological approaches. *Behav. Brain. Sci.* Accepted for publication.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Res.*, 31:1957–1968.

O'Brien, V. (1958). Contour perception, illusion, and reality. *J. Opt. Soc. Am.*, 48:112–119.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain. Res.*, 34:171–175.

Omurtag, A., Kaplan, E., Knight, B. W., and Sirovich, L. (2000a). A population approach to cortical dynamics with an application to orientation tuning. *Network: Comput. Neural Syst.*, 11(4):247–260.

Omurtag, A., Knight, B. W., and Sirovich, L. (2000b). On the simulation of large populations of neurons. *J. Comput. Neurosci.*, 8(1):51–63.

Oram, M. W., Földiàk, P., Perret, D. I., and Sengpiel, F. (1998). The 'ideal homunculus': decoding neural population signals. *Trends Neurosci.*, 21(6):259–265.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthal. (Suppl.)*, 6:1–103.

O'Toole, B. and Wenderoth, P. (1977). The tilt illusion: repulsion and attraction effects in the oblique meridian. *Vision Res.*, 17:367–374.

Palmer, L. A. and Davis, T. L. (1981). Receptive field structure in cat striate cortex. *J. Neurophysiol.*, 46:260–276.

Papoulis, A. (1965). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York.

Paradiso, M. A. and Nakayama, K. (1991). Brightness perception and filling-in. *Vision Res.*, 31:1221–1236.

Parent, P. and Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(8):823–839.

Parida, L. and Geiger, D. (1998). Junctions: detection, classification, and reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(7):687–698.

Parker, A. J. and Hawken, M. J. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *J. Opt. Soc. Am. (A)*, 5:598–605.

Pastore, N. (1971). *Selective history of theories of visual perception.* Oxford University Press, New York.

Perona, P. and Malik, J. (1987). Scale space and edge detection using anisotropic diffusion. In *Proc. IEEE Computer Society Workshop on Computer Vision (Miami Beach, Nov. 30–Dec. 2, 1987)*, pp. 16–22. IEEE Computer Society Press, Washington.

Perona, P. and Malik, J. (1990a). Detecting and localizing edges composed of steps, peaks and roofs. Tech. Rep. UCB/CSD 90/590, UC Berkeley.

Perona, P. and Malik, J. (1990b). Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639.

Pessoa, L. (1996). Mach bands: how many models are possible? recent experimental findings and modeling attempts. *Vision Res.*, 36(19):3205–3227.

Pessoa, L., Mingolla, E., and Neumann, H. (1995). A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Res.*, 35(15):2201–2223.

Pessoa, L. and Neumann, H. (1998). Why does the brain fill in? *Trends Cogn. Sci.*, 2(11):422–424.

Pessoa, L., Thompson, E., and Noë, A. (1998). Finding out about filling-in: a guide to perceptual completion for visual science and the philosophy of perception. *Behav. Brain. Sci.*, 21(6):723–802.

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

Peterhans, E. and von der Heydt, R. (1990). Neurons with end-stopped receptive fields detect occlusion cues. *Soc. Neurosci. Abstr.*, 16(1):293.

Peters, A. and Payne, B. R. (1993). Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb. Cortex*, 3(1):69–78.

Petry, S., Harbeck, A., Conway, J., and Levey, J. (1983). Stimulus determinants of brightness and distinctness of subjective contours. *Percept. Psychophys.*, 34:169–174.

Pettet, M. W., McKee, S. P., and Grzywacz, N. M. (1998). Constraints on long range interactions mediating contour detection. *Vision Res.*, 38(6):865–879.

Pikaz, A. and Dinstein, I. (1994). Using simple decomposition for smoothing and feature point detection of noisy curves. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):808–813.

Poggio, T. A., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317(26):314–319.

Polat, U. (1999). Functional architecture of long-range perceptual interactions. *Spat. Vis.*, 12(2):143–162.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391:580–584.

Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res.*, 33(7):993–999.

Polat, U. and Sagi, D. (1994). The arcitecture of perceptual spatial interactions. *Vision Res.*, 34(1):73–78.

Pollen, D. A. and Ronner, S. F. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. Syst. Man Cybern.*, SMC-13(5):907–916.

Posner, M. I. and Gilbert, C. D. (1999). Attention and primary visual cortex. *Proc. Natl Acad. Sci. USA*, 96(6):2585–2587.

Pouget, A. and Zhang, K. (1997). Statistically efficient estimation using cortical lateral connections. *Advances Neural Inform. Proc.*, 9:97–103.

Prazdny, K. (1983). Illusory contours are not caused by simultaneous brightness contrast. *Percept. Psychophys.*, 34(4):403–404.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numercial Recipies in C*. Cambridge University Press, New York. Also available online from http://www.nr.com.

Prewitt, J. M. S. (1970). Object enhancement and extraction. In Rosenfeld, A. and Lipkin, B. S. (eds.), *Picture Processing and Psychophysics*, pp. 75–149. Academic Press, New York.

Pugh, M. C., Ringach, D. L., Shapley, R., and Shelley, M. J. (2000). Computational modeling of orientation tuning dynamics in monkey primary visual cortex. *J. Comput. Neurosci.*, 8(2):143–159.

Purghé, F. and Russo, A. (1996). Illusory figures produced in the absence of abrupt changes in inducer contour direction. *Perception (Suppl.)*, 25:134b. Proc. 19th Europ. Conf. Visual Perception (ECVP-96).

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., and McNamara, J. O. (eds.) (1997). *Neuroscience*. Sinauer Assoc., Inc., Sunderland, MA.

Raizada, R. D. S. and Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: a unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, 8(3):431–466.

Ramachandran, V. S. (1992). Blind spots. *Sci. Am.*, 266(5):86–91.

Ramachandran, V. S. and Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350:699–702.

Ramachandran, V. S., Gregory, R. L., and Aiken, W. (1992). Perceptual fading of visual texture border. *Vision Res.*, 33(5/6):717–721.

Rao, R. P. N. and Ballard, D. H. (1991). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neurosci.*, 2(1):79–87.

Reid, R. C. and Alonso, J. M. (1995). Specifity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378:281–284.

Reid, R. C. and Alonso, J. M. (1996). The processing and encoding of information in the visual cortex. *Curr. Opin. Neurobiol.*, 6(4):475–480.

Reynolds, R. I. (1981). Perception of an illusory contouras a function of processing time. *Perception*, 10:107–115.

Riedel, K. O., Hansen, T., and Neumann, H. (2001). Filling-in for brightness perception: theoretical investigations and computational findings. Submitted. An earlier version of the manuscript has been published as Report 1999/11, SFB 527, Univ. Ulm, Germany.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., and Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Am.*, 43:495–501.

Ringach, D. L., Hawken, M. J., and Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387:281–284.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In Tippet, J. T. et al. (eds.), *Optical and Electro-Optical Information Processing*, pp. 159–197. MIT Press, Cambridge, MA.

Rock, I. and Palmer, S. (1990). The legacy of Gestalt psychology. *Sci. Am.*, 263(6):84–90.

Rockland, K. S. and Lund, J. S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. *Brain. Res.*, 169:19–40.

Rockland, K. S. and Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. *J. Comp. Neurol.*, 216:303–318.

Rockland, K. S. and Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain. Res.*, 179:3–20.

Rockland, K. S. and Virga, A. (1989). Terminal arbors of individual "feedback" axons projecting from area V2 to V1 in the macaque monkey: a study using immunohistochemistry of anterogradely transported Phaseolus vulgaris-lencoagglunitin. *J. Comp. Neurol.*, 285:54–72.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cells response to visual stimuli. *Vision Res.*, 5:583–601.

Rogers-Ramachandran, D. C. and Ramachandran, V. S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Res.*, 38:71–77.

Rohr, K. (1994a). Localization properties of direct corner detectors. *J. Math. Imag. Vision*, 4:139–150.

Rohr, K. (1994b). Recognizing junctions by fitting parametric models. *Int. J. Comput. Vision*, 9(3):213–230.

Ross, W. D., Grossberg, S., and Mingolla, E. (2000). Visual cortical mechanisms of perceptual grouping: interacting layers, networks, columns, maps. *Neural Networks*, 13(6):571–588.

Rossi, A. F. and Paradiso, M. A. (1999). Neural correlates of perceived brightness in retina, lateral geniculate nucleus, and striate cortex. *J. Neurosci.*, 19(14):6145–6156.

Rossi, A. F., Rittenhouse, C. D., and Paradiso, M. A. (1996). The representation of brightness in primary visual cortex. *Science*, 273:1104–1107.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, 30(3):339–366.

Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: effect of signal density and spatial-frequency. *Vision Res.*, 30:1377–1388.

Sagi, D. and Kovács, I. (1993). Long range processes involved in the perception of Glass patterns. *Invest. Ophthalmol. Visual Sci. (Suppl.)*, 34(4):1084.

Salin, P.-A. and Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiol. Rev.*, 75(1):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(1):38–48.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R. (1999). Contrast's effects on spatial summation in macaque V1 neurons. *Nature Neurosci.*, 2(8):733–739.

Scharr, H., Körkel, S., and Jähne, B. (1997). Numerische Isotropieoptimierung von FIR-Filtern mittels Querglättung. In Paulus, E. and Wahl, F. M. (eds.), *Mustererkennung*, pp. 367–374. Springer, Braunschweig.

Schiller, P. H. (1982). Central connections of the retinal on and off pathways. *Nature*, 297:580–583.

Schiller, P. H., Finlay, B. L., and Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. ii. orientation specifity and ocular dominance. *J. Neurophysiol.*, 39:1321–1333.

Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *Int. J. Comput. Vision*, 37(2):151–172.

Schmidt, K. E., Goebel, R., Löwel, S., and Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *Europ. J. Neurosci.*, 9:1083–1089.

Schneeweis, D. M. and Schnapf, J. L. (1995). Photovoltage of rods and cones in the macaque retina. *Science*, 268:1053–1055.

Schwarz, H. R. (1993). *Numerische Mathematik*. Teubner, Stuttgart, Germany.

Sclar, G. and Freeman, R. D. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.*, 46(3):457–461.

Sepp, W. and Neumann, H. (1999). A multi-resolution filling-in model for brightness perception. In *Proc. ICANN*. Edinburgh, UK.

Sha'ashua, A. and Ullman, S. (1988). Structural saliency: the detection of globally salient structures using a locally connected network. In *Proc. 2nd Int. Conf. Computer Vision (ICCV 1988)*, pp. 321–327. IEEE Computer Society Press, Los Alamitos, CA.

Shao, Z. and Burkhalter, A. (1996). Different balance of excitation and inhibition in forward and feedback circuits in rat visual cortex. *J. Neurosci.*, 15:7353–7365.

Shao, Z., Harding, G. W., and Burkhalter, A. (1996). $GABA_B$ mediated inhibition in feedback circuits is weaker than in feedforward circuits in rat visual cortex. *Soc. Neurosci. Abstr.*, 22:490.

Shapley, R. and Enroth-Cugell, C. (1984). Visual adaption and retinal gain controls. *Progress in Retinal Research*, 3:263–346.

Shapley, R. and Gordon, J. (1983). A nonlinear mechanism for the perception of form. In *Investigative Ophthalmology and Visual Science 24 (Supplement)*, p. 238.

Shapley, R. and Gordon, J. (1987). The existence of interpolated illusory contours depends on contrast and spatial separation. In Petry, S. and Meyer, G. E. (eds.), *The Perception of Illusory Contours*. Springer, New York.

Shipley, T. F. and Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Percept. Psychophys.*, 48(3):259–270.

Sigman, M., Cecchi, G. A., Gilbert, C. D., and Magnasco, M. O. (2001). On a common circle: natural scenes and Gestalt rules. *Proc. Natl Acad. Sci. USA*, 98(4):1935–1940.

Sillito, A. M. (1975). The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *J. Physiol.*, 250(2):305–329.

Sillito, A. M., Jones, H. E., Gerstein, G. L., and West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369:479–482.

Sillito, A. M., Kemp, J. A., Milson, J. A., and Berardi, N. (1980). A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain. Res.*, 194(2):517–520.

Sincich, L. and Blasdel, G. G. (1995). Lateral connections and orientation preference in layers II/III of squirrel monkey striate cortex. *Soc. Neurosci. Abstr.*, 21:393.

Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, 18:555–586.

Sinha, P. and Adelson, E. H. (1993). Recovering reflectance in a world of painted polyhedra. In *Proc. Fourth International Conference on Computer Vision (ICCV' 93), Berlin, Germany*, pp. 156–163. IEEE.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., and Freeman, R. D. (1987). The effect of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J. Neurophysiol.*, 57(3):773–786.

Smith, A. and Over, R. (1979). Motor aftereffect with subjective contours. *Percept. Psychophys.*, 25:95–98.

Smith, S. and Brady, J. (1997). SUSAN – a new approach to low level image processing. *Int. J. Comput. Vision*, 23(1):45–78.

Sobel, I. E. (1970). *Camera Models and Machine Perception*. Ph.D. thesis, Stanford Univ.

Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, 15(8):5448–5465.

Sompolinsky, H. and Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Curr. Opin. Neurobiol.*, 7(4):514–522.

Southall, B., Buxton, B. F., Marchant, J. A., and Hague, T. (2000). On the performance characterisation of image segmentation algorithms: a case study. In Vernon, D. (ed.), *Computer Vision – ECCV 2000, 6th European Conference on Computer Vision, Dublin, Ireland, Vol. II*, vol. 1843 of *LNCS*, pp. 351–365. Springer, Berlin Heidelberg New York.

Sperling, P. (1990). Retina. In Shepherd, G. M. (ed.), *The Synaptic Organization of the Brain*, chap. 6, pp. 170–213. Oxford University Press, New York, 3 edn.

Stork, D. G. and Wilson, H. R. (1990). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *J. Opt. Soc. Am. (A)*, 7(8):1362–1373.

Strang, G. (1993). *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, MA.

Stryker, M. P., Chapman, B., Miller, K. D., and Zahs, K. R. (1990). Experimental and theoretical studies of the organization of afferents to single-orientation columns in visual cortex. *Cold Spring Harb. Symp. Quant. Biol.*, 55:515–527.

Swets, J. A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Invest. Radiol.*, 14(2):109–121.

Szeliski, R. (1990). Bayesian modeling of uncertainty in low-level-vision. *Int. J. Comput. Vision*, 5(3):271–301.

Tanaka, K. (1983). Cross-correlation analysis of geniculostriate neuronal relationships in cats. *J. Neurophysiol.*, 49(6):1303–1318.

Terzopoulos, D. (1983). Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics, Image Proc.*, 24(1):52–96.

Terzopoulos, D. (1986). Regularization of inverse visual problems involving discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(4):413–424.

Thornber, K. K. and Williams, L. R. (1996). Analytic solutions of stochastic completion fields. *Biol. Cybern.*, 75:141–151.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-Posed Problems*. V. H. Winston & Sons, Washington D. C.

Todorović, D. (1987). The Craik-O'Brien-Cornsweet effect: new varieties and their theoretical implications. *Percept. Psychophys.*, 42(6):545–560.

Todorović, D. (1997). Lightness and junctions. *Perception*, 26(4):379–394.

Tolhurst, D. (1972). On the possible existence of edge detector neurons in the human visual system. *Vision Res.*, 12:797–804.

Torre, V. and Poggio, T. A. (1980). On edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(2):147–163.

Toyama, K., Kimura, M., and Tanaka, K. (1981). Cross-correlation analysis of interneuronal connectivity in cat visual cortex. *J. Neurophysiol.*, 46(2):191–201.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends Neurosci.*, 24(5):295–300.

Troy, J. B. and Enroth-Cugell, C. (1993). X and Y ganglion cells infor the cat's brain about contrast in the retinal image. *Exp. Brain Res.*, 93:383–390.

Troyer, T. W., Krukowski, A. E., Priebe, N. J., and Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: thalomocortical input tuning and correlation-based intracortical connectivity. *J. Neurosci.*, 18(15):5908–5927.

Ts'o, D. and Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.*, 8:1712–1727.

Ts'o, D., Gilbert, C. D., and Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *J. Neurosci.*, 6:1160–1170.

Tsumoto, T., Eckart, W., and Creutzfeldt, O. D. (1979). Modification of orientation sensitivity of cat visual cortex neurons by removal of GABA-mediated inhibition. *Exp. Brain Res.*, 34(2):351–363.

Ullman, S. (1976). Filling-in the gaps: the shape of subjective contours and a model for their generation. *Biol. Cybern.*, 25:1–6.

Ullman, S. (1980). Against direct perception. *Behav. Brain. Sci.*, 3:373–415.

Ullman, S. (1995). Sequence of seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex*, 2:310–335.

van Tres, H. L. (1968). *Detection, Estimation and Modulation Theory*, vol. 1. John Wiley and Sons, New York.

Victor, J. D. (1999). Temporal aspects of neural coding in the retina and the lateral geniculate. *Network: Comput. Neural Syst.*, 10:R1–R66.

Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276.

Vogels, R. (1990). Population coding of stimulus orientation by striate cortical cells. *Biol. Cybern.*, 64:25–31.

von Campenhausen, C. (1993). *Die Sinne des Menschen*. Georg Thieme Verlag, Stuttgart.

von der Heydt, R. (1987). Approaches to visual cortical functions. *Rev. Physiol. Biochem. Pharmacol.*, 108:69–150.

von der Heydt, R. (1995). Form analysis in visual cortex. In Gazzaniga, M. S. (ed.), *The Cognitive Neurosciences*, pp. 365–382. MIT Press, Cambridge, MA.

von der Heydt, R. and Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J. Neurosci.*, 9(5):1731–1748.

von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224:1260—1262.

von Helmholtz, H. E. F. (1911). *Handbuch der physiologischen Optik*. Voss, Hamburg. An English translations has been published under the title "Treatise on physiological optics", J. P. C. Southall (Ed. and Trans.), Dover, New York 1962.

Wald, G. (1968). The molecular basis of visual excitation. *Nature*, 219:800–807.

Walls, G. (1954). The filling-in process. *Am. J. Optomet.*, 31:329–340.

Wässle, H. (1996). Das Auge als Fenster zum Gehirn. In *Fachtagung Informationstechnik und Biotechnologie*, pp. 67–79. Projektträger des BMBF für Informationstechnik bei der DLR, Wissenschaftszentrum, Bonn.

Wässle, H. and Boycott, B. B. (1991). Functional architecture of mammalian retina. *Physiol. Rev.*, 71(2):447–480.

Wässle, H., Grünert, U., Röhrenbeck, J., and Boycott, B. B. (1989). Cortical magnification factor and the ganglion cell density in the primate retina. *Nature*, 341:643–646.

Wässle, H., Grünert, U., Röhrenbeck, J., and Boycott, B. B. (1990). Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Res.*, 30:1897–1911.

Wässle, H., Peichl, L., and Boycott, B. B. (1981a). Morphology and topography of on- and off-alpha cells in the cat retina. *Proc. R. Soc. London (B)*, 212:157–175.

Wässle, H., Peichl, L., and Boycott, B. B. (1981b). Morphology and mosaic of on- and off-beta cells in the cat retina and some functional considerations. *Proc. R. Soc. London (B)*, 212:177–195.

Watanabe, T. and Cavanagh, P. (1991). Texture and motion spreading, the aperture problem, and transparency. *Percept. Psychophys.*, 50(5):459–464.

Watanabe, T. and Cavanagh, P. (1993). Transparent surfaces defined by implicit X junctions. *Vision Res.*, 33(16):2339–2346.

Watt, R. J. and Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Res.*, 25(11):1661–1674.

Weber, E. H. (1846). Tastsinn und Gemeingefühl. In Wagner, R. (ed.), *Handwörterbuch der Physiologie*. Reprinted in W. Ostwald (1905), *Klassiker der exakten Wissenschaften*, vol. 149. Engelmann, Leibzig.

Webster, M. A. and De Valois, R. L. (1985). Relationship between spatial frequency and orientation tuning of striate cortex cells. *J. Opt. Soc. Am. (A)*, 2:1124–1132.

Wehmeier, U., Dong, D., Koch, C., and Van Essen, D. C. (1989). Modeling the visual system. In Koch, C. and Segev, I. . (eds.), *Methods in Neural Modeling*, pp. 335–359. MIT Press, Cambridge, MA.

Weickert, J. (1995). Multiscale texture enhancement. In Hlaváč, V. and Šára, R. (eds.), *Computer Analysis of Images and Patterns*, vol. 970 of *LNCS*, pp. 230–237. Springer, Berlin Heidelberg.

Weickert, J. (1996a). Theoretical foundations of anisotropic diffusion in image processing. *Comput. Suppl.*, 11:221–236.

Weickert, J. (1996b). Theoretical fundations of anisotropic diffusion in image processing. In Kropatsch, W., Klette, R., and Solina, F. (eds.), *Theoretical Fundations of Computer Vision*, pp. 221–236. Springer, Wien.

Weickert, J. (1997a). Recursive separable schemes for nonlinear diffusion filters. In ter Haar Romeny, B. M., Florack, L., Koenderink, J., and Viergever, M. (eds.), *Scale-Space Theory in Computer Vision*, vol. 1252 of *LNCS*, pp. 260–271. Springer, Berlin Heidelberg.

Weickert, J. (1997b). A review of nonlinear diffusion filtering. In ter Haar Romeny, B. M., Florack, L., Koenderink, J., and Viergever, M. (eds.), *Scale-Space Theory in Computer Vision*, vol. 1252 of *LNCS*, pp. 3–28. Springer, Berlin Heidelberg.

Weickert, J. (1998). *Anisotropic Diffusion in Image Processing.* ECMI Series. Teubner-Verlag, Stuttgart, Germany.

Weickert, J. (2000). Anwendungen partieller Differentialgleichungen in der Bildverarbeitung. Online manuscript http://www.cvgpr.uni-mannheim.de/weickert.

Weickert, J., Ishikawa, S., and Imiya, A. (1997a). On the history of Gaussian scale-space axiomatics. In Sporring, J., Nielsen, M., Florack, L., and Johansen, P. (eds.), *Gaussian Scale-Space Theory*, pp. 45–59. Kluwer, Dordrecht.

Weickert, J., ter Haar Romeny, B. M., and Viergever, M. A. (1998). Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Trans. Image Proc.*, 7(3):398–410.

Weickert, J., Zuiderveld, K. J., ter Haar Romeny, B. M., and Niessen, W. J. (1997b). Parallel implementation of AOS schemes: a fast way of nonlinear diffusion filtering. In *Proc. IEEE Int. Conf. Image Proc. (ICIP-97, Santa Barbara. Oct. 26–29, 1997)*, vol. 3, pp. 396–399.

Wennekers, T. (2001). Orientation tuning properties of simple cells in area V1 derived from an approximate analysis of nonlinear neural field models. *Neural Comput.*, 13(8):1721–1747.

Wersing, H. (2000). *Spatial feature binding and learning in competitive neural layer architectures.* Ph.D. thesis, Technische Fakultät, Univ. Bielefeld, Germany.

Wersing, H. and Ritter, H. (1999). Feature binding and relaxation labeling with the competitive layer model. In *Proc. Europ. Symp. Artific. Neural Networks (ESANN '99), Bruges, Belgium*, pp. 295–300.

Wersing, H., Steil, J. J., and Ritter, H. (2001). A competitive-layer model for feature binding and sensory segmentation. *Neural Comput.*, 13(2):357–387.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychol. Forsch.*, 4:301–350. A translation has been published as "Laws of oganization in perceptual forms" in Ellis, W., *A Source Book of Gestalt Psychology*, pp. 71–88, Routledge & Keagan Paul, London, 1938. Also available online from http://psychclassics.yorku.ca.

Wertheimer, M. (1938). *Laws of Organization in Perceptual Forms.* Hartcourt, Brace, Jovanovich, London.

Williams, L. R. (2001). Fruit and texture images. http://www.cs.unm.edu/~williams/saliency.html.

Williams, L. R. and Jacobs, D. W. (1997a). Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Comput.*, 9(4):837–858.

Williams, L. R. and Jacobs, D. W. (1997b). Local parallel computations of stochastic completion fields. *Neural Comput.*, 9(4):859–881.

Williams, L. R. and Thornber, K. K. (1998). A comparison of measures for detecting natural shapes in cluttered backgrounds. In Burkhardt, H. and Neumann, B. (eds.), *Computer Vision – ECCV '98, 5th European Conference on Computer Vision, Freiburg, Germany*, vol. 1407 of *LNCS*, pp. 432–448. Springer, Berlin Heidelberg New York.

Williams, L. R. and Thornber, K. K. (1999). A comparison of measures for detecting natural shapes in cluttered backgrounds. *Int. J. Comput. Vision*, 34(2/3):81–96.

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J., and De Valois, R. L. (1990). The perception of form: retina to striate cortex. In Spillmann, L. and Werner, J. S. (eds.), *Visual Perception: The Neurophysiological Foundations*, pp. 231–272. Academic Press, New York.

Wilson, R. and Bhalerao, A. H. (1992). Kernel design for efficient multiresolution edge detection and orientation estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):384–390.

Witkin, A. P. (1983). Scale-space filtering. In *Proc. 8. Int. Joint Conf. Art. Intell. (IJCAI)*, vol. 2, pp. 1019–1022.

Wörgötter, F. and Koch, C. (1991). A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *J. Neurosci.*, 11(7):1959–1979.

Würtz, R. P. and Lourens, T. (2000). Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image Vision Comput.*, 18(6–7):531–541.

Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press, New York.

Yen, S.-C. and Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Res.*, 38(5):719–741.

Young, R. A. (1985). The Gaussian derivative theory of spatial vision: Analysis of cortical receptive field line-weighting profiles. Publication GMR-4920. General Motors Research Labs, Computer Science Dept., 30500 Mound Road, Box 9055, Warren, Michigan 48090-9055.

Zaidi, Q., Spehar, B., and Shy, M. (1997). Induced effects of backgrounds and foregrounds in three-dimensional configurations: the role of T-junctions. *Perception*, 26(4):395–408.

Zeki, S. M. (1969). Representation of central visual fields in prestriate cortex of monkey. *Brain. Res.*, 14:271–291.

Zeki, S. M. (1976). The functional organization of projections from striate to prestriate visual cortex in the rhesus monkey. *Cold Spring Harb. Symp. Quant. Biol.*, 40:591–600.

Zeki, S. M. (1993). *A Vision of the Brain*. Blackwell, Oxford.

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Comput.*, 10(2):403–430.

Zemel, R. S. and Hinton, G. E. (1994). Developing population codes by minimizing description length. *Neural Comput.*, 7(3):549–564.

Zetzsche, C. and Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Res.*, 30(7):1111–1117.

Zipser, K., Lamme, V. A. F., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.*, 16(22):7376–7389.

Zucker, S. W. (1985). Early orientation selection: tangent fields and the dimensionality of their support. *Computer Vision, Graphics, Image Proc.*, 32:74–103.

Zucker, S. W., David, C., Dobbins, A., and Iverson, L. A. (1988). The organization of curve detection: coarse tangent fields and fine spline covering. In *Proc. 2nd Int. Conf. Computer Vision (ICCV 1988)*, pp. 577–586. IEEE Computer Society Press, Los Alamitos, CA.

Zucker, S. W., Dobbins, A., and Iverson, L. A. (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Comput.*, 1:68–81.

Zuniga, O. A. and Haralick, R. M. (1983). Corner detection using the facet model. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Washington, D. C.*, pp. 30–37.

Zweck, J. W. and Williams, L. R. (2000). Euclidean group invariant computations of stochastic completion fields using shiftable-twistable functions. In Vernon, D. (ed.), *Computer Vision – ECCV 2000, 6th European Conference on Computer Vision, Dublin, Ireland, Vol. II*, vol. 1843 of *LNCS*, pp. 100–116. Springer, Berlin Heidelberg New York.

# Author Index

# Subject Index