CONTRIBUTED ARTICLE

Mechanisms of Neural Architecture for Visual Contrast and Brightness Perception

HEIKO NEUMANN

Universität Ulm

(Received 21 January 1994; revised and accepted 2 February 1996)

Abstract—A neural architecture is proposed that serves as a framework for further empirical as well as theoretical investigations for a unified theory for contrast and brightness perception. The work further extends the brightness perception model developed by Grossberg and Todorovic. The proposed new computational architecture utilizes a (retinal) preprocessing stage with center-surround antagonisms of both polarities. The preprocessed data are shown to multiplex contrast as well as luminance information that can be de-multiplexed subsequently using a scheme of cross-channel interaction. Based on a hypothesized luminance-related channel, a three-stage process is suggested for brightness reconstruction. The separate channel for the representation of luminance-related information provides a key mechanism to assign the reconstructed brightness to an absolute reference level. The architecture provides a framework for the analysis of processes in brightness perception.


1. INTRODUCTION

Physiological and anatomical findings show that visual processing starts with antagonistic center-surround processing based on receptive fields (RF) at various layers of the retina. For further processing,

the axons of retinal ganglion cells pass the lateral geniculate nuclei (LGN). The LGN cells in turn project to the primary visual cortical area (V1, area 17). Whereas LGN RF profiles show similar antagonistic response profiles to those recorded in the retina, cortical RF properties include selectivity for orientation and polarity of contrast as well as odd- and even-symmetric varieties. These RF types are assumed to contribute to the functional basis of form perception [e.g., Hubel & Wiesel (1977); Pollen & Ronner (1983)]. Various (linear) feature detector models have been proposed for detection of intrinsically one-dimensional (1-D) signal variations. Starting from so-called simple and complex cells, processing units of successively higher order and complexity have been identified and given such names as end-stopped, hypercomplex, and higher-order hypercomplex cells [e.g., Barlow (1983)]. The neural, anatomical and computational basis for brightness perception has not been investigated in similar detail. Only a few neural computational architectures have been proposed. Simply from introspection, it is obvious that the brain machinery at some stage must contain a set of mechanisms that enable a living organism to have brightness available in real time for the control of behavior. By far the most fully elaborated model that incorporates primary stimulus features in a coherent way is the BCS/FCS model as part of the FACADE theory developed by Grossberg

Acknowledgements: The work has been supported in part by the German Ministry of Research and Technology (BMFT), grant 413-5839-01 IN 101 C/1.

The work described has been performed while the author was a visiting scholar at Boston University, Center for Adaptive Systems and Department of Cognitive & Neural Systems, Boston, MA. During this time the author was on leave from Universität Hamburg, Fachbereich Informatik, Hamburg (Germany).

The author wishes to express his gratitude to Professors Ennio Mingolla (Boston University) and H. Siegfried Stiehl (Universität Hamburg) for stimulating discussions during different periods of this research. He is also indebted to Luiz Pessoa (Boston University) for his generous help and for spending hours in running simulations of the full single scale architecture. Michael Nagel (Universität Hamburg) helped with the software development for a first simulation testbed. Ian Billups (University of Birmingham) and Heiko Dobrick (Universität Hamburg) carefully analyzed the first version of the manuscript and provided critical comments and useful suggestions. The critical comments of the anonymous reviewers on different versions of the paper as well as their suggestions for improvement are also gratefully acknowledged.

Requests for reprints should be sent to H. Neumann, Universität Ulm, Fakultät für Informatik, Abteilung Neuroinformatik, Oberer Eselsberg, D-89069 Ulm, Germany, Fax: +49 (731) 502-4156; E-mail: hneumann@neuro.informatik.uni-ulm.de.

921
and coworkers [e.g., Grossberg (1990)]. This theory provides a research framework that takes into account a comprehensive set of empirical data and suggests a neural architecture that utilizes different computational principles for the parallel processing of contrast and brightness information.

In this paper, specific attention is given to the neural network model proposed by Grossberg and Todorovic (1988) (G-T model for short). The original architecture (see Section 2 for a brief description) used only the ON- (B) processing stream for brightness reconstruction, whereas the parallel OFF- (D) stream is only fed forward to orientation selective model simple cells. Here, we begin with an analysis of the properties of cell responses in a shunting network. The network used here consists of parallel local on-center/off-surround and off-center/on-surround processing mechanisms. It will be shown that both processing streams (or channels) can be viewed as multiplexing local contrast as well as luminance data [see also Neumann (1993) for a brief description]. Furthermore, it is also demonstrated that a simple interaction scheme, namely cross-channel inhibition and pooling, can be used to segregate activation according to basic stimulus attributes. As a result, a functional theory is presented that suggests the two parallel retinal processing streams to provide a minimal architecture for subsequently segregated representation of contrast information (of both polarity) and luminance information. Based on the principal processing scheme that allows the 2-to-3 channel mapping or preprocessed input data, an outline of a new processing architecture is described. The two channels with encoded contrast information feed into the first layer of a subsystem for contrast and form perception. The activation distributions in the segregated contrast channels are also used to enhance the contrast of the compressed and blurred luminance distribution. Following the suggested interaction between parallel subsystems of BCS and FCS, the activity distribution of the contrast pathway controls the activity diffusion for filling-in of brightness in the segregated brightness and darkness (B&D) channel. The new framework allows the formulation of different model architectures for center-surround processing as special cases of the model proposed here. The G-T model can also be set in relation to the new model. The architecture proposed in this paper therefore unifies previously suggested approaches and mechanisms.

2. CONTRAST AND BRIGHTNESS PROCESSING IN FACADE THEORY

During the last two decades mechanisms have been developed in order to define a neural network architecture for general-purpose preattentive vision. This ongoing work culminated in the FACADE ("form-and-color-and-depth") theory described in, e.g., Grossberg (1987a,b, 1994). In this section, a brief overview of the full theory is given. It emphasizes the principles in monocular visual information processing. In particular, the G-T model introduced in Grossberg and Todorovic (1988) will be highlighted, since its functional principles serve as a reference and key motivation for the investigation reported in this paper.

The FACADE theory stresses composite and coherent appearance of specific visual stimulus features. This is based on empirical evidence on, for example, the complex response of cells in V4 that suggest multiplexed representations of spatial and spectral information (Desimone et al., 1985). Through the analysis of a number of possibly contradicting observations, it is attempted to identify the basic constraints and mechanisms involved in biological information processing. This biological-evolutionary approach contrasts with other major contributions to the understanding and modeling of central mechanisms in vision. In particular, it denies the approach of functional modularism [see, e.g., Marr, (1982)] where specialized, functionally autonomous processes transform representations in a hierarchy of successive stages. Based on a huge amount of empirical data from psychophysics, physiology, and anatomy, the existence of two interacting but functionally autonomous perceptual subsystems has been proposed:

- The first subsystem (boundary contour system, BCS), which is primarily responsible for the processing of luminance contrasts, generates stimulus-dependent segmentations of form and the perceptual organization of visual entities.
- The second subsystem (feature contour system, FCS) is responsible for the generation of homogeneous perceptual region and surface qualities.

The functionality of the BCS and FCS hierarchically compensates for a number of perceptual uncertainties. Based on noisy and fragmented local measurements, object/surface outlines must be segmented that may be partly obscured by other objects in the scene. The measurements themselves are influenced by uncertainty of a finite spatial aperture. The detection of these surface regions in turn must be invariant against variable illumination conditions. In order to account for these constraints, both subsystems differ in their underlying functional principles. Contrast-related phenomena, e.g., on spatial grouping and illusory contours, suggest local orientation selectivity, insensitivity to local contrast polarity and inwardly directed grouping between inducing elements. A competitive-cooperative loop
combines fragmented measurements of local oriented stimulus contrast in order to generate an emergent context-sensitive boundary (or form) segmentation [e.g., Grossberg & Mingolla (1985)]. Feature quantities, such as brightness, will be reconstructed by utilizing mechanisms of isotropic activation spreading that is specific to local contrast polarity. The form-sensitive segmentations generated in the BCS are signalled to the FCS. The topographically organized BCS activations structure the FCS processes through compartmentalization of the representation space. The orientation insensitive activation spreading is thereupon modulated by the activity patterns in the BCS.

In order to evaluate the functionality of the hypothesized architecture, an instance of the BCS/FCS has been realized and described in Grossberg and Todorovic (1988). Its scope was to demonstrate its capacity to predict classical and new phenomena in brightness perception. In this study, it was sufficient to deal only with a simplified version of the BCS. The reduced version has been realized only as a strictly feedforward hierarchical scheme of contrast processing [see, e.g., Grossberg & Mingolla (1985) for the complete feedback system]. Figure 1 shows a sketch of the major processing steps. Levels 1 and 2 identify initial front-end preprocessing of luminance stimuli by center-surround interaction. Activation feeds forward to the (simplified) BCS that consists of levels 3–5. This cascade begins with orientation- and polarity-sensitive processing that generates orientation fields at the simple cell level. Activation of opposite polarity is pooled for each orientation (complex cell level). Finally, local energy that is generated by individual cell activities is accumulated by the local sum of activity over all orientation fields at each spatial location. Only the ON-channel activation of level 2 feeds into level 6 to inject local filling-in generators (FIGs). The lateral spreading (diffusion) of activation in the FCS stage (level 6) is controlled by activities generated at the output stage of the BCS (level 5). Pooled activation in the BCS generates barriers of low conductivities in the FCS (filling-in barriers, FIBs). The barriers define the borders of compartmentalizations in the visual stimulus field.

For the functional modeling of the dynamics of visual processes in FACADE theory, the framework of first-order differential equations has been used for formal description. Based on the analysis of general properties of feedforward networks, shunting center-surround interaction has been proved to ensure the required network properties, such as bounded network activation, suppression of featureless activity distributions, and relative reflectance processing [see, e.g., Grossberg (1980)].

3. CENTER-SURROUND ANTAGONISM, ON/OFF-CHANNELS AND BRIGHTNESS PERCEPTION

It has been demonstrated that receptor cells build pairwise excitatory and inhibitory synaptic contacts with bipolar cells in the retina. This is the physiological basis of the two parallel pathways that can be characterized by the “light-ON” (B) and “light-OFF” (D) channels [see, e.g., Jung (1973) for an early description]. The response in these two parallel data streams can be characterized at the retinal ganglion cell level by local antagonistic center-surround interactions of respective opposite polarity. These individual processing paths remain separated at least up to the innervations into visual cortical area 17. This proposition has been strongly supported by selective blocking of ON-channel activities (Schiller, 1984, 1992; Fiorentini et al., 1990). The response properties of individual ganglion cells have been rigorously investigated (Enroth-Cugell & Robson, 1984). An important result for our modelling approach is that ganglion cells also respond (at least

1 Separation, in any case, does not rule out interactions in the sense of cross-channel talk for, e.g., activity normalization, noise reduction, etc. It does, however, denote a segregation of processing machinery that is uniquely accessible, functionally autonomous and, furthermore, considered as functionally significant in an evolutionary sense.
minimally) to homogeneous illumination, i.e., in cases where no contrast occurs in the luminance distribution [see also, Li et al. (1992)²].

Shunting equations of the general type

\[ \dot{x}_i = -Ax_i + (B - Cx_i)\text{net}_i^+ - (D + Ex_i)\text{net}_i^- \]  

account well for various phenomena (see Section 2). In the equation, parameters \( A, B, \ldots, E \) are constants, \( \text{net}_i^\pm \) defines the excitatory/inhibitory net input activations and \( x_i \) is the activation of a cell or cell population. Based on this type of equation, one can describe the response properties at the ganglion cell layer using center and surround activity summation based on spatial weighting functions of different extent. It is assumed here that the spatial extents of on/off and off/on RF profiles are symmetric. The weighting functions \( \lambda_{ij} \) that denote the strength of spatial couplings enter into both the B-(ON-) and the D-(OFF-) system³. Therefore, the conditions

\[ \text{ON} \lambda_{ij} = \text{OFF} \lambda_{ij} \equiv \lambda_{ij}^L \]

and

\[ \text{ON} \lambda_{ij}^- = \text{OFF} \lambda_{ij} \equiv \lambda_{ij}^I, \]  

are fulfilled (superscripts "+" and "−" denote excitatory and inhibitory weights for spatial coupling functions). The filtered feedforward inputs are thereupon defined as

\[ \text{net}_i^F \equiv \text{ON net}_i^+ = \text{OFF net}_i^- = \sum_j L_j \cdot \lambda_{ij} \]

and

\[ \text{net}_i^I \equiv \text{ON net}_i^- = \text{OFF net}_i^+ = \sum_j L_j \cdot \lambda_{ij}^I, \]  

respectively, where \( L_j \) denotes discretized input luminance values. Thus, the activity distributions \( \text{net}^F \) and \( \text{net}^I \) are the result of convolution of the input

\[ \text{net}^f = \lambda_i x_i + \text{net}^i \]

luminance with weighting functions defined by \( \lambda_{ij}^L \) and \( \lambda_{ij}^I \), respectively. In addition to the input generation for center and surround mechanisms, according to Grossberg and Todorovic (1988), it is assumed that both channels have the same upper and lower saturation levels, \( \text{ON} B = \text{OFF} B \equiv B \) and \( \text{ON} C = \text{OFF} C \equiv C \), respectively⁴.

The two systems that generate the ON- and OFF-channel activation are assumed to relax quickly to an equilibrium state. The resulting steady-state equations can be written in a format that keeps the antagonistic structure of the RF interaction similar to a linear center-surround model. We get

\[ x_i^+ = \gamma(L) \cdot [B \cdot (\text{net}_i^F - \text{net}_i^I) + (B - C) \cdot \text{net}_i^I]^+ \]

and

\[ x_i^- = \gamma(L) \cdot [B \cdot (\text{net}_i^I - \text{net}_i^F) + (B - C) \cdot \text{net}_i^I]^+, \]  

with \( \gamma(L) = (A + \text{net}_i^F + \text{net}_i^I)^{-1} \), \( L \) to denote the luminance distribution and \( [x]^\pm = \max[x, 0] \) (half-wave rectification). Both channels therefore encode a scaled version of a difference of low-pass (DoLP) filtered input stimulus that does not contain any DC component. In addition to that, both channels also encode a scaled version of the low-pass (LP) filtered stimulus. The superimposed signal components are non-linearly scaled by a function \( \gamma(\cdot) \) which in turn depends on the local luminance level. The shunting properties of a center-surround antagonism, as compared to linear additive DoLP transforms, therefore, not only provide the additional properties of, for example, saturation levels but, at the same time, also encode a scaled and compressed low-pass filtered version of the luminance distribution. The spatial frequency characteristics of the low-pass component in the ON- and OFF-channel directly relate to the inhibitory mechanisms of the DoLP component in the respective channel. This finding provides the basis for a functional theory of the coexisting ON- and OFF-channels. It demonstrates that the pair of model retinal ON- and OFF-processing pathways provides a scheme for transmission of multiplexed contrast (polarity) and luminance information. Such a pair of independent channels is then a necessary prerequisite for de-multiplexing information into separate representations for (i) local ON-, (ii) local OFF-contrast, and (iii) (compressed low spatial frequency) luminance-related (B&D) information.

² Note also that Marr and Hildreth (1980) mentioned response on homogeneously illuminated ganglion cell RFs. In their work, which proposes the Laplacian of Gaussian as an adequate model for retinal isotropic filtering, they report that the response amplitude also scales with the intensity of diffuse illumination.

³ This assumption contrasts the propositions stated in, e.g., Gerrits and Vendrik, 1970. In this approach asymmetries with slightly increased RF sizes for the OFF-channel have been suggested. Since no clear evidence for one of these propositions has been published so far (Fiorentini et al., 1990), the assumption made here is less strong thus proposing the existence of pairwise processing channels with balanced weighting functions (see Krüger & Fischer, 1975).

⁴ For the following model descriptions, by setting \( C = E = 1 \) and renaming \( D \rightarrow C \), a simplified version of eqn (1) is used, such that \( \dot{x}_i = -Ax_i + (B - x_i)\text{net}_i^+ - (C + x_i)\text{net}_i^- \).
given in eqn (4) can now be used to segregate local contrast and luminance-related information. These specialized processing channels provide individual representations as input to different processes such as contrast detection and brightness reconstruction. The postulated mechanisms are:

1. ON-OFF dipole competition (reciprocal cross-channel inhibition) that eliminates the DC-components from both channels. It generates separate representations of DoLP filtered luminance for contrast detection. We get the following equations:

\[ y^+_i = [x^+_i - x^-_i]^+ = \gamma (L) \cdot (B + C) [net^+_i - net^-_i]^+ \]
\[ y^-_i = [x^-_i - x^+_i]^+ = \gamma (L) \cdot (B + C) [net^-_i - net^+_i]^+. \] (5)

The cross-channel inhibition implements a simplified version of the scheme proposed by Maffe \\& Fiorentini (1972) for hypothesized center-surround interaction in a competitive retino-geniculate mechanism. Here the suggested spatial interaction between opponent channel activation is restricted to single locations. However, the suppression of DC-activation also contrast enhances the initial retinal ganglion cell responses as suggested in Maffe \\& Fiorentini (1972).

The resulting (steady-state) \( y \)-activations feed forward to segregated contrast channels. These activations represent contrast responses that are not biased by any additive DC-level component, and are therefore comparable with linear models of center-surround mechanisms. However, as a consequence of the shunting formalism, these responses are non-linearly scaled by the function \( (B + C) \cdot \gamma (L) \) that—as a function of the local luminance level—generates values of scaling coefficients in the range

\[ 0, \frac{B + C}{A}. \]

2. Base level activity is generated in the B&D channel by additive pooling the ON- and OFF-channel responses. As a result, one gets a blurred and compressed version of the input luminance distribution\(^5\)

\[ s_i = x^+_i + x^-_i = \gamma (L) \cdot (B - C) [net^+_i + net^-_i]. \] (6)

The resulting activity distribution defines a reference level that modulates with the luminance intensity. Rewriting eqn (6) as a function of sum-of-low-pass filtered luminance, we get

\[ h (net^+_i) = m \cdot net^+_i / (A + net^+_i) \]

with \( net^+_i = net^+_i + net^-_i \) and \( m = B - C \). For \( m > 0 \), \( h (\cdot) > 0 \) and \( h'(\cdot) < 0 \), the activity distribution in the segregated channel corresponds to the shape of the input luminance distribution. The compression follows a first-order Naka–Rushton non-linearity [Naka \\& Rushton (1966), see the discussion in Section 5].

In summary, a functional theory has been proposed for parallel ON- and OFF-processing streams. These physiologically identified pathways that signal light-increment and light-decrement, respectively, have been modeled at the ganglion cell level using a network with shunting interaction. It is demonstrated here, that each channel not only encodes local contrast information but also a component that is related to the blurred and compressed luminance distribution. It is hypothesized that the layout of two parallel pathways (ON and OFF) is the necessary prerequisite for the segregation of individual channels for the representation of local contrast of any polarity and luminance-related information. This basic anatomy is a powerful scheme to support further processing stages. In particular, it provides input to processes of contrast and form perception as well as to brightness and the representation of surface properties, layout and illumination.

### 4. FURTHER PROCESSING IN THE SEGREGATED CHANNELS

The segregated channels of contrast and luminance information relate to the primary inputs of the
cortical interblob and blob system, respectively (Livingstone & Hubel, 1988). The general outline of the abstract data flow is sketched in Figure 2. The segregated ON- and OFF-contrast channels provide input to simple and complex cells in striate cortex (see Hubel & Wiesel, 1977; Ferster, 1988). The pooled activities of the B&D channel define the second major pathway. The B&D channel at this stage signals a transformed version of the input luminance distribution as a base level and may therefore be labeled as luminance channel (following Cavanagh, 1988). Further stages of the hierarchical processing in this channel are suggested to involve non-linear interaction with activations in both contrast channels and subsequent filling-in to generate the final brightness percept.

In the following, we sketch hypothesized mechanisms necessary for the new proposed architecture to achieve a functionality that compares with the BCS/FCS. It will be shown that the new framework not only allows one to assemble previous models, such as the G-T model, but also provides a basis for more detailed analysis and quantitative prediction. The functionality of selected processing steps is demonstrated by some simulation results.

4.1. Brightness and Darkness (B&D) Channel

4.1.1. Contrast Enhancement of B&D Channel Activity. The activity distribution $s_i$ is a compressed low-pass filtered, i.e., blurred, version of the input. The activities are therefore hypothesized to be contrast enhanced in order to serve for the generation of sharply bounded region outlines. Such a compensatory mechanism for modulation of B&D channel activity can be realized by excitatory ON-channel and inhibitory OFF-channel interactions. Smoothed activities, $y_i^+$ and $y_i^-$, contrast enhance the activity in the B&D channel via a shunting interaction:

$$u_i = -Du_i + s_i + Ey_i^+ - Fu_iy_i^-.$$  (7)
Again, due to its hypothesized fast convergence, a steady-state equation
\[ u_i = \frac{s_i + E\gamma_i^+}{D + \gamma_i^-} \]
(8)
is used instead. The smoothing of activity distributions on the ON- and OFF-channels is described in Section 4.2 below. Figure 3 shows the activity distributions in subsequent stages of the B&D channel for an input luminance step.

4.1.2. Activity Diffusion and Filling-in. Empirical findings on image stabilization and, more recently, brightness-masking experiments suggest the existence of mechanisms of permanent diffusive activity within the B&D channel (e.g., Gerrits & Timmerman, 1969; Paradiso & Nakayama, 1991). Basically, this diffusion process actively fills in quantities generated at the borders of a circumscribed perceptual region. It has been postulated that the local permeability that regulates lateral activity spreading is modulated by the activation in the parallel topographically organized BCS (see Grossberg, 1987a).

The process of geometry-driven diffusion is formalized by making use of the discrete scheme first described in Cohen and Grossberg (1984) [see Grossberg & Todorovic (1988) for the two-dimensional (2-D) extension]. The dynamics of the activity diffusion between network lattice sites of this final brightness representation stage is described by
\[ \dot{v_i} = -Gv_i + \sum_{j \in N_i} (v_j - v_i)P_{ij} + u_i, \]
(9)
where \( N_i \) defines the spatial neighbourhood of lattice site \( i \). The coefficient \( P \) that regulates the local permeability is modulated by BCS (z-) activities
\[ P_{ij} = \frac{\rho}{1 + f(z_{ij})}, \]
(10)
where \( \rho \) denotes the effective diffusion coefficient when BCS activation is absent. In general, \( f(\cdot) \) is a monotonically increasing function of input activities. Thus we obtain low diffusion coefficients at locations of high probability for the presence of a perceptually salient boundary. Simple versions of this function are (for the 1-D case) \( f(z_{ij}) = e \cdot (z_i < op > z_j) \), with \( < op > \in \{+, -\} \) (for generation of z-activities, see Section 4.2 below).

4.2. ON- and OFF-Contrast Channels

4.2.1. Simple and Complex Cell Responses from ON- and OFF-Contrast Channels. Electrophysiological findings support the view that simple cells exist with contrast and orientation sensitive RF profiles of almost linear behavior and selective specificity to odd and even luminance variations (e.g., Hubel & Wiesel, 1977; Pollen & Ronner, 1983). Furthermore, based on Hubel and Wiesel’s sequential hierarchical model, the activations of individual simple cells are hypothesized to combine generating complex cell responses. Whereas the simple cells have been identified to be sensitive to contrast polarity and specific to the local luminance profile, complex cells do not have these properties. The segregated ON- and OFF-contrast channels in this model architecture provide a simple representation scheme to realize a set of mechanisms that account for the basic empirical findings described above.
It is postulated here that smoothing (blurring) of the activities in the separate channels realizes a scheme of divergent activity propagation. This can be formalized as a convolution with a spatial low-pass filter (e.g., a Gaussian). The activation spreading is assumed to fulfill the self-similarity constraint proposed in Grossberg (1987b) such that the smearing is based on the inhibitory weighting function in the simulated retinal ON/OFF RF. Due to the separation of the ON- and OFF-channels, selected activities in the response distribution show properties ascribed to simple cell behavior. This distributed representation implicitly encodes various stimulus attributes such as, for example, contrast polarity, type (odd/even specificity) of local luminance variations, and 2-D orientation. Since it is not a major concern here to model simple cells that make information explicit in a localized representation (see Pollen & Ronner, 1983; Ferster, 1988), only the activation streams in the segregated channels will be considered.

Based on separate ON- and OFF-channels, the response of these linear subfield components of simple cells is generated by the additive equations

\[ \dot{y}_{e}^{+} = -y_{e}^{+} + \sum_{j} y_{j}^{+} \cdot \lambda_{ij} \]

and

\[ \dot{y}_{e}^{-} = -y_{e}^{-} + \sum_{j} y_{j}^{-} \cdot \lambda_{ij}, \]

respectively. The blurring function \( \lambda_{ij} \) is based on the circular weighting function of the retinal surround RF, \( \lambda_{ij}^{r} \), with the additional 2-D feature of an (at least slight) elongation in the direction \( e \). This accounts for the orientation selectivity of simple cells observed for 2-D stimuli presentation. If a spatial frequency scale is added to the framework, the spatial blurring kernels \( \lambda_{ij} \) must be applied for different spatial extent controlled by an additional parameter \( \omega \), thus generating a family of kernels \( \lambda_{ij}^{r} \). So far, however, the architecture is outlined only for a single spatial frequency channel.

The properties of complex cells can now be synthesized easily by smearing and additive combination (spatial pooling) of the ON- and OFF-channel simple cell responses. The self-similar blurring for the generation of \( y_{e}^{+} \) and \( y_{e}^{-} \) activities is the necessary basis for the generation of single maxima in the combination of ON- and OFF-activities (for details of the mathematical derivation, see Neumann (1994)). This is of importance in the case of odd luminance variations (e.g., step edges, see Figures 4 and 5). The complex cell activations show no specificity to contrast polarity and selectivity to local phase (i.e., responsiveness to odd/even symmetry). The activity of the cells is described again by the additive equation

\[ \dot{y}_{e} = -y_{e} + \sum_{j} \left( y_{j}^{+} + y_{j}^{-} \right) \cdot \lambda_{ij}. \]  

Both equations that describe the generation of the components of model simple and complex cell responses (11) and (12) are solved at equilibrium, i.e., for \( y_{e}^{+} = y_{e}^{-} = y_{e} = 0 \). For the processing of static input activation the linear operations involved can be combined into one processing stage. The effective blurring is achieved by variation of the utilized space constant to \( \tau^{*} = \left( \tau_{1}^{2} + \tau_{2}^{2} \right)^{1/2} \). Thus the steady-state complex cell response is

\[ y_{e} = \sum_{j} \left( y_{j}^{+} + y_{j}^{-} \right) \cdot \lambda_{ij}. \]

In case of stimulation with an even symmetric luminance profile (bar, line) of small extent as compared to the RF sizes in the initial (retinal) processing stage, the location of maximum activation coincides with the center of the bar/line profile (see Figure 4). For spatial extents bigger in size, the response profile becomes bimodal; thus the appearance of the bar function splits into two separate odd symmetric contrast variations of inverse polarity. In the case of odd symmetric contrast profiles, e.g., a local contrast step or ramp, the location of maximum activity is always shifted towards the darker side of the local luminance distribution (see Figure 5). This bias is due to the multiplicative factor \( \gamma(L) \) which is a function of the intensity level in the input luminance distribution [see eqn (4)]. The proposed mechanism thus produces a categorical change of edge vs line responses without any predefined mechanism such as specific contrast and line detectors.

4.2.2. Boundary Contour System (BCS). The \( y^{e} \)-activities are suggested to feed into a nonlinear feedback network for which basic architecture has been developed in, e.g., Grossberg and Mingolla (1985). The principal processing stages have been briefly outlined in Section 2. For the sake of simplicity only a rather simplified approximation of the BCS is considered here. The functionality of this subsystem in the overall processing architecture is to generate localized high amplitude diffusion barriers. Such barriers will be generated by a local scheme of non-
maximum suppression (or local winner-take-all mechanism). This mechanism allows one to sharpen a local maximum response of a cell, whereas local non-maxima will be suppressed. The following gating-like interaction realizes the suggested mechanism:

$$z_i = \text{sgn} \left( \left[ y_i - y_{i-1} \right]^+ \cdot \left[ y_i - y_{i+1} \right]^+ \right) \cdot y_i. \quad (13)$$

The generated $z$-activities enter as inhibitory components into the local mechanisms that regulate the conductivities of the diffusion layer in the B&D system [see eqn (10)].

Figure 6 gives an overview of the overall architecture introduced in this paper.

5. FUNCTIONAL PROPERTIES

The introduction of the B&D channel is a central feature of the architecture outlined here. The brightness-from-luminance processing is suggested to consist of three primary stages. In this section, some functional properties of the mechanisms will be analyzed. First, the characteristics of basic B&D channel activities that follow the pooling of initial ON-OFF streams will be analyzed. The results allow for a quantitative prediction of the transmission of varying luminance levels in static stimuli that are encoded in the B&D channel. Second, the two subsequent processing stages in the B&D channel will be analyzed. It is shown that the proposed
architecture of (i) a contrast enhancement stage followed by (ii) the diffusion stage can be simplified so that both stages can be merged into one. For both cases a unique solution for the activity diffusion can be identified. A key motivation for the investigations was the unification of approaches to the modeling of contrast and brightness perception. Therefore, in the last paragraph of this section, different approaches will be embedded in the framework outlined here.

5.1. Response of the B&D Channel

The base level activity of the B&D channel is determined by the sum of x-activities in the initial ON- and OFF-data streams [see eqn (6)]. In order to analyze the response properties of the B&D channel static homogeneous illumination is assumed for a spatial region which is at least covered by the surround weighting function of the RF. This required feature is fulfilled by, for example, intensity functions of constant level as well as ramps of constant slope. Furthermore, a unit integral weighting function, e.g., a spatial Gaussian or any other symmetric low-pass filter function, is assumed for λ_{ij}. Under these conditions, the s-activity results in a function of (mean) input luminance L:

\[ s_i(L) = \left\{ \frac{(B - C)(\lambda^i + \lambda') \otimes L}{A + (\lambda^i + \lambda') \otimes L} \right\}_i = \left\{ \frac{2(B - C)L}{A + 2L} \right\}_i, \]

where \( L \) denotes the local luminance distribution, and '\( \otimes \)' denotes the convolution operator. The lower and
ranges of the normalized input. The compressive nonlinearity for the transfer is of the form described by Naka and Rushton (1966). The equation corresponds to the version with unit exponent value having a slope of \(2/A^*\) for \(L^* \approx 0\), approaching a value of zero as \(L^* \gg A^*\). It is not intended here to link this specific parametrization to perceptual data. However, it provides a reference to further investigate the luminance-to-brightness mapping for different background levels. Figure 7 shows graphs of the function \(s^*\) for different values of \(A^*\).

5.2. Filling-In Processes

We assume that the diffusion stage (see Section 4.1) utilizes an ensemble of cells (each site indicated by \(i \in \{1, \ldots, N\}\)) with local nearest neighbor coupling. The neighborhood \(\mathcal{N}_i\) therefore can be defined as the set of discrete locations \(\mathcal{N}_i = \{i - 1, i, i + 1\}\). With this neighborhood set \(\mathcal{N}_i\), the steady-state solution of eqn (9), \(\tilde{v}_i = 0\), can be written as [see also Arrington (1993)] for the first description of the approach to the solution

\[
(G + P_{i,j-1} + P_{i,j+1})v_i - P_{i,j-1}v_{i-1} - P_{i,j+1}v_{i+1} = u_i, \quad (15)
\]

with \(u_i = (s_i + E\tilde{v}_i)/(D + F\tilde{v}_i)\) [eqn (8)]. For the complete network’s lattice sites eqn (15) results in a linear system of equations

\[
\begin{pmatrix}
G + P_{1,0} + P_{1,2} & -P_{1,2} & 0 \\
-P_{2,1} & G + P_{2,1} + P_{2,3} & -P_{2,3} \\
0 & -P_{3,2} & G + P_{3,2} + P_{3,4} \\
\vdots & \vdots & \vdots \\
0 & \ldots & v_1 \\
0 & \ldots & v_2 \\
-P_{3,4} & \ldots & v_3 \\
G + P_{4,3} + P_{4,5} & \ldots & v_4 \\
\vdots & \vdots & \vdots \\
& & v_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4 \\
u_5 \\
\end{pmatrix} \quad (16)
\]

or, in abbreviated form, \(M \cdot v = u\). As we can see

\[
\text{7\footnote{In order to keep the characteristic line sensitive to variations over a wide range of luminance levels, we learn that \(A\) should have a magnitude of \(A > L_{\text{max}}/4\). This raises the issue of an adaptation of \(A\) in cases of significantly varying luminance levels. Possibly, \(A\) itself could be a function of average luminance measured within a large spatial neighborhood such as the outer-surround region reported in Li et al. (1992). In order to also account for temporal effects, one has to consider photoreceptor dynamics as well as processing at the bipolar and horizontal cell level (see Gaudiano, 1992).}}
\]
from eqn (16), M is a tri-diagonal matrix with \( N \times N \) elements. For non-vanishing decays in the diffusion equation [eqn (9)], M is always strictly diagonally dominant and therefore invertible (Young & Gregory, 1988). The unique solution is given by \( v = M^{-1} \cdot u \). The corresponding circuit of local interactions for contrast enhancement and diffusive filling-in (as part of the overall architecture depicted in Figure 6) is shown in Figure 8a.

In a variation of the description of the investigated processing stages the separated hierarchical steps of contrast enhancement and diffusion are merged into one stage. Therefore, the Laplacian-like diffusion term in eqn (9) is added to the equation that denotes the contrast enhancement [eqn (7)]. The deduced circuit is sketched in Figure 8b.

This results in\(^8\)

\[
\hat{v}_i = -Dv_i + s_i + E\gamma_i^+ v_i - Fv_i \gamma_i^- + \sum_{j \in N_i} (v_j - v_i)P_{ij}. \tag{17}
\]

Again, at equilibrium and for the discrete neighborhood \( N_i \), we can rewrite the equation to get

\[
(D + F\gamma_i^- + P_{i,j-1} + P_{i,j+1})v_i - P_{i,j-1}v_{i-1} - P_{i,j+1}v_{i+1} = s_i + E\gamma_i^+. \tag{18}
\]

The corresponding linear system of equations then results in

\[
\begin{pmatrix}
  b_1 + P_{1,0} + P_{1,2} & -P_{1,2} & 0 \\
  -P_{2,1} & b_2 + P_{2,1} + P_{2,3} & -P_{2,3} \\
  0 & -P_{3,2} & b_3 + P_{3,2} + P_{3,4} \\
  0 & 0 & -P_{4,3} \\
  \vdots & \vdots & \vdots \\
  0 & \cdots & 0 \\
  0 & \cdots & \cdots \\
  -P_{3,4} & \cdots & v_3 \\
  b_4 + P_{4,3} + P_{4,5} \cdots & \cdots & v_4 \\
  \vdots & \vdots & \vdots 
\end{pmatrix}
\begin{pmatrix}
  v_1 \\
  v_2 \\
  \vdots \\
  v_3 \\
  v_4 \\
  \vdots \\
  \vdots 
\end{pmatrix}
= \begin{pmatrix}
  a_1 \\
  a_2 \\
  \vdots \\
  a_3 \\
  a_4 \\
  \vdots \\
  \vdots 
\end{pmatrix}
\]

\( b_i = D + F\gamma_i^- \) correspond to the numerator and the denominator of the steady-state solution of the former contrast enhancement step [see eqn (8)]. This matrix can also be inverted [as in the case of eqn (16)] to uniquely solve for \( v = (v_1, v_2, \ldots, v_N)^T \).

The second approach has the advantage that, by

\(^8\) The name of the variable has been changed from \( u \) to \( v \) in order to identically denote the diffusion stages in both models.
elimination of parameter $G$, the description is reduced by one degree of freedom. However, a thorough comparative evaluation of the two alternative descriptions (including the differences in the temporal evolution of the dynamics) is beyond the scope of this paper, and is therefore left for further analysis.

5.3. Relation to Other Models of Isomorphic Brightness Reconstruction

5.3.1. The G-T Model Considered within the New Framework. The development of the new model was motivated by the FCS architecture proposed by Grossberg and Todorovic (1988). In their approach, initial antagonistic center-surround processing feeds into parallel ON- and OFF-pathways identical to the initial processing stage used here [see eqn (4)]. However, among these two parallel pathways, only the ON-pathway signal (directly) enters into the brightness diffusion stage of the G-T model (see Section 2). The different contributions generated by local contrast information and baseline activity are represented in segregated, thus independent, channels in the new model (Figure 6). Brightness prediction in the G-T model is based on the initial ON-cell responses that have now been identified to multiplex a baseline as well as a local contrast activation of both polarities. Therefore, in addition to the greater flexibility, the new architecture can be used to evaluate brightness prediction generated by the G-T model. For simplicity, the $x^+_i$-signal [eqn (4)] is assumed to be positive, so that the rectification can be neglected (see also the discussion in Section 3). Based on the observation, that $\max[x, 0] + \max[-x, 0] = x$, we get

$$x^+_i = u_i = B \gamma(L)(\text{net}^+_i - \text{net}^-_i) + (B - C) \gamma(L) \text{net}^-_i = s^+_i + \frac{B}{2(B + C)}(x^+_i - x^-_i)$$  \hspace{1cm} (19)$$

where $s^+_i = s_i - (B - C) \gamma(L) \cdot \text{net}^-_i \approx \frac{1}{2} s_i$. This shows that the ground level activities in the B&D channel can be related to those generated by the $x^+_i$-signal in the G-T model. The compressed, low-pass filtered luminance distribution is also enhanced by local contrast information. However, in the G-T model this enhancement is generated additively, whereas in the new architecture a shunting inhibition is involved. The latter guarantees the B&D activity [$u_i$, eqn (7)] to be always positively bounded, whereas this cannot be guaranteed in the G-T model. The hypothesized shunting interaction furthermore introduces an additional degree of freedom to control the balance between contributions of ON- and OFF-contrast channel activities to generate the final percept. The major contribution of the new proposed architecture certainly is that the reciprocal inhibition and pooling interaction of initial ON- and OFF-pathways allows the segregation of a separate channel to represent the level of input luminance. It is argued that the visual system must have access to an absolute reference level. Otherwise, if the reconstruction mechanism is based on local contrast alone, the generation would be restricted to produce only relative brightness levels and the assignment to an arbitrary quantitative reference would be necessary. Due to the similarities in the logic of the brightness reconstruction process, the new architecture is able to reproduce all the findings and behavior generated with the original G-T model [see Grossberg & Todorovic (1988) for static 1-D and 2-D stimuli.

\[9\] The distribution of $s$-activity is given by $s_i = \{g - B \gamma(L)(x^+_i + \lambda') \odot L_i\}$. For the corresponding $s^+$-activity, one gets $s^+_i = \{g - \gamma(L) x^+_i \odot L_i\}$. The difference between the two profiles is that the scaled and compressed input luminance is filtered by a sum-of-low-pass and a low-pass filter, respectively. Within homogeneous regions, the identity $s_i = 2s^+_i$ is fulfilled. At discontinuities, the activation profile gets slightly less blurring for $s_i$ as compared to $s^+_i$. \[10\]
and Arrington (1993) for simulations of temporal phenomena.

5.3.2. Comparison with other Models of Brightness Diffusion. In the general outline of the FACADE theory the existence of pairs of parallel diffusion layers for black and white as well as opponent colors (red/green and blue/yellow) has been hypothesized. For brightness perception, therefore, two FCS copies (or B&D channels) for ON- and OFF-activation will be assumed. Based on this theoretical framework, different contributions have been published that include a stage of brightness processing using multiple diffusion layers. In Grossberg and Wyse (1991) a processing scheme for figure-ground separation of target objects from noisy background was proposed. The definition of the ON-channel in this architecture again is the same as the one used here. In contrast to the architecture proposed here, the upper and lower saturation levels in the opponent channels have been interchanged and the OFF-channel has been defined to realize a type of image inversion. Using the notation introduced in Section 3, the steady-state equations read

\[ x_1^+ = \gamma(L) \cdot \left[ (\text{Bnet}_1^+ - \text{Cnet}_1^+) + x_1^- \right] \]

and

\[ x_1^- = \gamma(L) \cdot \left[ (\text{A1} + \text{Cnet}_1^- - \text{Bnet}_1^-) \right] \]

where \( I \) denotes a tonic activity and \( \gamma(L) = \left( A + \text{net}_1^+ + \text{net}_1^- \right)^{-1} \). The activity pooling yields \( s_1 = x_1^+ + x_1^- = \gamma(L) \cdot \text{A1} \). In comparison with the architecture proposed in this article, the pooling there results in a different functional behavior. The overall sensitivity curve, in comparison with the one of the B&D channel, shows the opposite sign for the slope. In the limits, for homogeneous luminance distributions, we find \( \lim_{L \to 0} s_1 = I \) and \( \lim_{L \to \infty} s_1 = 0 \).

The work reported in Arrington (1993) and Gove (1994) is directly concerned with the modeling of brightness phenomena. Two parallel ON- and OFF-parallel have been assumed to coexist, both having zero baseline activity. This can be directly accomplished by setting the upper and lower saturation levels of the shunting equations to equal magnitude, \( B = C \). The activities in the ON- and OFF-contrast channels then become

\[ y_1^+ = 2B \gamma(L) \left[ \pm \left( \text{net}_1^+ - \text{net}_1^- \right) \right] \]

and the base-level activity in the B&D channel is simply \( s_1 = 0 \).

A further issue is the incorporation of a pair of diffusion layers into the architecture proposed here. The incorporation of a more elaborate stage for non-linear contrast detection at the simple cell level and a recurrent field for boundary processing extends the architecture to successfully account for a wide variety of perceptual data. The extended model can account for a wide variety of data such as high and low contrast phenomena, gradual luminance transitions, repetitive contrast patterns, and Mach band effects. The mechanisms and model evaluation have been described in Pessoa et al. (1995).

Meanwhile, the models defined by Grossberg and Todorovic (1988), Gove (1994), as well as Pessoa et al. (1995) have been successfully applied to real world image processing tasks. The comparative study of performance of the different approaches can be found in Grossberg et al. (1995).

6. SUMMARY

A framework for a neural architecture for contrast and brightness processing has been proposed. A cornerstone of this proposal is that retinal ON- and OFF-channels not only encode contrast information but also compressed low-pass filtered versions of the raw luminance distribution. Activations in the retinal ON- and OFF-pathways are segregated into three parallel channels for ON- and OFF-contrast and brightness and darkness information, respectively. This suggests a functional theory of retinal ON- and OFF-processing pathways to enable a 2-to-3 mapping of channel representations. Based on this, a three-stage process of brightness reconstruction is postulated. The activations within the two parallel contrast channels provide input to model cortical simple and complex cells at the front-end stage of the BCS. The segregated contrast activities are also hypothesized to contrast enhance the base level activity in the B&D channel. The excitatory and inhibitory interaction for the enhancement act as filling-in generators for the subsequent diffusion stage.

The work reported here provides a number of new insights. The initial 2-to-3 mapping has been proved to also apply to other processing schemes that utilize parallel ON- and OFF-processing pathways. By this, the new proposal allows one to embed different approaches in this general scheme. As part of a neural instruction set, it suggests a mechanism for the processing luminance in combination with contrast information. The segregated luminance-related, or B&D, channel provides a reference level for brightness reconstruction. In this framework, it is also possible to analyze the quantitative prediction of the level of final brightness reconstruction. The analysis of response properties of the B&D channel, furthermore, allows for the quantitative analysis of other related models, in particular the G-T model. The pooling of ON-/OFF-contrast channel activity suggests a simple scheme for contrast detection. It shows an adaptive functional behavior of categorical change in response to odd and even symmetric luminance variations.

REFERENCES

Mechanisms for Contrast and Brightness Perception

Boston University, Graduate School, Dept. of Cognitive and Neural Systems.


**NOMENCLATURE**

- $x_i$: activity of model cell at lattice site $i$ for feedforward shunting interaction
- $N$: number of model cells
- $\lambda^{x,i}$: isotropic spatial coupling functions for initial center-surround processing that denote effective weights between sites $i$ and $j$ (superscripts "x" and "i" denote center and surround kernels, respectively)
luminance distribution used as input stimulus
input activation described by the spatial convolution of a weighting function with the raw luminance or an activity distribution of the previous cell layer (superscripts “c” and “s” denote center and surround activation, respectively)

\[ \langle \rangle^+ \] half-wave rectification of cell activation, \( \max[x, 0] \)

\( \lambda() \) non-linear scaling function that depends on the local luminance level

\( x_i^+ \) activity of model cell for on-center/off-surround processing (ON-channel, denoted by "\(+\)") and for off-center/on-surround processing (OFF-channel, denoted by "\(-\)")

\( A, B, C \) parameters of feedforward center-surround processing to generate \( x \)-activity

\( y_{i+}^\pm \) segregated contrast channel activation (reciprocal cross-channel inhibition to generate ON-contrast, "\(+\)", and OFF-contrast, "\(-\)", activation)

\( s_i \) base level activation in the segregated B&D channel (pooling of initial ON- and OFF-channel activities)

\( u_i \) activity in the B&D channel after contrast enhancement of \( s_i \)

\( D, E, F \) parameters of shunting equation for contrast enhancement, \( u \)-activity

\( v_i \) activity in the diffusion layer

\( G \) decay parameter of diffusion equation

\( P_{ij} \) effective membrane permeability for diffusion between cell sites \( i \) and \( j \) modulated by the local BCS-activation (reduction of the diffusion coefficient \( \rho \))

\( y_{iu}^{\pm} \) activations of the orientation selective components of model simple cells at site \( i \) and orientation \( \varepsilon \) (superscript "s" denotes the stage of simple cell processing with segregated ON- ("\(+\)") and OFF- ("\(-\)") contrast activation as input)

\( \lambda_{\varepsilon} \) anisotropic spatial coupling function at the simple cell stage with elongation along orientation \( \varepsilon \) (an additional superscript indicates different spatial frequency channels)

\( y_{iu}^c \) activations of complex cells (superscript "c") that integrate simple cell responses

\( z_i \) activations generated by the (reduced) boundary contour system (BCS)

\( M \) \( N \times N \) matrix that represents steady-state diffusion activity