Biol. Cybern. 85, 355–369 (2001)

Biological Cybernetics

Visual filling-in for computing perceptual surface properties

Heiko Neumann[1], **Luiz Pessoa**[2], **Thorsten Hansen**[1]

[1] Universität Ulm, Fakultät für Informatik, Abteilung Neuroinformatik, Oberer Eselsberg, 89069 Ulm, Germany
[2] Universidade Federal do Rio de Janeiro, Programa de Engenharia de Sistemas e Computação – COPPE Sistemas, Ilha do Fundao, Rio de Janeiro, RJ 21945-970, Brazil

Received: 11 March 1999 / Accepted in revised form: 14 March 2001

Abstract. The visual system is constantly confronted with the problem of integrating local signals into more global arrangements. This arises from the nature of early cell responses, whether they signal localized measures of luminance, motion, retinal position differences, or discontinuities. Consequently, from sparse, local measurements, the visual system must somehow generate the most likely hypothesis that is consistent with them. In this paper, we study the problem of determining achromatic surface properties, namely brightness. Mechanisms of brightness filling-in have been described by qualitative as well as quantitative models, such as by the one proposed by Cohen and Grossberg [Cohen and Grossberg (1984) Percept Psychophys 36: 428–456]. We demonstrate that filling-in from contrast estimates leads to a regularized solution for the computational problem of generating brightness representations from sparse estimates. This provides deeper insights into the nature of filling-in processes and the underlying objective function one wishes to compute. This particularly guided the proposal of a new modified version of filling-in, namely *confidence-based* filling-in which generates more robust brightness representations. Our investigation relates the modeling of perceptual data for biological vision to the mathematical frameworks of regularization theory and linear spatially variant diffusion. It therefore unifies different research directions that have so far coexisted in different scientific communities.

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; Rogers-Ramachandran and Ramachandran 1998; see also Grossberg and Mingolla 1985). Several perceptual completion phenomena (see Pessoa et al. 1998) suggest that, on a functional level, regions inherit local border contrast information by means of "spreading mechanisms" or "filling-in" (Paradiso and Nakayama 1991; Caputo 1998). 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.

In this paper, we first review recent empirical findings about perceptual filling-in and surface property determination. We show that the filling-in of brightness through a diffusion process can be understood in terms of the general framework of feature reconstruction through regularization that is widely used in computer vision. In general, 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 ill-posed. Given sparse contrast estimates at image contours, the problem of finding a corresponding brightness or texture representation has no unique solution nor is it 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. It will be demonstrated that the steady-state representation of filled-in brightness, as proposed by Cohen and Grossberg (1984), provides a regularized solution of the inverse problem of generating a dense representation from local estimates.

Based on our formalization of filling-in within the framework of regularization theory, we introduce a new version of brightness filling-in. As shown below, given sparse contrast estimates at image contours, diffusive filling-in generates a brightness representation that is driven by both a smoothness constraint and the input data (i.e., the contrast estimates). This representation should not be confounded, however, by data terms at

Correspondence to: H. Neumann

locations where they are not available – such as inside uniform luminance regions. Accordingly, our proposed scheme employs a *confidence* measure to ensure that data terms are only considered where available. As demonstrated through simulations, this allows filling-in to more readily account for extended regions of uniform brightness. Finally, we also demonstrate the behavior of our new version of filling-in with classic stimuli, such as the Craik-O'Brien-Cornsweet (COC) stimulus (Cornsweet 1970).

2 Empirical evidence for neural filling-in mechanisms

Among those perceptual surface qualities that relate to photometric quantities, *brightness* is defined as perceived illumination and *lightness* relates to perceived reflectance (Beck 1972; Fiorentini et al. 1990). In the following, we consider only luminance stimuli depicting flat matte 2-D surface layouts that are devoid of any scenic attributes depending on surface orientation, shadowing, and depth variations. Under those conditions the reflected light of different regions and their reflectivities are directly related, and the neural mechanisms affect brightness and lightness in similar ways. We therefore make no particular distinction between the two attributes *brightness* and *lightness* (compare Fiorentini et al. 1990). The luminance-to-brightness mapping is not unique but depends on the figural arrangement that surrounds the target region (for a detailed discussion, see Todorović 1987; Pessoa et al. 1998). Consequently, the processes of brightness perception are not primarily reconstructive in their nature but are mainly guided by invariance properties that help to generate robust percepts under variable scene conditions. As a result, percepts are created which can systematically deviate from the pattern of input intensities. The goal is to reveal the neural computational mechanisms underlying these processes of brightness perception.

Paradiso and Nakayama (1991) used a visual masking paradigm to investigate two issues: the role of edge information in determining the brightness of homogeneous regions, and the temporal dynamics of brightness perception. They reasoned that if brightness perception relied on some form of activity spreading, it would be possible to demonstrate its existence by interrupting it. If boundaries interrupt spreading, what happens when new borders are introduced? Is this filling-in process affected before it is complete? Figure 1 shows the paradigm they used as well as the basic result. The target is presented first and is followed at variable intervals by a mask. For intervals of the order of 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. Moreover, contours interrupt the propagation. Thus, for example, for a circular-shaped

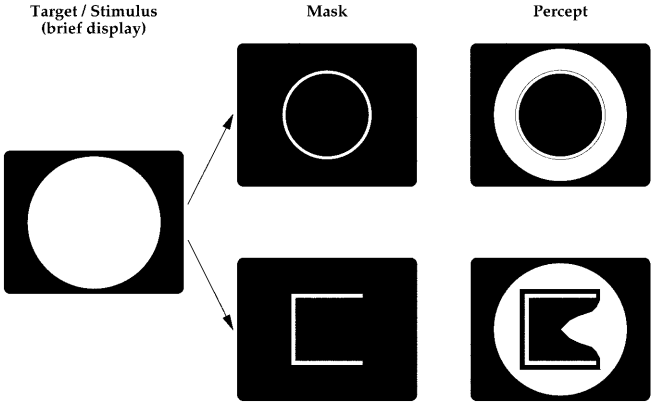

Fig. 1. Masking paradigm in the temporal dynamics of brightness study by Paradiso and Nakayama (1991). Brightness suppression of a disk-shaped target by a mask. The target and mask are each presented for 16 ms. Optimizing the temporal delay between target and mask yields a percept in which the brightness in a large central area of the disk is greatly suppressed. Brightness suppression is highly dependent on the arrangement of the contours in the mask

mask, brightness signals originating from the target border seem to be entirely "blocked" (hence a dark middle disk). An altogether different outcome results for larger delays between target and mask stimuli. If the mask is presented after 100 ms, the brightness of the central region is largely unaffected. Corroborating the hypothesis that the propagation of brightness signals is involved, Paradiso and Nakayama (1991) showed that brightness suppression depended on the distance between target and mask. In particular, for larger distances maximal suppression occurred at later times, revealing a propagation rate of 110–150 deg/s (6.7–9.2 ms/deg); see also Stoper and Mansfield (1978). Recently, Caputo (1998) employed a similar masking paradigm to investigate brightness filling-in within texture patterns. Again the spreading could be mainly blocked by the mask if the stimulus-to-mask interval was in accordance with the propagation rate required to travel the distance between boundary and mask position.

De Valois et al. (1986) employed center-surround standard (reference) and matching (variable) stimuli, similar to the ones used in classic simultaneous contrast studies. They compared the results of direct changes in brightness where the center of the standard pattern was explicitly modulated in luminance (as was the matching pattern), to the induced changes that occurred when the surround was modulated sinusoidally while the center was kept constant at the mean level (i.e., a temporal version of simultaneous contrast). Their studies revealed two main findings: (1) the temporal modulation at frequencies ranging from 0.5 to 8 Hz had little effect on the apparent brightness change in the *direct condition*; and (2) in the *induced condition*, the amount of brightness change fell drastically as the temporal frequency increased (around 2.5 Hz). (Note that these frequencies are much lower than the ones usually revealed in flicker studies, which have cutoff frequencies of more than 30 Hz and peak around 4–6 Hz.) These results can be interpreted in terms of a spreading mechanism of

induction that occurs over time, one that would provide a spatially continuous representation for filling-in. Rossi and Paradiso (1996) replicated the brightness induction results of De Valois et al. (1986) and studied the role of pattern size. Recently, Davey et al. (1998) investigated the temporal properties of brightness induction in COC patterns (cf. Fig. 11). Their main finding was that induced brightness in the COC gratings was stronger and persisted until higher temporal frequencies for higher spatial frequencies (see Pessoa and Neumann (1998) for a discussion of these results).

In summary, these studies are suggestive of active neural filling-in processes that are initiated at region edges. In brightness filling-in, the brain seems to be generating a spatially organized representation, and it seems to be doing so through a roughly continuous propagation of signals, a process that takes time (see Pessoa et al. 1998). Below we show how the use of standard techniques for the solution of inverse problems is useful in elucidating the computational and mathematical properties of the putative filling-in brain mechanisms.

3 Filling-in for the dense representation of surface properties

Many of the objects we perceive have roughly uniform regions of surface color, brightness, and depth. At the same time, cells in the visual cortex in general do not respond to uniform regions, but rather to discontinuities (Hubel and Wiesel 1962, 1968). In other words, many neurons respond more strongly to *boundaries* than to regions or surfaces. In the vicinity of object borders, contrast signals are produced by the visual system. How should the appearance of inner regions be determined given the absence of direct neural support? To introduce concepts, we consider the task of generating a continuous representation of surface layout as one of *painting* (or *coloring* – Mumford 1994) an empty region that is bounded by the local measures at region boundaries. The task thus consists of generating an internal representation of surface properties from sparse data. Individual surfaces occur at different sizes and with various shapes. Therefore, any such mechanism has to be insensitive to such size and shape differences. Filling-in models suggest that bounded local contrast measures are used in the determination of surface appearance through a process of lateral spreading, or diffusion (Gerrits and Vendrik 1970; Cohen and Grossberg 1984; Hamada 1984).

Models of brightness perception were among the first to explore the dichotomy of boundary and surface subsystems. Based on stabilized image studies, Gerrits and Vendrik (1970) proposed that the perception of brightness depends on filling-in processes that occur within separate ON and OFF channels. The two channels would be involved in "brightness" (B) and "darkness" (D) processes, respectively (Jung 1973). Figure 2 (left) sketches the main elements of their model. These basic ideas were formalized by Cohen and Grossberg (1984) and Grossberg and Mingolla (1985), who proposed a

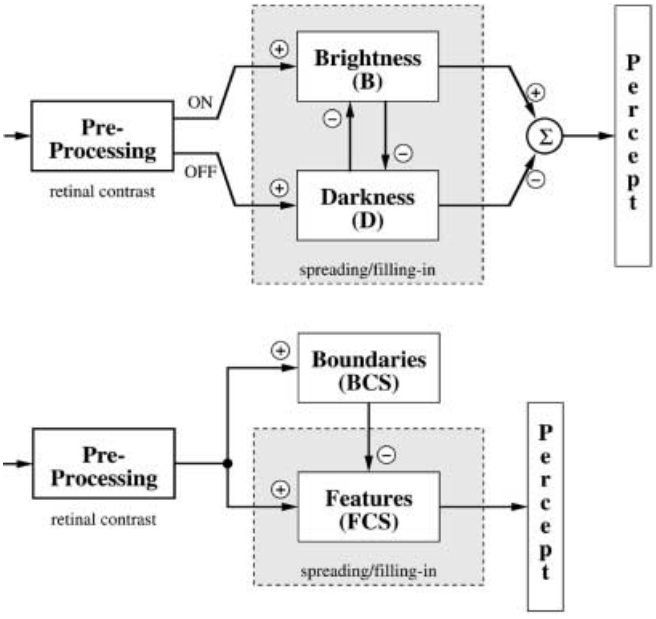

Fig. 2. Macroscopic components of filling-in models. *Left*: The raw stimulus is preprocessed utilizing a center-surround processing stage. The resulting antagonistic contrast responses feed to separate filling-in stages that generate "brightness" (*B*) and "darkness" (*D*) activation. Juxtaposed B–D activations form local barriers which in turn prevent filling-in to spread into bounded regions. Superposition of equilibrated B- and inverted D-activity on a reference, or Eigengrau, level determines the final brightness percept (Gerrits and Vendrik 1970). *Right*: Contrast signals from the preprocessing stage are input to boundary detection (*BCS*) and filling-in (*FCS*) stages. Filling-in is regulated by boundary signals that constrain signals to occur within bounded regions. The activity within the filling-in stage is the model's correlate of perceived brightness (Grossberg and Todorović 1988)

model of complementary boundary and surface systems (boundary contour system/feature contour system, BCS/FCS). In a nutshell, BCS/FCS processing occurs as follows. The input distribution is initially processed by center-surround mechanisms, analogous to retinal ganglion cells that code luminance ratios at image edges. Contrast signals are then used in two ways. The BCS extracts boundaries (as in other edge-detection algorithms) and further groups them into surface boundary arrangements. Contrast signals are also fed to a filling-in stage (in the FCS) where they undergo a process of lateral spreading. The FCS computations determine the appearance of the incoming stimulus, such as *brightness*, hue, and depth (see Fig. 2, right). The two systems are involved in complementary computations and, as indicated by image stabilization studies (Krauskopf 1963; Yarbus 1967), the final boundary signals produced by the BCS are used to regulate a filling-in process that occurs within the FCS system. Grossberg and Todorović (1988) showed how this proposal is capable of qualitatively accounting for several brightness phenomena, including simultaneous contrast, brightness assimilation, the COC effect, the Hermann grid, and Mondrian displays. Figure 2 (right) contains a sketch of the main elements of the model proposed by Grossberg and Todorović. Recent simulations of the model by

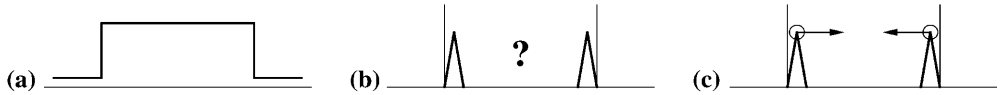

Fig. 3. a Profile of an example stimulus arrangement with a light region in the center and a darker background. **b** Processing of such stimulus utilizing initial center-surround interaction produces local activations in the vicinity of the contrast. For the generation of a dense representation of surface quality the empty inner region must be somehow filled with activation. **c** Filling-in mechanisms take the initial contrast responses near the luminance step and paint the rest of the region by a process of lateral spreading, or diffusion

Arrington (1994) have also shown that it is capable of accounting for data on the temporal dynamics of brightness perception, as studied by Paradiso and Nakayama (1991) (see above). A recent extension of the model motivated by phenomena exhibiting brightness gradients (Pessoa et al. 1995) was used to account for stimuli such as trapezoidal and triangular Mach bands (see Pessoa 1996a,b), low- and high-contrast missing fundamental stimuli, and sinusoidal waves.

For concreteness, let us consider how a simple BCS/FCS scheme can explain brightness perception in a simplified figure–ground scene (see Fig. 3). For simplicity we employ a 1-D representation of the stimulus and assume only an ON channel is present. The luminance difference between the disk and the surround produces contrast signals associated with the dark-to-light transitions. Contrast signals are then used in two ways. They generate boundary signals that determine the regions of influence of the initial contrast measurements. Contrast signals (or a blurred version of them) are also sent to the filling-in stage. There, filling-in signals undergo lateral diffusion. Initially, filling-in signals are equivalent to (blurred) contrast signals. With time, the distribution of activity of filling-in signals changes as they spread laterally. Spreading occurs as long as signals are not stopped by a boundary. Eventually, after the brightness signals spread, a brightness plateau ensues defining a center region as lighter than the background. Overall, the BCS/FCS in particular, as well as other filling-in proposals (Hamada 1984; Arrington 1996) may be viewed as attempts to bridge the gap from local contrast responses (ratios) to more continuous spatial representations (which in the BCS/FCS comprise the equilibrated filling-in signals).

The task of integrating local measurements into more global percepts is not limited to brightness perception. In fact, it is a central problem in early vision (Poggio et al. 1985). For instance, the domains of depth and motion perception are confronted with similar challenges. Using random dot stereograms, Julesz (1971) showed that the visual system strives to find a smooth surface layout compatible with the disparity information. This holds even in cases where only a minor fraction of points gives rise to localized disparity signals (so-called 5% stereograms; Julesz 1971, p. 122). Instead of perceiving individual dots floating in depth, the subject perceives a *surface* in depth that is consistent with the local estimates. Recent evidence suggests that depth is assigned to the interiors of bounded homogeneous regions in which only the *vertically* oriented boundaries provide a source of local disparity (Nakayama and Shimojo 1992). The depth assignment revealed by these and other studies may be mediated by a filling-in mechanism similar to the one for the determination of brightness. Starting from sparsely localized disparity estimates, a smooth surface in depth would be generated (see Anstis and Howard (1978) and Brookes and Stevens (1989) for other depth-brightness-related illusions; compare also the discussion in Howard and Rogers (1995) on depth- and motion-induced contrast illusions similar to brightness).

The visual system is thus constantly confronted with the issue of how to integrate local signals into more global arrangements. Mathematically, this can be interpreted as solving an *inverse problem*. From sparse, local measurements, the visual system must somehow generate the most likely hypothesis that is consistent with them. This reconstruction process is in general not well defined (or well-posed), as there may be an infinite number of possible configurations consistent with the sparse measurements. Constraints are thus required in order to define a unique solution that would correspond to the percept. Below, we investigate filling-in as a mechanism solving the inverse problem of generating dense representations from sparse estimates.

4 Filling-in as a mechanism to solve an inverse problem

4.1 The basic filling-in equation

The filling-in equation analyzed in this section has been used as a building block in several models of early vision (e.g., Gove et al. 1995; Pessoa et al. 1995; Grossberg and McLoughlin 1997). Within the empty space of bounded regions, activity generated by initial measurements spreads laterally, whereas at boundaries the diffusion stops. Cohen and Grossberg (1984) – see also Grossberg and Todorović (1988) for an extension to the 2-D case – suggested a nonconservative filling-in mechanism which utilizes a steady (or clamped) source of local input signals, c. The resulting filling-in activity will be denoted by v. The lateral spreading of activity v is controlled by an auxiliary boundary system which generates local activities w. In addition, in order to take into account a neurally plausible implementation based on leaky cell compartments, a passive decay of activity (with rate K) is also incorporated. In all, the discretized equation for the filling-in mechanism reads

$$\dot{v}_i = -Kv_i + c_i + \sum_{j \in \mathcal{N}_i} (v_j - v_i)\rho_{ij} \ . \qquad (1)$$

In this equation i and j denote spatial locations, and the sum over a nearest-neighbor coupling, $\mathcal{N}_i$, corresponds to the discretized diffusion component that is controlled by a spatially varying permeability $\rho(w)$. [The scalar function which denotes the efficacy of lateral spreading is termed *diffusivity* in the (nonlinear) diffusion literature (e.g., Weickert 1997). In our problem domain, this relates to the *permeability* of an assembly of laterally interacting model cells, or compartments. We therefore use the second term throughout the paper.] The permeability is defined by a monotonically decreasing function of boundary signals w, for example $\rho_{ij} = \rho/(1 + a(w_i + w_j))$, which is independent of v.

4.2 Filling-in, diffusion and regularization theory

Filling-in in continuous form. The discretized filling-in equation, (1), contains a diffusion and a linear reaction term, the latter to bias the equilibrium solution towards the contrast input at region boundaries. The diffusion term is described by the sum of weighted differences between activations at neighboring network sites. Each component represents a numerical first-order difference term whose efficacy is modulated by ρ_{ij}. Each difference can therefore be considered as a discretized version of the spatially continuous form $\rho(w)\nabla v$, where $w = w(\mathbf{x})$, such that the term $\sum_{j\in\mathcal{N}_i}(v_j - v_i)\rho_{ij}$ is a first-order discretization of $\nabla \cdot (\rho(\cdot) \cdot \nabla v(\mathbf{x}))$ with

$$\rho(v_i - v_{i-1}) \approx \rho\nabla v \equiv y_l, \quad \rho(v_{i+1} - v_i) \approx \rho\nabla v \equiv y_r,$$
$$\text{and } (y_r - y_l) \approx \nabla \cdot y$$

for the 1-D case (compare with Fig. 4). Overall we can formulate the filling-in equation in *continuous* form as

$$\partial_t v(\mathbf{x};t) = \nabla \cdot (\rho(w) \cdot \nabla v(\mathbf{x};t)) + g(v,c) \ , \qquad (2)$$

with $g(v,c) = c(\mathbf{x};t) - Kv(\mathbf{x};t)$ $(\partial_t \equiv \partial/\partial t)$. Here $\mathbf{x} = (x,y)$ denotes the spatial coordinates and ∇ is the generalized differentiation operator $(\partial_x \ \partial_y)^{\mathrm{T}}$. By this mechanism the diffusion is biased by a source, $c(\mathbf{x};t)$, associated with local contrast measurement signals, and a sink, $-Kv(\mathbf{x};t)$, that, in the absence of any input, produces a decay of activation at rate K [compare Eq. (1)]; $\rho(\cdot)$ is the spatially variant function that controls the permeability. The steady-state solution of (2) identifies a state in the diffusion system in which the driving input and the spontaneous decay are balanced without further lateral spreading of activation.

Continuous filling-in and diffusion mechanisms. The continuous filling-in equation, (2), reveals many similarities with mechanisms used in diffusion models of computer vision (see Weickert (1997) for a recent review). Without the bias term $g(v,c) = c(\mathbf{x};t) - Kv(\mathbf{x};t)$ in (2), the filling-in mechanism results in a standard *spatially variant*, or *inhomogeneous*, *linear diffusion equation*. The permeability coefficient is locally controlled by the activities $w(\mathbf{x})$ of the segregated topographically organized boundary system, which leads to an *inhomogeneous* process. The diffusion process itself is *linear* since the permeability function $\rho(\cdot)$ is independent of v and the bias term $g(\cdot)$ is a linear function of v. In addition, the filling-in process is *isotropic* since $\rho(\cdot)$ is a scalar-valued function, and consequently, the resulting flow $\Phi = \rho(\cdot)\nabla v$ is oriented parallel to the gradient of the filled-in activity. (Please note that in an anisotropic scheme, the *direction* of maximum spread varies locally, e.g., pointing along the direction of iso-intensity lines or the tangent of a boundary.) In all, the filling-in mechanism displayed in (2) is a linear isotropic, geometry-driven process in which the local graduation of lateral spreading is controlled by the external form-sensitive boundary system. The filling-in equation also contains an additional source and a sink component. Taken together

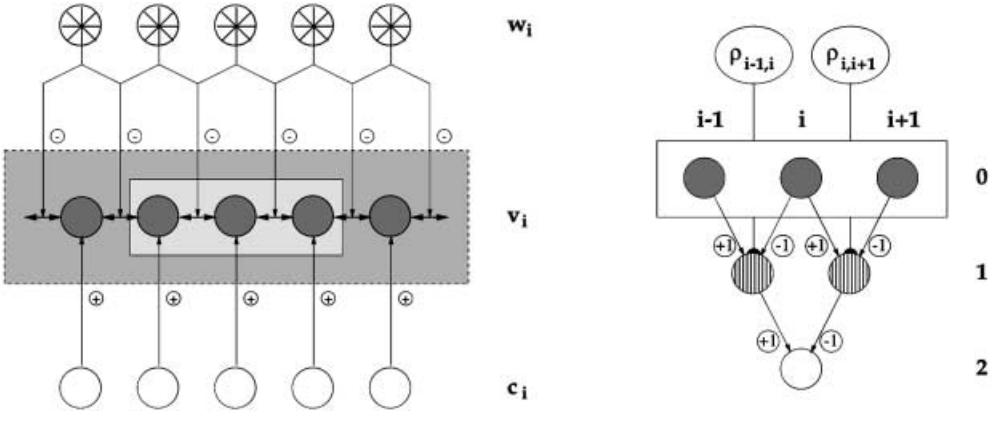

Fig. 4. Sketch of the discretized network for filling-in. *Left*: Scheme of lateral interaction based on nearest neighbor coupling in a 1-D grid ($\mathcal{N}_i = \{i-1, i+1\}$) to generate filling-in activities v_i. Each node in the filling-in network is fed by excitatory feedforward projections from sparse contrast activation c_i. Lateral coupling between neighboring lattice sites is modulated by inhibitory external inputs generated in the boundary system (ρ_{ij}; the wheel icons indicate pooled activation from filters tuned to different orientations). *Right*: Numerical approximation of spatial derivatives (difference scheme for the central site coupled to its direct neighbors; efficacies are denoted by signed multipliers). The three-level cascade realizes the spatially variant diffusion component $\nabla \cdot (\rho\nabla v)$ in 1-D, the hatching highlights those nodes whose contribution is modified by the spatially variant permeability

they define a linear *reaction term* which guarantees that the equilibrium solution will be close to the input. Such reaction terms have been used in a number of computational vision models for non-linear diffusion and reaction-diffusion processes (e.g., Schnörr and Sprengel 1994; for an overview, see Weickert 1997).

ON and OFF contrast systems. Modeling brightness reconstruction from local contrast information must take into account the non-negativity of cell responses. Consider a luminance contrast such as in Fig. 3: ON contrast cells signal the luminance *increment* at the brighter region whereas OFF contrast cells signal the luminance *decrement* for the darker region in the vicinity of the edge (not shown). For the filling-in mechanism to function properly, the diffusion must be based on separate representations for ON and OFF contrast activation, respectively (see Gove et al. 1995; Pessoa et al. 1995; Arrington 1996). Therefore, the computational architecture includes two segregated ON and OFF filling-in networks that generate activities $v^+(\mathbf{x};t)$ and $v^-(\mathbf{x};t)$, following the dynamics

$$\begin{aligned}\partial_t v^+(\mathbf{x};t) &= c^+(\mathbf{x};t) - Kv^+(\mathbf{x};t) + \nabla\cdot(\rho(w)\cdot\nabla v^+(\mathbf{x};t)),\\ \partial_t v^-(\mathbf{x};t) &= c^-(\mathbf{x};t) - Kv^-(\mathbf{x};t) + \nabla\cdot(\rho(w)\cdot\nabla v^-(\mathbf{x};t))\ .\end{aligned} \tag{3}$$

In both systems the permeability is controlled by the activation of the auxiliary boundary system. The corresponding discretized filling-in equations $\dot{v}_i^+$ and $\dot{v}_i^-$ are defined in accordance to the scheme shown in (1).

Computing brightness representations and regularization theory. Computing dense brightness representations from local contrast estimates can be formulated more generally as finding a solution for v (the brightness) in the space X given the measured contrast data c in some space Y. The mapping between elements in the pair of spaces (X, Y) is denoted by the operator A, thus $c = Av$. In our case A merges together the initial center-surround filtering (see Neumann 1996). The existence and uniqueness of a solution and its continuous dependence on the data cannot be guaranteed since the measurements may be noisy and are sparse. The inverse problem is therefore classified as ill-posed in the sense of Hadamard (Tikhonov and Arsenin 1977; Poggio et al. 1985; Bertero et al. 1988). The solution to the problem has to be regularized such that proper constraints are imposed on the possible set of candidates in the function space of solutions. We select the function $\hat{v}$ as the solution that minimizes the norm $\|A\hat{v} - c\|_Y$ subject to the additional constraint of smoothness of $\hat{v}$, where smoothness is characterized by a minimized derivative (e.g., of first order). The overall goal is to minimize the constraint given by the local differences between the measured data and the reconstructed function values (data term) and the stabilizing functional imposed on the function (smoothness term). This results in the goal of minimizing the quadratic functional

$$\|Av - c\|^2 + \lambda\|Pv\|^2 \rightarrow \min\ , \tag{4}$$

where P denotes a "constraint operator" (a mapping $P : X \rightarrow Z$) that stabilizes the solution for the inverse problem and $\|\cdot\|$ represents a proper norm in the spaces Y and Z (Poggio et al. 1985). For a detailed discussion of the formal mathematical background we refer to Bertero et al. (1988) and Baumeister (1987).

In order to solve the inverse problem we minimize a quadratic functional $E(v) = \int_\Omega E_d(v,c) + \lambda E_p(v) \rightarrow \min$ to yield a solution function that realizes a compromise between the data term ("similarity" or "closeness" to the data) and the model term ("smoothness" of the solution) (Gelfand and Fomin 1963). If we consider only first-order variations in the smoothing term, the resulting functional for each contrast channel reads

$$E(v^\pm) = \iint_\Omega F\left[v^\pm(x,y), v_x^\pm(x,y), v_y^\pm(x,y), x, y\right]\mathrm{d}x\,\mathrm{d}y \rightarrow \min\ , \tag{5}$$

where $F = \kappa(\mathbf{x})(Kv^\pm(\mathbf{x}) - c^\pm(\mathbf{x}))^2 + \rho(w)(v_x^\pm(\mathbf{x}) + v_y^\pm(\mathbf{x}))^2$, with $w = w(\mathbf{x})$ and $\mathbf{x} = (x,y)$. Again, $c^\pm$ represents the local estimated contrast data in either ON or OFF channel, $v^\pm$ is the brightness function we search for in the function space for the given inverse problem, $v_x^\pm(\cdot), v_y^\pm(\cdot)$ are its first-order partial derivatives, and κ is a function that represents local confidence measures at a sparse set of locations. The permeability $\rho(w)$ serves as a spatially variant regularization parameter that allows the modulation of the magnitude of the contribution of the smoothness term (compare Terzopoulos 1986).

The necessary condition for the existence of a solution for this minimization problem is denoted by the Euler–Lagrange (E–L) equation, which for F results in (Gelfand and Fomin 1963)

$$\frac{\partial}{\partial v^\pm}F - \left(\frac{\partial}{\partial x}\frac{\partial}{\partial v_x^\pm}F + \frac{\partial}{\partial y}\frac{\partial}{\partial v_y^\pm}F\right) = 0\ . \tag{6}$$

In the functional defined for the given F we seek a solution close to the data, while at the same time minimizing the first-order derivatives of the approximating function [see Eq. (5)]. The corresponding E–L equation for ON and OFF contrast activation finally reads

$$K\kappa(\mathbf{x})(Kv^\pm(\mathbf{x}) - c^\pm(\mathbf{x})) - \nabla\cdot\left(\rho(w)\cdot\nabla v^\pm(\mathbf{x})\right) = 0\ . \tag{7}$$

4.3 Analysis and predictions

Above we have derived an equation for the continuous filling-in mechanism [see Eq. (3)] and the E-L equation for the membrane-regularized solution for generating a brightness representation [see Eq. (7)]. This indicates that the ill-posed problem of constructing a dense representation of surface property is regularized by the

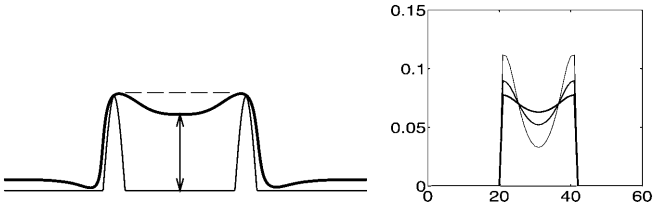

Fig. 5. *Left*: Generation of a brightness representation using standard filling-in. Given a constant confidence value everywhere ($\tilde{\kappa}(\mathbf{x}) = 1$), the zero contrast values for inner regions (*thin line*) force the model terms to approach this reference value (*thick line*). *Right*: Simulations of standard filling-in for different permeability values, $\rho \in \{15, 45, 135\}$ (*normal*, *bold*, and *extra bold*, respectively). The decay parameter was set constant to $K = 1$. The process of standard filling-in produces a "bending" in the final activity distributions where the amount of bending increases for lower values of ρ (see text)

mechanism of diffusive filling-in utilizing the network depicted in Fig. 4. An important difference between the above-mentioned equations remains, however. In (7) the data compatibility term $(Kv^{\pm}(\mathbf{x}) - c^{\pm}(\mathbf{x}))$ denoting the similarity between the signal and the measured input data is multiplied by an explicit confidence measure, or "data availability" coefficient, $K\kappa(\mathbf{x}) \equiv \tilde{\kappa}(\mathbf{x})$. Thus in the process of generating a representation from sparse data, the contribution of a data compatibility term should only be effective at those locations where input data is available. If a confidence measure is omitted, any data point in the otherwise sparse input field would be used.

The situation represented by (3) may be interpreted as if a confidence measure of constant value, e.g., $\tilde{\kappa}(\mathbf{x}) = 1$, exists at all spatial positions, irrespective of the distribution of input measurements. In this case, a missing input measurement (corresponding to object regions away from borders) is treated as a *zero-amplitude* data point that should be approached by the function that represents the dense surface property. Thus uniform regions must always, at equilibrium, exhibit "bowed" signal distributions. Figure 5 (left) illustrates the situation. Associated with a luminance pedestal, contrast signals will be generated in the vicinity of the edges. Given that $\tilde{\kappa}(\mathbf{x}) = 1$ everywhere and that the local measures for inner regions value zero, a bow in the signal distribution is generated such that at inner positions the model term is driven to zero. The parameter ρ in the permeability function $\rho(w) = \rho/(1 + aw)$ regulates the flatness of the filled-in signal: decreasing the value of ρ decreases the contribution of the smoothness term and, therefore, increases the relative contribution of the input data. As a result, the signal approaches more closely the available input data and – with the non-vanishing confidence measure – gets more bowed. Figure 5 (right) shows the results of computer simulations of a luminance pedestal stimulus using (1). The actual amount of bending depends on the permeability ρ of the diffusion equation. For smaller permeability values the bending is quite considerable, for larger values it can be reduced, but it is always present.

The amount of "bowing" in the distribution of a brightness signal also depends on the size of a target region which is to be filled-in by local contrast responses at the edges. For illustration purposes, consider a bounded region (corresponding to the pedestal in Fig. 5) in which input contrast is available only near its left boundary. (This restriction is feasible since in the filling-in process any contribution of input contrast superimposes in a linear fashion. Therefore, the effects of "bowing" can be studied in this reduced scenario.) Based on our analysis above, assuming constant unit-value confidence values filling-in approaches zero input data; however, always keeping the balance between the closeness to the data and the imposed smoothness constraint of the underlying energy functional (5). Consequently, if the ratio between the number of sites having zero input and the total number of sites increases, the tendency of the filled-in signal to approach zero amplitude is stronger. Figure 6 sketches the layout and the predicted amplitudes of filling-in response. This observation predicts that without an additional mechanism to control the influence of the input data, filling-in is not invariant against the size of a surface region.

5 Confidence-based filling-in

We have shown above how to relate diffusive filling-in to the regularization of the inverse problem of generating a brightness representation from sparse data. Based on this analysis, we propose a new extended scheme of *confidence-based* filling-in. Our proposal utilizes a confidence function to validate the availability of input contrast that contributes to the data compatibility term in the corresponding regularization functional. We now discuss how model complex-cell responses can be effectively used to compute confidence values.

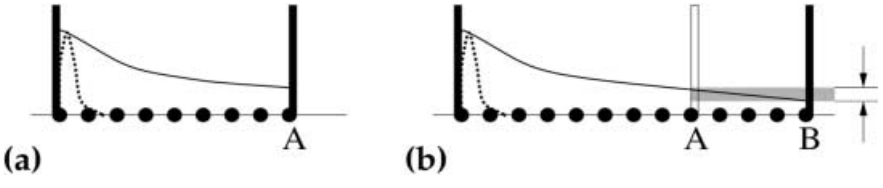

Fig. 6a,b. Sketch of the "bowing" effect in the filled-in signal distribution for a mechanism not utilizing any confidence measure: **a** A region bounded by two contours with contrast input at one side (*dashed bold*). Filled-in activation (*continuous line*) tends to bend downward as to approach the level of zero input activation. **b** A larger region but with the same input conditions as in case **a**. The difference in the decline of activity is highlighted by the *gray rectangle*

5.1 Complex cell response and confidence measure

Input contrast, and simple- and complex-cell responses. How should confidence be defined in the domain of contrast measures and the generation of a brightness representation? Consider again the structure of the hierarchical processing of input luminance distributions. Initial center-surround interaction by unoriented filters generates contrast responses, $c^{\pm}$, in separate ON and OFF channels that flank any sharp luminance transition (edge) (the computational stages involved in the initial center-surround processing are described in the Appendix). Contrast responses are then integrated by oriented weighting functions which define the subfields of cortical simple cells. Simple-cell responses in turn feed into polarity-insensitive model complex cells.

Simple-cell responses are computed by integrating juxtaposed contrast responses from elongated offset subfields. In order to generate more localized responses in case of a localized luminance transition (edge), we "boost" the response by using an additional component of multiplicative combination of both sub-field responses (see Neumann et al. 1999). In particular, for a cell sensitive to dark–light (DL) luminance transitions (polarity) for orientation θ at location $\mathbf{x}$, we get the steady-state response $S_\theta^{\mathrm{DL}}(\mathbf{x})$, and the response $S_\theta^{\mathrm{LD}}(\mathbf{x})$ for a cell with light–dark (LD) sensitivity.

Complex-cell responses are generated by pooling the simple-cell responses of opposite contrast polarity. Here we utilize the self-normalizing properties of shunting interactions proposed by, for example, Grossberg (1980) which has recently been adopted by Heeger (1992) to explain empirical data of cortical cell responses. At steady state, the mechanism reads

$$Z_\theta(\mathbf{x}) = \beta_c \frac{S_\theta^{\mathrm{DL}}(\mathbf{x}) + S_\theta^{\mathrm{LD}}(\mathbf{x})}{\alpha_c + \{(S^{\mathrm{DL}} + S^{\mathrm{LD}}) \otimes \Lambda\}_\theta(\mathbf{x})} \quad , \tag{8}$$

where α_c and β_c are constants and Λ denotes an isotropic weighting function to integrate activations in a space/orientation neighborhood.

Confidence measure. The shunting mechanism contributes a divisive inhibitory component such that the complex-cell responses are compressed in a nonlinear fashion (compare Heeger 1992). The steady-state responses can now be directly utilized to derive a spatially continuous field of confidence measures to gate, or modulate, the input contrast responses [compare Eq. (7)]. We get

$$\tilde{\kappa}(\mathbf{x}) = \frac{K}{\beta} Z(\mathbf{x}) + \varepsilon \quad , \tag{9}$$

where $Z(\mathbf{x}) = \max_\theta(Z_\theta(\mathbf{x}))$, and ε is a small tonic input which is needed to achieve the well-posedness of the functional in (5). K corresponds to the decay parameter and β is a scaling constant; both are utilized to achieve a proper mapping of confidences in a range of $[0, 1]$.

The key observation is that simple-cell responses are pooled to form a field of complex-cell responses. After spatial pooling, these cell responses are more uncertain with respect to edge position than the initial input contrast responses. Thus, the inherent *positional uncertainty* always spans the spatial extent covered by the distribution of center-surround filters. Of central importance here is the fact that the representations of local ON and OFF responses generated by center-surround filters are used as the input signals for both filling-in and model complex cells – the latter via the intermediate stage of simple cell integration. Both representations are thus spatially related. Simple- and complex-cell responses define a continuous representation that implicitly encodes the presence and validity of contrast estimates that are used for filling-in. Hence, complex-cell responses can be used to define confidence measures. The stages of this scheme with their different computational roles are depicted in Fig. 7.

5.2 Discretized confidence-based filling-in

Similarly, we can also derive a discretized version of the new filling-in equation. We first sample the confidence

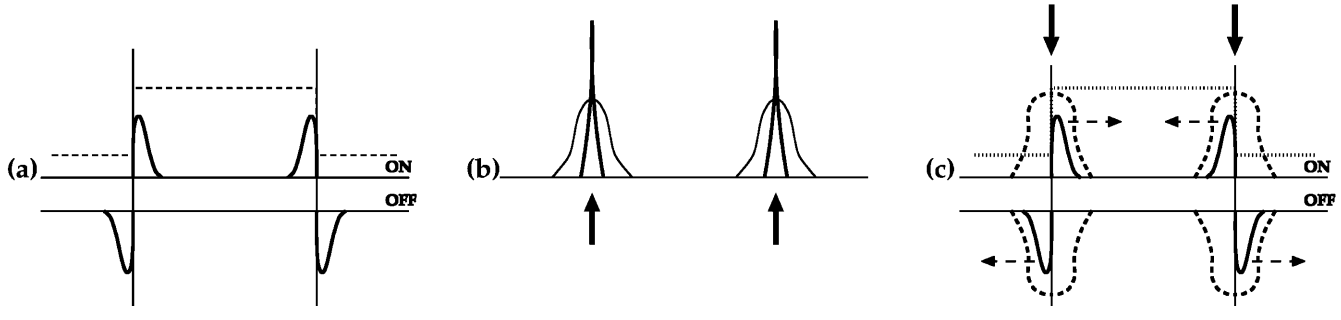

Fig. 7a–c. Computational principles involved in confidence-based filling-in. **a** Input profile of a bright region on a darker background (*dashed lines*). ON and OFF contrast responses near luminance edges (*thick lines*). **b** Spatially offset ON and OFF contrast responses are integrated into opponent subfields of cortical simple cells which are pooled to form complex-cell responses (*thin lines*). High amplitude responses (*bold arrows*) are sharpened to generate localized boundary activations (*thick lines*). **c** ON and OFF contrast responses (*thick lines*) feed into separate filling-in systems. Boundary responses generate barriers at locations of luminance edges (*bold arrows*) to control the lateral spreading in the filling-in systems (*dashed arrows*). Confidence signals derived from broadly tuned complex-cell responses (*dashed lines*) control the contribution of the data term in the regularization functional

activity $\tilde{\kappa}(\mathbf{x})$ (9) at discrete positions, $\tilde{\kappa}_i$. In accordance with (1) we get

$$\dot{v}_i^{\pm} = \left(c_i^{\pm} - K v_i^{\pm}\right)\tilde{\kappa}_i + \sum_{j \in \mathcal{N}_i} \left(v_j^{\pm} - v_i^{\pm}\right)\rho_{ij} \ . \tag{10}$$

The steady-state response for $\dot{v}_i^{\pm} = 0$ (for all i) is given by the linear system $\mathbf{M} \cdot \mathbf{v}^{\pm} = \mathbf{c}^{\pm}$, in which $\mathbf{M}$ denotes the matrix of lateral interactions on the discrete grid, $\mathbf{v}^{\pm}$ denotes the steady-state filling-in activations, and $\mathbf{c}^{\pm}$ represents the contrast inputs weighted by the gain of the spatially varying confidence-values. For parameters $K > 0$ and using the definition of the confidence function $\tilde{\kappa}_i$, the matrix $\mathbf{M}$ is guaranteed to be invertible. The solution for the filled-in activities is thus defined by

$$\mathbf{v}^{\pm} = \mathbf{M}^{-1} \cdot \mathbf{c}^{\pm} \ . \tag{11}$$

6 Simulations

The properties of the discretized filling-in mechanisms of (1) and (10) have been investigated on the basis of various types of stimuli. First, we demonstrate the dependency of the results of standard filling-in (1) on the parametrization and the brightness predictions on the actual region size. Confidence-based filling-in is shown to be more robust against these degrees of freedom. The mechanism of confidence-based filling-in is then applied to psychophysical stimuli. In order to demonstrate the model's invariance properties and its capacity to deal with real world data, we also show results of processing real camera images. All the resulting brightness predictions shown below were generated by taking the difference between activities from ON and OFF filling-in networks. The reference level is set to the "Eigengrau", such that we get

$$b(\mathbf{x}) = \text{gray} + (v^{+}(\mathbf{x}) - v^{-}(\mathbf{x})) \ . \tag{12}$$

6.1 Invariance properties

Our first investigation focuses on the properties of the filling-in mechanisms and their dependency on the parameter settings and the size of the region to be filled-in. We start with simple luminance patterns each with a central light square of different size on a dark background (see Fig. 8). The brightness signals generated by the standard filling-in mechanism (1) tend to bow depending on the strength of the permeability coefficient. An increase in the permeability helps the generation of flat signals (see Fig. 8, bottom, left panel in first and second pair). The corresponding brightness patterns generated by the confidence-based filling-in mechanism remain invariant against these parameter changes and are always flat (Fig. 8, bottom, right panel in first and second pair). The results of the confidence mechanism remain stable for the different sizes of the square figure.

Next, the same mechanisms have been applied to another test image that contains shapes of different form and size but with the same luminance level. The results reveal the potential weaknesses of standard filling-in: depending on the size or diameter of a pattern (which is unknown), the brightness signals appear at a different amplitude and show different amounts of bowing (Fig. 9). With the confidence-based filling-in mechanism,

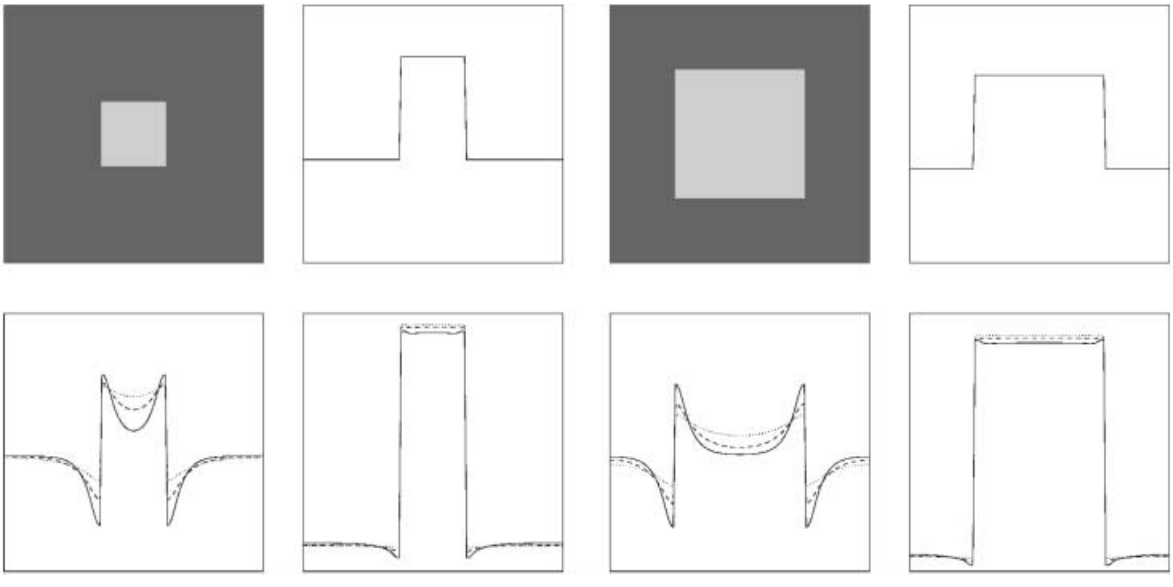

Fig. 8. Generation of brightness appearances for rectangular test patterns of different sizes utilizing mechanisms of standard and confidence-based filling-in. *Top row*: Pairs of luminance pattern (*left*) and profile (*right*), each for a small and a large test square, respectively. Input size was 128×128 pixels with bright regions of 32×32 (small) and 64×64 (large) pixels. *Bottom row*: Simulation results for both pattern sizes shown in pairs of standard filling-in (*left*) and confidence-based filling-in (*right*) for each pattern size. The parameter settings in both models are $K = 0.5$ and $\rho \in \{15, 45, 135\}$ (*solid*, *dashed*, and *dotted lines*, respectively)

the brightness patterns appear homogeneous and of almost the same brightness. (Note that the small variations in the brightness amplitude are mainly due to different contrast input levels that are measured at the shape boundaries. Since the contrast is measured using an isotropic center-surround mechanism, the responses vary as a function of local contour curvature.) This property holds true for virtually arbitrary large homogeneous regions, since zero confidence values switch off the data term in the interior of such regions. We conclude that confidence-based filling-in helps to generate a brightness representation that is largely invariant against shape size and, therefore, improves the robustness of filling-in mechanisms.

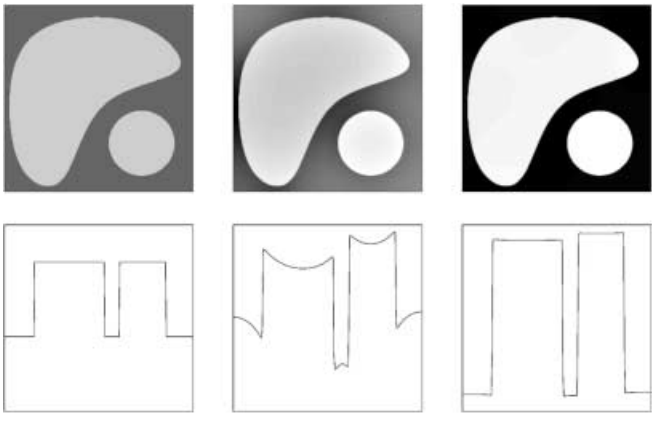

Fig. 9. Filled-in brightness signals for shapes of different size but same luminance. Signals are generated by the filling-in mechanisms using the parameter settings that achieved proper results for the "square" test pattern above. *Top row*: Input luminance pattern (*left*), and brightness signal generated by standard filling-in (*middle*) and by confidence-based filling-in (*right*). *Bottom row*: Corresponding profile sections of the luminance function and the brightness patterns taken along the diagonal (from *upper left* to *lower right corner*)

6.2 Brightness data

In this section we demonstrate the ability of the confidence-based mechanism of filling-in to process classical luminance patterns that have been investigated in brightness perception. The first simulation demonstrates that the model is able to produce brightness contrast effects, as have previous filling-in schemes (Grossberg and Todorović 1988; Pessoa et al. 1995; Neumann 1996). Figure 10 shows that gray patches of identical luminance appear darker or lighter depending on the luminance of the adjacent surround. This effect is known as simultaneous contrast (see Fiorentini et al. 1990).

With the other stimuli, we particularly focus on remote border contrast effects and their creation of brightness differences. These cases provide examples of the crucial role that edges have on brightness appearance. For example, two regions of equal uniform luminance separated by a "cusp edge" appear of different brightness – the COC effect (see Cornsweet 1970; Todorović 1987). These types of stimuli have been identified as the most challenging ones for alternative theories of brightness perception such as, for example, filter theories. In fact, as yet only filling-in models appear to properly predict the brightness appearance for COC stimuli and their variants (compare Blakeslee and McCourt 1997, 1999).

Figure 11 (left) shows the input luminance distribution of a standard COC stimulus where a cusp edge appears in the center. The lower-luminance side of the cusp is associated with a uniformly darker region, and the higher-luminance side of the cusp with a uniformly brighter region. Figure 11 (right) shows the prediction generated by the confidence-based filling-in mechanism. As with previous filling-in proposals (e.g., Grossberg and Todorović 1988), it correctly predicts the effect.

A sequence of cusp edges having pairwise opposite contrast polarities generates a brightness pattern of a

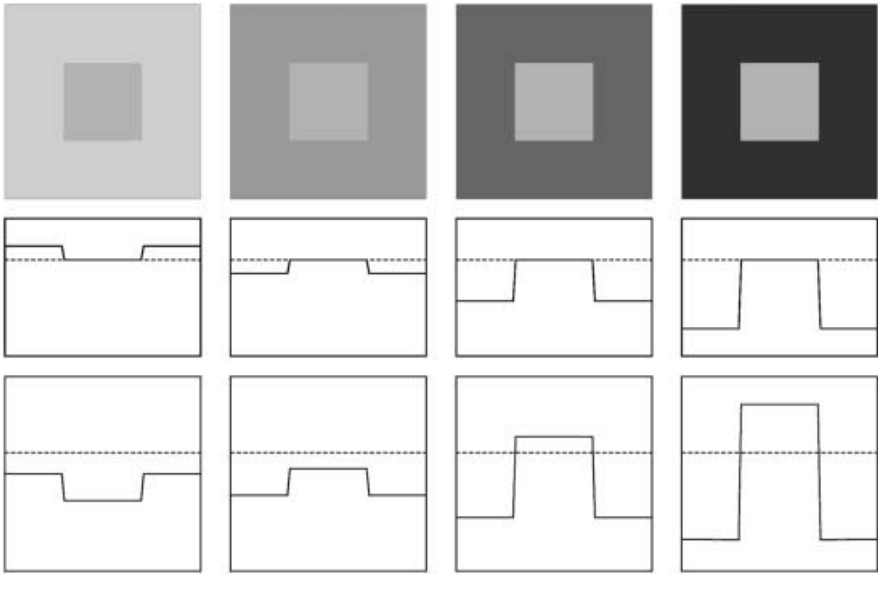

Fig. 10. Simultaneous contrast patterns. *Top row*: 2-D luminance stimuli of a constant gray patch embedded in surrounds of decreasing luminance. *Central row*: Luminance profiles from central 1-D horizontal cross sections through the stimuli shown in the *top row*. The constant gray level of the central square is highlighted by the *dashed line*. *Bottom row*: Profile sections from corresponding 2-D simulation results utilizing confidence-based filling-in. The *dashed line* serves as a reference level generated by the mean brightness activity measured at the central regions of all four patterns

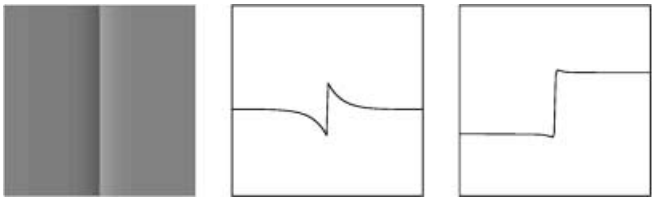

Fig. 11. Processing two regions of identical luminance level separated by a luminance cusp (the Craik-O'Brien-Cornsweet stimulus). *Left*: 2-D input stimulus. *Center*: Profile of central cross section through the input luminance pattern. *Right*: Profile section from the corresponding 2-D simulation result utilizing confidence-based filling-in

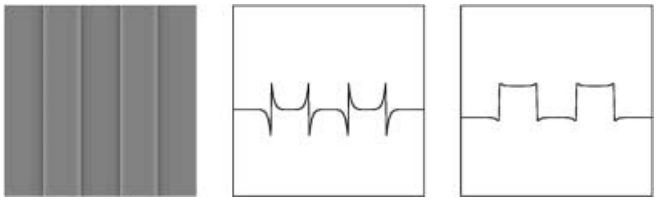

Fig. 12. Filled-in brightness signals for a square wave pattern made of Craik-O'Brien-Cornsweet cusps of opposite contrast polarities. *Left*: Input luminance pattern. *Middle*: Horizontal profile. *Right*: Profile of the 2-D brightness pattern generated by confidence-based filling-in

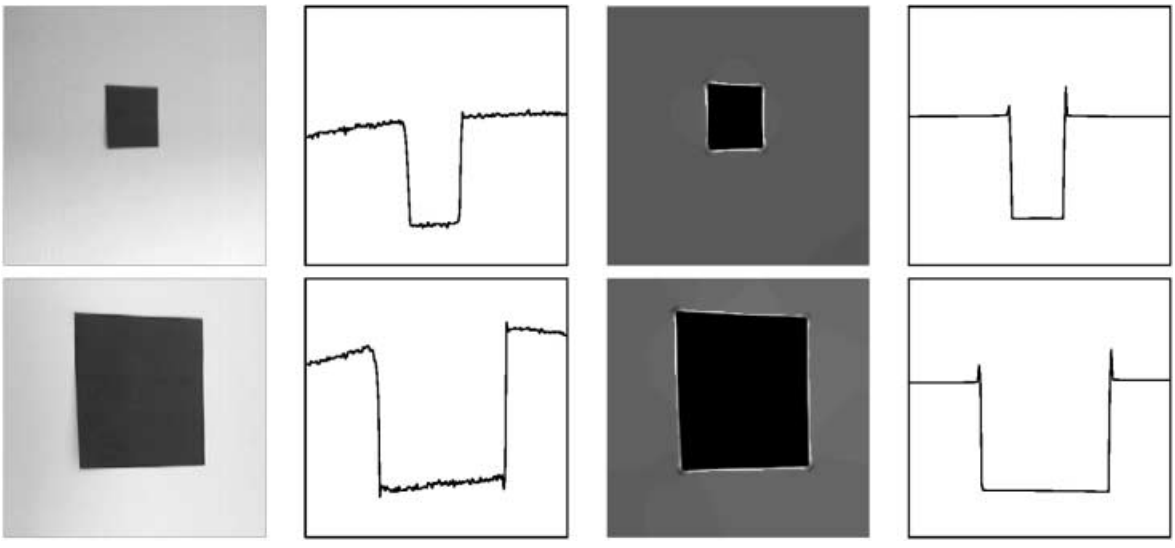

Fig. 13. Processing camera images of a flat 3-D object acquired from different distances. The target object of low reflectance on a lighter background surface is illuminated by a primary light source that generates a visible illumination gradient. *Top* (*left* to *right*): Input intensity image of the object at a larger viewing distance and a profile section (*left pair*); the corresponding result of confidence-based filling-in is shown together with a profile section (*right pair*). *Bottom* (*left* to *right*): Corresponding input representations and processing results for the object at a closer viewing distance

series of alternating dark and bright patches (stripes) similar to a rectangular square wave. Such a stimulus is shown in Fig. 12. As discussed in Sect. 2, the temporal dynamics of brightness perception in such a COC arrangement is consistent with a filling-in mechanism. Confidence-based filling-in, at equilibrium, again correctly predicts the appearance of the final brightness square-wave pattern.

6.3 Real world application

In order to demonstrate the functional significance of the newly proposed mechanism, we show the processing results for a camera image of a real object. In order to exclude any possible influences from 3-D effects (e.g., by shadowing or variations in surface orientations), we used a card-board attached to a flat background surface. This intrinsically flat scene was directly illuminated by a point-like light source at a distance of approximately 2 m. This generates a significant intensity gradient in the original intensity image. The target surface has been imaged from distances of about 2 and 1 m.

The results of processing are shown in Fig. 13. This demonstrates that the mechanism of confidence-based filling-in is capable of generating a representation of homogeneous surface properties. The result is independent of the projected region size, thus showing the property of size invariance. Also, the illumination gradient is discounted and the noise is successfully suppressed.

7 Discussion and conclusions

7.1 Results

This article makes three main contributions. First, we provide a unification of filling-in models (that attempt to explain perceptual data) and regularization approaches widely used in computer vision. This investigation provides deeper insights into the computational mechanisms of filling-in processes and the underlying objective function one wishes to compute. In particular, we have identified an energy functional for the computation of a brightness representation that is minimized by the

employment of the filling-in mechanism. Second, the analysis of filling-in in the context of regularization theory helped to identify a potential weakness in the standard approach, namely the variable bowing of the representation of homogeneous surface properties. Third, the results of this analysis guided the proposal of *confidence-based* filling-in. The filling-in systems are fed by sparsely located ON and OFF contrast signals near the boundaries of regions whose perceptual brightness quality will be generated. Reliable local nonzero *confidence* measures are thus located in the vicinity of object/region boundaries. Since the stages of oriented contrast detection will also be fed by initial contrast responses, there exists a spatial coincidence of localized contrast data (the generating input for filling-in), significant responses from oriented contrast cells, and high confidence values to guide the filling-in process. This led us to propose that a spatial distribution of transformed contrast responses may act to gate the ON/OFF signals, thus emphasizing the data contribution for the filling-in process.

It is interesting to note that Paradiso and Nakayama (1991) suggested that filling-in processes could be related to the visual system's solution of the ill-posed problem of brightness reconstruction (Poggio et al. 1985). The present paper shows that diffusive filling-in as proposed by Cohen and Grossberg (1984), as well as other variants such as confidence-based filling-in, in fact provide a regularized solution to the problem of brightness reconstruction. However, the precise mechanisms by which the visual system implements such computations remains an open question (see the discussion associated with Pessoa and Neumann 1998; Pessoa et al. 1998).

7.2 Related work

Filling-in models of brightness perception. Our analysis showed that the filling-in mechanism proposed by Cohen and Grossberg (1984) and Grossberg and Todorović (1988) generates a regularized solution to the problem of creating a representation of surface brightness. The analysis further revealed a potential weakness in its original formulation: missing contrast measures at region interiors are treated as zero-amplitude inputs to the filling-in systems. As a consequence, the final filled-in representation will always be "bowed," a problem that is more or less severe depending on the diffusion parameters and the size of the bounded region. The model of Grossberg and Todorović (1988) employed a residual luminance activation (nonzero DC level) from the initial stage of retinal ON-center/OFF-surround processing. Consequently, an already dense input activity distribution – superimposed (compressed) luminance and contrast measures – feeds into the filling-in system. It has been shown in Neumann (1996) that Grossberg and Todorović's (1988) scheme can be generalized such that ON- and OFF-channels carry localized contrast activations and a parallel segregated luminance channel which encodes a blurred and nonlinearly compressed version of the original luminance signal. The filled-in contrast activation is finally combined with the luminance activity to generate the brightness appearance. Therefore, the visible amount of bending in the brightness signal depends on the relative proportion of the filled-in contrast responses in relation to the luminance activation.

More recent versions of filling-in models such as in Gove et al. (1995) and Pessoa et al. (1995) utilize ON and OFF-contrast channels that feed segregated filling-in mechanisms. The latter approach also employs a dense representation of compressed luminance information in a segregated channel. Our analysis predicts that the final brightness representation depends on the precise parameter setting and the size of the surface regions to be filled-in with the corresponding surface property. In the model of Pessoa et al. (1995), as in Grossberg and Todorović's (1988) model, these effects can be "hidden" by the relative contribution of the luminance activation. The steady-state activity distributions in the ON/OFF filling-in systems, however, is predicted to vary with parameter settings and with the sizes of the projected surface regions and may, therefore, be adjusted by proper parameter tuning.

The same insight also applies to the model of filling-in of brightness qualities in segregated depth planes proposed by Grossberg and McLoughlin (1997). Here, spatially localized disparity (or relative depth) information is measured at local luminance discontinuities, or contrasts, in order to find surface boundaries. Boundary segments initiate a first stage of *monocular* filling-in. The equilibrated filling-in activities pass on to a stage of center-surround processing, through which spatial contrasts in the filling-in representation are detected. Matchable ipsi- and contralateral input responses determine the generators of the *binocular* filling-in which, in turn, determines the final layout of surface brightness in depth. Again, the dynamics of the lateral spreading systems is governed by the same basic equation that has been analyzed above. This diffusion process tends to generate spatially inhomogeneous brightnesses for large surface regions to be filled-in. A confidence-based mechanism, on the other hand, will help to generate flat homogeneous layout of surface brightness, independent of the given constant parameter settings.

Processes of surface reconstruction in computer vision. The generation of a brightness representation using filling-in has several aspects in common with computational approaches for surface reconstruction in computer vision. In general the problem is that of generating a rich (or dense) representation given only sparse and possibly conflicting samples of the input pattern. In particular, a critical problem is the integration of sparse data provided by separate modules for shape recovery, such as in stereopsis, motion, and monocular shape-from-X processing (Grimson 1981; Marr 1982; Terzopoulos 1986). Poggio et al. (1985) identified several computational problems in early vision as instances of "inverse optics." These problems are classified as ill-posed since a unique solution does not exist. Surface reconstruction from sparse estimates is also ill-posed.

Initial approaches employed in surface reconstruction (Grimson 1981; see also Marr 1982) utilized stabilizing functionals in 1-D of the form

$$Pv = \int \sum_{m=0}^{q} \rho_m \left(\frac{\mathrm{d}^m}{\mathrm{d}x^m} v \right)^2 \mathrm{d}x$$

with constant coefficients (Tikhonov and Arsenin 1977). A regularized solution v minimizes the stabilizing functional such that it closely approximates the estimated data and is smooth in-between. The problem with constant coefficients ρ_m is that sharply localized discontinuities in the reconstructed function v tend to be smoothed out. Terzopoulos (1986) presented more elaborate computational approaches for surface reconstruction that involve discontinuities. He utilized controlled-continuity stabilizers to minimize the smoothness functionals. These spatially-variant functions are real-valued weighting functions $\rho_r(x)$ with a range of magnitudes in the interval $[0, 1]$. At depth discontinuities, $\rho_r(x) = 0$ should be achieved to break the continuity of the reconstructed surface (Terzopoulos 1986). At the same time, approaches to surface reconstruction have utilized confidence values at discrete locations to justify the contribution of data terms (Szeliski 1990). The measure of data compatibility uses a weighted Euclidean norm of the difference between the estimated data and the reconstructed function. A proper choice for the confidence measure is $\kappa_i = (1/\sigma_i^2) \cdot \delta_i$, available at sparsely distributed (discrete) image locations sampled by δ_i (the Kronecker delta), with σ_i^2 denoting the variances of signal measurements (Terzopoulos 1986; Szeliski 1990). Thus the confidence is increased for reliable measurements with only minor noise. These authors suggest that the confidence value should approach zero at locations where there is no input measurement available from previous processing stages.

We have identified the filling-in mechanism as one that corresponds to a diffusion equation that minimizes a first-order variational problem. The steady-state solution of a filled-in brightness representation corresponds to a smoothed approximation of estimated input data by a physical membrane with tension. Whereas Terzopoulos relied on the generation of a (usually incomplete and error prone) discontinuity map by some external processes, we discussed, in the domain of brightness, the close spatial relationship between contrast detection (by model complex cells) and center-surround filtering whose signals are used for filling-in (see Fig. 7).

The lateral spreading of activity in filling-in also shares some similarities with the conception of the recovery of *intrinsic images* (Barrow and Tenenbaum 1978). These intrinsic images are defined as 2-D arrays of scene properties such as depth, surface orientation, illumination, and reflectance. Each layer is assumed to be in spatial registration with the original luminance distribution. In this framework, a given input (image) is processed with three basic classes of mechanisms. So-called *vertical* processes ensure pointwise consistency of intrinsic properties based on a pre-defined imaging equation (which utilizes a simple Lambertian reflectance model). *Horizontal* processes ensure the continuity (or smoothness) of the representation within each intrinsic image layer. Finally, *edge-marking* mechanisms take an edge map of detected contrasts as input and manipulate these initial estimates by allowing the generation of new edge elements – which in turn break the continuity of the otherwise smooth intrinsic attribute(s) – or delete initial edge estimates. Filling-in relates to the category of horizontal processes in their nature of laterally propagating localized measurements of surface qualities to neighboring sites within a layered representation. Although the intrinsic image model was never formally specified, its original conception ensues from computations in a pointwise fashion as the imaging condition must be achieved at any spatial position. Therefore, by definition, information is not filled-in to otherwise empty spaces, which is the key aspect in the case studied here.

7.3 Further issues

A fundamental constraint that applies to filling-in models, but not necessarily to computer vision applications in general, comprises the nature of *perceptual data*. Data must be used to assess perceptual models, and as such they should also be used to evaluate perceptual models that make use of regularization theory. For instance, the minimization of first-order information dictated by membrane regularization needs to meet the data on the perception of brightness. We, therefore, propose the use of regularization theory as a mathematical tool to predict perceptual data for the generation of homogeneous surface qualities. In turn, data derived from psychophysical experiments should be used to test the model against stimuli, thus verifying or invalidating the suggested mechanisms involved in the generation of surface appearance. We emphasized in our description that the results of standard filling-in tend to "bend" generating concavities in a signal representation that attempted to be flat. This tendency can be circumvented by a mechanism of confidence-based filling-in. It remains, however, a topic of further research to investigate the perceptual representation and whether or not it contains similar "bendings". For example, early studies by von Békésy (1968) provide some evidence that brightness slightly erodes as one moves farther away from an edge. However, these studies utilized techniques and stimuli that are not directly compatible with the cases studied here. Also, that author's evaluation techniques do not allow the drawing of specific psychophysical references that could be compared to any of our computational predictions.

Our definition of *confidence-based* filling-in is mainly motivated by the theoretical argument of achieving a representation of dense brightness (or lightness) quantities that are measured at region boundaries. Whether this computational principle to achieve this invariance is also adopted by the visual system remains a topic of further detailed investigation. In any case, the utilization of

regularization theory has outlined a theoretical framework for the study of filling-in mechanisms to explain processes of perceptual completion. Finally, given the growing evidence for neural filling-in (see Pessoa and Neumann 1998; Pessoa et al. 1998), the elucidation of the underlying brain mechanisms will greatly help to bridge the gap between the perceptual and neural domains.

Acknowledgements. We are grateful to Christoph Schnörr and Joachim Weickert for their valuable suggestions that helped strengthen the theoretical arguments developed in this paper. Gregory Baratoff and the anonymous reviewers provided insightful comments on earlier versions of the manuscript. This research is supported by the German-Brazilian Academic Research Collaboration Program (DAAD-CAPES, Probral 1997, Project Nos. 70712 and 054/97). The research work of H.N. and T.H. is partly supported by the German Science Foundation (DFG, SFB-527). L.P. is also supported by the CNPq/Brazil grant 520419/96-0.

Appendix: Center-surround preprocessing

The raw input luminance distribution L is spatially processed, or filtered, by antagonistic *center-surround mechanisms* that resemble the opponent interaction of retinal ganglion cells. These mechanisms feed two parallel pathways selective to opposite contrast polarities, namely ON and OFF. The equilibrated responses are generated by

$$x_i^{\pm} = (A + \mathrm{net}_i^{+} + \mathrm{net}_i^{-})^{-1}\Big((B + C)\big[\mathrm{net}_i^{\pm} - \mathrm{net}_i^{\mp}\big]^{+} + (B - C)\big(\mathrm{net}_i^{\pm} + \mathrm{net}_i^{\mp}\big)\Big)\ , \qquad \text{(A1)}$$

where $[x]^{+} = \max(x, 0)$ (rectification) and $\mathrm{net}^{\pm}$ denote low-pass filtered versions of the input luminance L utilizing a small space constant for the center and a larger space constant for the surround; i is a spatial position index.

The x-responses undergo an additional stage of competitive interaction generating DC-level-free contrast responses to feed the *boundary system*. Activations of ON and OFF contrast channels are

$$c_i^{+} = \big[x_i^{+} - x_i^{-}\big]^{+}, \quad \text{and} \quad c_i^{-} = \big[x_i^{-} - x_i^{+}\big]^{+}\ . \qquad \text{(A2)}$$

A detailed description and analysis is presented in Neumann (1996).

References

Anstis SM, Howard IP (1978) A Craik-O'Brien-Cornsweet illusion for visual depth. Vision Res 18: 213–217

Arrington KF (1994) The temporal dynamics of brightness filling-in. Vision Res 34: 3371–3387

Arrington KF (1996) Directional filling-in. Neural Comput 8: 300–318

Barrow HG, Tenenbaum JM (1978) Recovering intrinsic scene characteristics from images. In: Hanson AR, Riseman EM (eds) Computer vision systems. Springer, Berlin Heidelberg New York, pp 3–26

Baumeister J (1987) Stable solution of inverse problems. Vieweg, Braunschweig

Beck J (1972) Surface color perception. Cornell University Press, Ithaca, N.Y.

Békésy G von (1968) Mach- and Hering-type lateral inhibition in vision. Vision Res 8: 1483–1499

Bertero M, Poggio T, Torre V (1988) Ill-posed problems in early vision. Proc IEEE 76: 869–889

Blakeslee B, McCourt ME (1997) Similar mechanisms underlie simultaneous brightness contrast and grating induction. Vision Res 37: 2849–2869

Blakeslee B, McCourt ME (1999) A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. Vision Res 39: 4361–4377

Bressan P, Mingolla E, Spillmann L, Watanabe T (1997) Neon color spreading: a review. Perception 26: 1353–1366

Brookes A, Stevens KA (1989) The analogy between stereo depth and brightness. Perception 18: 601–614

Caputo G (1998) Texture brightness filling-in. Vision Res 38(6): 841–851

Cohen M, Grossberg S (1984) Neural dynamics of brightness perception: features, boundaries, diffusion, and resonance. Percept Psychophys 36: 428–456

Cornsweet TN (1970) Visual perception. Academic, New York

Davey MP, Maddess T, Srinivasan MV (1998) The spatiotemporal properties of the Craik-O'Brien-Cornsweet effect are consistent with 'filling-in'. Vision Res 38: 2037–2046

De Valois RL, Webster MA, De Valois KK, Lingelbach B (1986) Temporal properties of brightness and color induction. Vision Res 26: 887–897

Elder RL, Zucker SW (1998) Evidence for boundary specific grouping. Vision Res 38: 143–152

Fiorentini A, Baumgartner G, Magnussen S, Schiller PH, Thomas JP (1990) The perception of brightness and darkness – relations to neuronal receptive fields. In: Spillmann L, Werner JS (eds) Visual perception – the neurophysiological foundations. Academic, San Diego, pp 129–161

Gelfand IM, Fomin SV (1963) Calculus of variations. Prentice-Hall, Englewood Cliffs

Gerrits H, Vendrik A (1970) Simultaneous contrast, filling-in process and information processing in man's visual system. Exp Brain Res 11: 411–430

Gove A, Grossberg S, Mingolla E (1995) Brightness perception, illusory contours, and corticogeniculate feedback. Vis Neurosci 12: 1027–1052

Grimson WEL (1981) From images to surfaces: a computational study of the human visual system. MIT Press, Cambridge, Mass

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

Grossberg S, McLoughlin N (1997) Cortical dynamics of three-dimensional surface perception: binocular and half-occluded scenic images. Neural Netw 10: 1583–1605

Grossberg S, Mingolla E (1985) Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. Psychol Rev 92: 173–211

Grossberg S, Todorović D (1988) Neural dynamics of 1-D and 2-D brightness perception: a unified model of classical and recent phenomena. Percept Psychophys 43: 723–742

Hamada J (1984) A multistage model for border contrast. Biol Cybern 51: 65–70

Heeger D (1992) Normalization of cell responses in cat striate cortex. Visual Neurosci 9: 184–197

Howard IP, Rogers BJ (1995) Binocular vision and stereopsis. Oxford University Press, Oxford

Hubel D, Wiesel T (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (Lond) 160: 106–154

Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol (Lond) 195: 215–243

Julesz B (1971) Foundations of cyclopean perception. University of Chicago Press, Chicago

Jung R (1973) Visual perception and neurophysiology. In: Jung J (ed) Central processing of visual information. (Handbook of Sensory Physiology, vol VII/3) Springer, Berlin Heidelberg New York, pp 1–152

Krauskopf J (1963) Effect of retinal image stabilization on the appearance of heterochromatic targets. J Opt Soc Am A 53: 741–744

Marr D (1982) Vision. Freeman, San Francisco

Mumford D (1994) Neuronal architectures for pattern-theoretic problems. In: Koch C, Davis JL (eds) Large-scale neuronal theories of the brain. MIT Press, Cambridge, Mass

Nakayama K, Shimojo S (1992) Experiencing and perceiving visual surfaces. Science 257: 1357–1363

Neumann H (1996) Mechanisms of neural architecture for visual contrast and brightness perception. Neural Netw 9: 921–936

Neumann H, Pessoa L, Hansen T (1999) Interaction of on and off pathways for visual contrast measurement. Biol Cybern 81: 515–532

Paradiso M, Nakayama K (1991) Brightness perception and filling-in. Vision Res 31: 1221–1236

Pessoa L (1996a) Mach band attenuation by adjacent stimuli: experiment and filling-in simulations. Perception 25: 425–442

Pessoa L (1996b) Mach bands: how many models are possible? Recent experimental findings and modeling attempts. Vision Res 36: 3205–3227

Pessoa L, Neumann H (1998) Why does the brain fill-in? Trends Cogn Sci 2: 422–424

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

Pessoa L, Thompson E, 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: 723–802

Poggio T, Torre V, Koch C (1985) Computational vision and regularization theory. Nature 317: 314–319

Rogers-Ramachandran DC, Ramachandran VS (1998) Psychophysical evidence for boundary and surface systems in human vision. Vision Res 38: 71–77

Rossi AF, Paradiso MA (1996) Temporal limits of brightness induction and mechanisms of brightness perception. Vision Res 36: 1391–1398

Schnörr C, Sprengel R (1994) A nonlinear regularization approach to early vision. Biol Cyber 72: 141–149

Szeliski R (1990) Bayesian modeling of uncertainty in low-level-vision. Int J Comput Vis 5: 271–301

Stoper A, Mansfield J (1978) Metacontrast and paracontrast suppression of a contourless area. Vision Res 18: 1669–1674

Terzopoulos D (1986) Regularization of inverse visual problems involving discontinuities. IEEE Trans Pattern Anal Mach Intell 8: 413–424

Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems, Winston, Washington, D.C

Todorović D (1987) The Craik-O'Brien-Cornsweet effect: new varieties and their theoretical implications. Percept Psychophys 42: 52–96

Weickert J (1997) A review of nonlinear diffusion filtering. In: ter Haar Romeny B, Florack L, Koenderink J, Viergever M (eds) Scale-space theory in computer vision. Springer, Berlin Heidelberg New York

Yarbus A (1967) Eye movements and vision. Plenum, New York