# A multi-resolution filling-in model for brightness perception


Wolfgang Sepp[†] [¶] , Heiko Neumann[‡] [¶]

Universität Ulm, Fakultät für Informatik
Abteilung Neuroinformatik
D-89069 Ulm, Germany



## Abstract

*We present a multi-scale neural filling-in model for brightness reconstruction of initial DoG filtered images. In contrast to the classical single-scale filling-in models it no longer requires an additional (luminance) signal to restore arbitrary images. Moreover, it substantially reduces the computational cost of the reconstruction process. We present a multi-layered hierarchical neural network comparable to a Laplacian pyramid in which contrast measures are filled-in in dedicated frequency domains. We show in simulations how this model operates on synthetic as well as on real-world images.*


## 1 Introduction

There is increasingly more evidence available that the brain may use active filling-in mechanisms for reconstruction of perceptual surface quantities such as, e.g., brightness [6]. Such mechanisms have been formalized as geometry-driven diffusion processes [2]. By investigating filling-in of contrast measurements we may be able to explain humans ability to perceive brightness irrespective of the lighting conditions.

The filling-in models presented so far are confined to lateral spreading mechanisms at a single spatial scale. They neglect structure information given at lower spatial resolutions. We propose a neurally plausible multi-resolution filling-in model which takes into account structure information given at different scales and thus may help to explain brightness/lightness phenomena which depend on global arrangements of objects.


[†]Wolfgang Sepp is now with the Institute of Robotics and System Dynamics, German Aerospace Research Establishment (DLR), Wessling (Germany); wolfgang.sepp@dlr.de

[‡]hneumann@neuro.informatik.uni-ulm.de



[¶]The research work was supported in part by the Collaborative Research Center Sfb 527 at the University of Ulm


## 2 Model

We present a multi-resolution extension of confidence-based filling-in [3] (see Sec. 2.1). The scheme consists of a hierarchy of grid layers which differ in the size of their meshes. The layers have cells placed at each node which are laterally connected to the next neighbors on the grid. The hierarchy of differently sized grids resembles a Gaussian/Laplacian pyramid structure similar to the one introduced by Burt & Adelson [1]. In addition, in our model a radial basis function is associated to each node whose size varies with the corresponding scale (the lower the spatial resolution the broader the basis function). This hyper basis function network [7] aims to reconstruct brightness from initial contrast measurements. This is accomplished through filling-in processes at each resolution scale.

In the following sections we present the hierarchical model in detail (see also Fig. 1). At first, the single-scale confidence based filling-in model of Neumann and Pessoa is outlined (see Sec. 2.1). Since the shape of the input signal and the constraints are crucial for the reconstruction process, we briefly sketch the preprocessing stage in section 2.2. The filling-in processes compete with each other to ensure that the filling-in processes at higher resolution layers do not interfere with the signals reconstructed at lower resolution scales. The neural circuit described

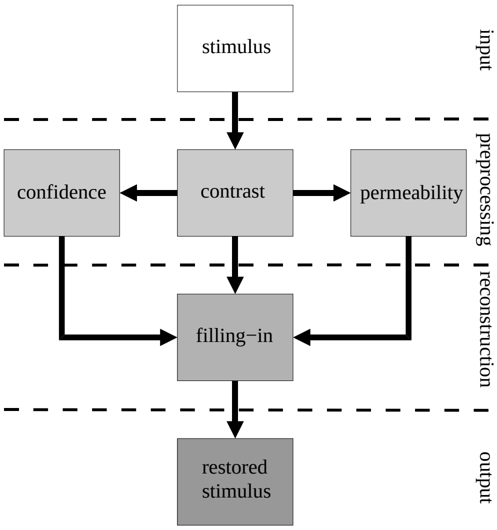


Figure 1: Module architecture.

in section 2.3 determines the winner of the competition. All these computational elements are used to explain the multi-scale filling-in model (see Sec. 2.4).

## 2.1 Confidence-based filling-in

Neumann and Pessoa proposed in [3] a single-scale, confidence-based filling-in model which is able to correctly reconstruct homogeneous regions irrespective of the size of the closed region. The filling-in process is defined for two separate channels, namely "lightness" (+, ON) and "darkness" (−, OFF), respectively. Both channels are

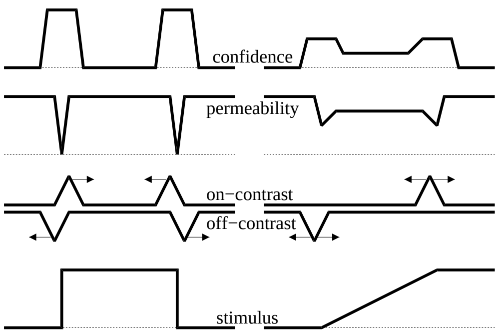


Figure 2: Contrast, permeability and confidence signal for luminance steps and ramps.

determined by an inhomogeneous diffusion equation which reads in the continuous form

$$\frac{\partial}{\partial t} v^{\pm} = \nabla\cdot \left(\rho\, \nabla v^{\pm}\right) + \kappa \left(c^{\pm} - K v^{\pm}\right) \;, \tag{1}$$

where $v$ is the filling-in signal, $c$ the input signal (i.e. contrast), $\rho$ the space variant permeability function, $\kappa$ the confidence measure and $K$ a constant decay rate. Here $v$, $c$, $\rho$ and $\kappa$ are functions defined over space and time, $\nabla\cdot$ denotes the div operator and $\nabla$ the gradient.

The above diffusion equation consists of two terms that are comparable to those used in regularization theory: a model term which accounts for smoothness and a data term which is responsible for the proximity to the data. The smoothness term is controlled by the space variant signal $\rho$ similar to the "controlled continuity stabilizer" defined by Terzopoulos [8]. The data term is gained by the fuzzy signal $\kappa$ which determines where initial data measurement are available and which, in turn, should be approximated in the filling-in reconstruction.

Figure 2 sketches the constraint signals $\rho$ and $\kappa$ and the feature signal $c^{\pm}$ involved in one-dimensional diffusion processes. The

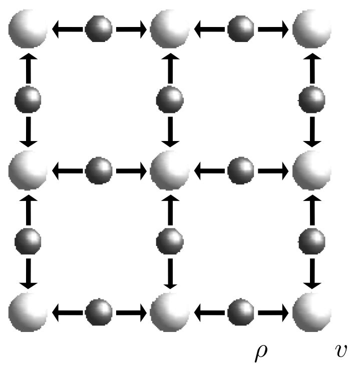


Figure 3: Diffusion on a regular grid: Each connection between filling-in cells $v$ is controlled by the permeability $\rho$.

filling-in process on a two-dimensional regular grid is described by a discretization of eqn. 1. Figure 3 shows the corresponding neural representation.

A problem of the single-scale filling-in model is that brightness cannot be accumulated over nested luminance steps because spreading of activity is blocked at sharp boundaries. We present in the following a hierarchical extension of the confidence-based filling-in which overcomes this problem and moreover accelerates significantly the reconstruction process.

## 2.2 Preprocessing stage

The preprocessing module computes the contrast and the constraint signals for filling-in processes at different resolution scales. These signals are determined separately for all spatial resolutions of the original image (see Fig. 4). Thus, the first step

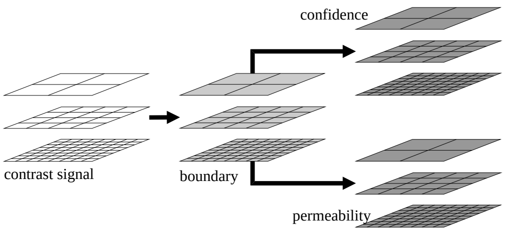


Figure 4: Sketch of the preprocessing model for the contrast, permeability and the confidence signal at different grid-layers.

consists of down-sampling the initial stimulus $I_0$. The corresponding equation reads for a Gaussian pyramid of height $N$:

$$I_i = (I_{i-1} * B_i) \cdot \text{Ш}_{2^i} \ , \quad \forall\, 0 < i < N \quad (2)$$

where $B$ denotes a proper low-pass filter whose variance depends on the scale $i$ (see Fig. 5), $\text{Ш}_{2^i}$ is a down-sampling operator with resolution dependent step-size, and $*$ denotes the spatial convolution operator. Each signal $I_i$ is filtered with a center-surround mechanism analogous to receptive fields of retinal ganglion cells. Two signals

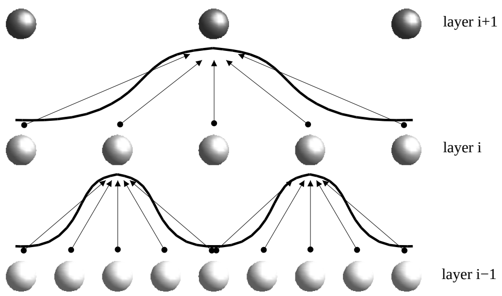


Figure 5: Sub-sampling with Gaussian radial basis functions.

are computed: one that responds for light center and dark surround, the other for dark center and light surround. A simplified computational scheme is given by

$$c_i^+ = [I_i * \mathrm{DoG}]^+ \ , \quad (3)$$
$$c_i^- = [I_i * (-\,\mathrm{DoG})]^+ \ , \quad (4)$$

where $[\cdot]^+$ denotes a half-wave rectification operator. The next step is to determine a boundary signal for each resolution scale. Boundary signals could be generated in different ways. A simple approach may directly utilize the raw contrast estimates generated by elongated filters (simple and complex cells, e.g. [4]). Proper signals are generated by the recurrent network used in [5] which takes into account context effects and better scenic groupings for boundary representation.

Given the boundary signal $w$ at scale $i$, the confidence signal $\kappa$ can be computed by smoothing the boundaries with a Gaussian

$$\kappa_i = w_i * \mathrm{Gauss} \ . \quad (5)$$

The diffusivity signal $\rho$ can be determined by inverting $w$ according to

$$\rho_i = \frac{1}{1 + A w_i} \ , \quad (6)$$

where $A$ is a proper scaling factor.

## 2.3 Competence controlling circuit

A small artificial neural circuit is placed between the filling-in layers. It ensures that coarse signal components are filled-in at higher layers and that the details of the final brightness percept are restored on a finer grid utilizing narrow basis functions.

The input to the circuit are the signals of the filling-in layers of grid $i+1$ expanded to the next lower layer $i$:

$$\hat{v}_i^{\pm} = \left(v_{i+1}^{\pm} * B_{i+1}\right) \cdot \text{Ш}_{2^i} \ , \ \forall\, 0 \le i < N-1 \quad (7)$$

where $B_i$ is the basis function already used for down-sampling (see eqn. 2). We argue that responses in ON- and OFF-channel which spatially overlap indicate that there is a degree of uncertainty in the signal that is reconstructed in this area. Whenever one channel is inactive, no conflicts may have occurred.

In the first stage of the artificial neural network (see Fig. 6) two signals are determined. The one denominated as $n^o$ reflects the similarity between ON- and OFF-channel responses of a filling-in layer. The other signal $n^d$ is an energy measure of these

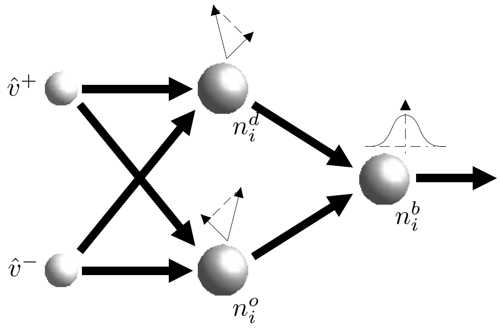


Figure 6: Circuit, which identifies those regions which have to be refined on subsequent resolution scales.

input signals. The activation rule for these signals is:

$$n_i^o = \frac{\hat{v}_i^+ - \hat{v}_i^-}{D + \hat{v}_i^+ + \hat{v}_i^-} , \quad (8)$$

$$n_i^d = \frac{\hat{v}_i^+ + \hat{v}_i^-}{D + \hat{v}_i^+ + \hat{v}_i^-} , \quad (9)$$

where $\hat{v}^+$ and $\hat{v}^-$ denote the up-sampled ON- and OFF-filling-in layers, $D$ is a constant. Finally, the responses are combined according to:

$$n_i^b = n_i^d \cdot \exp\left(-\frac{1}{2}\left(\frac{n_i^o}{\sigma_b}\right)^2\right) , \quad (10)$$

where $\sigma_b$ is constant for all layers and each spatial location. The resulting signal $n^b$ is a fuzzy signal which rates the similarity between the input signals $\hat{v}^+$ and $\hat{v}^-$. The circuit is invariant with respect to absolute activity levels in that it responds to weak brightness signals equally well as to strong signals. Zero energy signals are not considered by the circuit due to the factor $n^d$. This behavior meets exactly our requirements because the purpose of the circuit is to mark regions of overlapping ON- and OFF-responses.

## 2.4 Multi-Scale filling-in

Given the contrast and constraint signals determined by the preprocessing stage, filling-in processes are initiated on all resolution scales in parallel. The diffusion equation is the same for all scales and reads

$$\frac{\partial}{\partial t} v_i^{\pm} = \nabla \cdot \left(\rho_i^* \nabla v_i^{\pm}\right) + \kappa_i^* \left(c_i^{\pm} - K v_i^{\pm}\right) , \quad (11)$$

where $\rho_i^*$ and $\kappa^*$ denote the hierarchically modified permeability and confidence signals, respectively. They are computed by linear interpolation between the constraint values which suppress diffusion and the initial (unchanged) constraint signals. The equation reads for all layers, except the top layer:

$$\rho_i^* = \rho_i \cdot n_i^b , \quad (12)$$

$$\kappa_i^* = 1 + (\kappa_i - 1) \cdot n_i^b . \quad (13)$$

The activation of the constraint signals at the top layer is identical to the distribution computed at the preprocessing stage.

Whenever the fuzzy signal $n_i^b$ evaluates to 1 then the constraint signals of the preprocessing stage are passed unchanged to the filling-in processes. An inactive signal $n_i^b$ sets the permeability to 0 and the confidence measure to 1. In this case the diffusion process is locally suppressed and an already active filling-in signal decays to the level given by the input signal $c_i$ (see eqn. 1).

When all filling-in processes have reached steady-state then the overall filling-in response is generated. Radial basis functions are positioned on each node of the different layers and activated by the filling-in signal that has arrived at the particular grid point. The signal reconstruction assumes a zero-level Eigengrau as a reference. Brightness *increments* are generated by additive contribution from ON-filling-in activities, whereas brightness *decrements* are generated by subtraction of OFF-filling-in activities. In all, we have

$$v = \sum_{i=0}^{N} v_i^+ * B_i - v_i^- * B_i , \quad (14)$$

where $B_i$ is the basis function used in eqn. 2.

# 3 Simulation Results

In the following, simulations results are presented to demonstrate the functionality and capability of our hierarchical filling-in model. We utilize a simplified scheme for boundary computation that utilizes simple and complex cell computations at multiple scales [4].

**Ellipse:** This simple stimulus demonstrates how regions of uniform brightness are restored with hierarchical filling-in employing three resolution scales (see Fig. 7). Plain regions are restored at the lowest resolution scale, while sharp boundaries are refined at higher resolution layers (Fig. 8).

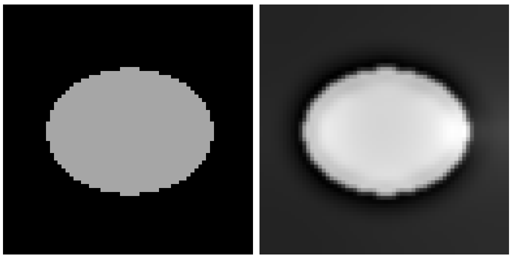

Figure 7: Left: Original stimulus of size $64 \times 64$. Right: Reconstructed stimulus

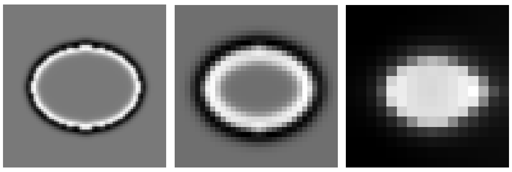

Figure 8: From left to right: Activation of filling-in layers at steady-state from the finest grid to the broadest.

**Brightness pyramid:** Single-scale filling-in is not able to reconstruct nested brightness steps but hierarchical filling-in does (see Fig. 9 and 10).

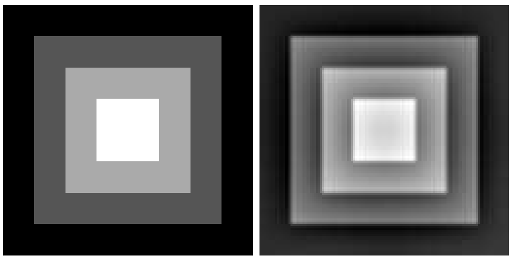

Figure 9: Left: Original stimulus of size $128 \times 128$. Right: Reconstructed stimulus

**The "Lena" image:** This simulation demonstrates that our model can be employed on real-world images, too (see Fig. 11).

**Performance measurement:** Multi-scale filling-in reconstructs images significantly faster than single-scale filling-in. Confidence-based filling-in requires 2195 iterations at the average in both channels to fill the ellipse of Fig. 7 which corresponds to 57 seconds on a Sun-Ultra-1 Workstation

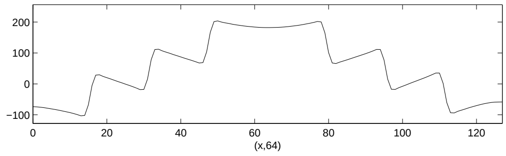


Figure 10: Cross section of the reconstructed brightness pyramid.

Figure 11: Top: Original stimulus of size $256 \times 256$. Bottom: Reconstructed stimulus

(167 MHz UltraSparc CPU). The hierarchical filling-in model restores the same stimulus already in 2.8 seconds. Table 1 shows the number of iterations that were needed in the individual layers.

| | layer 0 | layer 1 | layer 2 |
|---|---|---|---|
| iterations | 59 | 60 | 457 |
| seconds | 1.74 | 0.36 | 0.56 |

Table 1: Time measurement for the reconstruction process at single resolution scales.

## 4 Discussion

The purpose of filling-in models is to decipher the way how humans perceive brightness. This may help to build image processing systems which are able to reliably recognize objects under different lighting conditions.

First, we observe that although man perceives absolute brightness levels our retinal ganglion cells are sensitive to *contrasts* in luminance of images and not to *absolute* luminance. Therefore, there has to be a mechanism which reconstructs the absolute levels from initial contrast measurements.

The model we present is fed by a sparse representation of the original signal, that is the contrast signal which resembles the signal supplied by retinal ganglion cells. The neural plausibility of our model goes further in that the basis functions used for expanding and collapsing the Gaussian pyramid can be interpreted as receptive-field structures of the human brain.

Our hierarchical filling-in mechanism restores brightness by focusing on those parts of the frequency spectrum which got lost on the initial contrast measurement. Low frequencies are more affected than higher frequencies. The neural circuit which manipulates the filling-in constraints ensures that the filling-in processes are confined at all resolution scales to the corresponding frequency domains. Finally, the equilibrated filling-in activity distribution of all layers resembles a Laplacian pyramid.

Hierarchical filling-in differs from numerical multi-grid techniques used for fast approximation of differential equations. It does not reproduce the effects of single-scale filling-in but it extends the confidence-based filling-in model to more resolution scales. The architecture belongs to the class of hyper basis function networks [7] which are well suited for surface reconstruction.

# 5 Conclusion

We developed a hierarchical filling-in model which is able to correctly reconstruct synthetic as well as natural input stimuli from a sparse, DoG filtered signal. It is up to 20 times faster than a single-scale confidence-based filling-in process what makes the model interesting for real-time application.

Although, we did not rely on high quality boundary signals in our simulations, the hierarchical filling-in mechanism was able to reliably reconstruct the original signal.

Multi-scale filling-in is neurally plausible and may help to explain several lightness/brightness effects.

# References

[1] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. on Comm.*, 31(4):532–540, 1983.

[2] M. Cohen and S. Grossberg. Neural dynamics of brightness perception: Features, boundaries, diffusion and resonance. *Perception & Psychophysics*, 36:428–456, 1984.

[3] H. Neumann and L. Pessoa. Visual filling-in and surface property reconstruction. Technical Report 98-04, Univ. Ulm, Fak. f. Informatik, Februar 1998. submitted.

[4] H. Neumann, L. Pessoa, and T. Hansen. Interaction of On and Off pathways for visual contrast measurement. *Biol. Cybernetics*, 1999. accepted for publication.

[5] H. Neumann and W. Sepp. Recurrent V1-V2 interaction in early visual boundary processing. *Biol. Cybernetics*, 1999. in print.

[6] L. Pessoa and H. Neumann. Why does the brain fill in? *Trends in Cognitive Sciences*, 2(11):422–426, 1998.

[7] T. Poggio and F. Girosi. A theory of networks for approximation and learning. Technical Report 1140, M.I.T., July 1989.

[8] D. Terzopoulos. Regularization of inverse problems involving discontinuities. *IEEE Trans. on PAMI*, 8(4):413–424, 1986.