ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

Recurrent V1 – V2 Interaction for Early Visual Information Processing

Heiko Neumann
Abteilung Neuroinformatik
Universität Ulm
89069 Ulm, Germany

Wolfgang Sepp
Abteilung Neuroinformatik
Universität Ulm
89069 Ulm, Germany

Abstract

A majority of cortical areas are connected via feedforward and feedback fiber projections. The computational role of the descending feedback pathways at different processing stages remains largely unknown. We suggest a new computational model in which normalized activities of orientation selective contrast cells are fed forward to the next higher processing stage. The arrangement of input activation is matched against local patterns of curvature shape to generate activities which are subsequently fed back to the previous stage. Initial measurements that are consistent with the top-down generated context-dependent responses are locally enhanced. In all, we present a computational theory for recurrent processing in visual cortex in which the significance of measurements is evaluated on the basis of priors that are represented as contour code patterns. The model handles a variety of perceptual phenomena, such as e.g. bar texture stimuli, illusory contours, and grouping of fragmented shape outline.

1 Motivation

The brain is steadily confronted with a massive information flow that arrives via several sensory channels. In vision, pattern arrangements that signal coherent surface quantities must somehow reliably be detected and grouped into significant items. Such a grouping enables the segregation of figural components from cluttered background as well as the adaptive focussing of processing capacities while suppressing unimportant parts of the scene (e.g. Grossberg, 1980). A characteristic feature of the cortical architecture is that the majority of (visual) cortical areas are linked bidirectionally by feedforward and feedback fiber projections. So far, the precise computational role of the descending feedback pathways at different stages of processing remains largely unknown. Recent empirical evidence supports the view that top-down projections primarily serve as a modulation mechanism to control the responsiveness of cells in primary visual cortex (Lamme, 1995; Salin & Bullier, 1995).

Based on these findings our model proposes abstract computational principles of feedforward and feedback interaction between a pair of cortical areas. Along the *bottom-up* stream localized features are detected and subsequently integrated by matching them against coarse model shape outline. In the *top-down* stream activations are fed back to selectively enhance elements of salient contour arrangements via a gain control mecha-

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

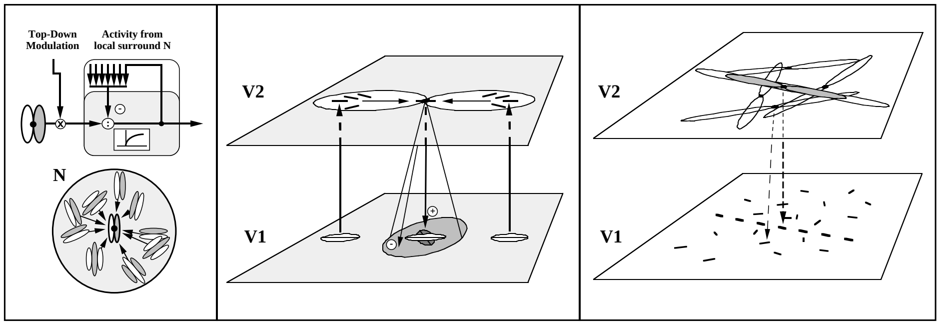

Figure 1: Model components. Left: Feeding input is generated by responses of localized oriented receptive fields which are normalized by a shunting competition in a local neighborhood $\mathcal{N}$ (compare with Heeger *et al.*, 1996); Center: Model V2 cells resembling 'curvature templates' integrate V1 activations from elongated branches. Feedback enhances those activated V1 contrast cells of matching orientation; Right: Recurrent network interaction selectively enhances colinear arrangements of items to pop out.

nism. As such, recurrent interaction provides a key mechanism for the segmentation of surface layout and figure-ground segregation. The model links physiology and psychophysics, incorporating empirical data from both research directions, and provides a common framework for distinct perceptual phenomena.

2 Computational Model

Functionality and Computational Mechanisms. We suggest that a variety of empirical findings can be explained on the basis of a set of underlying computational mechanisms. In the ascending processing stream local contrast orientation is initially measured by cells with oriented receptive fields (RF), such as cortical simple and complex cells. Thus, for a pair of bidirectionally connected cortical areas (V1 and V2 in our case) the "lower" area serves as a stage of feature measurement and signal detection. This activity from local measurement is normalized by a mechanism of divisive inhibition (see Heeger *et al.* (1996) for a summary of findings and a model). Figure 1 (left) sketches the basic mechanisms involved. The resulting activations are fed forward to the "higher" area where they are integrated by oriented cells utilizing long-range RF (Grossberg & Mingolla, 1985). Due to their increased RF size such an integration along an oriented path enables to bridge gaps including those corresponding to perceived illusory contours. The strength of contribution to the integration is based on stimulus features, such as spatial position and local orientation. Arrangements of items support an individual contrast element at a target location in a graded fashion. The spatial weights of V2 cell RF represent "models" of visual entities that frequently occur. The support from activities in a space-orientation neighborhood thus encodes the probability of occurrence of stimulus shape segments. The "higher" area thus locally matches expected "model templates" (or priors; see Mumford, 1994) of visual structure against the incoming data carried by the ascending pathway.

The matching process generates an activity pattern in the higher area that is propagated backwards via the descending feedback pathway. By way of the feedback pathway those

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

activities that match position and orientation of V2 contour cells are enhanced, those that do not are inhibited. Thus, the action of feedback provides a confidence measure on the basis of the stimulus-defined context to selectively enhance those signal patterns that are consistent with the model expectations and to inhibit those that do not (Fig. 1, center). A gain control mechanism, that is accompanied by competitive interactions, realizes a "soft gating" mechanism that filters activities corresponding to salient input arrangements while suppressing spurious signals that are inconsistent with the top-down priors or shape templates. Figure 1 (right) illustrates the V2 integration stage and the enhancement of an arrangement of local V1 measurements through feedback activation and local competition.

Description of Mechanisms. We have implemented a version of the above sketched model. All network levels are modeled as consisting of single compartment cells with gradual saturation-type first-order activation dynamics.

The initial measurement stage consists of processing the input luminance stimulus by masks of local contrast sensitivity, such as simple and complex cells. Responses for different orientations are synthesized by interpolation of outputs from two mutually orthogonal filters utilizing a steerability equation. Outputs of pooled opposite contrast directions, $c_{i\varepsilon}$ (with location i and orientation ε), are fed to a sequence of competitive interactions in model *area V1*, the first stage of which combines output of oriented contrast detection with feedback activation generated by V2 'curvature template' cells. This recurrent interaction generates activities $l_{i\varepsilon}^{(1)}$ by

$$\frac{\partial}{\partial t} l_{i\varepsilon}^{(1)} = -\alpha_1 l_{i\varepsilon}^{(1)} + \left(\beta_1 - \gamma_1 l_{i\varepsilon}^{(1)}\right) c_{i\varepsilon} \left(1 + C \left\{h^{(2)} \star \psi^+\right\}_{i\varepsilon}\right) - \zeta_1 l_{i\varepsilon}^{(1)} \left\{h^{(2)} \star \psi^- \star \lambda^-\right\}_{i\varepsilon} . \tag{1}$$

The constants $\alpha_1, \beta_1, \gamma_1$ and ζ_1 define the activity decay, α_1, and the magnitudes of excitatory and inhibitory saturation levels, γ_1/β_1 and ζ_1, respectively. The constant C represents the gain factor of top-down excitatory contour template activation. Positionally invariant weighting of activities is denoted by a convolution ($\star$) utilizing space and orientation kernels, λ and ψ. Excitatory and inhibitory interactions are indicated by '$+$' and '$-$'. Activities $h_{i\varepsilon}^{(2)}$ denote feedback activations along the ascending pathway.

The enhancement via feedback activation is only effective at those positions with non-zero V1 contrast activity. The multiplicative excitation is similar to the linking mechanism proposed by Eckhorn *et al.* (1990). In contrast to approaches, such as the BCS of Grossberg & Mingolla (1985), in our model no activity spreading or completion occurs for locations between inducing elements of a salient perceptual contour arrangement. In the BCS, for example, the action of feedback recurrency acts to *complete* an otherwise fragmented representation of boundary activity. In the model presented here, activity in the higher area is used to *assess* the validity and significance of measurements at the lower area. Thus, here the computational competence of feedforward and feedback interaction is the context-sensitive selection and enhancement of early measurements.

The top-down gated activities subsequently undergo a second stage of shunting ON-center/OFF-surround competition between activities in a space-orientation neighborhood to contrast enhance and normalize activations through divisive inhibition (compare Heeger *et al.*, 1996). In all, both competitive processing stages in model V1 implement a *soft*-gating mechanism: V1 activities that are selectively enhanced by matching V2 contour template activation in turn provide more inhibitory energy in the normalization stage. Thus, salient contrast arrangements will be enhanced while at the same time

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

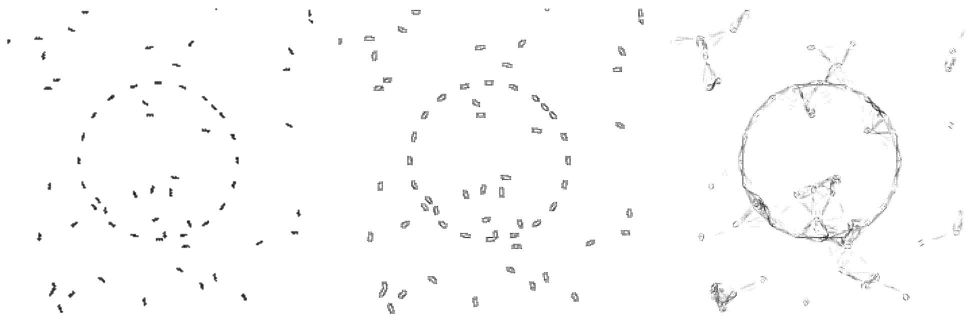

Figure 2: Processing of a noisy fragmented shape outline. Left: Input image; Center: Equilibrated normalized model V1 cell responses; Right: Equilibrated model V2 cell responses utilizing 'curvature templates'. Saturation after four cycles of iteration.

spurious and perceptually irrelevant responses will we suppressed by way of inhibition.

The arrangement of V1 (model stage 2) activities, $l_{i\varepsilon}^{(2)}$, is fed forward to orientation selective cells in model ***area V2***. The effective weighting function of a V2 cell RF can be considered as a (static) filter that represents templates of typical shape outline patterns with varying curvature (Mumford, 1994) to be matched against the structure of input measurement. The matching is realized by utilizing collinearly aligned pairs of oriented one-sided weighting functions (lobes) which sample a segment of the spatial neighborhood. In order to combine matching input from opposite half-spaces a subsequent non-linear accumulation stage integrates the activities from a collinear pair of lobes (Grossberg & Mingolla, 1985; Peterhans & von der Heydt, 1989). Activation of such a cell requires input activation from *both* branches. This is consistent with findings about the non-linearities in V2 contrast cell responses (e.g. Peterhans & von der Heydt, 1989). Our V2 RF model is based on the bipole concept of long-range interaction first suggested by Grossberg & Mingolla (1985). However, in our model the corresponding $h_{i\varepsilon}^{(1)}$-activity is generated by mechanisms of self-inhibition of individual lobes and disinhibition of activation among both sub-field branches, denoted by $l_{i\varepsilon}^{L}$ and $l_{i\varepsilon}^{R}$, respectively (compare Fig. 1, right). This guarantees that the target cell generates a response only when both branches get activated simultaneously. Furthermore, the weighting function consists of ON- and OFF-subfield components. The net effect of a lumped representation of cell response results in a multiplicative, or gating-like, combination of activity from both branches

$$h_{i\varepsilon}^{(1)} \propto l_{i\varepsilon}^{L} l_{i\varepsilon}^{R} \left(\frac{2}{\zeta_3} + l_{i\varepsilon}^{L} + l_{i\varepsilon}^{R} \right), \tag{2}$$

where ζ_3 determines the shape of the compressive non-linear transfer function.

Similar to model V1 the $h_{i\varepsilon}^{(1)}$ activation generated by the stage of long-range contrast integration undergoes a shunting center-surround interaction in the space-orientation domain for contrast enhancement and normalization of activities. The resulting $h_{i\varepsilon}^{(2)}$-activation is fed back via the descending pathway to enhance the activities of initial measurements by V1 oriented contrast cells via non-linear on-center/off-surround interaction (see Eqn. 1).

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

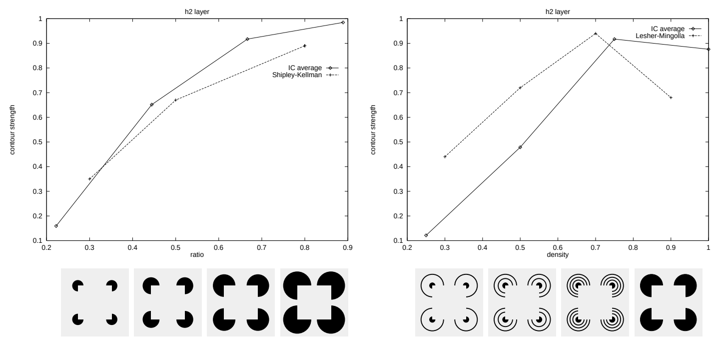

Figure 3: *Predictions for illusory contour strength after grouping (model V2 cell responses, $h^{(2)}$) for Kanizsa (left) and Varin figures (right). See text.*

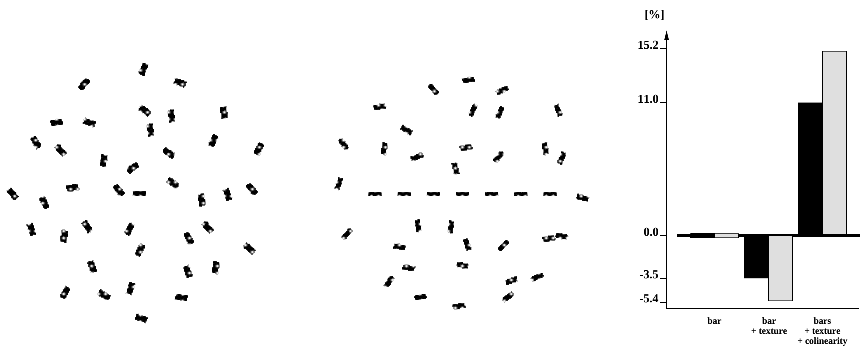

Figure 4: Results for a texture stimulus composed of oriented bars. See text.

3 Simulation Results

Simulation results and model predictions are consistent with a broad range of experimental data. Both local processes and the bidirectional interaction between layers (model areas) achieve adaptive behaviors of V1 cells such as the generation of sharp orientation selectivities of cells, the context-sensitivity of cell tuning, and the enhancement of line endings. Model V2 generates the grouping of fragmented perceptually salient contour segments (see Fig. 2). The initial estimates of local contrast orientation in model area V1 are selectively sharpened. In Fig. 3 we show the strength of model V2 illusory contour responses generated for Kanizsa (left) and Varin figures (right). For the Kanizsa figures the strength is a function of the ratio between inducer radius and total contour length. In the Varin figures strength is a function of the density of evenly spaced circular arcs. In both cases data from psychophysical measurements by Shipley & Kellman (1992) and Lesher & Mingolla (1993) are displayed for comparison. Figure 4 demonstrates the context dependency of responses utilizing texture bar patterns that have been used for physiological investigations by Kapadia *et al.* (1995). Relative responses for a target item have been investigated utilizing stimulus patterns with an isolated bar (not shown), the bar embedded in a texture with random oriented bars, and a texture in which the central bar is supplied by oriented and aligned bars (Fig. 4, left & center). Individual

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 165-170

responses based on local measurements should appear different in varying contexts of visual stimulation (see Mumford, 1994). The computational experiment shows a drop in V1 cell response as a bar item appears as part of a random texture arrangement. The response raises even beyond the level of the reference if the target bar item is flanked by an arrangement of aligned bars with colinear orientation (columns in Fig. 4 (right) show results for different values of the gain control factor $C = 5$ (black), $C = 10$ (grey)).

4 Summary

We have suggested a computational theory for the bidirectional interaction between pairs of areas in the visual pathway. Within such a pair the “lower area” is viewed as a stage of signal measurement whereas the “higher area” evaluates the significance of arrangements of local activity patterns on the basis of local context information. This more global, or coarse-scale, activation pattern is in turn used to selectively enhance those initial measurements that are consistent with the broader context.

Acknowledgements. This research is performed in the Collaborative Research Center SFB-527, which is located at the Univ. of Ulm and funded by the German Sci. Found.

References

Eckhorn, R., Reitboeck, H.J., Arndt, M. & Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comp.*, **2**, 293-307.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psych. Rev.*, **87**, 1-51.

Grossberg, S. & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentation. *Perc. & Psychophys.*, **38**, 141-171.

Heeger, D.J., Simoncelli, E.P. & Movshon, J.A. (1996). Computational models of cortical visual processing. *PNAS*, **93**, 623-627.

Lesher, G.W. & Mingolla, E. (1993). The role of edges and line-ends in illusory contour formation. *Vis. Res.*, **33**, 2253-2270.

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

Lamme, V.A.F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.*, **15**, 1605-1615.

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

Peterhans, E. & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J. of Neurosci.*, **9**, 1749-1763.

Salin, P.-A. & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiol. Rev.*, **75**, 107-154.

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