Robust contour extraction and junction detection by a neural model utilizing recurrent long-range interactions

Thorsten Hansen[1, 2] & Heiko Neumann[2]

[1]Abteilung Allgemeine Psychologie, Justus-Liebig-Universität Gießen
[2]Abteilung Neuroinformatik, Universität Ulm
Thorsten.Hansen@psychol.uni-giessen.de
hneumann@neuro.informatik.uni-ulm.de

Abstract. In the traditional view, early visual processing is basically the extraction of local features by a bank of predefined filters such as Gabor filters. Recent results from physiology and psychophysics however stress the importance of nonlocal, contextual influences and recurrent interactions. We present a model of primary visual cortex V1 for oriented contrast processing utilizing recurrent long-range interactions. The core mechanisms of the model include 1) localized receptive fields, 2) cooperative nonlocal long-range integration, 3) competitive short-range interactions, and 4) interlaminar feedforward and feedback processing. In contrast to other models of visual contour grouping, the present model uses long-range interactions which are confined to colinear and near-colinear elements, in accordance with empirical findings (e.g., Bosking et al., 1997). The model can account for empirical data on contextual facilitation by colinear fragmented contours and contextual suppression by random texture stimuli (Kapadia et al., 1995). With same parameters, we evaluate the competencies of the model for the processing of camera images. Detection performance of the new model is compared to a basic linear feedforward model. When applied to the processing of synthetic and natural images, contours can be more robustly extracted compared to a simple feedforward scheme. Further we show that intrinsically twodimensional features such as corners and junctions can be extracted with high accuracy from the resulting contour information. In particular, we show that localization of generic junction configurations (T, L, X, Y, W, Ψ) is improved. To analyse the junction detection performance, we employ a ROC analysis for a threshold-free evaluation of the different methods. Overall, the model shows how biological principles of recurrent processing and nonlocal colinear integration result in a more robust extraction of elementary image features such as contours and junctions.

1 Introduction

The robust extraction of elementary image features such as contours and junctions are among the first processing steps in most artificial vision systems. Traditional computer vision approaches try to solve each problem separately by local, highly specialized filters. In contrast to these schemes, we present an integrated approach which is based on biological motivations and allows a coherent, robust extraction of contours and junctions within a single architecture.

2 Model

In this section we give a brief overview of the model. For a detailed description of model equations and parameters the reader is referred to Hansen, 2003 or Hansen & Neumann, 2004.

The model architecture is defined by a sequence of preprocessing stages and a recurrent loop of long-range interaction, realizing a simplified architecture of V1 (Fig. 1). The computational model incorporates a number of properties which are based on empirical findings, such as cooperative nonlocal long-range integration, which selectively link cells with colinear aligned RFs (Bosking et al., 1997, Gilbert & Wiesel, 1983), competitive short-range interactions (Bosking et al., 1997), and modulating feedback, where feedback alone is not sufficient to drive cell responses (Hupe et al., 1998).

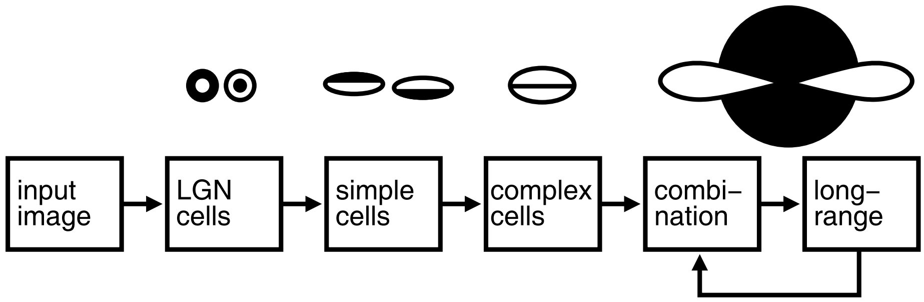

Fig. 1. Overview of the model stages. A sketch of the receptive fields for a sample orientation of 0 deg is shown above each stage.

2.1 Feedforward preprocessing

In the feedforward path, the initial luminance distribution is processed by isotropic LGN-cells, followed by orientation-selective simple and complex cells. The interactions in the feedforward path are governed by basic linear equations to keep the processing in the feedforward path relatively simple and to focus on the contributions of the recurrent loop. In our model, complex cell responses provide an initial local estimate of contour strength, position and orientation which is used as bottom-up input for the recurrent loop.

2.2 Recurrent long-range interaction

The recurrent loop has two stages, namely a combination stage where bottom-up and top-down inputs are fused, and a stage of long-range interaction. At the combination stage, feedforward complex cell responses and feedback long-range responses are added and subject to a nonlinear compression of high amplitude activity following the Weber-Fechner law. At the long-range stage the contextual influences on cell responses are modeled. Orientation-specific, anisotropic long-range connections provide the excitatory input. These long-range connections are modeled by a filter which is narrowly tuned to the preferred orientation, reflecting the highly significant anisotropies of long-range fibers in visual cortex. The spatial layout of the filter is similar to a bipole filter as proposed by Grossberg & Mingolla, 1985. The inhibitory input is given by isotropic interactions in the spatial and the orientational domain. To implement the modulating feedback, the excitatory long-range input is gated by the feedforward activity. Modulating feedback does not generate illusory contours, which are generated in higher visual areas (Neumann & Sepp, 1999). In the model, feedback selectively enhances bottom-up activities which are consistent within a more global context (Carpenter & Grossberg, 1988).

2.3 Junction Detection

Corners and junctions are characterized by points in the visual space where responses for multiple orientations are present and high overall activity exists within a hypercolumn. We use a measure of circular variance to signal multiple orientations. The overall activity is given by the sum across all orientations within a hypercolumn. Thus, the junction map for a distributed hypercolumnar representation is given by the multiplication of the circular variance with the overall activity. To visualize the data, single junction points are marked as local maxima in the junction map.

3 Simulations

All simulation results shown in this section are based on the same set of parameters. Long-range results are shown after 12 recurrent interactions, after which the activities have saturated, and are compared to the bottom-up activity at the complex cell stage. The edge maps show the sum across all orientations within each hypercolumn.

3.1 Contour Extraction

In the first simulation we show that the proposed model can account for empirical data on contextual interactions (Kapadia et al., 1995). In good agreement with the empirical data, the model responses to a central bar element is enhanced by colinear flankers and suppressed by noisy, random textures (Fig. 2).

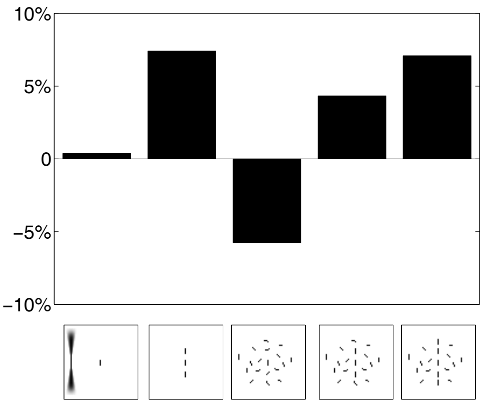

Fig. 2. Simulation of empirical data by Kapadia et al., 1995 shows colinear facilitation and contextual suppression by random textures compared to the response to a single bar element.

Next we employ the model for the processing of synthetic and camera images. The recurrent long-range interaction leads to more robust representation of contours (Fig. 3). Also, small gaps in the contour are closed as long as initial bottom-up activity is present.

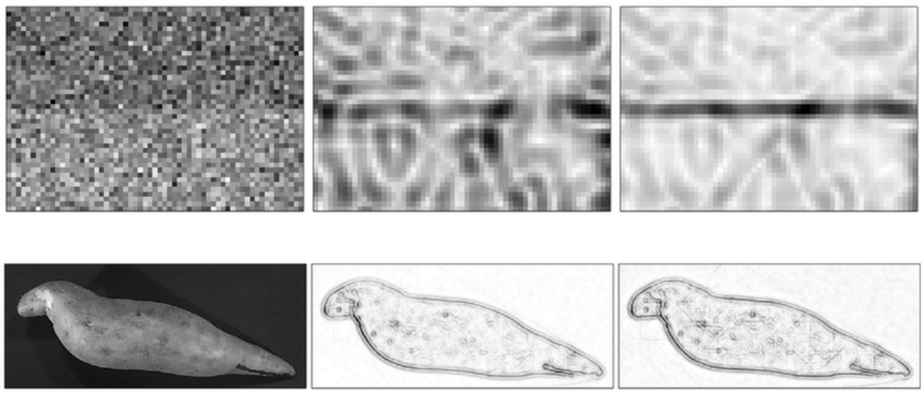

Fig. 3. Contour enhancement for synthetic and a natural image. Left to right in each row: Input image, feedforward complex cell response, and long-range response.

3.2 Junction Detection

In the first simulation we evaluate the localization accuracy of junction responses for L-, T-, Y-, W- and Ψ-junctions. We measure the Euclidean distance between the ground-truth location and the location as computed by the detection scheme which is based on feedforward complex cells alone and recurrent long-range interactions. For all junction types, the localization is considerably better for the method based on recurrent long-range interactions (Fig. 4).

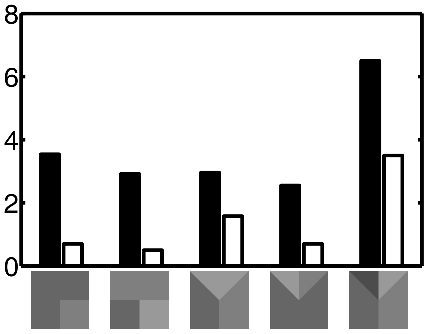

Fig. 4. Localization of generic junction configurations. For each junction type (L-, T-, Y-, W, and Ψ), the distance in pixels from the ground-truth location is considerably smaller for the long-range interactions (open bars) than for the complex cell responses (solid bars).

Finally, we evaluate the junction detection scheme for the processing of camera images. Receiver operator characteristic (ROC) analysis is used for a threshold-free evaluation of the two approaches (Fig. 5). The ROC curve obtained from the junction detection based on recurrent long-range interaction lies above the curve obtained for the feedforward complex cell responses. Thus, detection accuracy is increased by the recurrent long-range processing.

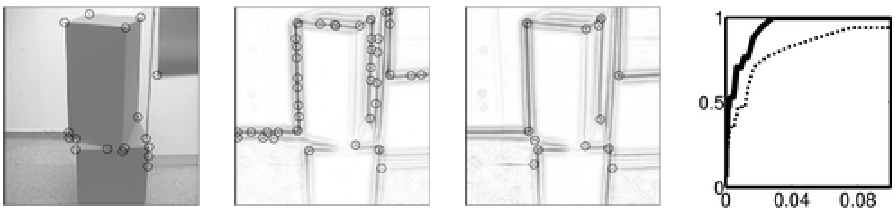

Fig. 5. Evaluation of the junction detection scheme for a camera image. Left to right: Input image, detected junctions based on the complex cell responses, and on the long-range responses, and the corresponding ROC curves (complex dashed, long-range solid). For better visualization, only a cut-out of the left part of the ROC curves is shown.

Overall, the results obtained based on the long-range responses are superior to the results based on the purely feedforward complex cells responses. However, the results are not perfect in the sense that every corner and junction is detected by the new method. The focus of this work is not to propose an ideal junction detector, but to show how mechanisms of recurrent long-range processing in V1 lead to a coherent representation of contours and junctions.

4 Conclusion

We have proposed a novel method for corner and junction detection based on a distributed representation of orientation information within a hypercolumn (Zucker et al., 1989). The explicit representation of a number of orientations in a cortical hypercolumn is shown to constitute a powerful and flexible, multi-purpose scheme for feature extraction. This scheme can be used to extract and represent intrinsically 1D signal variations like contours as well as 2D variations like corners and junctions.

References

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J. Neurosci., 17(6):2112–2127.

Carpenter, G. A. & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. Computer, 21:77–88.

Gilbert, C. D. and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. J. Neurosci., 3:1116–1133.

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

Hansen, T. (2003). A neural model of early vision: Contrast, contours, corners and surfaces. Doctoral dissertation, Ulm Univ., Faculty of Computer Science, Dept. of Neural Information Processing.
Available on-line: http://vts.uni-ulm.de/query/longview.meta.asp?documentid=3022.

Hansen, T. & Neumann, H (2004). Neural mechanisms for the robust representation of junctions. Neural Computation 16(45):1013–1037.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature, 394:784–787.

Kapadia, M. K., 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(4):843–856.

Neumann, H. and Sepp, W. (1999). Recurrent V1-V2 interaction in early visual boundary processing. Biol. Cybern., 81:425–444.

Zucker, S. W., Dobbins, A., & Iverson, L. A. (1989). Two stages of curve detection suggest two styles of visual computation. Neural Computation, 1:68–81.