A Computational Approach to Boundary Detection *

B. S. Manjunath and R. Chellappa

Signal and Image Processing Institute
Department of Electrical Engineering–Systems
University of Southern California
Los Angeles, CA 90089-MC0272

Abstract

This paper presents a unified approach to boundary perception. The model consists of a hierarchical system which extracts and groups salient features in the image at different spatial scales. In the first stage a Gabor wavelet decomposition provides a representation of the image which is orientation selective and has optimal localization properties, and provides a good model for early feature detection. Following this, local competitive interactions are introduced which help in reducing the effects of noise and illumination variations. Scale interactions help in localizing line ends and corners, and play an important role in boundary perception. The final stage groups similar features aiding in boundary completion. Experimental results on detecting edges, texture boundaries and illusory contours are provided.

1 Introduction

In this paper we suggest a simple biologically motivated approach for detecting image boundaries. Biological vision systems, especially those of mammals and in particular human's, are extremely adept at processing the vast amount of intensity data projecting from the 3-D external world on to the 2-D retina. Intensity discontinuities in 2-D images are a result of physical phenomena such as occlusion in the 3-D world, depth discontinuities, changes in surface orientation, reflectance properties or illumination. In computer vision most edge detection algorithms have been developed for finding step edges [1, 2], and consequently their performance is poor on other edges such as lines (or bars) and ramps. In general no linear filtering operation will be able to detect and localize composite edges accurately [3, 4, 5]. Textures form another important class of natural scenes, and like intensity edges provide useful information regarding shape and motion. In computational vision, many models ranging from stochastic to structural ones have been used in analyzing textures. Computational models for human texture perception have also been extensively studied [6, 7]. In addition to intensity edges and textures, we also consider the detection of subjective contours. The perception of these contours is a consequence of mechanisms involved in interpreting incomplete and noisy information. We restrict here to simple cases of such contours induced by line ends, which can be recovered without using high level knowledge or context information.

We discuss here a unified framework in which all these different types of boundaries can be detected. The basic steps include (a) feature extraction using oriented band pass filters at various scales (modeled by Gabor wavelet decomposition), (b) local feature interactions consisting of competition between spatial neighbors in each frequency channel,competition between orientations at each spatial location and scale interactions to generate end-inhibition, and (c) Grouping of features and locating boundaries.

2 Multiscale representation and Gabor Wavelets

The multiscale approach provides an elegant hierarchical framework for image analysis. There has been a growing interest in the use of wavelets for multiscale representation of the image data [8]. Wavelets are families of basis functions generated by dilations and translations of a *basic wavelet*. The wavelet transform is thus a decomposition of the function in terms of these basis functions. One of the objectives of such a transformation is to provide a simultaneous description of data in frequency and spatial domains. The

*Partially supported by AFOSR under grant 90-0133

basic wavelet used in our work is a Gabor function and we refer to the resulting transform as the Gabor wavelet transform.

Gabor functions are Gaussians modulated by complex sinusoids. In its general form, the 2-D Gabor function can be written as [9],

$$g(x,y;u_0,v_0) = e^{(-[\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}]+2\pi i[u_0\ x+v_0\ y])} \quad (1)$$

σ_x and σ_y define the widths of the Gaussian in the spatial domain and (u_0, v_0) is the frequency of the complex sinusoid. A well known property of these functions is that they achieve the minimum possible joint resolution in space and frequency domains [9]. Gabor functions form a complete but non-orthogonal basis set and any given function $f(x,y)$ can be expanded in terms of these basis functions. Such an expansion provides a localized frequency description and has been used in texture analysis [10]. Local frequency analysis of this nature, however, is not suitable for feature representation as it requires a fixed window width in space and consequently the frequency bandwidth is constant on a linear scale. In order to optimally detect and localize features at various scales, filters with varying support rather than a fixed one are required. This would suggest a transformation similar to wavelet decomposition rather than a local Fourier transform. We now consider such a wavelet transform where the *basic wavelet* is a Gabor function of the form

$$g_\lambda(x,y,\theta) = e^{(-[x^2\lambda^2+y^2]/2+i\pi[x\cos\theta+y\sin\theta])} \quad (2)$$

where λ is the spatial aspect ratio, θ is the preferred orientation. To simplify the notation, we drop the subscript λ and unless otherwise stated assume that $\lambda = 1$. The corresponding family of wavelets is

$$(g(\alpha(x-x_0, y-y_0),\theta)), \alpha \in \mathcal{R}, \theta \in [0,\pi] \quad (3)$$

The Gabor wavelet transform is then defined by

$$W_{f,g}(\alpha,\theta,x_0,y_0) = \int f(x,y)\ g^*(\alpha(x-x_0, y-y_0),\theta)\ dx\ dy \quad (4)$$

For practical applications, discretization of the parameters is necessary. The discretized parameters must cover the entire frequency spectrum of interest. Let the orientation range $[0,\pi]$ be discretized into N intervals and the scale parameter α be sampled exponentially as $\alpha^j, j \in \mathcal{Z}$. This results in the wavelet family

$$g_k(\alpha^j(x-m, y-n)) = e^{-\alpha^{2j}\frac{x^2+y^2}{2}+i\pi\alpha[x\cos\frac{k\pi}{N}+y\sin\frac{k\pi}{N}]} \quad (5)$$

3 The Model

We now present a simple model for feature processing based on the Gabor wavelet decomposition of the image and local feature interactions. We begin with a brief analysis of the Gabor wavelets as edge detectors.

3.1 Line and Edge detectors

The wavelet decomposition using Gabor functions has an important physical interpretation. The even and odd components of the complex Gabor function respond maximally to *line edges (or bars)* and *step* edges respectively, in the image. These features can be detected at the local maxima in their *energy* [3, 5]. If R_s and I_s represent the response from the even and odd symmetric feature detectors at a position s, then the local energy E_s at s is given by $E_s = \sqrt{R_s^2 + I_s^2}$. The limitations of conventional edge detectors are well known [4, 5]. Some important observations are that feature detection/localization by any type of linear filtering operation is not adequate, and in particular zero crossings of the result of applying any linear operator to the image do not capture all significant features in the image. Secondly, there is a need for directionally selective quadrature filter pairs. The outputs of these filters cannot be analyzed separately (except when either the step edges alone or the line edges alone are present) and one way to combine the output is to consider the total energy in them as discussed above. These energy feature detectors not only respond to, and localize simple line and step edges, but also other perceptually significant features such as roof edges and Mach bands [3]. A discussion on alternate means of combining information from the quadrature filters and their relative merits can be found in [4].

3.1.1 Performance Analysis

In the following we assume 1-D functions for simplicity and give an analysis of signal to noise ratio (SNR) and localization properties analogous to the one given in [2] (see also [5] for a similar analysis). The Gabor wavelet function in 1-D can be written as $g(x) = \exp(-\alpha^2x^2/2 + i\pi\alpha x)$. Let the true location of the edge $e(x)$ be at $x = 0$ and due to presence of noise $\eta(x)$ the observed maximum in the energy $E(x)$ is at x_0. Let $y(x)$ be the response of the complex filter $g(x)$ to a noisy input $e(x)+\eta(x)$, and $f(x)$ and $n(x)$ denote the signal and noise terms respectively in the output. For a white noise process, the output noise energy is

$$\int_{-\infty}^{\infty} g(x)g^*(x)dx = \int_{-\infty}^{\infty} e^{-\alpha^2x^2}dx = \sqrt{\pi}/\alpha \quad (6)$$

SNR at the true location of the edge is given by the ratio of the signal power to the noise power at the origin:

$$SNR = \frac{|f(0)|}{\sqrt{\sqrt{\pi}/\alpha}} \tag{7}$$

The edges are located at the local maxima in the energy $E(x) = y(x)y^*(x)$ and $E'(x_0) = 0$. Further, assuming that the noise power is small compared to the signal, $E(x) \approx ff^*(x) + 2(f_r(x)n_r(x) + f_i(x)n_i(x))$, where the subscripts r and i denote the real and complex components. As suggested in [2], the inverse of the variance of x_0, $(\mathcal{E}(x_0^2))^{-1}$ provides a measure of localization, and is given by [11]

$$\mathrm{L} = [\mathcal{E}(x_0^2)]^{-\frac{1}{2}} \approx \frac{|[(ff^*)''(0)]|}{2\sqrt{\mathcal{E}[(f_r n_r + f_i n_i)'(x_0)]^2}} \tag{8}$$

We computed the SNR and localization for the complex Gabor filter [11] for the cases of line and step edges:

$$\mathrm{SNR(step)} = 0.2774/\sqrt{\alpha},\ \mathrm{L(step)} \approx 1.93\sqrt{\alpha} \tag{9}$$

$$\mathrm{SNR(line)} = 0.7511\ \sqrt{\alpha}, \mathrm{L(line)} \approx 0.33\alpha\sqrt{\alpha} \tag{10}$$

In general (except for the special case of line edge), the SNR improves with increasing filter width whereas localization deteriorates. The SNR (for step edges) is poor compared to the first derivative of Gaussian used in [2], but its overall performance in the presence of composite edges is better. Note that both the Marr-Hildreth [1] and Canny [2] operators will fail to detect line edges at their true locations. In fact it is easy to show that at the true locations of line edges the SNR is zero for these operators.

3.2 Local Interactions

Following feature extraction using Gabor wavelets, we now consider local competitive interactions and grouping of these features. Competitive interactions help in noise suppression and reducing the effects of illumination. These interactions are modeled by non-linear lateral inhibition between features. Two types of such interactions are identified. The first type includes competition between different orientations at each spatial position and the second between spatial neighbors within each orientation. The following notation is used in explaining the interactions: The output of a cell at position s in a given frequency channel with a preferred orientation θ is denoted by $Y_{s,\theta}$, with $I_{s,\theta}$ being the excitatory input to that cell from the previous processing stage. Let N_s be the local spatial neighborhood of $s = (x, y)$. The competitive dynamics is represented by

$$\dot{X}_{s,\theta} = -a_{s,\theta}X_{s,\theta} + I_{s,\theta} - \sum_{s' \in N_s} b_{s,s'}Y_{s',\theta} - \sum_{\theta' \neq \theta} c_{\theta,\theta'}Y_{s,\theta'} \tag{11}$$

$$Y_{s,\theta} = g(X_{s,\theta}) \tag{12}$$

and (a, b, c) are positive constants. In our experiments we have used a sigmoid non-linearity of the form $g(x) = 1/(1+\exp(-\beta x))$. The dynamics of (11) can be visualized as follows : At each location within a single frequency channel, the corresponding cell receives an excitatory input ($I_{s,\theta}$) from a similarly oriented feature detector (of the same spatial frequency). Further it also receives inhibitory signals from the neighboring cells. We assume that all these interactions are symmetric ($b_{s,s'} = b_{s',s}$ and $c_{\theta,\theta'} = c_{\theta',\theta}$). The competitive dynamics of the above system can be shown to be stable. The Lyapunov function for the system [12] can be written as

$$\begin{aligned} E(Y) \;=\; & \frac{1}{2}\sum_{s,s'} b_{s,s'}Y_{s,\theta}Y_{s',\theta} + \frac{1}{2}\sum_{\theta,\theta'} c_{\theta,\theta'}Y_{s,\theta}Y_{s,\theta'} \\ & + \sum_{s,\theta}\int_0^{Y_{s,\theta}} \left(a_{s,\theta}g_{s,\theta}^{-1}(y) - I_{s,\theta}\right)\, dy \end{aligned} \tag{13}$$

Under the assumptions that the interactive synapses are symmetric and that $g(\cdot)$ is monotone non-decreasing, the time derivative of E is negative and the system represented by (11) always converges.

3.2.1 Scale Interactions and End-Inhibition

End-inhibition is a property peculiar to *hypercomplex* cells in the visual cortex of mammals. These cells respond to small lines and edges in their receptive field, and their response decreases as the length of lines/edges increase (hence these are often referred to as end detectors). These cells appear to play an important role in localizing line-ends and texture boundaries. We model here the response of these end-detector cells by simple scale interactions. If $Q_{s,\theta}^{(i)}$ denotes the response of such an end-detector at position s in the frequency channel i and orientation θ, then

$$Q_{s,\theta}^{(i)} = g(\alpha_1 I_{s,\theta}^{(i)} - \alpha_2 I_{s,\theta}^{(j)}) \tag{14}$$

At line ends, cells with shorter receptive fields will have a stronger response than those with larger fields, and consequently will be able to excite the hypercomplex cells. At other points along the line, both small and large receptive field cells are equally excited and in the process the response of the hypercomplex cells is inhibited. It appears that such scale interactions to

generate end inhibition do exist in the visual cortex. In [13] it is observed that layer 6 cells in the cat's striate cortex, which have large receptive fields, influence the end-inhibition property of the hypercomplex cells in layer 4 having relatively shorter receptive fields. Inactivating layer 6 cells results in a loss of end-inhibition in layer 4 cells, without affecting their other attributes such as orientation selectivity. The original idea of using such interactions dates back to the early work of Hubel and Wiesel [14]. An alternate model for end-inhibition using local competitive interactions is discussed in [15]. The hyper-complex cells in turn activate cells whose orientation is orthogonal to those of hypercomplex cells and with larger receptive fields to initiate the grouping process. A similar model for grouping the responses of hypercomplex cells has been proposed in [16].

3.2.2 Grouping

The final stage involves grouping similar orientations. The grouping process receives inputs from the competitive stage and from end-detectors described in section 3.2.1. Note that the orientation of the activating end-detector is orthogonal to the actual orientation of the grouping process. If $Z_{s,\theta}$ represents the output of this process, then

$$Z_{s,\theta} = g\left(\int d(s-s',\theta)(Y_{s',\theta} + Q_{s',\theta'})\, ds'\right) \quad (15)$$

$$d(s,\theta) = e^{(-(2\sigma^2)^{-2}[\lambda^2(x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2])} \quad (16)$$

$d(s,\theta)$ represents the receptive field of $Z_{s,\theta}$, θ is the preferred orientation, θ' is the corresponding orthogonal direction, and λ is the aspect ratio of the Gaussian. The Z cells thus integrate the information from similar oriented cells within each frequency channel and from appropriately oriented end-detectors. Since the various frequency channels are sampled, the effective standard deviation of the Gaussian is σ/α^i, where α^i is the scale parameter for channel i.

3.3 Locating Boundaries

Intensity edges and Subjective contours

In section 3.1 the usefulness of energy detectors in localizing image features was discussed. In detecting the intensity edges in the image we used the energy features as input to the competitive stage. Thus the input to a cell in the competitive stage at a position (x,y) in the ith frequency channel is given by

$$I^i_{(x,y),\theta} = ||W_{f,g}(\alpha^i,\theta,x,y)|| \quad (17)$$

where $W(\cdot)$ is as in (4), and $i = \{0,-1,-2,-3,\ldots\}$ and $\theta = k\pi/N$, $k = \{0,1,\ldots,N-1\}$, N is the number of discrete orientations. The edges are located at the local maxima of the $Z(\cdot)$ in (15). The same energy features are also used in our experiments to detect line-ends through scale interactions and the perceptual boundaries for the examples in Figure 4 are also marked at the local maxima of $Z(\cdot)$.

Texture boundaries

The information extracted by the wavelets can be used in several ways to detect textures, though the results reported here are obtained using the energy measure. Texture boundaries are located at the local maxima of the gradient of Z.

To summarize, boundary detection consists of three distinct steps (a) Feature extraction using Gabor wavelets, (b) Local interactions between features, including local scale interactions for end-inhibition, and (c) Grouping of features and identifying boundaries.

4 Experimental Results

The performance of the model is illustrated on several images. The following parameter values were used in all the experiments described here: $\beta = 4.0$ in the transfer function $g(\cdot)$. The strengths of the inhibitory synapses in (11) are $b_{s,s'} = 1/||N_s||$ and $c = 1/N$, where $||N_s||$ is the cardinality of the neighborhood set and N is the number of discrete orientations used. The aspect ratio of the Gaussians in both the Gabor wavelets (2) and in the receptive field of Z cells (16) is set to 0.5.

Regarding implementing the dynamics of competition, we used a simple gradient descent on the corresponding energy function instead of solving the set of differential equations. The equilibrium points in general for these two methods will be different. The results of edge detection are shown in Figure 1. The model was also tested on different texture images containing both real and synthetic textures. We give one example for each in Figure 2. Use of scale interactions in texture boundary detection is illustrated in Figure 3 and Figure 4 gives the result of perceiving subjective contours.

5 Conclusions

In this paper we have a developed a common framework for detecting perceptually significant features such as edges and texture boundaries in images. We

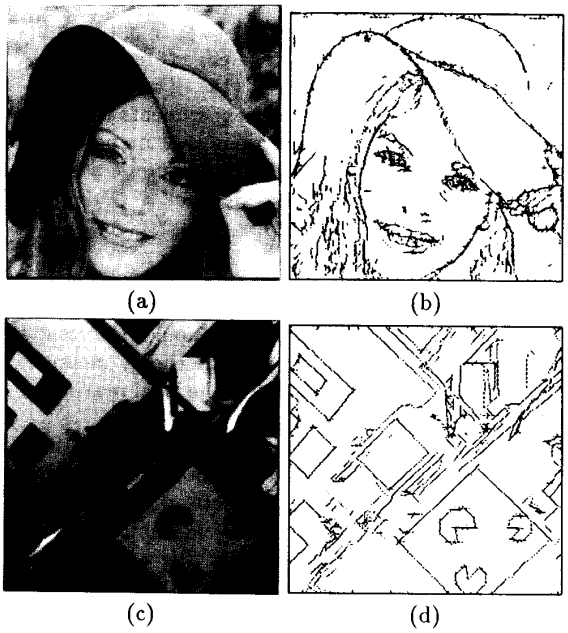

Figure 1: (a) and (c) show two 256×256 images and the corresponding edges detected are shown in (b) and (d). In (b) the edges are from two channels $\alpha^i = \{1/\sqrt{2}, 1/2\}$ and in (d) $\alpha^i = 1/\sqrt{2}$. For both examples $\sigma = 1$.

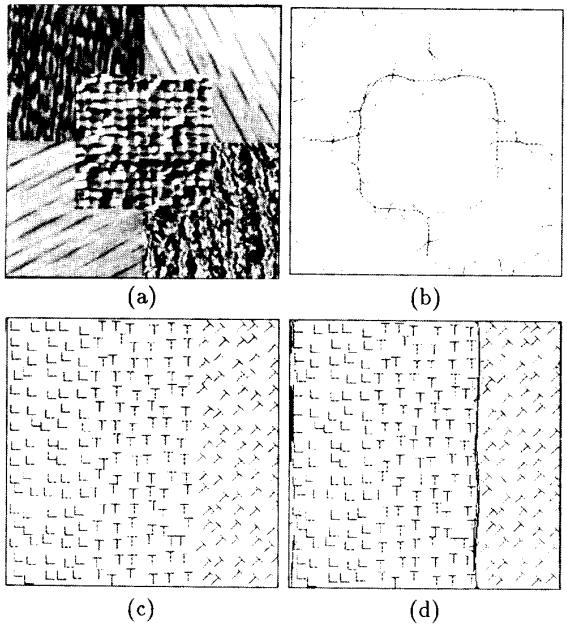

Figure 2: (a) Image consisting of four natural textures, water, wood, raffia and grass, (b) texture boundary detected, (c) texture consisting of three regions, L, T and tilted-Ts. The boundary between L and T s can not be easily detected. However the orientation difference between the two T regions is enough to discriminate between the two regions in almost all frequency channels, and (d) boundary detected is superimposed on the original. For both examples the results shown correspond to the combined output from channels $\alpha = \{1/2, 1/2\sqrt{2}, 1/4\}$ and $\sigma = 5$ pixels.

have suggested a simple model based on detecting oriented features at different spatial scales and on local interactions between features. Interactions between frequency channels is used in generating end-inhibition which plays an important role in boundary perception. Several examples are provided to illustrate the performance of this approach in detecting different types of boundaries.

Acknowledgements

We wish to thank Professor von der Malsburg for his suggestions and encouragement during the course of this work. We also would like to thank Joachim Buhmann, Anand Rangarajan, Sundereshan Chandrashekar and Qinfen Zheng for valuable discussions.

References

[1] D. Marr and E. Hildreth, "Theory of edge detection", *Proceedings of Royal Society of London (B)*, pp. 187–217, 1980.

[2] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, 1986.

[3] M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase dependent energy model", *Proceedings of Royal Society of London (B)*, 235, pp. 221–245, 1988.

[4] C. Ronse, "A twofold model of edge and feature detection", pre-print, Sept 1990.

[5] P. Perona and J. Malik, "Detecting and localizing edges composed of steps, peaks and roofs ", In *Proc. Intl. Conf. Computer Vision 90*, pp. 52–57, Tokyo,Japan, December 1990.

[6] J. Beck, K. Prazdny, and A. Rosenfeld, "A Theory of textural segmentation", In J. Beck, B. Hope,

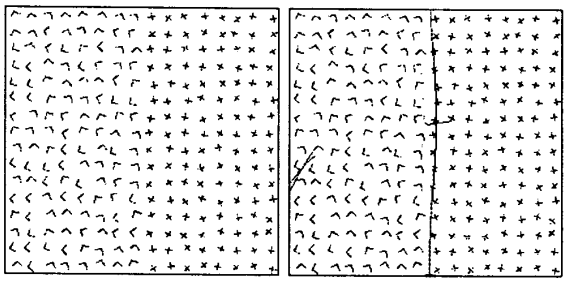

Figure 3: Texture consisting of randomly oriented L and +. The line segments are all 7 pixels wide and the image is 256 × 256 pixels. The two regions differ in the distribution of line-ends, intersections and corners, and do not offer any specific orientation discriminability. The boundary shown in (b) (superimposed on the original texture) is detected using the output of the scale interactions with $\sigma = 16$. The scales used in this example are $\alpha^i = \{1/2, 1/4\}$.

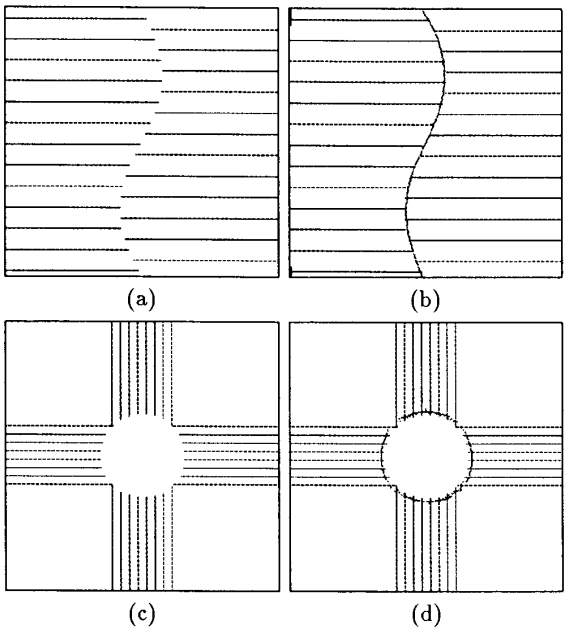

Figure 4: Two examples of illusory contours formed by line terminations and the detected contours superimposed on the original images. In (b) $\alpha^i = \{1/2, 1/4\}$ and $\sigma = 8$ and in (c) $\alpha^i = \{1/\sqrt{2}, 1/4\}$ and $\sigma = 2$.

and A. Rosenfeld, editors, *Human and Machine Vision*, pp. 1–38, Academic Press, 1983.

[7] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms", *J. Opt. Soc. Am. A*, vol. 7, pp. 923–932, 1990.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, July 1989.

[9] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Am. A*, 2, pp. 1160–1169, 1985.

[10] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters", *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-12, pp. 55–73, January 1990.

[11] B. S. Manjunath and R. Chellappa, "A unified approach to boundary perception: edges, textures and illusory contours", Technical Report SIPI 91-167, Dept. of Elec Engg., Univ. of Southern California, January 1991.

[12] M. A. Cohen and S. Grossberg, "Absolute stability of global pattern formation and parallel memory storage by competitive neural networks", *IEEE Trans. Syyst. Man Cybern.*, pp. 815–825, 1983.

[13] J. Bolz and C. D. Gilbert, "Generation of end-inhibition in the visual cortex via interlaminar connections", *Nature*, vol. 320, pp. 362–365, March 1986.

[14] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas(18 and 19) of the cat", *Journal of Neurophysiology*, 28, pp. 229–289, 1965.

[15] S. Grossberg and E. Mingolla, "Neural dynamics of surface perception: Boundary webs, illuminants, and shape-from-shading", *Computer Vision, Graphics, and Image Processing*, pp. 116–165, 1987.

[16] E. Peterhans and R. von der Heydt, "Mechanisms of contour perception in monkey visual cortex. II. Contour bridging gaps", *Journal of Neuroscience*, 9(5), pp. 1749–1763, 1989.