A hierarchical model of peripheral vision

Leyla Isik, Joel Z. Leibo, Jim Mutch, Sang Wan Lee,
and Tomaso Poggio

CSAIL

A hierarchical model of peripheral vision

Leyla Isik[1,2], Joel Z. Leibo[1,3], Jim Mutch[1,3], Sang Wan Lee[1,3], and
Tomaso Poggio[1,3]

[1]CBCL, McGovern Institute for Brain Research, Massachusetts Institute of Technology
[2]Computational and Systems Biology, Massachusetts Institute of Technology
[3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

June 17, 2011

Abstract

We present a peripheral vision model inspired by the cortical architecture discovered by Hubel and Wiesel. As with existing cortical models, this model contains alternating layers of simple cells, which employ tuning functions to increase specificity, and complex cells, which pool over simple cells to increase invariance. To extend the traditional cortical model, we introduce the option of eccentricity-dependent pooling and tuning parameters within a given model layer. This peripheral vision system can be used to model physiological data where receptive field sizes change as a function of eccentricity. This gives the user flexibility to test different theories about filtering and pooling ranges in the periphery.

In a specific instantiation of the model, pooling and tuning parameters can increase linearly with eccentricity to model physiological data found in different layers of the visual cortex. Additionally, it can be used to introduce pre-cortical model layers such as retina and LGN. We have tested the model's response with different parameters on several natural images to demonstrate its effectiveness as a research tool. The peripheral vision model presents a useful tool to test theories about crowding, attention, visual search, and other phenomena of peripheral vision.

1   Introduction

Limited acuity in the periphery creates several interesting perceptual effects that are important for reading, visual search and attention. Limited peripheral vision has been attributed to uneven distribution of photoreceptors and to different amounts of cortical magnification in the periphery versus the fovea. Receptive field sizes and cortical magnification have been mapped in V1 [1], V2 [3], V3 and V4 [4], and shown to be increasing with eccentricity. Despite this knowledge, these peripheral parameters have not yet been incorporated into a hierarchical model of the visual system.

We present a hierarchical model, inspired by Hubel and Wiesel, which can account for limited discrimination in the periphery. The model is based on the feed-forward architecture of the visual cortex; it consists of alternating layers of simple S cells that employ tuning functions to increase selectivity and complex C cells that pool over S cells to increase invariance as in [5]. We introduce the ability to vary tuning and pooling parameters in S and C layers in an eccentricity-dependent way.

2 The hierarchical peripheral vision model

2.1 The HMAX model

We have adapted HMAX model [5], shown in Figure 1, for the visual periphery. As in the original HMAX model, the peripheral vision model contains alternating layers of simple cells (feature detectors) and complex cells (which pool over simple cells to build invariance).
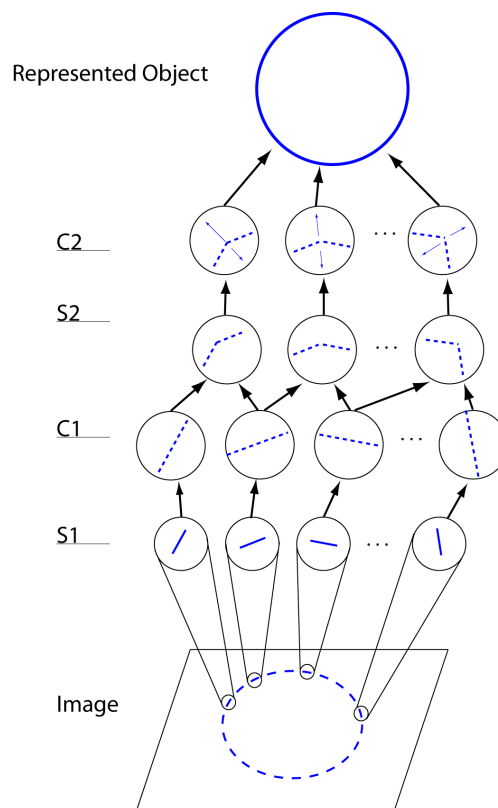

Figure 1: The HMAX model.

2

In its simplest form the model has 4 layers:

S1 - This layer contains simple orientation-tuned Gabor filters, which are analogous to simple cells found in the primary visual cortex. These cells can be described by the following equations.

$$T(x,y) = \exp\left(-\frac{(x_0^2 + \gamma^2 y_0^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right) \text{ s.t.,} \tag{1}$$

$$x_0 = x\cos\theta + y\sin\theta \text{ and } y_0 = -x\sin\theta + y\cos\theta \tag{2}$$

For an input image patch $z = z_1, ..., z_m$, the S1 response to a given Gabor filter template $T = t_1, ..., t_m$ is the normalized dot product between the image patch and template.

$$r = \sum_{j=1}^{m} \frac{z_j t_j}{\sqrt{\sum_{i=1}^{m} z_i^2}} \tag{3}$$

The filter's aspect ratio γ, orientation θ, effective width σ, and wavelength λ can all be tuned to fit physiological data.

C1 - The units in this layer perform max pooling locally over the same orientation (feature) S1 cells in a given neighborhood, and are analogous to complex cells in the primary visual cortex. The response r of a C1 unit with afferents $(z_1, ..., z_m)$ is:

$$r = \max_{j=1,...,m} z_j \tag{4}$$

S2 - The next simple cell layer performs template matching between C1 responses from an input image and the C1 responses of stored prototypes.

Template matching can be performed with a radial basis function, where the responses have a Gaussian-like dependence on the Euclidean distance between the C1 response of an input image patch and a stored prototype. The response r of an S2 unit to a given input image patch $z = z_1, ..., z_m$ (from the previous C1 layer) and a prototype or template $T = t_1, ..., t_m$ is:

$$r = \exp\left(-\beta \sum_{j=1}^{m} (z_j - t_j)^2\right) \tag{5}$$

The β parameter defines the sharpness of S2 tuning.

Alternatively, template matching can be performed with a normalized dot product between the template and the input C1 units (as in the S1 layer). The response r of an S2 unit to a given input image patch $z = z_1, ..., z_m$ (from the previous C1 layer) and a prototype or template $T = t_1, ..., t_m$ will then be given by:

$$r = \sum_{j=1}^{m} \frac{z_j t_j}{\sqrt{\sum_{i=1}^{m} z_i^2}} \tag{6}$$

3

C2 - In the final model layer, global spatial pooling over all S2 cells of a given feature is performed.

For a more detailed description, please see [5].

2.2 Cortical Network Simulator

The HMAX model can be implemented with Cortical Network Simulator (CNS), a GPU-based framework for fast simulation [6]. The peripheral vision extensions described below, are implemented as additional cell types that integrate easily with the HMAX CNS package.

2.3 Peripheral vision extension

To model the broader tuning in the visual periphery, S1 cells' Gabor filter width parameters can be set to increasing values as you move to the periphery. To model larger pooling ranges in the periphery of the visual cortex, the C cells can pool (either max pooling or averaging) over increasing numbers of cells as a function of eccentricity. A more detailed descriptions of the model parameters is given in Table 1.

Parameter	Model Layer	Interpretation
λ	S1 (tuning)	The wavelength of the sinusoidal component of the Gabor filter.
σ	S1 (tuning)	The effective width of the the Gaussian component of the Gabor filter.
γ	S1 (tuning)	The spatial aspect ratio of the Gabor filter.
m	C1 (pooling)	The pooling range of each C1 unit (number of S1 afferents to complete max pooling over).

Table 1: Adjustable parameters.

2.4 Modeling physiology

In a specific instantiation of the model, pooling and tuning parameters in the first S and C layers can increase linearly with eccentricity. Physiology data shows this trend occurs throughout the visual system [1, 2, 3, 4]. The eccentricity-dependent C1 cells can also be used to model retinal or LGN layers before the visual cortex. Averaging an input image with these C1 cells yields an eccentricity-dependent blurring that is similar to the effect of decreased photoreceptor density in the visual periphery.

When these eccentricity-dependent operations are implemented in the hierarchical framework, it gives us the ability to map the response at any model layer to its corresponding point in the input image or to points in other cortical layers. We can then fit physiological parameters to the model at different cortical layers and answer questions about connectivity and eccentricity-dependence

between these layers. With this framework, we can then test whether physiological parameters in visual cortex can reproduce various perceptual effects in the periphery, such as crowding.

3   Demonstration

Below is a visual demonstration of the eccentricity-dependent S and C cells' response to a checkerboard image - the original image is shown in Figure 3(a). To test the effect of C cells and demonstrate their retinotopic mapping abilities, they were applied directly to the checkerboard image (with no S layer before). Using C cells whose average pooling ranges increase with eccentricity creates a fisheye effect shown in Figure 3(b). This direct application of eccentricity-dependent averaging can be used to simulate pre-cortical layers of vision as mentioned above. The results of C max cells with the same increased pooling ranges are shown in Figure 3(c).

A normal (foveal) S unit will have a constant response to the checkerboard shown in Figure 4(a) for four orientations of Gabor filters. The peripheral model employs increasingly wider Gabor filters in the periphery, simulating physiological effects of broader tuning in the periphery. The effects of the eccentricity-dependent S cells can be observed in Figure 4(b) as the Gabor filter responses change from sharp oriented bars to broad blocks at the edge of the image.

The peripheral vision model allows us to test theories about higher levels of the visual cortex and to what extent they employ additional eccentricity-dependence or only inherit it from layers below. Figure 5(a) shows the response of one S layer and one C layer (with max pooling) on the checkerboard image with the original model (no eccentricity-dependence), Figure 5(b) with eccentricity-dependence in the S layer only, and Figure 5(c) with eccentricity-dependence in both S and C layers.

In Figures 4-6, the above demonstrations are repeated on a natural image (the MIT dome). Additionally, eccentrically-increasing average pooling complex cells are applied to a protest image in Figure 6(b). With appropriate pooling ranges, the center sign clearly reads "End segregated rules in public schools," while the peripheral signs are increasingly blurred and illegible. This is a toy demonstration of the model's potential uses in studies of reading and attention.

4   Preliminary results

Since the pooling and tuning operations can be separately manipulated, this model allows to test theories about sources of visual errors in the periphery. Crowding and low acuity in the periphery are often attributed to increased pooling ranges [8]. Our model allows us to test this theory directly, by varying only pooling, only tuning, or both, and observing the effects on letter acuity in the periphery.
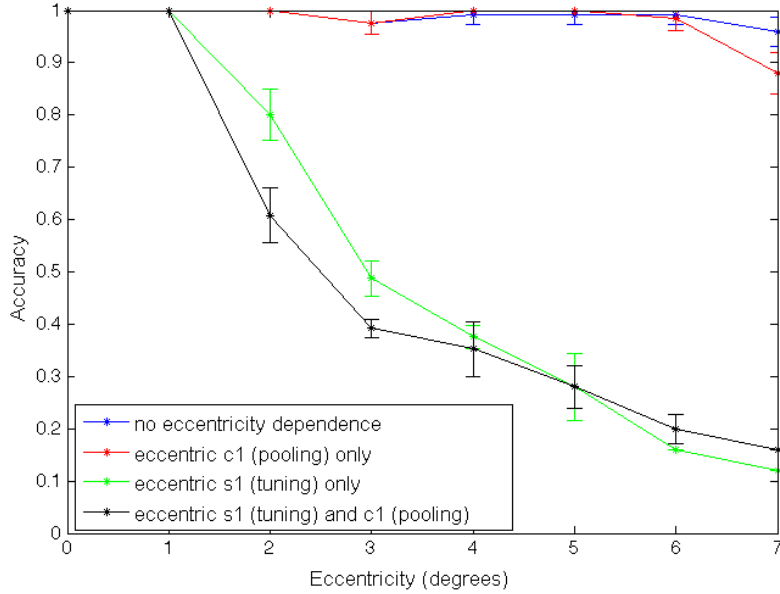
Figure 2: Letter acuity as a function of eccentricity for different pooling/tuning eccentricity-dependences. S2 templates were tuned to natural images. Classification was performed with a nearest neighbor classifier, using the correlation between the C2 vectors of centered letters and the C2 vectors of test images (with letters at various eccentricities). Model parameters were set to those outlined in [5], for increasing receptive field sizes.

The preliminary results, shown in Figure 2, indicate that increasing pooling ranges has little effect on acuity, while increasing tuning in the periphery decreases acuity with a similar trend as observed in psychophysics [9].

This hierarchical peripheral vision model presents a useful tool to fit physiological parameters and model perceptual effects due to eccentricity-dependence in the visual cortex.
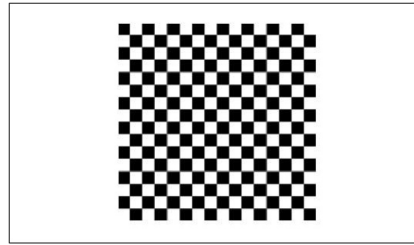
(a) Original checkerboard image
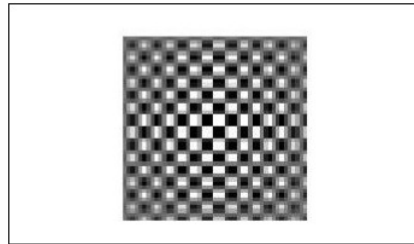
(b) Eccentrically increasing average pooling
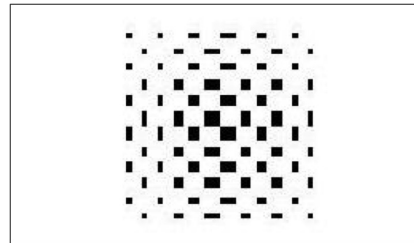
(c) Eccentrically increasing max pooling
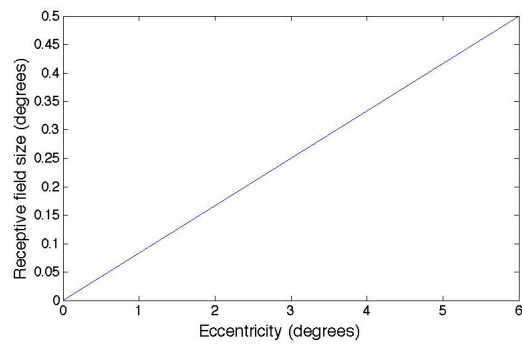
(d) Eccentricity-dependence of pooling ranges, where 1 degree = 20 pixels. Similar to values reported in [1]

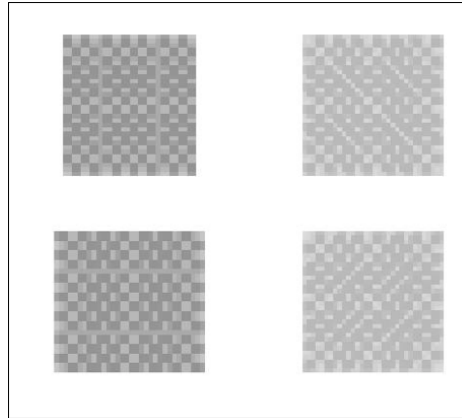Figure 3: C layer response of the peripheral vision model
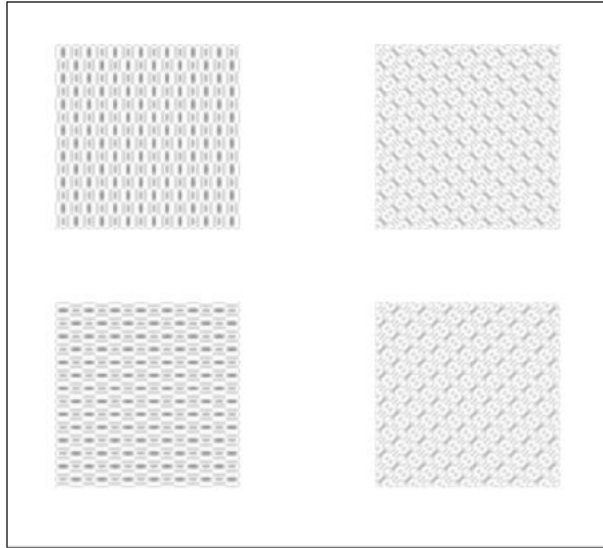
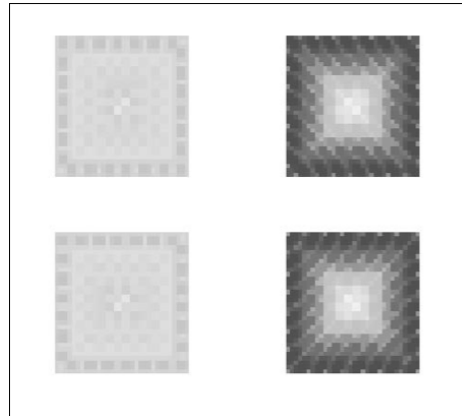(a) Normal S layer response for four orientations
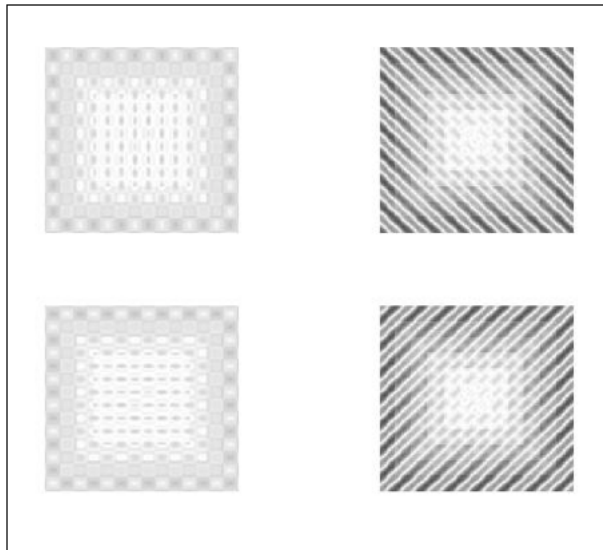
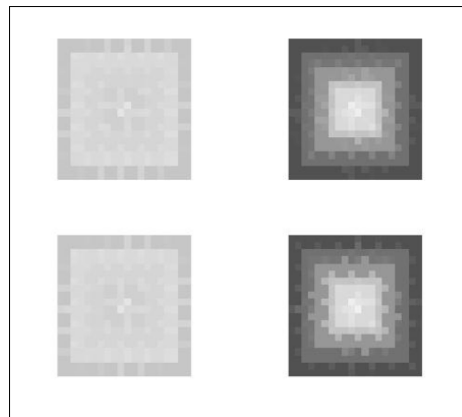(b) Eccentricity-dependent S layer response for four orientations

Figure 4: S response to the above checkerboard image.

8

(a) Normal S and C layers

(b) Eccentricity-dependent S layer, normal C layer
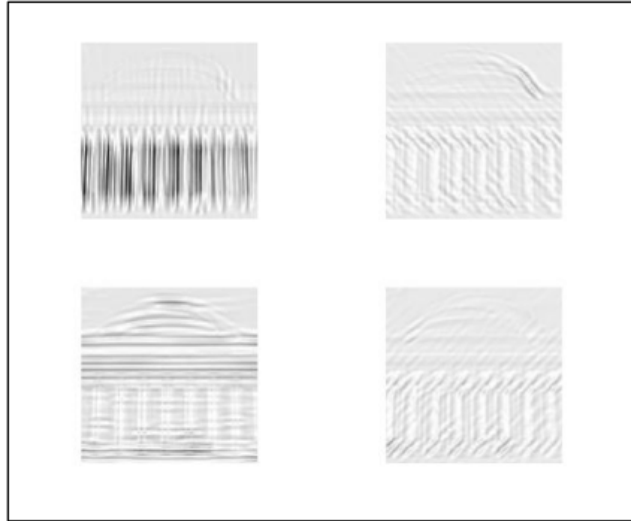

(c) Eccentricity-dependent S and C layers

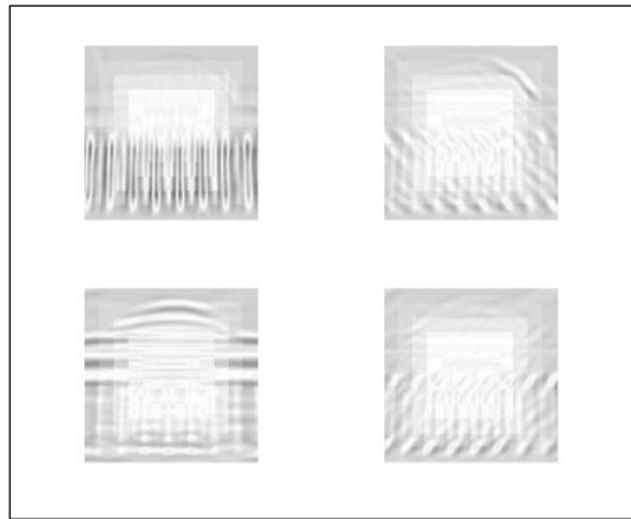Figure 5: C responses after S response to checkerboard image.

(a) The MIT dome

(b) 1964 Civil Rights march. The center poster clearly reads "End segregated rules in public schools," while the off-center posters are illegible.

Figure 6: "Retinal" outputs of two natural images.
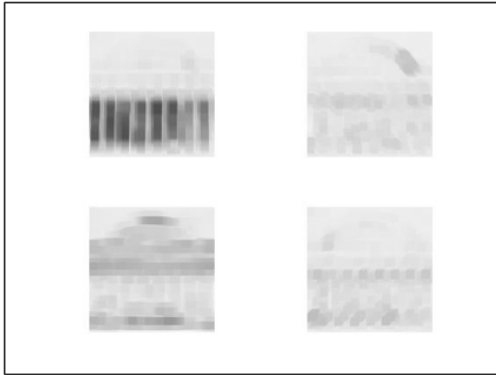
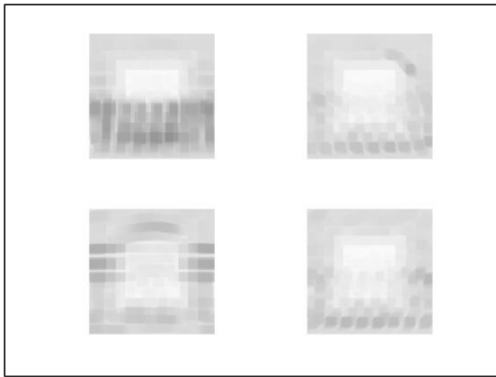(a) Normal S layer response for four orientations

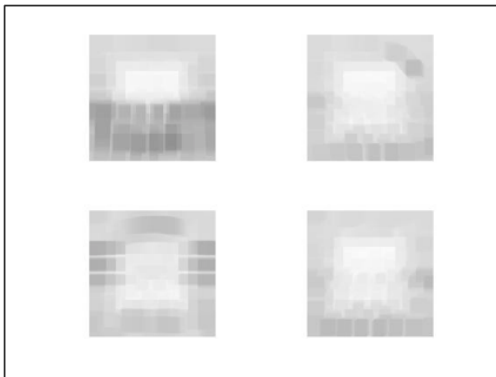(b) Eccentricity-dependent S layer response for four orientations

Figure 7: S response to MIT dome image above.

(a) Normal S and C layers

(b) Eccentricity-dependent S layer, normal C layer

(c) Eccentricity-dependent S and C layers

Figure 8: C responses after S response to the above image of the MIT dome .

5 Acknowledgements

References

[1] Hubel & Wiesel. Uniformity of Monkey Striate Cortex: A Parallel Relationship between Field Size, Scatter, and Magnification Factor. *The Journal of Comparative Neurology* 158.3 (1974).

[2] Duncan & Boynton. Cortical Magnification within Human Primary Visual Cortex Correlates with Acuity Thresholds. *Neuron* 38 (2003).

[3] Gattass, Gross, & Sandell, J. H. Visual topography of V2 in the macaque. *J Comp Neurol* 201, 519-539 (1981).

[4] Gattass, Sousa, & Gross. Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8, 1831-1845 (1988).

[5] Serre, Wolf, Bileschi, Riesenhuber, & Poggio. Robust Object Recognition with Cortex-Like Mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29.3 (2007).

[6] Mutch, Knoblich, & Tomaso Poggio. CNS: a GPU-based framework for simulating cortically-organized networks. *MIT-CSAIL-TR-2010-013* (2010).

[7] Hubel, & Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the CatÕs Visual Cortex. *J. Physiology* 160, 106-154 (1962).

[8] Whitney & Levi. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences* 15.4, 160-168 (2011).

[9] Bouma. Interaction Effects in Parafoveal Letter Recognition. *Nature* 226, 177-178 (1970).