

Presented at: NIPS-98, Denver CO, 1-3 Dec 1998.
Published in: Advances in Neural Information Processing Systems
eds. M. S. Kearns, S. A. Solla, and D. A. Cohn
volume 11, pages 153--159
MIT Press, Cambridge, MA 1999.

Modeling Surround Suppression in V1 Neurons with a Statistically-Derived Normalization Model

Eero P. Simoncelli
Center for Neural Science, and
Courant Institute of Mathematical Sciences
New York University
eero.simoncelli@nyu.edu

Odelia Schwartz
Center for Neural Science
New York University
odelia@cns.nyu.edu

Abstract

We examine the statistics of natural monochromatic images decomposed using a multi-scale wavelet basis. Although the coefficients of this representation are nearly decorrelated, they exhibit important higher-order statistical dependencies that cannot be eliminated with purely linear processing. In particular, rectified coefficients corresponding to basis functions at neighboring spatial positions, orientations and scales are highly correlated. A method of removing these dependencies is to *divide* each coefficient by a weighted combination of its rectified neighbors. Several successful models of the steady-state behavior of neurons in primary visual cortex are based on such “divisive normalization” computations, and thus our analysis provides a theoretical justification for these models. Perhaps more importantly, the statistical measurements explicitly specify the weights that should be used in computing the normalization signal. We demonstrate that this weighting is qualitatively consistent with recent physiological experiments that characterize the suppressive effect of stimuli presented outside of the classical receptive field. Our observations thus provide evidence for the hypothesis that early visual neural processing is well matched to these statistical properties of images.

An appealing hypothesis for neural processing states that sensory systems develop in response to the statistical properties of the signals to which they are exposed [e.g., 1, 2]. This has led many researchers to look for a means of deriving a model of cortical processing purely from a statistical characterization of sensory signals. In particular, many such attempts are based on the notion that neural responses should be statistically independent.

The pixels of digitized natural images are highly redundant, but one can always find a linear decomposition (i.e., principal component analysis) that eliminates second-order cor-

Research supported by an Alfred P. Sloan Fellowship to EPS, and by the Sloan Center for Theoretical Neurobiology at NYU.

relation. A number of researchers have used such concepts to derive linear receptive fields similar to those determined from physiological measurements [e.g., 16, 20]. The principal components decomposition is, however, not unique. Because of this, these early attempts required additional constraints, such as spatial locality and/or symmetry, in order to achieve functions approximating cortical receptive fields.

More recently, a number of authors have shown that one may use higher-order statistical measurements to uniquely constrain the choice of linear decomposition [e.g., 7, 9]. This is commonly known as *independent components analysis*. Vision researchers have demonstrated that the resulting basis functions are similar to cortical receptive fields, in that they are localized in spatial position, orientation and scale [e.g., 17, 3]. The associated coefficients of such decompositions are (second-order) decorrelated, highly kurtotic, and generally more independent than principal components.

But the response properties of neurons in primary visual cortex are not adequately described by linear processes. Even if one chooses to describe only the mean firing rate of such neurons, one must at a minimum include a rectifying, saturating nonlinearity. A number of authors have shown that a gain control mechanism, known as *divisive normalization*, can explain a wide variety of the nonlinear behaviors of these neurons [18, 4, 11, 12, 6]. In most instantiations of normalization, the response of each linear basis function is rectified (and typically squared) and then divided by a uniformly weighted sum of the rectified responses of all other neurons. Physiologically, this is hypothesized to occur via feedback shunting inhibitory mechanisms [e.g., 13, 5]. Ruderman and Bialek [19] have discussed divisive normalization as a means of increasing entropy.

In this paper, we examine the joint statistics of coefficients of an orthonormal wavelet image decomposition that approximates the independent components of natural images. We show that the coefficients are second-order decorrelated, but *not* independent. In particular, pairs of rectified responses are highly correlated. These pairwise dependencies may be eliminated by dividing each coefficient by a *weighted* combination of the rectified responses of other neurons, with the weighting determined from image statistics. We show that the resulting model, with all parameters determined from the statistics of a set of images, can account for recent physiological observations regarding suppression of cortical responses by stimuli presented outside the classical receptive field. These concepts have been previously presented in [21, 25].

1 Joint Statistics of Orthonormal Wavelet Coefficients

Multi-scale linear transforms such as wavelets have become popular for image representation. Typically, the basis functions of these representations are localized in spatial position, orientation, and spatial frequency (scale). The coefficients resulting from projection of natural images onto these functions are essentially uncorrelated. In addition, a number of authors have noted that wavelet coefficients have significantly non-Gaussian marginal statistics [e.g., 10, 14]. Because of these properties, we believe that wavelet bases provide a close approximation to the independent components decomposition for natural images. For the purposes of this paper, we utilize a typical separable decomposition, based on symmetric quadrature mirror filters taken from [23]. The decomposition is constructed by splitting an image into four subbands (lowpass, vertical, horizontal, diagonal), and then recursively splitting the lowpass subband.

Despite the decorrelation properties of the wavelet decomposition, it is quite evident that wavelet coefficients are *not* statistically independent [26, 22]. Large-magnitude coefficients (either positive or negative) tend to lie along ridges with orientation matching that of the subband. Large-magnitude coefficients also tend to occur at the same relative spatial locations in subbands at adjacent scales, and orientations. To make these statistical relationships

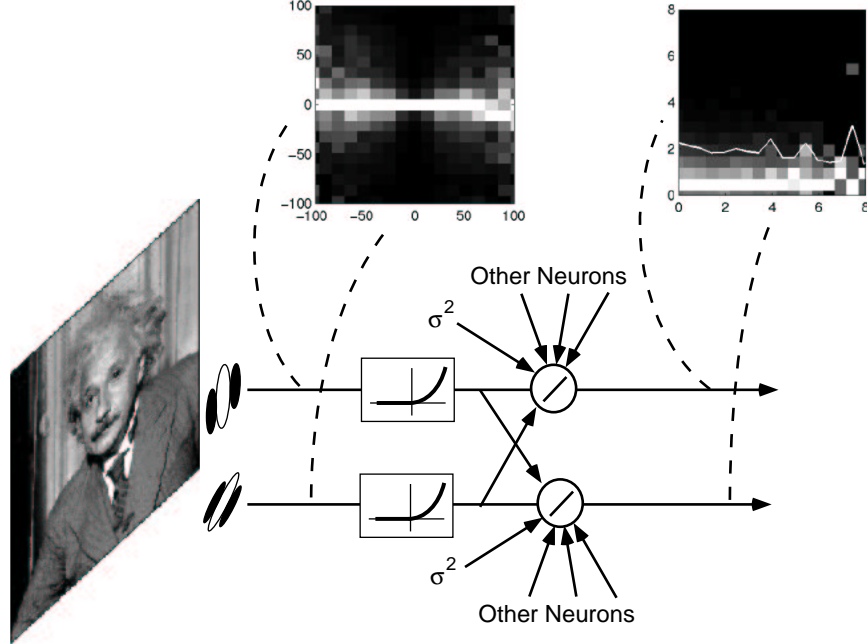


Figure 1: Illustration of image statistics as seen through two neighboring receptive fields. Left image: Joint conditional histogram of two linear coefficients. Pixel intensity corresponds to frequency of occurrence of a given pair of values, except that each column has been independently rescaled to fill the full intensity range. Right image: Joint histogram of divisively normalized coefficients (see text).

more explicit, the left panel of Fig. 1 shows a conditional histogram of coefficients associated with two neighboring receptive fields. Assuming stationarity, the statistics are gathered over all spatial positions of a single natural image. First, we see that the coefficients are well decorrelated: The expected value of the ordinate coefficient is approximately zero, independent of the value of the abscissa. But the *variance* of the ordinate clearly increases with the absolute value of the abscissa.

We have observed this type of dependency in pairs of coefficients at neighboring spatial positions, orientations and scales, and for a wide variety of imagery. We have previously used these relationships in applications of image compression, denoising, and synthesis [e.g., 22]. We have also shown that this dependency may be eliminated by *dividing*. Specifically, the squared coefficient, C^2 may be divided by a weighted sum of the neighboring squared coefficients plus a constant:

$$R = C^2 / \left[\sum_k w_k P_k^2 + \sigma^2 \right]. \quad (1)$$

The parameters $\{w_k\}$ and σ are chosen to minimize squared prediction error:

$$\{\hat{w}, \hat{\sigma}\} = \arg \min_{\{\tilde{w}, \tilde{\sigma}\}} \mathbf{E} \left[C^2 - \sum_k w_k P_k^2 - \sigma^2 \right]^2,$$

where the P_k are the values of coefficients at adjacent spatial positions, orientations and scales, and $\mathbf{E}[\cdot]$ indicates expected value (computed by integrating over the full spatial

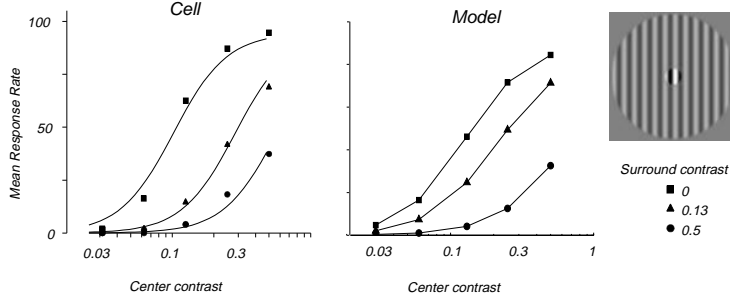


Figure 2: Response vs. center contrast, in the presence of a parallel surround stimulus of varying contrast. Physiological data from [8].

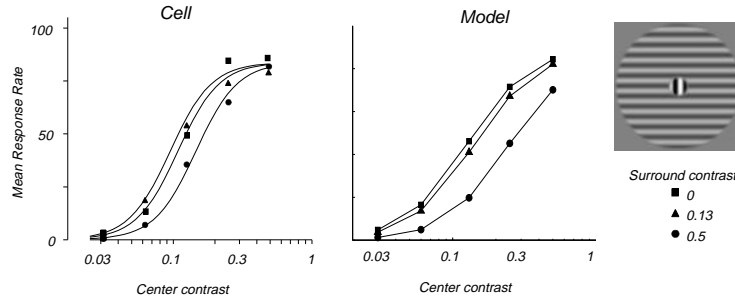


Figure 3: Response vs. center contrast in the presence of a perpendicular surround stimulus. Physiological data from [8].

extent of a set of images). A joint histogram of the square roots of two normalized coefficients is shown in the rightmost panel of Fig. 1, indicating that the resulting normalized components are nearly independent.

2 Physiological Comparisons

In this section, we examine predictions of the normalization model with weights determined from image statistics. The ability of normalization models to account for non-specific suppression within the classical receptive field has been documented [e.g., 12, 6]. Here we consider the influence of stimuli presented outside of the classical receptive field.

We examine electrophysiological data obtained from recordings of simple cells in area V1 of an anesthetized Macaque monkey in two different labs [8, 15]. In each example, an optimal drifting sinusoidal grating is presented in the classical receptive field of the neuron. Simultaneously, another drifting sine grating is presented in a large annular region surrounding the classical receptive field. Each experiment examines the effect of varying one parameter of the surround stimulus on the mean firing rate of the neuron.

For comparison, we show the normalized response, R , of a vertical basis function at the second recursion level of a wavelet pyramid, as specified by Eq. (1). Responses are averaged over all phases of the sinusoidal input, and scaled by a fixed constant α to produce response levels comparable to physiological responses. The normalization signal is a weighted combination of squared coefficients at two scales, all three orientations, and a spatial neighborhood of diameter 65 pixels (roughly 7 receptive fields). The normalization weights are optimized for the statistics of a set of three 512×512 images (“Goldhill”,

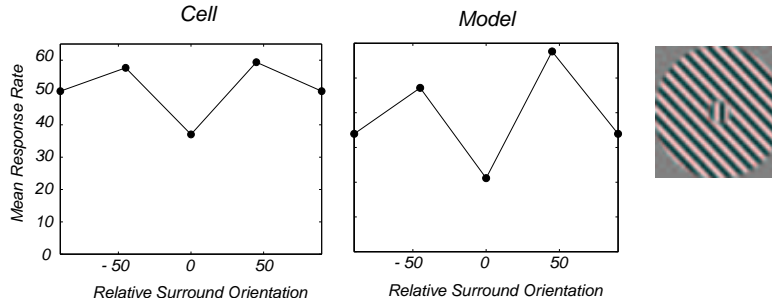


Figure 4: Response as a function of surround orientation. Physiological data from [8].

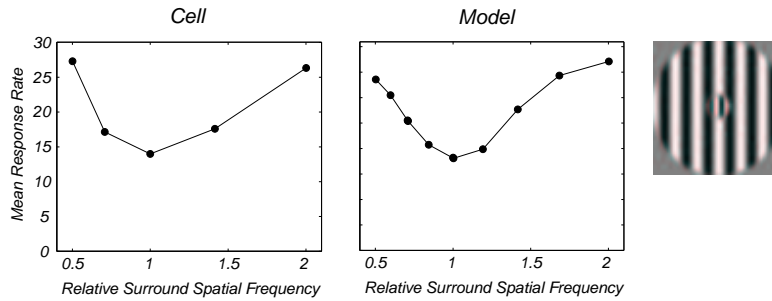


Figure 5: Response as a function of surround spatial frequency. Physiological data from [15].

“Boats”, and “Lena”).

The first two examples (Figs. 2 and 3) show response as a function of center stimulus contrast. Each curve corresponds to a different surround contrast. The physiological data are fit with a function of the form $R(c) = \alpha c^p / (c^p + \sigma^p)$. The model curves are less steep than those of the neurons, since the model uses a fixed exponent of two. As surround contrast increases, the entire curve shifts to the right (on a log scale), indicative of divisive suppression. The parallel surround stimulus produces a significantly larger shift than the perpendicular surround. In the model, this behavior is a direct consequence of the statistically-chosen normalization weights.

Figure 4 summarizes the suppressive effect of the surround (at the highest contrast) as a

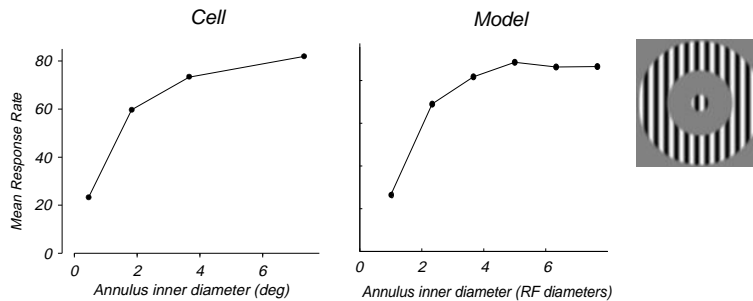


Figure 6: Response as a function of surround inner diameter. Physiological data from [8].

function of orientation. Figure 5 shows a similar behavior with respect to surround spatial frequency. The largest suppression is observed when the surround spatial frequency is the same as the center (i.e., the preferred spatial frequency of the cell). Figure 6 shows the effect of spatial proximity. As the surround stimulus is moved away from the receptive field, the suppressive effect is reduced.

3 Conclusions

We have presented a weighted normalization model for early visual processing. Both the form and the parameters of the model are specified by statistical measurements from natural images. Although the comparisons we have presented are somewhat anecdotal, we find the ability of this model to mimic physiological suppression behaviors quite remarkable.

Nevertheless, there are many tough issues to be resolved. Some of these are statistical. A fundamental question is whether statistical independence is a reasonable goal for neural processing. Additionally, one might ask if there is a statistical justification for cascaded sequences of normalization computations, as has been proposed in some models of cortical processing [e.g., 24]. More practically, the normalization procedure needs to be examined in the context of a proper independent components basis. The orthonormal wavelet approximation that we have used here has the disadvantage that the diagonal bands contain a mixture of orientations. Preliminary tests indicate substitution of a better basis produces qualitatively similar results.

An essential feature that is missing from our description is *time*. In particular, our normalization computation is simultaneous and instantaneous, and we have only modeled steady-state firing rates. A more realistic implementation would involve normalization of a population of neurons in parallel using feedback or lateral connections (necessarily delayed), and would thus introduce temporal dynamics as well as higher-order effects such as disinhibition. Furthermore, our modeling and implementation are based on still images and static receptive fields. This should be augmented to include spatio-temporal behaviors such as direction-selectivity: we suspect that these properties may be derived from statistics of image sequences.

Finally, an interesting issue is the plasticity of the normalization weights: Are these fixed (i.e., globally optimized over all images), or are they modified according to the statistics of the recent visual context? Our preliminary investigations indicate that such plasticity may account for adaptation effects that have been observed physiologically.

References

- [1] F Attneave. Some informational aspects of visual perception. *Psych. Rev.*, 61:183–193, 1954.
- [2] H B Barlow. Possible principles underlying the transformation of sensory messages. In W A Rosenblith, editor, *Sensory Communication*, page 217. MIT Press, Cambridge, MA, 1961.
- [3] A J Bell and T J Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [4] A B Bonds. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.
- [5] M Carandini and D J Heeger. Summation and division by neurons in primate visual cortex. *Science*, 264:1333–1336, 1994.
- [6] M Carandini, D J Heeger, and J A Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neuroscience*, 17:8621–8644, 1997.

- [7] J F Cardoso. Source separation using higher order moments. In *ICASSP*, pages 2109–2112, 1989.
- [8] J R Cavanaugh, W Bair, and J A Movshon. Orientation-selective setting of contrast gain by the surrounds of macaque striate cortex neurons. In *Neurosci Abstracts*, volume 23, page 227.2, 1997.
- [9] P Common. Independent component analysis, a new concept? *Signal Process.*, 36:387–314, 1994.
- [10] D J Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.
- [11] W S Geisler and D G Albrecht. Cortical neurons: Isolation of contrast gain control. *Vision Research*, 8:1409–1410, 1992.
- [12] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.
- [13] D J Heeger. Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiology*, 70(5):1885–1898, 1993.
- [14] S G Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pat. Anal. Mach. Intell.*, 11:674–693, July 1989.
- [15] J Müller, J Krauskopf, and P Lennie, 1998. Center for Visual Science, University of Rochester. Unpublished data.
- [16] E Oja. A simplified neuron model as a principal component analyzer. *J. Mathematical Biology*, 15:267–273, 1982.
- [17] B A Olshausen and D J Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7:333–339, 1996.
- [18] W Reichardt and T Poggio. Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.*, 35:81–100, 1979.
- [19] D L Ruderman and W Bialek. Statistics of natural images: Scaling in the woods. *Phys Rev. Letters*, 73(6), 1994.
- [20] T D Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473, 1989.
- [21] E P Simoncelli. Normalized component analysis and the statistics of natural scenes. In *Natural Scene Statistics Meeting*, Hancock, MA, September 11-14 1997.
- [22] E P Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conf Signals, Systems and Computers*, pages 673–678, Pacific Grove, CA, November 1997. Available from <http://www.cns.nyu.edu/~eero/publications.html>
- [23] E P Simoncelli and E H Adelson. Subband transforms. In John W Woods, editor, *Subband Image Coding*, chapter 4, pages 143–192. Kluwer Academic Publishers, Norwell, MA, 1990.
- [24] E P Simoncelli and D J Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, 1998.
- [25] E P Simoncelli and O Schwartz. Derivation of a cortical normalization model from the statistics of natural images. In *Investigative Ophthalmology and Visual Science Supplement (ARVO)*, volume 39, pages S–424, May 1998.
- [26] B Wegmann and C Zetsche. Statistical dependence between orientation filter outputs used in an human vision based image code. In *Proc SPIE Visual Comm. and Image Processing*, volume 1360, pages 909–922, Lausanne, Switzerland, 1990.

Acknowledgement: Special thanks to James Cavanaugh and James Müller for providing physiological data shown in this paper.