Notes on anisotropic multichannel representations for early vision

Silvio P. Sabatini

Department of Biophysical and Electronic Engineering, University of Genoa

1 Introduction

Although the basic ideas underlying early vision appear deceptively simple and their computational paradigms are known for a long time, early vision problems are difficult to quantify and solve. Such a difficulty is often related on the representation we adopt for the visual signal, which must be capable of capturing, through proper "channels", *what* is *where* in the visual signal, that is the structural ("what") and the positional ("where") information from the images impinging the retinas. Ever since the initial formulation of the channel concept, the problem arises of jointly handling the existence of spatial frequency channels on the one hand, and of orientation channels on the other. At a local operator level, the two-dimensional (2D) Gabor filter (proposed by J. Daugman [Daugman, 1985] and S. Marcelja [Marčelja, 1980], as an extension of its one-dimensional (1D) counterpart [Gabor, 1946]) retains the optimal joint information resolution in both the domains and meets thoroughly this demand, by underlining the 2D nature of the frequency representation and thus being isomorphic to the 2D character of the spatial manifold of the visual/retinal image. In this way, 2D Gabor filters reconciled the "atomistic" description of early vision, based on local feature detection in the space domain with the "undulatory" interpretation, based on a Fourier-like decomposition into spatial-frequency components.

2 Visual features as measures in the harmonic space

The goal of early vision is to extract as much information as possible about the structural properties of the visual signal. As pointed out by [Koenderink and van Doorn, 1987] and [Adelson and Bergen, 1991], an efficient internal representation is necessary to guarantee all potential visual information can be made available for higher level analysis. The measurement of specific, significant visual "elements" in a local neighborhood of the visual signal has led to the concept of “feature” and of “feature extraction”. An image feature can be defined in terms of attributes related to the visual data. Though, in practice, many features are also defined in terms of the particular procedure used to extract information about that feature, and, in more general terms, on the specific scheme adopted for the representation of the visual signal. At an early level, feature detection occurs through initial local *quantitative* measurements of basic image properties (e.g., edge, bar, orientation, movement, binocular disparity, color) referable to spatial differential structure of the image luminance and its temporal evolution (cf. linear visual cortical cell responses, see e.g. [Jones and Palmer, 1987] [De Angelis *et al.*, 1993] [Carandini *et al.*, 2005]). Later stages in vision can make use of these initial measurements by combining them in various ways, to come up with categorical *qualitative* descriptors, in which information is used in a non-local way to formulate more global spatial and temporal predictions (e.g., see [Krüger *et al.*, 2004]).

The receptive fields of the cells in the primary visual cortex have been interpreted as fuzzy differential operators (or local *jets* [Koenderink and van Doorn, 1987]) that provide regularized partial derivatives of the image luminance along different directions and at several levels of resolution, simultaneously. The jets characterize the local geometry in the neighborhood of a given point $\mathbf{x} = (x, y)$. The order of the jet determines the amount of geometry represented. Given the 2D nature of the visual signal, the spatial direction of the derivative (i.e., the orientation of the corresponding local filter) is an important “parameter”. Within a local jet, the directionally biased receptive fields are represented by a set of similar filter profiles that merely differ in orientation.

Alternatively, considering the space/spatial-frequency duality [Gabor, 1946, Daugman, 1985], the local jets can be described through a set of independent spatial-frequency channels, which are selectively sensitive to a different limited range of spatial frequencies. These spatial-frequency channels are equally apt as the spatial ones. From this perspective, it is formally possible to derive, on a local basis, a complete harmonic representation (amplitude, phase, and orientation) of any visual stimulus, by defining the associated analytic signal in a combined space-frequency domain through filtering operations with complex-valued band-pass kernels. Since spatial information is being linearly transformed from the space domain at the level of pixels, into a combined space-frequency domain at a cortical-like representation level, no actual analysis is taking place at this level, and the information is merely being put into another (presumably more useful) form [De Valois and De Valois, 1990].

Formally, due to the impossibility of a direct definition of the analytic signal in two dimensions, a 2D spatial frequency filtering would require an association between spatial frequency and orientation channels. Basically, this association can be handled ‘separately’, for each orientation channel, by using Hilbert pairs of band-pass filters that display symmetry and antisymmetry about a steerable axis of orientation[1].

For each orientation channel θ, an image $I(\mathbf{x})$ is filtered with a complex-valued filter:

$$f_A^\theta(\mathbf{x}) = f^\theta(\mathbf{x}) - i f_{\mathcal{H}}^\theta(\mathbf{x}) \tag{1}$$

where $f_{\mathcal{H}}^\theta(\mathbf{x})$ is the Hilbert transform of $f^\theta(\mathbf{x})$ with respect to the axis orthogonal to the filter's orientation:

$$f_{\mathcal{H}}^\theta(\mathbf{x}) = f_{\mathcal{H}}(x_\theta, y_\theta) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{f(\xi, y_\theta)}{\xi - x_\theta} \mathrm{d}\xi$$

with x_θ and y_θ the principal axes of the energy distribution of the filter in the spatial domain.

This results in a complex-valued *analytic image*:

$$Q_A^\theta(\mathbf{x}) = I * f_A^\theta(\mathbf{x}) = C_\theta(\mathbf{x}) + i S_\theta(\mathbf{x}) \ , \tag{2}$$

where $C_\theta(\mathbf{x})$ and $S_\theta(\mathbf{x})$ denote the responses of the quadrature filter pair. For each spatial location, the amplitude $\rho_\theta = \sqrt{C_\theta^2 + S_\theta^2}$ and the phase $\phi_\theta = \mathrm{atan2}(S_\theta, C_\theta)$ envelopes measure the harmonic information content in a limited range of frequencies and orientations to which the channel is tuned.

2.1 Compact band-pass filtering

In the harmonic space it is in general an important requirement to have both the spatial width of the filters and the spatial frequency bandwidth small, so that good localization and good approximation of the harmonic information is realized simultaneously. Gabor functions reaching the maximal joint resolution in space and spatial frequency domains are specifically suitable for this purpose and are extensively used in computational vision [Daugman, 1985]. Different band-pass filters have been proposed as an alternative to Gabor functions, on the basis of specific properties of the basis functions [Young, 1985, Watson, 1987, Hawken and Parker, 1987, Field, 1987, Martens, 1990, Stork and Wilson, 1990, Yang, 1992, Klein and Beutter, 1992], or according to theoretical and practical considerations of the whole space-frequency transform [Mallat, 1989, Reed and Wechsler, 1990, Freeman and Adelson, 1991, Perona, 1992, Simoncelli *et al.*, 1992, Felsberg and Sommer, 2004].

[1] As an alternative a 2D isotropic generalization of the analytic signal: the monogenic signal [Felsberg and Sommer, 2001], has been proposed, which allows us to build isotropic harmonic representations that are independent of the orientation (i.e., omnidirectional). By definition, the monogenic signal is a 3D phasor in spherical coordinates and provides a framework to obtain the harmonic representation of a signal respect to the dominant orientation of the image that becomes part of the representation itself

A detailed comparison of the different filters evades the scope of these notes and numerous comparative reviews can be already found in the literature (e.g., see [Jacobson and Wechsler, 1988] [Wechsler, 1990] [Navarro *et al.*, 1996]).

Let us consider a discrete set of oriented (i.e., anisotropic) Gabor filters and the classical steerable filter approach [Freeman and Adelson, 1991] that allows a continuous steerability of the filter with respect to any orientation. Hence, it is possible in principle to steer the filter with respect to the dominant orientation of the signal, which, yet, has to be known in advance and cannot be gained from the representation itself.

For all the filters considered, we chose the design parameters to have a good coverage of the space-frequency domain and to keep the spatial support (i.e., the number of taps) to a minimum, in order to cut down the computational cost. Therefore, we determined the smallest filter on the basis of the highest allowable frequency without aliasing, and we adopted a pyramidal technique [Adelson *et al.*, 1984] as an economic and efficient way to achieve a multi-resolution analysis (see also "Notes on band-pass spatial filters"). Accordingly, we fixed the maximum radial peak frequency (ω_0) by considering the Nyquist condition and a constant relative bandwidth β around one octave, that allows us to cover the frequency domain without loss of information. The result was an 11×11 filter mask capable of resolving sub-pixel phase differences. For Gabor and steerable filters, we should also consider the minimum number of oriented filters to guarantee a uniform orientation coverage. This number depends on the filter bandwidth and it is related to the desired orientation sensitivity of the filter (e.g., see [Daugman, 1985, Fleet and Jepson, 1990]); we verified that, under our assumptions, it is necessary to use at least eight orientations. To satisfy the quadrature requirement all the even symmetric filters have been "corrected" to cancel the DC sensitivity. Where possible, the filters have been expressed as sums of x-y separable functions to implement separate 1D convolutions instead of 2D convolutions in a similar way that [Nestares *et al.*, 1998], with a consequent further drop of the computational burden.

3 Filter design specification

Here we present a detailed description of the filters used.

Gabor filters - A Gabor oriented filter along an angle θ with respect to the horizontal axis is defined by:

$$f^{\theta}_{\mathrm{Gabor}}(x,y) = \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}\,\mathrm{e}^{j\omega_0(x\cos\theta+y\sin\theta)}$$

where ω_0 is the peak frequency of the filter and σ determines its spatial extension. The spatial window has been chosen as four times σ. At the highest scale $\omega_0 = \pi/2$ and $\sigma = 2.67$. Following

[Nestares *et al.*, 1998], we implemented the oriented filters as sums of separable filters. By exploiting symmetry considerations, all eight even and odd filters (see Fig. 1) can be constructed on the basis of twentyfour 1D convolutions. The 1D filters are modified by enforcing zero DC sensitivity on the resulting 2D filters in which they take part, and by minimizing the difference with the theoretical 2D Gabor filters. Specific care have been paid to adjust the coefficients of each filter function so that the even and odd symmetry is respected. To this purpose, a constrained non-linear multivariable minimization is adopted.

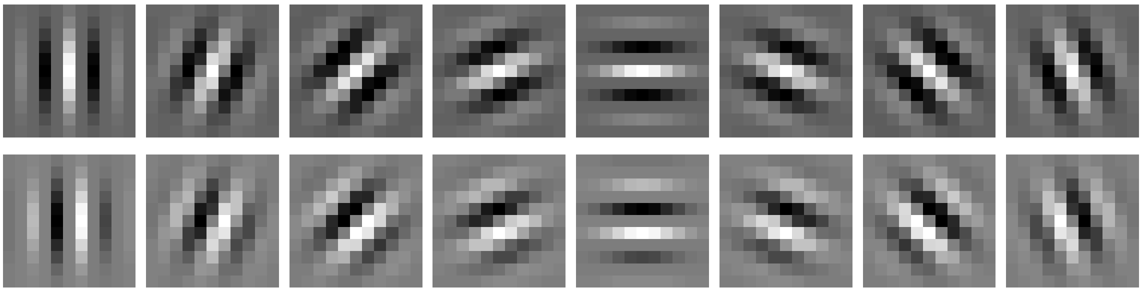

Figure 1: The resulting 11×11 quadrature pair of Gabor filters for $\omega_0 = \pi/2$ and 8 orientations.

Steerable filters - Following [Freeman and Adelson, 1991], an approximation of a complex-valued Gabor filter of arbitrary orientation θ can be synthesized by taking a linear combination of steerable quadrature pairs of 2D Gaussian directional derivatives, along the cardinal axes:

$$g_0(x,y) = \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$g_L(x,y) = \frac{\partial^{L-l}}{\partial x^{L-l}} \frac{\partial^l}{\partial y^l} g_0(x,y) \quad l = 0, 1, \ldots, L-1$$

$$f^{\theta}_{\mathrm{Steer}}(x,y) = g_0(x) \sum_{l=1}^{L} b_l(\theta) P_{l,\sigma}(x) Q_{l,\sigma}(y)$$

where $b_l(\theta)$ are the interpolation functions:

$$b_l(\theta) = (-1)^l \begin{pmatrix} L \\ l \end{pmatrix} \cos^{L-l}\theta \sin^l \theta$$

L is the order of differentiation, and P_l and Q_l are polynomial functions defined as:

$$P_{l,\sigma}(x) Q_{l,\sigma}(y) = \left(\frac{x^{L-l}}{\sigma^{2(L-l)}} + \cdots \right) \left(\frac{y^l}{\sigma^{2l}} + \cdots \right).$$

Gaussian derivatives asymptotically coincide to a Gabor function with a radial peak frequency $\omega_0 = \sigma^{-1}\sqrt{L+1}$ and an absolute bandwidth $\Delta\omega = \sigma^{-1}\sqrt{2}/2$ [Koenderink and van Doorn,

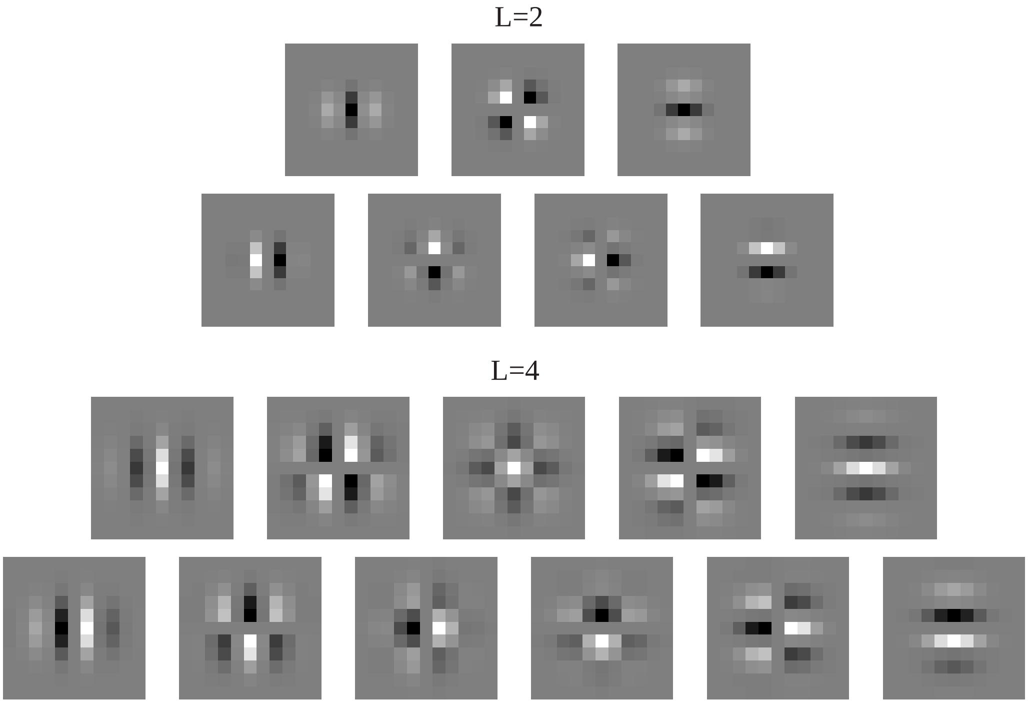

Figure 2: The 11×11 x-y separable, steerable quadrature pair basis filters for two different orders of differentiation. The width of the Gaussian function has been adjusted to have, for both cases, a resulting $\omega_0 = \pi/2$: $\sigma = 0.90$ for $L = 2$ and $\sigma = 1.27$ for $L = 4$.

1987]. Since the peak frequency and the bandwidth are jointly defined by σ, it is not possible to design banks of steerable filters with an arbitrarily constant relative bandwidth. We therefore adjusted the spatial extension of the Gaussian function (σ) in order to have the same peak frequency of the Gabor filters, and deduced as a consequence the relative bandwidth. The number of basis kernels to compute the oriented outputs of the filters depends on their derivative order. The quadrature pair of these filters has been obtained by approximating their Hilbert transform as a the least square fit to a polynomial times a Gaussian described in [Freeman and Adelson, 1991]. The basis filters corresponding to Gaussian derivatives of 2nd- or 4th-order (see Fig. 2) turned out as an acceptable compromise between the representation efficacy (i.e., optimality in terms of the Heisenberg-Weyl uncertainty principle) and the computational efficiency.

All the filters have been normalized prior to their use in order to have constant energy. The corresponding rosette-like frequency representation of the filters used is shown in Fig. 3, for three different scales (octaves).

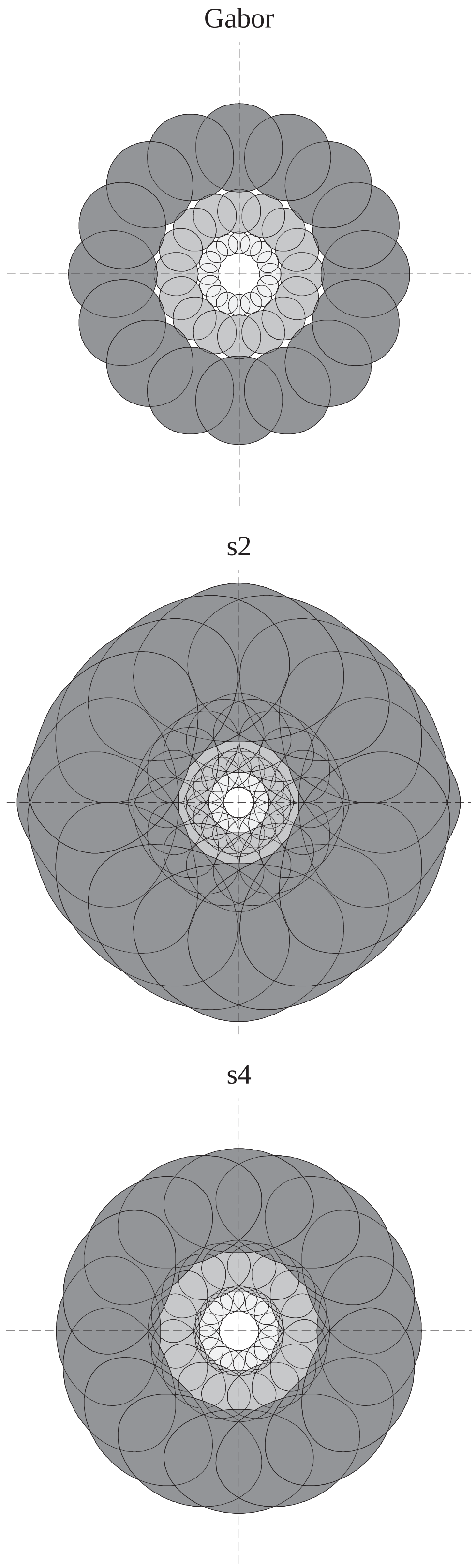

Figure 3: Rosette-like diagrams of the multichannel frequency representation for the Gabor, and the steerable filters s2 ($L = 2$) and s4 ($L = 4$), respectively. It is worth noting that the orientation bandwidth of the steerable filters is larger than that obtained with Gabor filters. Contours correspond to half-width cut-off frequencies, and each corona is separated by an octave scale.

References

[Adelson and Bergen, 1991] E.H. Adelson and J.R. Bergen. The plenoptic and the elements of early vision. In M.S. Landy and J.A. Movshon, editors, *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[Adelson *et al.*, 1984] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, and J.M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6):33–41, 1984.

[Carandini *et al.*, 2005] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, A. Olshausen, J.L. Gallant, and N.C. Rust. Do we know what the early visual system does? *J. Neuroscience*, 25:10577–10597, 2005.

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

[De Angelis *et al.*, 1993] G.C. De Angelis, I. Ohzawa, and R.D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J. Neurophysiol.*, 69:1118–1135, 1993.

[De Valois and De Valois, 1990] R.L. De Valois and K.K. De Valois. *Spatial Vision*. Oxford University Press, New York, 1990.

[Felsberg and Sommer, 2001] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 48(12):3136–3144, 2001.

[Felsberg and Sommer, 2004] M. Felsberg and G. Sommer. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision*, 21:5–26, 2004.

[Field, 1987] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Amer. A*, 4:2379–2394, 1987.

[Fleet and Jepson, 1990] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 1:77–104, 1990.

[Freeman and Adelson, 1991] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13:891–906, 1991.

[Gabor, 1946] D. Gabor. Theory of communication. *J. Inst. Elec. Eng.*, 93:429–459, 1946.

[Hawken and Parker, 1987] M.J. Hawken and A.J. Parker. Spatial properties of neurons in the monkey striate cortex. *Proc. Roy. Soc. Lond. B*, 231:251–288, 1987.

[Jacobson and Wechsler, 1988] L.D. Jacobson and H. Wechsler. Joint spatial/spatial-frequency representation. *Signal Processing*, 14:37–68, 1988.

[Jones and Palmer, 1987] J. Jones and L. Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neuroscience*, 58:1187–1211, 1987.

[Klein and Beutter, 1992] S.A. Klein and B. Beutter. Minimizing and maximizing the joint space-spatial frequency uncertainty of Gabor-like functions: comment. *J. Opt. Soc. Amer. A*, 9(2):337–340, 1992.

[Koenderink and van Doorn, 1987] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. *Biol. Cybern.*, 55:367–375, 1987.

[Krüger *et al.*, 2004] N. Krüger, M. Lappe, and F. Wörgötter. Biologically motivated multi-modal processing of visual primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5):417–428, 2004.

[Mallat, 1989] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7):674–693, 1989.

[Martens, 1990] J.B. Martens. The Hermite transform - Theory. *IEEE Trans. Acoust., Speech, Signal Processing*, 38:1595–1606, 1990.

[Marčelja, 1980] S. Marčelja. Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Amer. A*, 70:1297–1300, 1980.

[Navarro *et al.*, 1996] R. Navarro, A. Tabernero, and G. Cristobal. Image representation with gabor wavelets and its applications. In P. W. Hawkes, editor, *Advances in Imaging and Electron Physics*, pages 1–84. Academic Press, San Diego CA, 1996.

[Nestares *et al.*, 1998] O. Nestares, R. Navarro, J. Portilla, and A. Tabernero. Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *Journal of Electronic Imaging*, 7(1):166–173, 1998.

[Perona, 1992] P. Perona. Steerable-scalable kernels for edge detection and junction analysis. *Image Vis. Comput.*, 10:663–672, 1992.

[Reed and Wechsler, 1990] T. Reed and H. Wechsler. Segmentation of textured images and gestalt organization using spatial/spatialfrequency representations. *IEEE Trans. Pattern Analysis Mach. Intell.*, 12:1–12, 1990.

[Simoncelli *et al.*, 1992] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger. Shiftable multiscale transforms. *IEEE Trans. on Information Theory*, 38(2):587–607, 1992.

[Stork and Wilson, 1990] D.G. Stork and H.R. Wilson. Do Gabor functions provide appropriate decriptions of visual cortical receptive fields? *J. Opt. Soc. Amer. A*, 7(8):1362–1373, 1990.

[Watson, 1987] A.B. Watson. The cortex transform: rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing*, 39:311–327, 1987.

[Wechsler, 1990] H. Wechsler. *Computational Vision*. Academic Press, 1990.

[Yang, 1992] J. Yang. Do Gabor functions provide appropriate decriptions of visual cortical receptive fields?: comment. *J. Opt. Soc. Amer. A*, 9(2):334–336, 1992.

[Young, 1985] R.A. Young. The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles. Technical Report GMR-4920, General Motors Research, 1985.