From filters to features: Scale–space analysis of edge and blur coding in human vision

Mark A. Georgeson School of Life & Health Sciences, Aston University, Birmingham, UK

Keith A. May Department of Optometry, University of Bradford, Bradford, UK

Tom C. A. Freeman School of Psychology, Cardiff University, Cardiff, UK

Gillian S. Hesse Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK

To make vision possible, the visual nervous system must represent the most informative features in the light pattern captured by the eye. Here we use Gaussian scale–space theory to derive a multiscale model for edge analysis and we test it in perceptual experiments. At all scales there are two stages of spatial filtering. An odd-symmetric, Gaussian first derivative filter provides the input to a Gaussian second derivative filter. Crucially, the output at each stage is half-wave rectified before feeding forward to the next. This creates nonlinear channels selectively responsive to one edge polarity while suppressing spurious or "phantom" edges. The two stages have properties analogous to simple and complex cells in the visual cortex. Edges are found as peaks in a scale–space response map that is the output of the second stage. The position and scale of the peak response identify the location and blur of the edge. The model predicts remarkably accurately our results on human perception of edge location and blur for a wide range of luminance profiles, including the surprising finding that blurred edges look sharper when their length is made shorter. The model enhances our understanding of early vision by integrating computational, physiological, and psychophysical approaches.

Keywords: human vision, feature analysis, edges, blur, scale–space theory, spatial filters, Gaussian derivatives, half-wave rectification, visual cortex

Citation: Georgeson, M. A., May, K. A., Freeman, T. C. A., & Hesse, G. S. (2007). From filters to features: Scale–space analysis of edge and blur coding in human vision. *Journal of Vision, 7*(13):7, 1–21, http://journalofvision.org/7/13/7, doi:10.1167/7.13.7.

Introduction

Important structure in an image may occur at any spatial scale—from the sharp edges of objects to the smooth shading on curved surfaces or the very blurred boundaries of some shadows—and so image computations must be carried out at many spatial scales (ter Haar Romeny, 2003). Neurophysiological studies show that receptive fields (RFs) of visual cortical cells vary greatly in size and typically show a tuned, band-pass response to spatial sine waves (gratings) of different spatial frequencies (De Valois, Albrecht, & Thorell, 1982; Kulikowski, Marcelja, & Bishop, 1982; Movshon, Thompson, & Tolhurst, 1978). Psychophysical studies support the idea that a set of such cells can act as a multiscale bank of spatial filters (Blakemore & Campbell, 1969; De Valois & De Valois, 1990; Wilson, 1983), and natural image analysis shows that such filters provide an efficient initial coding of the image data that is well matched to image statistics (Field, 1987; Olshausen & Field, 1996; van Hateren & van der Schaaf, 1998). But to what end? The concept of spatial frequency filtering has been a driving force in vision research for 35 years, yet it remains unclear how the output of this multiscale population of cells or filters is used to locate and describe the key features in images, and despite much progress (Canny, 1986; Elder & Zucker, 1998; Georgeson, 1992, 1998; Marr & Hildreth, 1980; Morrone & Burr, 1988; Morrone & Owens, 1987; Peli, 2002; van Deemter & du Buf, 2000; Watt & Morgan, 1985), there is no adequate standard model of feature analysis for human vision (Hesse & Georgeson, 2005).

It is widely agreed that edges are key features in images (Geisler, Perry, Super, & Gallogly, 2001). But what is an edge? It is tempting to offer a priori definitions—for example, that an edge is a point of maximum luminance gradient, or a point of maximum phase congruence. In a brief passage, Helmholtz argued that when the gradient in

doi: 10.1167/7.13.7 Received December 13, 2006; published October 19, 2007 ISSN 1534-7362

the optical image of a step edge was very steep, "this sudden drop in intensity enables the eye to recognize the position of the edge," whereas with more blur "the falling off is more gradual, so that there is nothing to indicate exactly where the edge is" (Helmholtz, 1856/2000, p. 184). Rather than appealing to gradient magnitude, Ernst Mach (1865/1965), while citing Helmholtz's view, cautioned that "According to my experience I must also hold as very essential the transition from concave to convex and the point of inflection" in the light distribution. Mach thus anticipated by over a century Marr's notion that an edge is indicated by zero crossings in the second derivative (Marr & Hildreth, 1980). But, in the face of such uncertainty, instead of trying to prescribe what an edge is, we might rather adopt an inductive approach in which we systematically gather data on the perception of edges and then compare different models of edge coding. The simplest model that accounts well for a wide range of data may then offer a definition of what an edge is for human vision.

Here we develop a theoretical framework for the encoding of luminance edges. In earlier work (Georgeson & Freeman, 1997; Hesse & Georgeson, 2005), we found that points of phase congruence (Kovesi, 2000; Morrone & Burr, 1988; Morrone & Owens, 1987) did not, in general, predict the perceived locations of edges nor the Vernier alignment of contours, whereas an approach based on points of maximum gradient was more successful. We now expand that approach into a more comprehensive, multiscale framework and consider the potential role of derivative operators higher than the first. We apply it to human vision through psychophysical experiments on the perception of edge blur and edge location. One specific, nonlinear model emerges as strikingly successful in predicting all the perceptual results, and we find that the "channels" for encoding edges have two stages of spatial filtering that are analogous respectively to simple cells and to complex cells in the visual cortex.

Theory

A principled approach is needed, and we draw on Gaussian scale–space theory (Koenderink, 1984; Lindeberg, 1998; ter Haar Romeny, 2003; Witkin, 1983). A key idea is that features should be encoded using filters whose scale is related to the scale (or blur) of the feature. Because the scale of the feature is unknown and variable across space, a method is needed for automatic, local, scale selection (Lindeberg, 1998; Morrone, Navangione, & Burr, 1995; van Warmerdam & Algazi, 1989). Scale–space theory shows that to make observations of an input signal at a given scale σ, without prior assumptions about the structure of the input, the natural (indeed unique) "window" of observation (ter Haar Romeny, 2003) is the Gaussian function

$$G(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\sigma^2}\right). \quad (1)$$

The basic scale–space representation of an input image $I(x)$ then is effectively a stack of images resulting from Gaussian smoothing of the input image at different scales. Computations about image structure are based on the spatial derivatives of these Gaussian-blurred images. For example, the Gaussian first derivative scale–space representation of the image gradients is given by

$$L_1(x;\sigma) = \frac{\partial}{\partial x}(I(x) \otimes G(x;\sigma)) = I(x) \otimes \frac{\partial G(x;\sigma)}{\partial x}, \quad (2)$$

which is the convolution of the input image $I(x)$ with a set of Gaussian derivative operators at different scales.

Higher derivatives of order n can be treated in the same way:

$$L_n(x;\sigma) = I(x) \otimes \frac{\partial^n G(x;\sigma)}{\partial x^n}, \quad (3)$$

where n = 1, 2, 3…. RFs of simple cells in the visual cortex can often be well described as nth order Gaussian derivative operators (Figure 1A), where the RF has $n + 1$ adjacent ON and OFF subregions (Young & Lesperance, 2001). Reports of one to four or even five lobes in the simple cell RF (Hubel & Wiesel, 1962; Movshon et al., 1978; Kulikowski & Bishop, 1981; Camarda, Peterhans, & Bishop, 1985; DeAngelis, Ohzawa, & Freeman, 1993; Ringach, 2002) suggest that derivative operators up to third or fourth order could play a role in spatial vision.

Figure 2B is a scale–space map showing the gradient response pattern L_1 across space (x) and scale (σ), given two Gaussian-blurred edges as input. How might the location, polarity, and blur of these edges be identified? The sign of the responses corresponds to the polarity of the edge (dark–light or light–dark), but the biggest response is always at the finest scale (1 pixel; top row of the map). This is true even for much larger blur (Figure 2G). The degree of blur is implicit in how far the response extends through scale, and across space, but further analysis would be needed to quantify it. The edge would be explicitly identified, however, if the response pattern had a unique peak at the scale and location of the edge. Lindeberg (1998) devised a powerful, general method of "scale selection" to achieve this goal. Lindeberg explored algorithms in which the response measures for localization and for scale selection could be different. We simplified his method by supposing that peaks in a single response surface (Equation 4) should identify the position

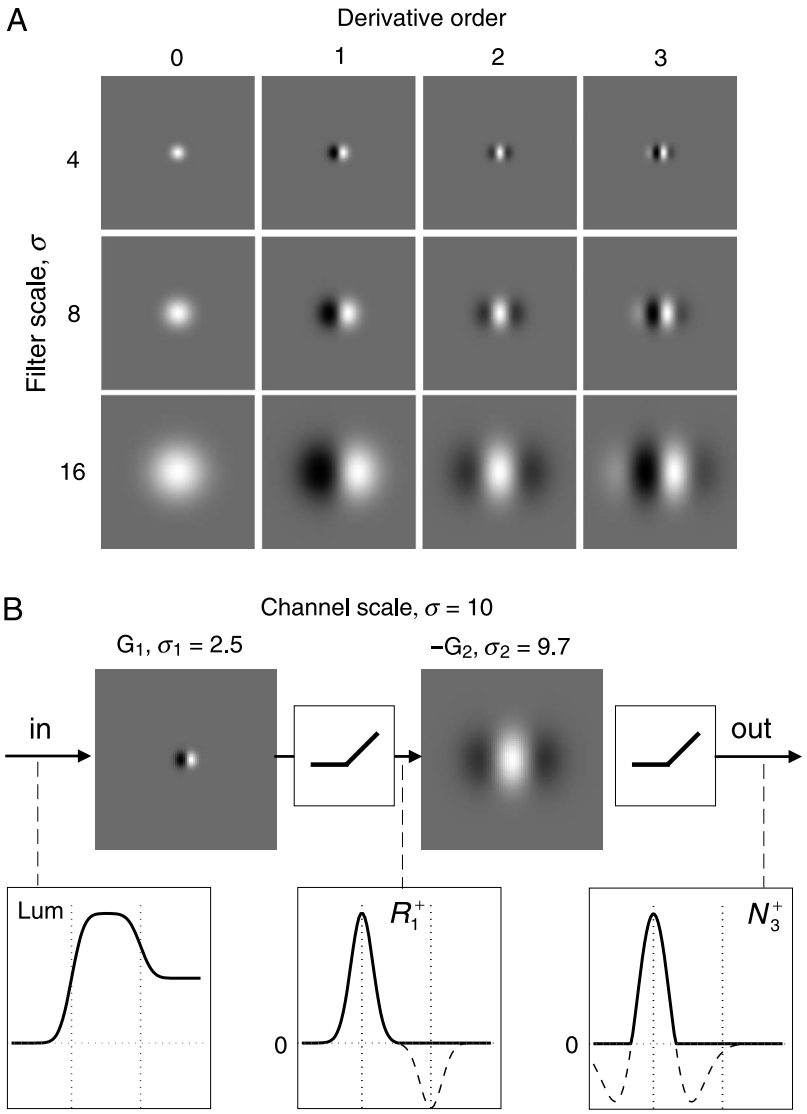

Figure 1. (A) Receptive fields (RFs) of Gaussian derivative spatial filters up to Order 3, at several scales. Sign (polarity) of RF has been inverted for Orders 2 and 3. (B) Proposed nonlinear, third derivative channel (N_3^+) for edge analysis. Channel scale is given by $\sigma = \sqrt{(\sigma_1^2 + \sigma_2^2)}$. Bottom row: input luminance profile (left) has two blurred edges of opposite polarity, but this channel responds only to the positive-going edge. The first half-wave rectifier suppresses the first filter's response (centre, dashed curve) to a negative edge. The second rectifier vetoes negative responses (dashed curves, right) introduced by the second filter, leaving an unambiguous response peak at the positive edge location.

and scale of the features. The Gaussian derivative operators are multiplied by a scale-dependent gain factor σ^α to give a normalized scale–space representation, N_n:

$$N_n(x;\sigma) = \sigma^\alpha I(x) \otimes \frac{\partial^n G(x;\sigma)}{\partial x^n}. \quad (4)$$

By differentiating N_n with respect to σ, it is straightforward to show that, for a given class of feature such as Gaussian edge, bar, or blob, one can choose the exponent α such that the response always peaks at the true location and scale of the feature (Lindeberg, 1998). For Gaussian-blurred edges (Figures 2A and 2F), $\alpha = n / 2$ (where n is an odd integer, implying odd-symmetric RFs). Figures 2C and 2H show that with this multiscale representation of gradients, N_1, there are unique response peaks at the correct locations and scales for the edges in the image. These edge features have been "made explicit" (Marr, 1982). Importantly, this rendering of edge finding as a simple peak-finding problem seems to reduce, or perhaps eliminate, the need for more elaborate algorithms that combine or "track" edge information across scales (Bergholm, 1987; Zhang & Bergholm, 1997). Figure A1 (Appendix A) shows how the normalization converts monotonic response profiles across scale into peaked ones

that identify the edge blur. Appendix A analyzes some of the scaling properties of these scale-normalized derivatives.

These examples used a first derivative scheme. Cortical RFs, however, often have more than two subregions, and the bandwidths of physiologically and psychophysically identified filters are often narrower than the Gaussian first derivative, but similar to the third derivative (Bruce, Green, & Georgeson, 2003). Hence, we should consider the possible role of higher, odd-order derivatives in edge finding. Some initial evidence for a third derivative edge operator comes from our recent finding of "Mach edges"—the edges of *Mach Bands* (Georgeson, 2006). Although the classic light and dark bands correspond to peaks and troughs in the second derivative (Ratliff, 1965; Watt & Morgan, 1985), we find that in triangle wave gratings, blurred to different extents, Mach edges are seen at peaks in the third derivative where there are no corresponding peaks in the first derivative nor zero crossings in second derivative, at any scale. This suggests that peaks in the third derivative are a cue to edge finding, but there is a crucial problem to be overcome.

At first sight, the third derivative representation N_3 appears unpromising because the additional differentiation introduces false-positives—two extra peaks or troughs in the response to each edge (Figures 2D and 2I). To avoid this problem in machine vision, Lindeberg (Lindeberg, 1998) used the first derivative to locate edges and the third derivative to identify their scale.

We have discovered (or rediscovered; Kovasznay & Joseph, 1955) that a simple, physiologically plausible modification to the linear N_3 scheme solves the multiple peaks problem and makes accurate predictions about perceived edge location and blur. The necessary modification is to split the differentiation into two stages and to transmit only the positive parts of the response at each stage (Figure 1B). We denote the half-wave rectified first stage output as R_1^+ and the normalized second stage output as N_3^+:

$$R_1^+(x;\sigma_1) = \max\left[I(x) \otimes \frac{\partial G(x;\sigma_1)}{\partial x}, 0\right], \quad (5)$$

$$N_3^+(x;\sigma) = \sigma^{1.5} \max\left[R_1^+(x;\sigma_1) \otimes \frac{-\partial^2 G(x;\sigma_2)}{\partial x^2}, 0\right], \quad (6)$$

where $\sigma = \sqrt{(\sigma_1^2 + \sigma_2^2)}$ and $\sigma_1 = \sigma/4$. The choice of $\sigma_1 = \sigma/4$ was an empirical one, but we show later that the choice is not critical provided $0 < \sigma_1 < \sigma/2$; hence, we chose a fixed value ($\sigma/4$) in the middle of that range. It was not adjusted to fit individual experiments. The sign reversal introduced in Equation 6 simply ensures a positive output for edges of positive gradient. This scheme creates a nonlinear filter or "channel" (Figure 1B) that is responsive to edges of only one polarity. The first rectifier vetoes regions of negative gradient, and the second rectifier eliminates the negative troughs of response flanking an edge with positive gradient. Thus, the response N_3^+ (Figures 2E and 2J) is a simple unimodal surface whose peak location and scale correctly identify the position and blur of a dark–light edge, but with no response to a light–dark edge. The problem of identifying the edge is reduced to simple peak finding, with no false-positives.

The N_3^+ response surface (e.g., Figure 2J) is much more compact than the N_1 surface (Figure 2H), implying better resolution for N_3^+ in both space and scale. Indeed, we found that N_1 failed to resolve separate peak responses when two edges of the same polarity and blur b were separated by less than $4.8b$, but N_3^+ resolved the two edges down to a separation of only $2.5b$ (Figure S1).

To identify dark–light edges a channel of the opposite polarity is obviously needed. This channel, denoted N_3^-, is simply obtained by reversing the sign of the first filter:

$$R_1^-(x;\sigma_1) = \max\left[I(x) \otimes \frac{-\partial G(x;\sigma_1)}{\partial x}, 0\right], \quad (7)$$

$$N_3^-(x;\sigma) = \sigma^{1.5} \max\left[R_1^-(x;\sigma_1) \otimes \frac{-\partial^2 G(x;\sigma_2)}{\partial x^2}, 0\right]. \quad (8)$$

We applied these Gaussian scale–space models to edge analysis in human vision. Previous psychophysical studies have examined and modelled discrimination of edge blur and edge position (Watt, 1988; Watt & Morgan, 1983), and the Viewprint model (Klein & Levi, 1985) offered a detailed scale–space framework for spatial discrimination (hyperacuity) thresholds, based on Cauchy filters rather than Gaussian derivatives. To study the perceptual representation of edges and their blur, we have found blur-matching experiments especially useful. The general aim is to find pairs of edge profiles that are physically different (e.g., sine wave vs. Gaussian integral) but perceptually matched in blur. Different models can then be tested to see how well or badly they predict the observed equivalence.

We should emphasize from the start that all our images were well above threshold (at contrasts around 30%) because our aim was to study the coding of visible features at high (but not saturated) signal levels, rather than the detection of edges at low signal–noise ratios. The

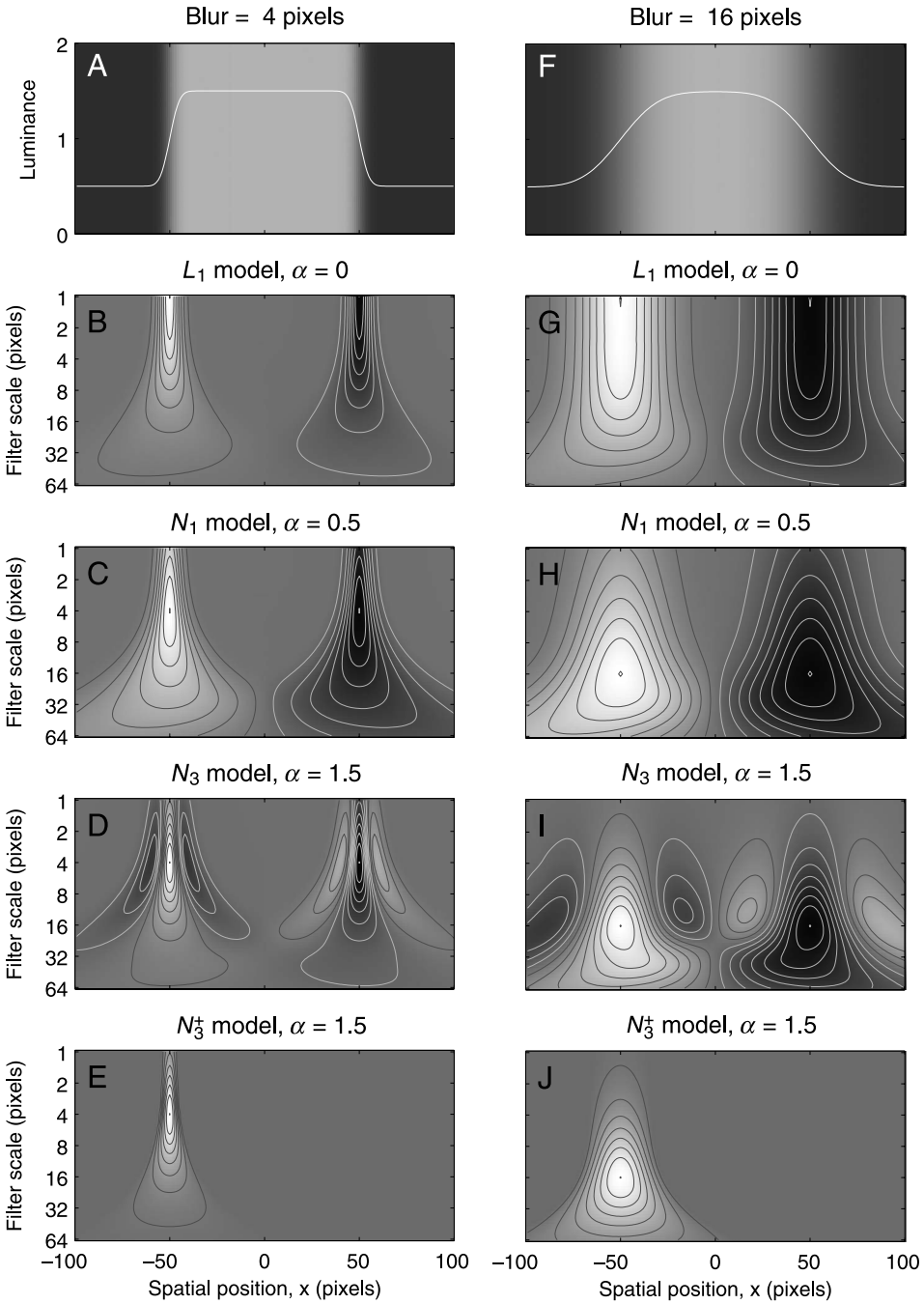

Figure 2. Multiscale Gaussian derivative models for edge analysis. (A) Input image is two Gaussian-blurred edges of opposite polarity. White curve is the luminance profile, $I(x)$. (B) Scale–space response map L_1: spatial distribution of responses from a set of Gaussian first derivative filters at different scales (σ = 1 to 64 pixels). Grayscale codes magnitude of response—positive (light) or negative (dark). Midgray is zero response. Smooth curves are level contours on the response surface at equally spaced heights. The filters were not "scale normalized" (i.e., α = 0); receptive fields (RFs) in one dimension were all derivatives of a unit-area Gaussian. Peak response to any edge occurs at the smallest filter scale. (C) As panel B, but filter output N_1 is "normalized" by the factor σ^{α} (see text). Peak response scale matches the edge blur (4 pixels). (D) As panel C, but for normalized third derivatives, N_3. The two additional derivative operations create two extra response peaks or troughs around each edge. (E) As panel D, but for the nonlinear channel N_3^{+} (Figure 1B). The first rectifier makes the channel sensitive only to positive edges; the second rectifier removes the flanking responses. (F–J) As panels A–E, but for input edges 4× more blurred. N_1 and N_3^{+} encode edge location and blur unambiguously by the scale–space position of the peak response, but L_1 and N_3 do not. Our psychophysical blur-matching experiments consistently favor the nonlinear third derivative mechanism, N_3^{+}.

Experiment no.	1, 2 (Figure 3)	3, 4, 5 (Figures 4, 5, and 6)
Computer graphics	Macintosh G4	Windows PC + VSG card
Software	NIH Image	Pascal + VSG
Grayscale monitor	Eizo 6600	Eizo 6500
Frame rate	75 Hz	60 Hz (90 Hz in Experiment 5)
Mean luminance	34 cd/m^2	75 cd/m^2
Test image size	256 × 256 pixels	512 × 512 pixels
Test image window	4.3° × 4.3° square with sharp edges	5° × 5° circle, with smoothed edges
Gray background field	17.2° (W) × 12.9° (H)	6.1° diameter disk
Test duration	300 ms	230 ms
Interstimulus interval	300 ms	580 ms
Interval order	Random order	Test first, comparison second
Polarity of test and comparison edges	Same polarity within trials; varied between trials	Same polarity within trials; varied between trials
Staircase rule	one-up, one-down	one-up, one-down
Staircase step size after first two reversals	1 dB (0.05 log unit)	1 dB
Order of trials within sessions	2 staircases per condition. All conditions randomly interleaved	2 staircases per condition. Conditions run in randomly ordered blocks of 20 trials, until all staircases completed.
No. of reversals used to estimate a match	8	10
No. of matches per subject per condition	4	4–6

Table 1. Summary of display and procedural details for blur matching in Experiments 1–2 and 3–5.

theoretical issues raised seem to us to be quite different in these two experimental settings. Hence, all our modelling assumes, for simplicity, a noise-free system.

Methods

Blur matching—general method

On each trial, a test image and a comparison image were successively presented for 230 or 300 ms. The observer judged which temporal interval contained the more blurred edge. The comparison image was always a single Gaussian-blurred edge, with the same polarity as the test edge, and its blur b_0 was adjusted by computer over a series of trials, using a standard, one-up, one-down double interleaved staircase procedure to find the comparison blur that appeared to match the test edge blur, estimated as the average of the last 8 or 10 reversal points in the staircase run. The edges were always vertical. Their contrast polarity randomly varied from trial to trial. If the test image contained multiple edges (e.g., a sine wave grating), the observer judged the edge that was in the centre of the screen. Screen luminance was calibrated with a digital photometer. Lookup tables in the graphics card were used to linearize (gamma correct) the display and control the contrast of the image. Further details of procedure are given in Table 1, with illustrations in the Supplementary methods.

Blur mixture experiment

The test edge I_{mix} consisted of two Gaussian edges superimposed at the same location, with the same polarity, but different blurs (b_1, b_2):

$$I_{\mathrm{mix}}(x) = I_0 \begin{pmatrix} 1 + c_1 \,.\, [2\Phi(x - x_0, b_1) - 1] \\ + c_2 \,.\, [2\Phi(x - x_0, b_2) - 1] \end{pmatrix}, \quad (9)$$

where I_0 is the fixed mean luminance, x_0 is the position of the edge, $\Phi(x, b)$ is the integral of the unit-area Gaussian $G(x, b)$ with space constant (blur) b, and c_1 and c_2 are the contrasts of the two component edges. Both edges were at the centre of the screen ($x_0 = 0$), and overall contrast $c_1 + c_2$ was constant (0.3). The ratio of contrasts of the two component edges $r = c_1/c_2$ was 0.1, 1, 3, 10, 30, or 100 and the component blurs (b_1, b_2) were (15, 5) or (30, 10) arcmin. Test intervals were 300 ms, separated by a gray (mean luminance) interval of 300 ms. The comparison

image in all experiments was a Gaussian edge of contrast $c = 0.3$, whose blur b_0 was varied by the staircase routine:

$$I_{\text{gauss}}(x) = I_0(1 + c.[2\Phi(x, b_0) - 1]). \quad (10)$$

Sharpened edge experiment

Procedure was similar to the blur mixture experiment. Test edges were modified ("sharpened") versions of a Gaussian edge whose original blur was $b = 10$, 20, or 30 arcmin. The local contrast function $C(x) = [I(x) - I_0]/I_0$ of the original Gaussian edge was defined by $C(x) = [2\Phi(x, b) - 1]$. The waveform $C(x)$ (range -1 to $+1$) was passed through a Naka–Rushton nonlinear transformation $C/(s + |C|)$, and after appropriate scaling of amplitude the luminance profile of the modified edge was

$$I_{\text{sharp}}(x; s) = I_0\left(1 + \frac{c.C(x)}{s + abs(C(x))} \cdot (s + \max(C(x)))\right). \quad (11)$$

Values of s were 0.1, 0.3, 1, 3, 10, 100, and 1000. The maximum gradient increased as s decreased, but Michelson contrast ($c = 0.3$) was held constant.

Sine wave edge experiment

Test waveforms were a half-period edge, single-period edge, and a sine wave grating filling the 5° display aperture—see Figure 4. Six spatial frequencies were tested, ranging from 0.354 to 2.0 cycles/deg in half-octave steps (with corresponding half-periods from 84.8 to 15 arcmin). Michelson contrast was 0.32, duration 230 ms.

Gaussian derivative experiment

Luminance profiles of the test images were defined as Gaussian derivative profiles of odd-order (-1, 1, 3, 5) in the x-direction with a flat profile in the y-direction at three scales (5.7, 11.3, and 22.6 arcmin). Subjects matched the blur of the central edge. Order -1 is a baseline condition where test and comparison edges are both Gaussian integrals. Michelson contrast was 0.32, duration 230 ms.

Length experiment

Test edge profiles were Gaussian integrals and four test blurs were used (2.8, 5.7, 11.3, and 22.6 arcmin). Length of the test edge was truncated sharply and symmetrically about the centre of the circular 5° test window. Comparison edge was a Gaussian integral, 5° long, filling the display window as usual. Michelson contrast 0.32, duration 230 ms.

Results

Blur mixture

When two edges with different blurs (b_1, b_2) and contrasts (c_1, c_2) are added together in the same place, how blurred does the resulting edge appear to be? Overall contrast was constant ($c_1 + c_2 = 0.3$), whereas the contrast ratio ($r = c_1/c_2$) was varied to determine what contribution the two component edges made to the mixture.

Symbols in Figures 3A and 3B plot the blur-matching values b_0 as the mixture ranged from mainly the smaller blur (b_2) through to mainly the larger blur (b_1). Not surprisingly, as more of the large blur entered the mixture, the perceived (matched) blur increased from b_2 to b_1 (Figure 3A: 5–15 arcmin; or Figure 3B: 10–30 arcmin). The interest lies in the manner of this transition. Firstly, vision does not simply average the blurs. Matched blur was not well predicted by a contrast-weighted average of the two component blurs $(c_1b_1 + c_2b_2)/(c_1 + c_2)$ (Figures 3A and 3B, black dash-dot curve). When the component edges had equal contrast ($r = 1$), the matched blur remained almost equal to the smaller blur value. By interpolation, the contrast of the more blurred edge had to be about five times greater than the less blurred edge (around $r = 5$) before the matched blur reached the average of the component blurs. Blur averaging then does not explain the perception of blur mixtures. (A referee asked whether compressive transformation of contrast before averaging might improve the fit. To test this, we applied a power function [exponent p] to the contrast values before calculating the weighted average blur. For p around 0.6, there was a small [9%] improvement in RMS error, but the fit to the data remained poor.)

A second simple (but incorrect) idea is the luminance template hypothesis. Suppose that, in selecting a blur match, the observer chose the Gaussian edge whose luminance profile was most highly correlated with the test edge. The predicted blur matches (Figures 3A and 3B, black dashed curve) were similar to blur averaging. In contrast to these poor fits, the predictions of the first derivative N_1 model (thin red curve) fell much closer to the experimental data.

The most accurate of all the models we explored was the third derivative model N_3 or N_3^+. (The two versions make identical predictions for these experiments, in which the luminance gradients across a given test edge are ≥ 0 everywhere [or, with reversed polarity, ≤ 0 everywhere].) The fit was good for one observer (K.A.M.) and excellent for the other (M.A.G.). There are no free parameters here,

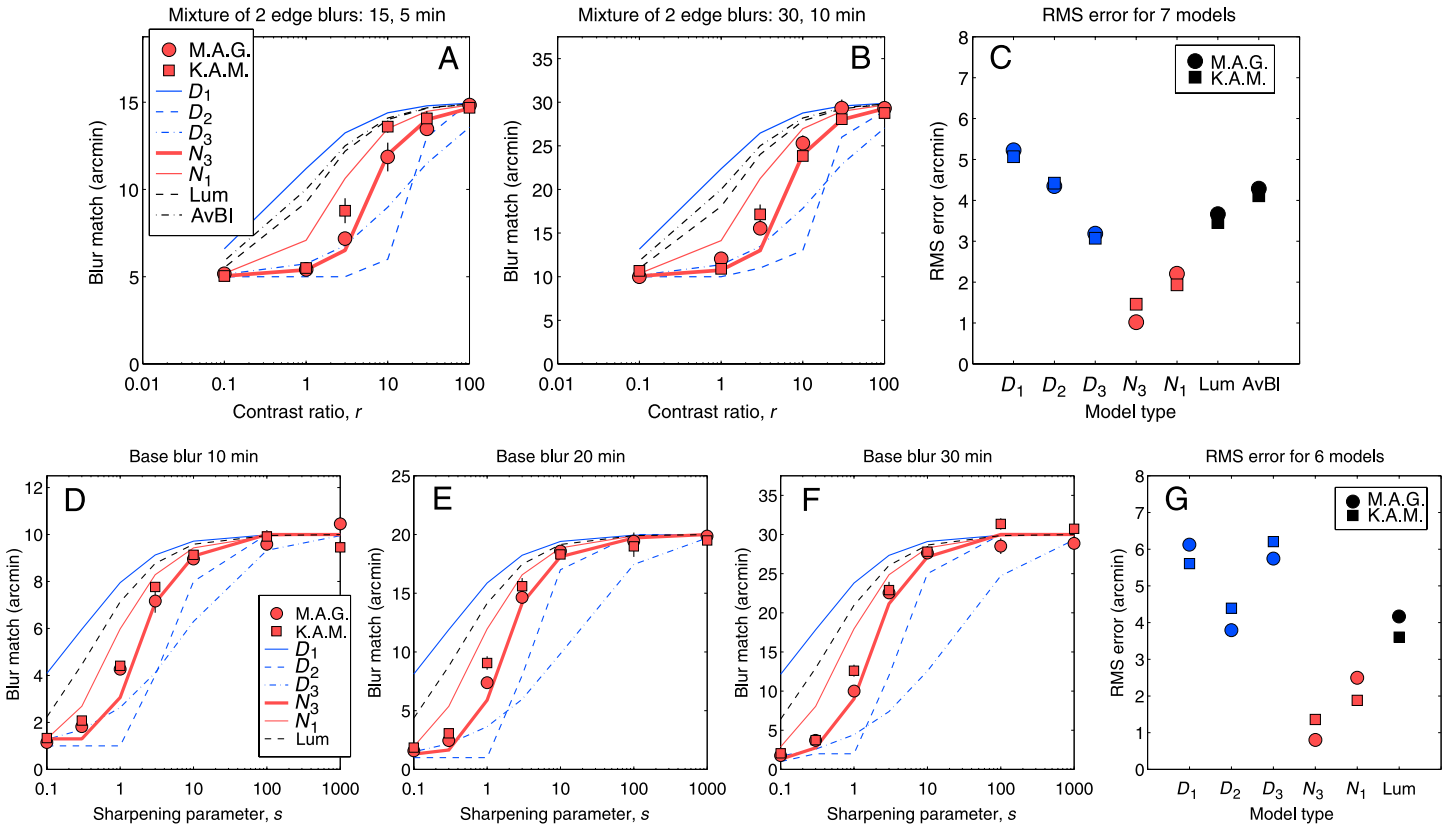

Figure 3. Perceived blur of non-Gaussian edges. (A, B) Blur mixture experiment. The blur of a Gaussian edge was adjusted to match the perceived blur of the sum of two Gaussian edges. Relative contrast (r) of the two component edges varied. Data for two observers (circles M.A.G., squares K.A.M. ±1 *SE*). Missing error bars are smaller than symbols. Curves are predictions of seven models: red—the two scale–space models; blue—three single-scale models; and black—the luminance template and average blur hypotheses. See text. (C) RMS error between models and data. (D, E, F) Similar to panels A and B but for the sharpened edge experiment. Test edge was formed by modifying a Gaussian edge whose blur was 10, 20, or 30 arcmin. Lower values of parameter s sharpen the waveform. (G) RMS errors for this experiment.

and nothing was adjusted to achieve this fit (Figures 3A and 3B, thick red curve). Taken over the 12 conditions and 2 observers, the RMS error between model and data (Figure 3C) was 4.2 arcmin for blur averaging, 3.6 for luminance template, 2.1 for N_1, but only 1.3 for N_3^+. In summary, blur matching in this experiment was fairly well predicted by the multiscale first derivative N_1 model but very well predicted by the multiscale third derivative N_3 or N_3^+.

Similar results were obtained when, instead of blur mixtures, the set of test edges was derived by starting with a Gaussian edge (10, 20, or 30 arcmin), then sharpening it to varying degrees by a nonlinear transformation (see Methods). As the sharpening parameter (s) decreased, the waveform was no longer Gaussian but had a steeper gradient and appeared sharper. Most importantly, the blur-matching values were well predicted by the N_3 model (Figures 3D, 3E, and 3F). The pattern of deviation for the other models was fairly similar to the blur mixture experiment. RMS error between model and data over 2 observers and 21 test conditions was 3.9 arcmin for luminance template, 2.2 for N_1, but only 1.1 for N_3 (Blur averaging was not definable in this case).

We also examined three plausible single-scale derivative-based algorithms for blur. Discrete first, second, and third derivatives were calculated at the smallest possible scale (1 pixel, corresponding to 1 arcmin) using the convolution operators [0.5 0 −0.5], [1 −2 1], and [1 −3 3 −1], respectively. Model D1 supposed that two edges would match in blur when the widths (standard deviations) of their gradient (first derivative) profiles were equal. Model D2 supposed that blur would match when the separations between peak and trough of the second derivative were equal (Watt & Morgan, 1983). Model D3 supposed a blur match when the ratios of first to third derivative were equal (Georgeson, 1994). It is clear from Figure 3 (blue curves) that none of these single-scale models fit the data well. RMS errors (rightmost panels) were three to six times higher than the N_3 model.

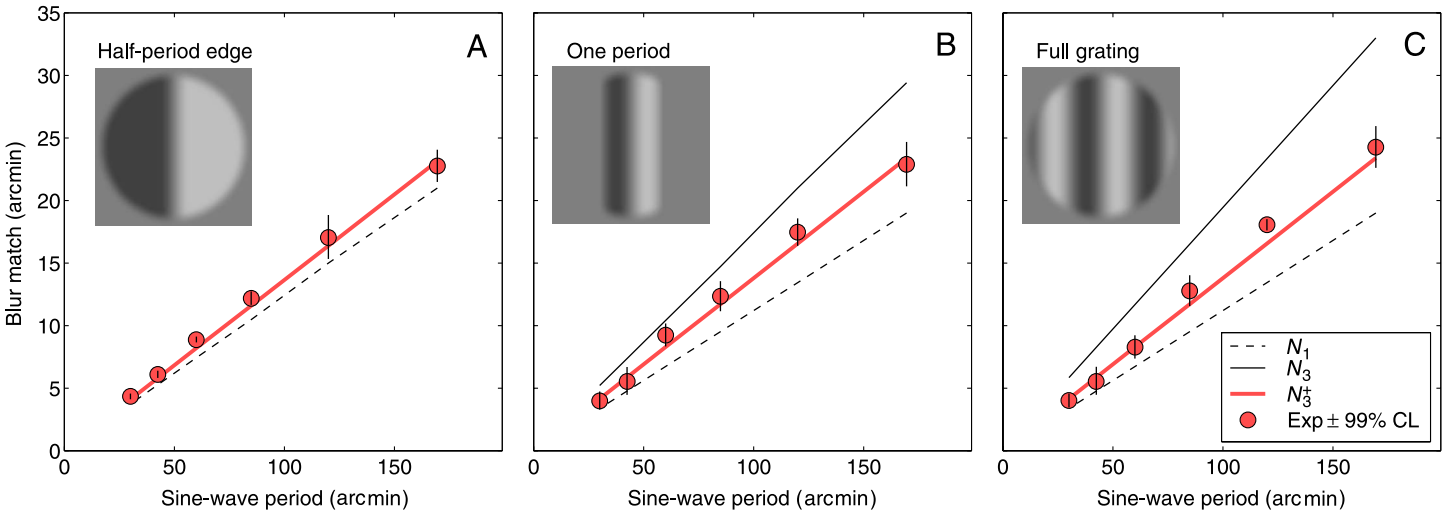

Figure 4. Blur matching for sine wave test edges assessed against a Gaussian comparison edge. Data points are geometric means of two subjects (M.A.G. and T.C.A.F.) with 99% confidence limits. Lines show the predictions of three scale–space models, N_1, N_3, and N_3^+. Only N_3^+ predicted the results accurately for all three types of test pattern.

Periodic edges

The first rectifier in the N_3^+ model plays no role when all gradients are positive. Hence, we further challenged that model and tested it for the presence of the first rectifier by introducing multiple edges with positive and negative gradients. For an existing data set (Georgeson, 1994), the test pattern was either a full sine wave grating (Figure 4C), a single period of a sine wave (Figure 4B), or a half-period sine wave edge (Figure 4A). These three test images were identical within the central half-period (the test edge) but different outside that region. They also have very different Fourier spectra—low-pass (single edge) versus narrowband (grating). Matched Gaussian blur was found to be directly proportional to the sine wave period (Figure 4) and was the same for all three test types. Thus, adding sine wave edges adjacent to the central test edge narrowed the Fourier spectrum and increased the total contrast energy but had no effect on perceived blur. Of the three multiscale models (N_1, N_3, and N_3^+), only N_3^+ predicted the data accurately in all three cases (Figures 4A, 4B, and 4C). The linear N_1 model underestimated the blur matches, whereas the linear N_3 model seriously overestimated them (Figures 4B and 4C). N_3^+ succeeds here because the first rectifier segments the image into regions of positive gradient, separated by zero-valued regions (where the gradient is zero or negative, vetoed by the rectifier). Responses in the neighborhood of an edge are then the same whether the image is periodic (a grating) or aperiodic (an isolated edge), leading to the same blur code in each case (see scale–space maps in Figure S2).

In a similar experiment, we used odd-order Gaussian derivatives as test stimuli. Successive differentiation increases the periodicity and the high spatial frequency content (Figure 5, insets). Perceptually, the higher derivatives looked progressively sharper (Figure 5, symbols). Predictions of the three models diverged in much the same way as they did for the sine wave experiment, and again the N_3^+ model was an accurate and clear winner (Figure 5).

Blurring reduces the high spatial frequency content of images. In two further blur-matching experiments, we varied the spatial frequency content of gratings by (a) progressively adding the odd harmonics to a sine wave to form successive approximations to a square wave grating and by (b) varying the relative contrast of the two components in a compound ($f + 3f$) grating. The N_3^+ model made very accurate predictions for both experiments (Figure S3). The linear N_1 model was fairly accurate for these two experiments, but the linear N_3 model was poor for the $f + 3f$ experiment. As in the sine wave experiment (Figure 4), it greatly overestimated the blur.

In summary, the Gaussian integral, Gaussian derivative, and sine wave edge profiles are all physically different, and so it is not obvious a priori how they should be matched. Subjects made matches with high reliability, and model N_3^+ was remarkably accurate in predicting the absolute values of blur matches, across a diverse set of conditions. Such unusual precision suggests that the model correctly captures some of the key processes in early visual coding of blur.

Two-dimensional filter shape

The stimuli and model have so far been expressed in one-dimensional form. To assess the length of the filters,

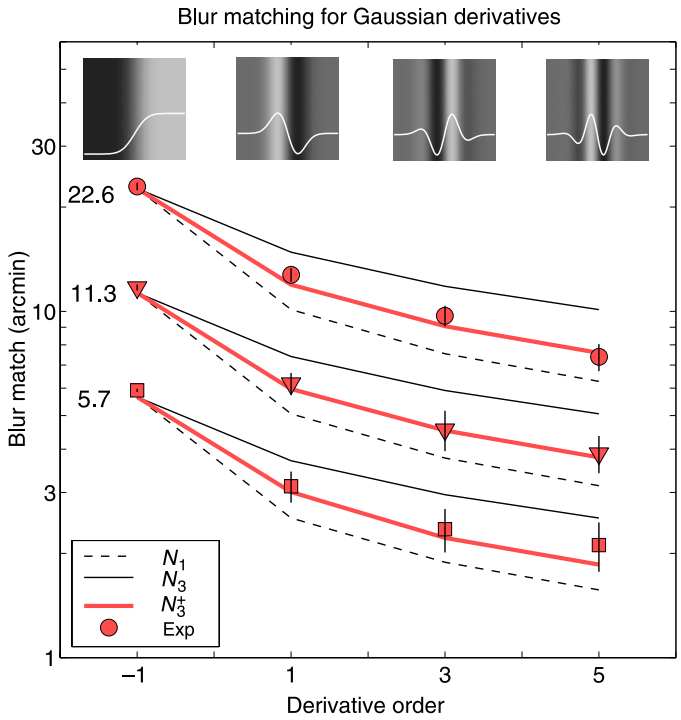

Figure 5. Blur matching for test stimuli defined as odd-order Gaussian derivative profiles (inset) at three scales (5.7, 11.3, and 22.6 arcmin). Symbols are geometric means of three subjects (M.A.G., T.C.A.F., and T.S.M.) ±1 *SE*. All three models (curves) predicted a scale-invariant decrease in blur with increasing derivative order, but the nonlinear model N_3^+ did so most accurately. (Note that in all experiments, stimulus polarity was randomized across trials; predictions for positive edges [using N_3^+] and negative edges [using N_3^-] are the same.)

we tested blur matching for Gaussian edges of different lengths. Surprisingly, shorter edges looked progressively less blurred as length was reduced (Figure 6), although the luminance profile (in the *x*-direction) was unchanged. As with our one-dimensional results, blur matching was scale invariant: The effects of edge length were nearly the same at all four test blurs (Figure 6) when the test length and the resulting blur match were expressed as a proportion of test blur. For an eightfold range of test blurs (b), the perceived blur started to decrease as test length fell below about $6b$, and edges looked about 50% less blurred when the length equalled b.

The scale–space model is inherently scale invariant. RFs of the large filters are both longer and wider than for small filters (Figure 1A), and this readily explains the effect of length on blur. As length is reduced, responses of the large filters drop because part of their input is removed. Smaller filters (at the centre of the edge) remain unaffected until the length is further reduced. Hence, length reduction induces a bias—a shift of the peak response in favor of the smaller filters (see scale–space maps in Figure S4)—and because it is the peak filter scale that represents blur, the shorter edges are seen as less blurred than long ones.

The decrease of blur with length depends on the aspect ratio of the filter kernels (RFs). Figure 6 (solid curve) shows the predictions of the N_3^+ (or N_3) model with filter kernels that were derivatives of an isotropic (circular) Gaussian (Figures 1A and 1B). The fit to the data was good. The N_1 model with the same isotropic assumption greatly overestimated the degree of sharpening with length reduction, although this would be improved by (perhaps implausibly) assuming an even shorter length–width ratio for the N_1 kernels.

Discussion

Finding edges and encoding their blur

There are many ways in which edges and their blur might be encoded by spatial filtering at multiple scales. Our analyses have shown that for human vision several plausible candidates can be distinguished because (a) they make different predictions about blur matching, and (b) experimental blur matching is sufficiently precise to allow a decision between them. On this basis, two scale–space

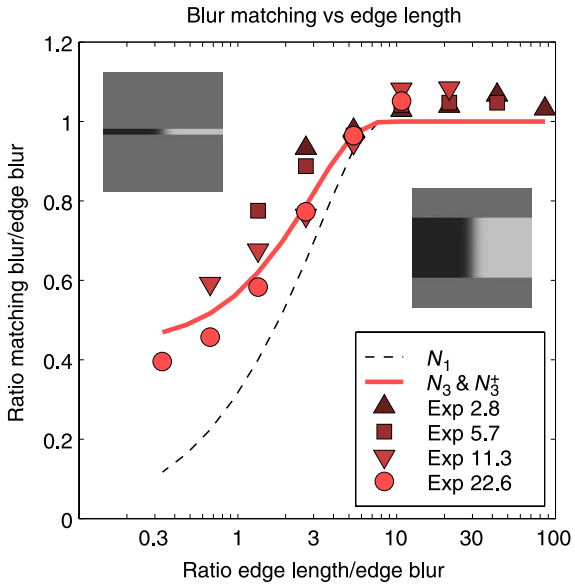

Figure 6. Truncating the length of an edge (inset) made it look sharper. Here, both the blur-matching values (geometric mean of two observers, M.A.G. and T.C.A.F.) and the test lengths are expressed as a proportion of the true test blur (2.8, 5.7, 11.3, and 22.6 arcmin). This reveals the scale invariance of the effect. Model N_3^+ (or N_3, equivalent for these stimuli) predicted this sharpening very well (solid curve), but N_1 over-estimated it (dashed curve). See text.

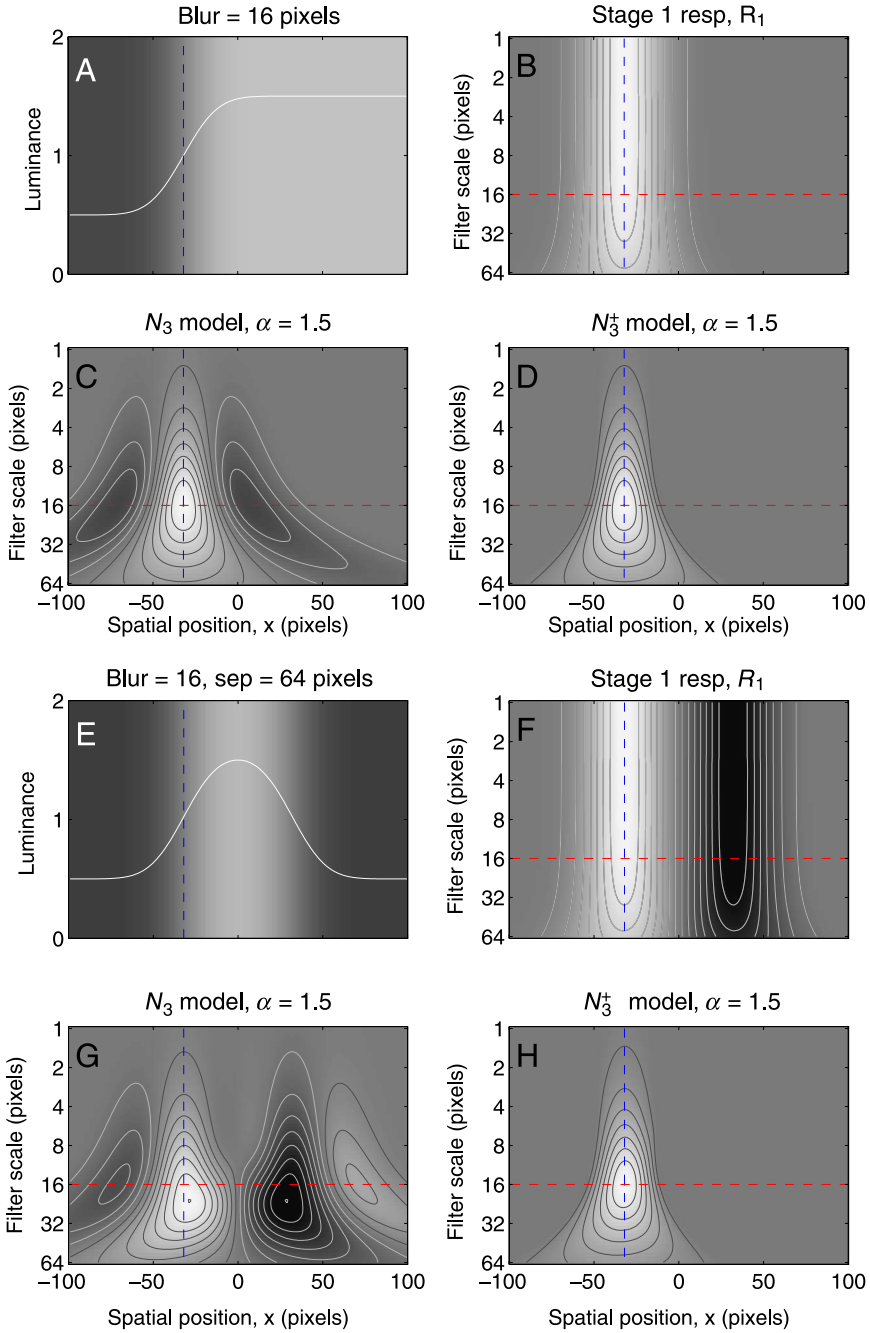

Figure 7. Comparison of the linear (N_3) and the nonlinear (N_3^+) scale–space responses to a single edge (A–D) and to a pair of edges (E–H). The two edges are separated by four times their blur. Dashed lines mark the true position and scale of the left-hand edge. Panel B shows the first-stage response R_1 to a single Gaussian-blurred edge (A); here half-wave rectification has no impact because $R_1 > 0$. Panel F shows response R_1 for the pair of edges (E) before rectification. The linear model has multiple response peaks (C) and distortion (G), but the nonlinear model (D, H) does not. See Discussion for details.

models (N_1, N_3^+) stood out from the others that we considered, and of these two, the nonlinear third derivative model, N_3^+, was consistently the better predictor of blur matching.

This model explained the appearance and matching of edge blur in (1) mixtures of two blurred edges, (2) edges that were "sharpened" by reshaping the luminance waveform, (3) Gaussian derivative waveforms, (4) compound (f + 3f) gratings and (5) sine wave gratings, as well as (6) the equivalence of blur in periodic and single sine wave edges, and (7) the striking finding that shorter edges look sharper. This entire raft of findings was predicted by

a scale–space model that has a very specific processing architecture (Figure 1B), and only one parameter was chosen to fit the data.

We now summarize the key properties of N_3^+ that make it a viable edge finder whereas the linear model N_3 is not. There are two filter stages, each followed by a half-wave rectifier. The first filter R_1 computes local gradients, at multiple scales. For a single edge (Figure 7A), the rectified output R_1^+ is the same as R_1 (Figure 7B)—a ridge in scale–space at the edge location. After differentiating twice more, inverting the sign, and applying the scale normalization, the linear response N_3 has a characteristic signature in scale–space—a central peak with two "wings" of opposite polarity (Figure 7C; also Figures 2D and 2I). Using a naive peak/trough rule, these wings could be falsely taken as additional negative-going edges. The first rectifier, however, guarantees that only positive peaks in the second-filter output correspond to edge locations. It does so because R_1^+ vetoes negative gradients. This veto ensures that negative parts of the second-filter response could not possibly arise from negative-going edges: They are already suppressed at Stage 1. Thus, the second rectifier can be routinely applied to exclude the negative wings and isolate the edge response at the correct location and scale (Figure 7D). Without the first rectifier, the negative troughs could not be safely excluded as candidate edges. In this way, the nonlinear structure of N_3^+ overcomes the "multiple-peaks problem" that occurs with narrowband linear spatial filtering (N_3).

In addition, the first rectifier eliminates interference from neighboring edges in blur coding. When two edges of opposite polarity are fairly close together (Figure 7E), the "wings" of the N_3 response to one edge interfere with the main response peak for the other edge (Figure 7G). There is some distortion of peak position, and the peak scale (blur code) is shifted by as much as half an octave. Thus, the N_3 model overestimated blur for all our experiments with periodic edges, but the human observers did not. The nonlinear channel neatly eliminates this interference: By excluding the negative gradients at Stage 1, the interfering "wings" are removed from Stage 2, and the response to the preferred edge (Figure 7H) is almost identical to that for an isolated edge (Figure 7D), with no distortion of scale or position. In short, the first rectifier plays two key roles: (1) It eliminates the irrelevant peaks that would be introduced by successive differentiation, and (2) it eliminates interference between neighboring edges.

Spatial location of edges

The nonlinear channel implements an important constraint: That in order to signal an edge, the gradient and the third derivative (at the appropriate scales) should be of opposite sign. To see this, consider that a simple, positive-going, blurred step edge has a peak in the first derivative, and a negative-going zero crossing (ZC) in second derivative (Marr & Hildreth, 1980). Hence, the third derivative (the slope of the ZC) is negative and of opposite sign to the gradient (see, e.g., Figure 1 of Georgeson & Freeman, 1997). For models based on ZCs, checking that the first and third derivatives are of opposite sign is a basic bit of calculus that enables real edges to be distinguished from "phantom" edges that also have a ZC, but which correspond to gradient minima rather than maxima (Clark, 1989). Together, the two rectifiers and the inverted sign of the second filter (Figure 1B and Equation 6) implement this constraint. A response gets through the first rectifier if the gradient is positive (Figure 7B), and through the second rectifier if the third derivative is negative (output from the second filter is positive; Figure 7D). Thus, a peak in the N_3^+ channel output, representing a positive-going edge, requires the gradient at that point to be positive while the third derivative is negative. The "wings" in the N_3 response (Figure 7C) do not satisfy this sign constraint and are rejected by N_3^+. Thus, the N_3^+ model does not explicitly identify ZCs, but it has much in common with models that do.

To compare the model's edge finding more rigorously with human perception, across a broad range of luminance waveforms, we reexamined the results of Hesse and Georgeson (2005). Predicted edge locations for the full model (N_3^+, N_3^-) were compared with data in which observers marked with a cursor the locations of perceived edges in a family of "phase-coherent" one-dimensional test images (see Supplementary methods for details).

Figure 8 shows how well the model predicts the observed edge locations across the entire set of images. At phases $\Phi = 0°, 45°, 135°$ (and their contrast-inversions at $\Phi = 180°, 225°, 315°$), observers and model reported a pair of edges flanking a light or dark bar, whereas at $\Phi = 90°$ they saw a single, central edge (triangles in Figure 8). As image blur increased (i.e., sharpness decreased), the model accurately captured the increasing perceived separation between pairs of edges (and it reported plausibly larger scale values). Although a simpler, single-scale consideration of gradient–peak locations gave a good account of these data (Hesse & Georgeson, 2005), it was important to establish that the multiscale, nonlinear N_3^+ model also does so and that it is well-behaved with no misses and no false-positives. The role of the rectifiers (discussed above) in suppressing "phantom" edges that would otherwise be introduced by the narrowband filtering (Clark, 1989) appears to be robust in this wide-ranging test.

Although the first rectifier may be thought to introduce harmonic distortion, it does not introduce spurious features in this or any other test that we have carried out. On the contrary, it is the (narrowband, N_3) linear filtering, without harmonic distortion, that does introduce spurious features (e.g., Figures 2I and 7G). This may seem paradoxical but begins to make more sense when we view

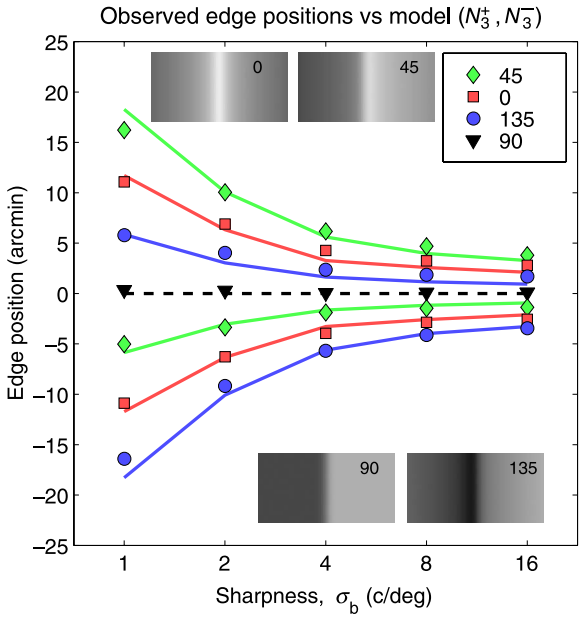

Figure 8. Subjects (mean of 6; data from Hesse & Georgeson, 2005) marked the perceived locations of edges across a family of images differing in phase (0°, 45°, 90°, 135°) and blur. Increasing σ_b represents sharper images. The N_3^+, N_3^- model (curves) accurately predicted the occurrence and location of all these edges. For this simulation only, the smallest channel scale was taken to be 2 arcmin, and the eye's optical blur was approximated by Gaussian blur of 1 arcmin.

the goal of these nonlinear channels as feature analysis, not frequency analysis.

By contrast, the influential, and also nonlinear, local energy model (Burr & Morrone, 1994; Morrone & Burr, 1988) does not explain these data (Figure 8) because it predicts a single feature (edge or bar) at a fixed position that is invariant with phase and blur, quite unlike the observed results (Hesse & Georgeson, 2005).

Links with cortical physiology

The filters in our model do not necessarily correspond to individual neurons in the visual system, yet there are some close parallels. We asked what the RFs of the N_3^+ channel might look like to a physiologist, and how they might correspond with those in V1. The first stage (R_1^+) is much like a simple cell with adjacent ON and OFF subregions. Like the standard model for simple cells, it is a linear spatial filter followed by half-wave rectification (Figure 1B). (We omit here the complications introduced by divisive gain controls in LGN (lateral geniculate nucleus) and cortex. This may be a reasonable simplification for our experiments where contrast was fixed at about 30% throughout. If the gain of the mechanisms is set by contrast, then at fixed contrast the system will be quasi-linear, which is what our model assumes.)

Overlapping ON and OFF subregions

The second stage output is more complex and depends critically on the relative scale (σ_1/σ) of the first stage. When σ_1 is relatively large (Figure 9D), the RF has nonoverlapping ON and OFF subregions separated by a small gap, much like a simple cell (cf. Figure 2A of Kagan, Gur, & Snodderly, 2002). When $\sigma_1 = \sigma_2$, the subregions abut (Figure 9C)—again like a simple cell. But when $\sigma_{1\ \mathrm{is}}$ relatively small (Figures 9A and 9B), the ON and OFF regions overlap considerably—a characteristic of complex cells (Hubel & Wiesel, 1962; Kagan et al., 2002; Mata & Ringach, 2005; Martinez et al., 2005). In this model, the separation or overlap of ON and OFF subregions is controlled by the size of the first filter.

All our simulations adopted $\sigma_1/\sigma = 0.25$. To see whether this was critical, we re-ran the model for several experiments with σ_1/σ ranging from 0.1 to 0.9. As expected, when the input had a single sign of gradient (an isolated positive edge), σ_1/σ was immaterial because the first rectifier then has no influence on the channel output. On the other hand, for periodic waveforms predicted blur remained close to the data for σ_1/σ up to 0.5 but was increasingly overestimated (by up to a factor of 2) as σ_1/σ increased from 0.5 to 0.9. We conclude that the success of N_3^+ in predicting perceived blur does require the first filter to be small compared with the second filter, and that this is associated with the complex-like RF of Figure 9B. The sequence of two nonlinear stages (filter–rectify–filter–rectify) is essential to its correct behavior in edge coding and has some parallel in recent physiological findings that simple cells in layer 4 of the cat cortex are prior to, and provide the input for, complex cells in layers 2 and 3 (Martinez & Alonso, 2001).

Spatial frequency tuning

Despite the nonlinearities, spatial frequency tuning of the N_3^+ mechanism (Figures 1B and 9B) was very similar to the band-pass tuning of the linear, Gaussian third derivative (N_3) filter. The only difference was a slight broadening of responses on the low frequency side of the peak (not shown). The spatial waveform of the response to a sine grating was fairly similar to a half-wave rectified sine wave, showing high response modulation ($F1/F0\sim$ = 1.6) that is thought to be more characteristic of simple cells than complex cells. We note, however, that the degree of modulation varies widely across cells, and that many cells classed as complex in the awake monkey were found to have high response modulation for drifting gratings (Kagan et al., 2002).

Filter–rectify–filter

Rectifying nonlinearities are ubiquitous in sensory physiology, and in vision, polarity specificity begins very

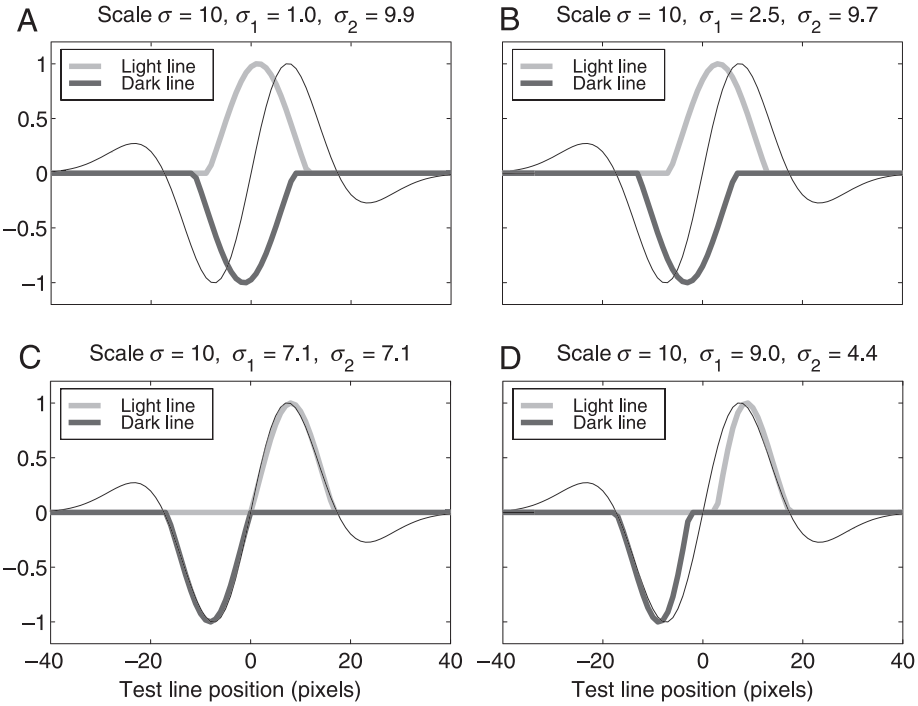

Figure 9. Receptive field (RF) analysis of the nonlinear channel (Figure 1B). ON and OFF subregions of the N_3^+ mechanism (gray and black curves) were computed in response to single light lines (plotted as positive) or dark lines (plotted as negative). Thin curve shows the corresponding linear N_3 filter kernel. Overall scale of the channel was fixed (σ = 10). When the first filter was relatively large (C, D), the channel behaved like a typical simple cell, with adjacent, nonoverlapping subregions. When the first filter was relatively small (A, B), the channel behaved like a complex cell, with substantial overlap of the ON and OFF regions (Kagan et al., 2002; Mata & Ringach, 2005). All our simulations adopted σ_1/σ = 0.25, as in panel B.

early with the separation into ON and OFF pathways at the retinal bipolar cells (Schiller, 1992). At higher levels of processing, the filter–rectify–filter (FRF) sequence of operations has been studied mostly in the context of "second-order vision"—e.g., the detection of static or moving texture boundaries or contrast modulation (Lu & Sperling, 1995; Schofield & Georgeson, 1999; Wilson & Kim, 1994). We naturally wondered whether the N_3^+ channel would respond to second-order structure as well as to luminance edges, but we found its computed response to contrast modulation to be weak whereas its response to the high frequency carrier was strong. There are two aspects of FRF channel design that can promote a strong response to second-order modulation while suppressing responses to the carrier. These are (1) full-wave rectification after the first filter and (2) little overlap in the orientation and/or spatial frequency tuning of the first and second filters (Bergen & Landy, 1991; Chubb & Sperling, 1988; Dakin & Mareschal, 2000; Wilson, Ferrera, & Yo, 1992). The N_3^+ channel, on the other hand, has half-wave rectification and considerable overlap in the filter tunings. These give it the interesting edge-coding properties discussed here but make it ill suited to second-order signal processing. Nonlinear FRF "sandwich" mechanisms can evidently be exploited for different purposes in first and second order vision, depending on the details of the FRF structure.

Natural images, spatial filters, the rectified contrast spectrum

Brady and Field (1995), Field (1987), and Field and Brady (1997) have argued that a good model for spatial filtering in early primate vision is a scheme in which there are self-similar RFs at all scales (as in Figure 1A). Crucially, it was proposed that these filters all have the same peak response to their own preferred spatial frequency. Contrast constancy (Georgeson & Sullivan, 1975) for gratings, Gabor patches, and band-pass noise is directly predicted by this scheme (Brady & Field, 1995). A theoretical benefit is that in a world where images on average have a 1/*f* spectrum, all filters carry the same information load—that is, all filters have the same expected variance in their outputs over space and time, and this allows information to be coded by neurons that have the same limited dynamic range in their responses.

Field and Brady (1997) extended this equal-amplitude filter model to propose an algorithm for coding the blur of

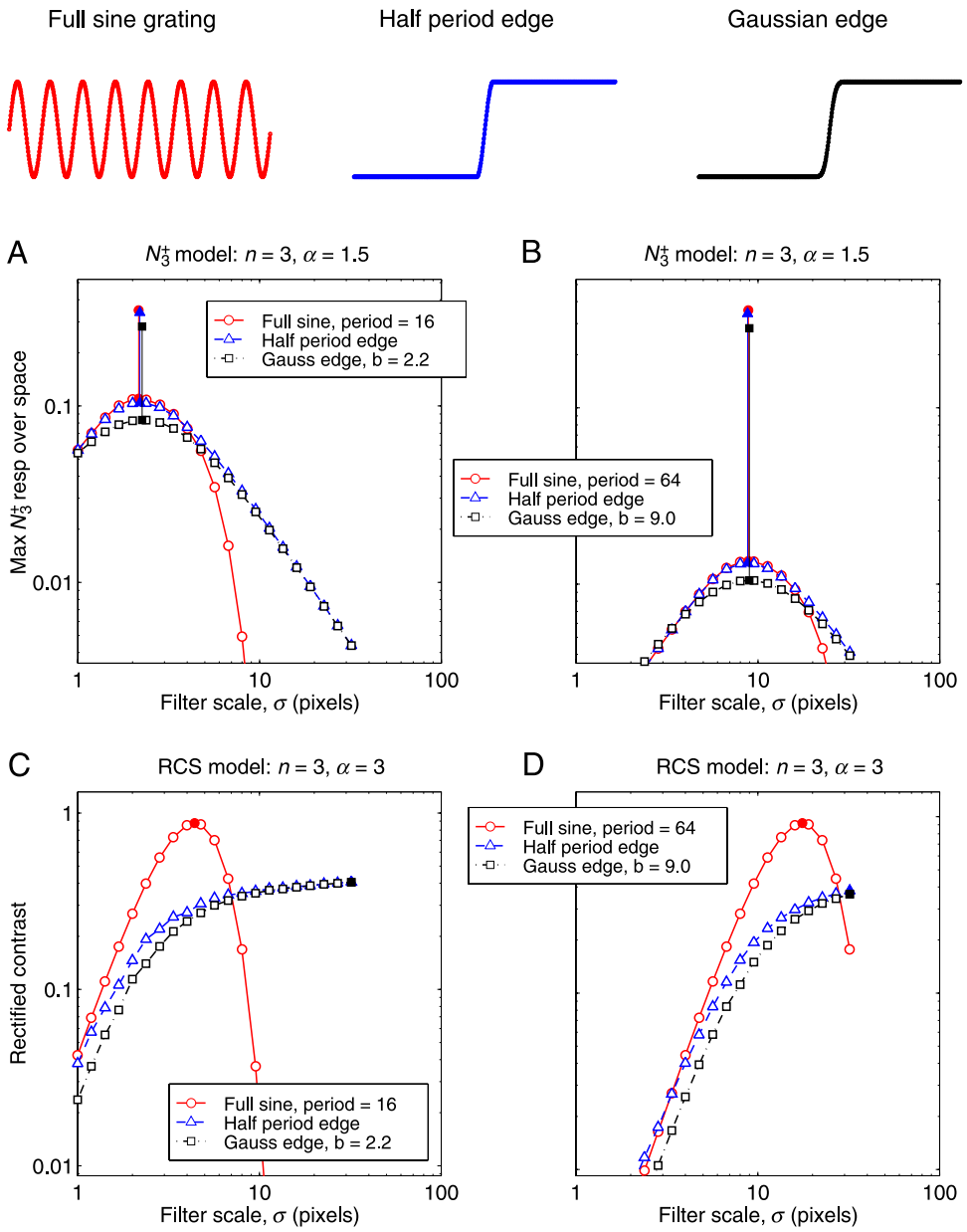

Figure 10. Comparison of two models for encoding blur. (A, B) N_3^+ model responses plotted over filter scale for the three waveforms illustrated (top). Grating period and Gaussian blur were chosen so that all three waveforms had the same perceived blur. Period was (A) 16 or (B) 64 pixels. Filled symbols identify the peak scale (found by parabolic interpolation), and those at the top of each vertical line show how edge amplitude can be recovered by rescaling the peak response value (see text, Appendix A). (C, D) The rectified contrast spectrum (RCS; Field & Brady, 1997) computed for the same three stimuli (top).

images. For an in-focus image with a $1/f$ spectrum, the output energy across filter scales will be constant. Thus, an image might be judged as in-focus when responses (aggregated across space) are equal across scale, but judged as blurred when responses decline at the smaller scales. To cope with the fact that spatial structure is sparsely distributed in some images, but dense in others, they introduced a nonlinear thresholding scheme—the rectified contrast spectrum (RCS)—in which the variance of each channel output was computed not over the whole image, but only over regions containing significant structure, where local responses exceeded a threshold ($s/2$, where s is the standard deviation of responses over the whole image). Finally, the slope of the RCS (on log–log axes) was taken as an index of image blur.

We asked how well the RCS scheme might explain the perception of edge blur. Setting $\alpha = n$ in Equation 4 produces an equal-amplitude filtering scheme of the required kind (see Appendix B), enabling us to compare our blur code (where $\alpha = n/2$) with that proposed by Field and Brady (1997). Figures 10A and 10B show the N_3^+ responses across filter scales computed for a sine

grating, a single (half-period) sine edge, and a Gaussian edge. Experiment 3 (Figure 4) found that Gaussian blur b would, on average, match a sine wave edge of period p when the ratio b/p was 0.14. This b/p ratio was therefore used in Figure 10 so that all three types of edge would have the same perceived blur. Figures 10A and 10B illustrate two key properties of N_3^+: that peaks of response exist for both periodic and aperiodic edges, and that they occur at the same filter scale when the waveforms are perceptually matched in blur. In short, peak scale predicts edge blur.

For comparison, Figures 10C and 10D show RCS profiles computed for the same three images, using the $s/2$ threshold (defined above). RCSs for the two single edges are similar, but they have little in common with the RCS for a grating. The grating response is scale tuned, whereas the single-edge response increases monotonically with filter scale. Similar divergence between the RCSs for periodic and aperiodic edges was seen for even-symmetric RFs ($n = 2$) and odd ones ($n = 1$ or 3) and at all threshold levels. These properties reflect the behavior of the underlying linear filters, seen in Figure A2. All the RCSs are curvilinear functions of scale, and so it is not easy to see any simple measure—such as slope—that would encode the blur or capture the equivalence of perceived blur between gratings and single edges. This contrasts with the success of the RCS in representing changes in spectral slope of natural images, textures, or two-dimensional noise via a single RCS slope measure. Indeed, Field and Brady (1997) anticipated that the RCS approach would have difficulty in encoding edge blur, partly because "altering the slope of the spectrum is not a good model of optical blur" (p. 3382) and partly because the RCS is still a global measure, whereas "a more accurate measure of blur will certainly involve local measures and will probably be best calculated on an edge by edge basis" (p. 3381). We agree with both points, and we propose peak finding in the N_3^+ scale–space as an effective and empirically supported algorithm for edge finding and local blur coding. It remains to be seen whether the sense of global blur obtained from an optically blurred image can be understood as some simple aggregate of these local blur measures (Dijk, van Ginkel, van Asselt, van Vliet, & Verbeek, 2003).

To summarize this section, Field and Brady's (1997) RCS serves well to encode the spectral slope of images, but (confirming Field and Brady's own suggestion) the RCS does not yet offer a simple measure of edge blur, nor does it capture the equivalence of perceived blur between periodic and single edges. However, Appendix B shows that if Brady and Field's (1995) account of filter gains and noise were correct at some level of processing, then a simple rescaling of those filter gains by $\sigma^{-n/2}$ would produce an output whose gains matched those needed for N_3^+. In short, the benefits of Field and Brady's scheme, and those of N_3^+, could coexist at successive stages of processing.

Conclusions

We asked how spatial filters serve to represent luminance edges in human vision. When the filter gains at different scales are set appropriately ("scale normalized," $\alpha = n/2$ in Equation 4), the problem of locating edges and determining their blur reduces to finding peaks in the scale–space map of responses. Predictions from a linear model (N_1) based on Gaussian first derivative filters were in fair agreement with our blur-matching data, but the nonlinear third derivative model was consistently more accurate. Each channel in the N_3^+ model has a two-stage structure analogous to the sequence from simple to complex cells in visual cortex, and the half-wave rectifying nonlinearities play a crucial role in enabling edge finding without false-positives. The N_3^+ model draws together three lines of thought about vision—computational, physiological, and psychophysical. It implements a principled, scale–space approach to the representation of key features in early vision and does so via physiologically plausible mechanisms, supported by some strikingly accurate predictions about human perception.

Appendix A

Some properties of the scale–space models

Following the approach of Lindeberg (1998), we analyze here some of the scaling properties of the family of scale–space models defined by Equation 4, in response to isolated Gaussian-blurred edges $I(x; b)$ of blur b and unit amplitude, where luminance I is the indefinite integral of the unit-area Gaussian $G(x;b)$.

Linear operations can be applied in any order, and so from Equation 4 we get

$$N_n(x,\sigma;b) = \sigma^\alpha \frac{\partial I(x;b)}{\partial x} \otimes \frac{\partial^{n-1} G(x;\sigma)}{\partial x^{n-1}}. \quad \text{(A1)}$$

The first derivative of I is simply $G(x; b)$, and because variances add under convolution, Equation A1 reduces to

$$N_n(x,\sigma;b) = \sigma^\alpha \frac{\partial^{n-1} G(x;s)}{\partial x^{n-1}}, \quad \text{(A2)}$$

where $s = \sqrt{(\sigma^2 + b^2)}$. When $n = 1$,

$$N_1(x,\sigma;b) = \frac{\sigma^\alpha}{\sqrt{2\pi(\sigma^2+b^2)}} \exp\left(\frac{-x^2}{2(\sigma^2+b^2)}\right). \quad \text{(A3)}$$

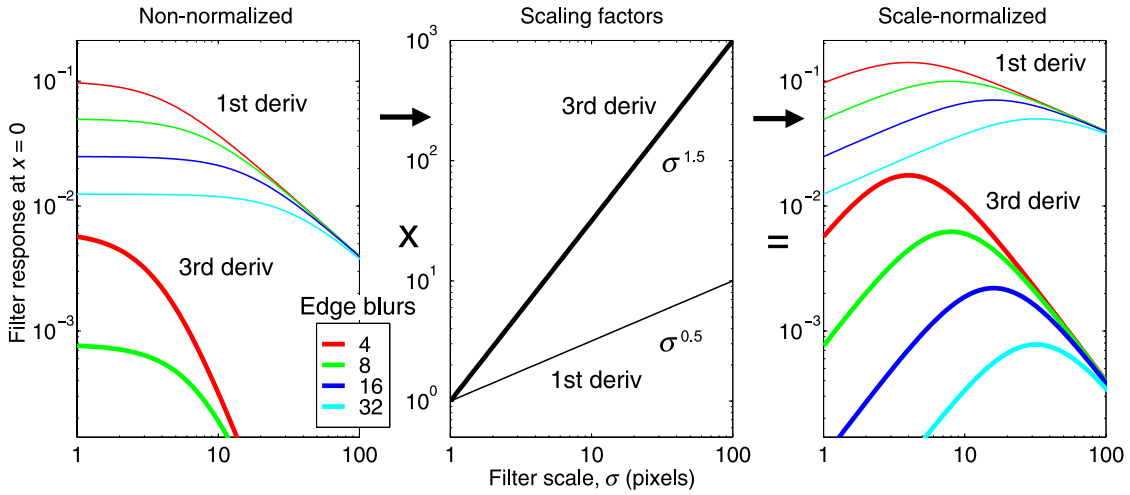

Figure A1. Properties of Gaussian derivative filter responses to single Gaussian-blurred edges, taken at the edge location and plotted as a function of filter scale. Responses to four different edge blurs are shown for linear first and third derivative filters (n = 1 or 3), either without scale normalization (left; α = 0) or with it (right; $\alpha = n/2$ in Equations A3 and A5). Centre panel shows the scale factors used to convert the nonnormalized responses (left) into the normalized ones (right). This scale normalization ensures that, for Gaussian edges, the filter scale at the peak response identifies the edge blur.

This shows that for N_1 the spatial profile of response to a Gaussian edge, at any filter scale σ, is a Gaussian whose spread increases with σ. These profiles peak at the edge location (x = 0). Response values at x = 0 are plotted across filter scales in Figure A1 (thin curves). As we saw in Figure 1, without scale normalization, responses are greatest at the smallest filter scale, but with the chosen normalization ($\alpha = n/2$), responses show a peak where $\sigma = b$. Responses to different edge blurs have a common asymptote at large filter scales. This means, as one might expect, that large-scale filters cannot distinguish between sharp edges and blurred ones.

The same general properties hold for the scale–space third derivative N_3, which (from Equation A2, with n = 3) becomes

$$N_3(x,\sigma;b) = \frac{\sigma^{\alpha}(x^2-\sigma^2-b^2)}{(\sigma^2+b^2)^{5/2}\sqrt{2\pi}}\ \exp\left(\frac{-x^2}{2(\sigma^2+b^2)}\right). \tag{A4}$$

At x = 0, this simplifies to

$$N_3(0,\sigma;b) = \frac{-\sigma^{\alpha}}{(\sigma^2+b^2)^{3/2}\sqrt{2\pi}}, \tag{A5}$$

plotted (dropping the negative sign) as thick curves in Figure A1.

Figure A1 also shows that the *shape* of the response curves (on log–log axes) is the same for all blurs. From Equations A3 and A4, for n = 1, 3 and for any α, response magnitude over scale at x = 0 is a function of relative scale, σ/b:

$$|N_n(0,\sigma;b)| = \frac{b^{\alpha-n}}{\sqrt{2\pi}}\frac{(\sigma/b)^{\alpha}}{\left(1+(\sigma/b)^2\right)^{n/2}}. \tag{A6}$$

This proves that response curve shape is invariant with blur, but amplitude scales as $b^{\alpha-n}$.

Information in the peak response: blur and contrast

It is straightforward to show by differentiating Equations A3 or A5 with respect to σ, that when $\alpha = n/2$, the peak response occurs at scale $\sigma_{max} = b$, as shown in Figure A1, right. This scaling property means that the scale σ_{max} of the most active filter identifies the edge blur b. This of course is the main goal of our model.

For n = 1, 3 and $\alpha = n/2$, with input edge amplitude c ($0 \le c \le 1$), the peak response values are proportional to c, but also vary with blur b:

$$N_1(0,b;b) = 0.5cb^{-1/2}/\sqrt{\pi}, \tag{A7}$$

$$N_3(0,b;b) = -0.25cb^{-3/2}/\sqrt{\pi}. \tag{A8}$$

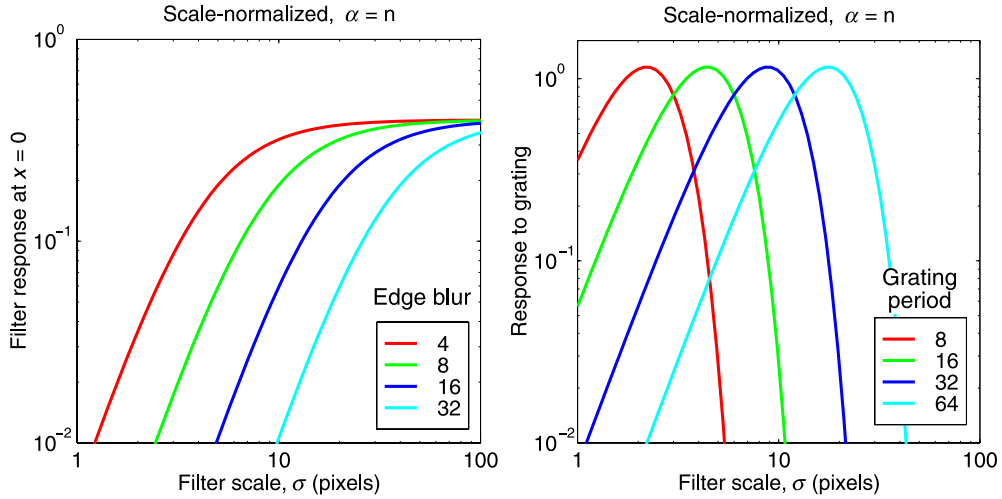

Figure A2. Gaussian derivative (n = 3) filter responses at x = 0, plotted over scale for edges at four different blurs (left) and for sine gratings at four spatial frequencies (right) when normalization exponent $\alpha = n$.

When $\alpha = n/2$, peak response amplitude falls as $b^{-\alpha}$ (Figure A1, right). Thus, there is no "contrast constancy" in the output of these filter sets. Nevertheless, these filters do encode edge contrast. Once the edge location and blur have been found, edge amplitude c can be recovered from the peak response value R_{max} and its corresponding scale σ_{max} by inserting these two values into Equation A7 or A8 and solving for c. It follows that for all edge blurs, the quantity $R_{max}(\sigma_{max})^{n/2}$ is directly proportional to contrast. We examine these contrast-coding ideas more closely elsewhere (May & Georgeson, 2007).

This appendix has considered the scale–space properties of linear Gaussian derivative filters in response to Gaussian-blurred edges. But all the equations and the conclusions developed here for N_3 also apply to the nonlinear mechanism N_3^+ because for single edges (and any other input with nonnegative gradients), the first rectifier is immaterial and the behaviors of N_3 and N_3^+ are identical.

Appendix B

Here we consider an important special case where all filters have the same peak sensitivity in the Fourier domain. This supports the Discussion section, where we analyze an alternative model for blur (Field & Brady, 1997) that is based on this property.

Filters with equal-amplitude spectra, $\alpha = n$

In general, the edge response (Figure A1) has a peak across scale, provided that $0 < \alpha < n$. An interesting consequence of Equation A6 is that when $\alpha = n$ (or $\alpha = 0$), the edge response amplitude no longer has a peak that can identify edge blur. See Figure S5 for scale–space maps illustrating this. When $\alpha = n$, the responses have no peak, but instead a common horizontal asymptote at large filter scales (Figure A2, left).

For sine gratings, however, responses remain "tuned" across filter scales even when $\alpha = n$ (Figure A2, right). We prove that result here for the case of linear filters, and we note that it is also true for the nonlinear N_3^+ channel. The response amplitude of the normalized nth Gaussian derivative operator of scale σ to a sine wave grating of spatial frequency f is easily obtained from the Fourier transform, F:

$$|F_n(f,\sigma)| = \sigma^{\alpha}(2\pi f)^n \exp\left(-2\pi^2 f^2 \sigma^2\right). \quad \text{(B1)}$$

Let the grating period $p = 1/f$, so that when $\alpha = n$,

$$|F_n(p,\sigma)| = (2\pi\sigma/p)^n \exp[-2\pi^2(\sigma/p)^2]. \quad \text{(B2)}$$

This implies that the response tuning and the amplitude are scale invariant because they depend only on the relative scale (σ/p).

To get a scale-tuned response to edges requires $0 < \alpha < n$. We used $\alpha = n/2$, which (uniquely) renders the response curve for Gaussian edges symmetrical about the peak, on a log scale axis (Figure A1, right). As α deviates from $n/2$, the response curves become more asymmetrical, and the correct (experimentally observed) equivalence between sine and Gaussian edges is gradually lost (not shown). Hence, there is a strong case for the specific model in which $\alpha = n/2$.

A general consequence of this scaling is that peak response values vary with edge blur, as $b^{\alpha-n}$ (see Appendix A). When $\alpha = n$, contrast constancy is obtained

directly, as in the Brady and Field (1995) filtering scheme. When $\alpha < n$, contrast constancy can be restored after peak finding simply through rescaling by the known factor $b^{n-\alpha}$ as discussed above (Appendix A) and illustrated by the filled symbols in Figures 10A and 10B.

One might worry that the low response values for large blurs would entail poor signal–noise ratios, but this is not necessarily so. Suppose that, at one level of processing, the filtering and the signal–noise ratios are correctly represented by a scheme like Brady and Field's (1995) (let's call it N_3^*, defined as N_3 but with $\alpha = n$). They proposed that the decline in contrast sensitivity at high SFs follows from the fact that the high frequency filters have a broader frequency bandwidth (in linear terms) and would collect more input noise when, as seems likely, the input noise has a fairly flat spectrum. Thus, at this level, the smaller scale filters have a poorer signal–noise ratio. A further processing step, rescaling all filter responses by $\sigma^{-n/2}$, would convert N_3^* to behave exactly as N_3 with $\alpha = n/2$. But if this final linear step—progressively attenuating the larger scale filters—adds no further noise, then signal–noise ratios in each filter remain unchanged, and the smaller scale filters still have the poorer signal–noise ratio. Despite their (now) low amplitude of response, the larger scale filters would continue to have the better signal–noise ratio.

We conclude that N_3 or N_3^+ cannot easily be rejected by signal–noise ratio arguments. We also see that Brady and Field's (1995) filtering scheme (like N_3^*), in which all filters have equal peak-response amplitude, could be easily transformed into N_3^+ to find edges and encode blur, provided that successive stages of processing are considered.

Acknowledgments

We thank the EPSRC (Ref. GR/G63582, GR/S07261/01) and Wellcome Trust (Ref. 056093/B/98) for grant support, Aston University for studentship support to K.A.M., and Tim Meese for critical reading of the MS.

Commercial relationships: none.
Corresponding author: Mark Georgeson.
Email: m.a.georgeson@aston.ac.uk.
Address: Vision Sciences Building, Aston University, Birmingham B4 7ET, UK.

References

Bergen, J. R., & Landy, M. S. (1991). Computational modelling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing*. Cambridge, MA: MIT Press.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9,* 726–741.

Blakemore, C. B., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology, 203,* 237–260. [PubMed] [Article]

Brady, N., & Field, D. J. (1995). What's constant in contrast constancy? The Effects of scaling on the perceived contrast of bandpass patterns. *Vision Research, 35,* 739–756. [PubMed]

Bruce, V., Green, P. R., & Georgeson, M. A. (2003). *Visual perception: Physiology, psychology and ecology*. Hove & New York: Psychology Press.

Burr, D. C., & Morrone, M. C. (1994). The role of features in structuring visual images. In G. R. Bock & J. A. Goode (Eds.), *Higher order processing in the visual system* (pp. 129–141). Chichester: Wiley.

Camarda, R. M., Peterhans, E., & Bishop, P. O. (1985). Spatial organization of subregions in receptive fields of simple cells in cat striate cortex as revealed by stationary flashing bars and moving edges. *Experimental Brain Research, 60,* 136–150. [PubMed]

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8,* 679–698.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, Optics and image science, 5,* 1986–2007. [PubMed]

Clark, J. J. (1989). Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 43–57.

Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research, 40,* 311–329. [PubMed]

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22,* 545–559. [PubMed]

De Valois, R. L., & De Valois, K. K. (1990). *Spatial Vision* (pp. 1–381). Oxford: Oxford University Press.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology, 69,* 1118–1135. [PubMed]

Dijk, J., van Ginkel, M., van Asselt, R. J., van Vliet, L. J., & Verbeek, P. W. (2003). A new sharpness

measure based on lines and edges. *Lecture Notes in Computer Science, 2756,* 149–156.

Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 699–716.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics and image science, 4,* 2379–2394. [PubMed]

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research, 37,* 3367–3383. [PubMed]

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711–724. [PubMed]

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society B: Biological Sciences, 249,* 235–245. [PubMed]

Georgeson, M. A. (1994). From filters to features: Location, orientation, contrast and blur. In G. R. Bock & J. A. Goode (Eds.), *Higher order processing in the visual system* (pp. 147–165). Chichester: Wiley.

Georgeson, M. A. (1998). Edge-finding in human vision: A multi-stage model based on the perceived structure of plaids. *Image and Vision Computing, 16,* 389–405.

Georgeson, M. A. (2006). Bars & edges: A multi-scale Gaussian derivative model for feature coding in human vision [Abstract]. *Journal of Vision, 6*(6):191, 191a, http://journalofvision.org/6/6/191/, doi:10.1167/6.6.191.

Georgeson, M. A., & Freeman, T. C. (1997). Perceived location of bars and edges in one-dimensional images: Computational models and human vision. *Vision Research, 37,* 127–142. [PubMed]

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *The Journal of Physiology, 252,* 627–656. [PubMed] [Article]

Helmholtz, H. v. (2000). *Treatise on physiological optics* (vol. 1, pp. 1–482). Bristol: Thoemmes Press. (Original work published 1856).

Hesse, G. S., & Georgeson, M. A. (2005). Edges and bars: Where do people see features in 1-D images? *Vision Research, 45,* 507–525. [PubMed]

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology, 160,* 106–154. [PubMed] [Article]

Kagan, I., Gur, M., & Snodderly, D. M. (2002). Spatial organization of receptive fields of V1 neurons of alert monkeys: Comparison with responses to gratings. *Journal of Neurophysiology, 88,* 2557–2574. [PubMed] [Article]

Klein, S. A., & Levi, D. M. (1985). Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validation. *Journal of the Optical Society of America A, Optics and image science, 2,* 1170–1190. [PubMed]

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics, 50,* 363–370. [PubMed]

Kovasznay, L. S. G., & Joseph, H. M. (1955). Image processing. *Proceedings of the Institute of Radio Engineers, 43,* 560–570.

Kovesi, P. (2000). Phase congruency: A low-level image invariant. *Psychological Research, 64,* 136–148. [PubMed]

Kulikowski, J. J., & Bishop, P. O. (1981). Linear analysis of the responses of simple cells in the cat visual cortex. *Experimental Brain Research, 44,* 386–400. [PubMed]

Kulikowski, J. J., Marcelja, S., & Bishop, P. O. (1982). Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics, 43,* 187–198. [PubMed]

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision, 30,* 117–154.

Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research, 35,* 2697–2722. [PubMed]

Mach, E. (1965). On the effect of the spatial distribution of the light stimulus on the retina. In F. Ratliff (Ed.), *Mach bands* (pp. 253–271). Holden-Day Inc. (Original work published 1865).

Marr, D. (1982). *Vision*. New York: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences, 207,* 187–217. [PubMed]

Martinez, L. M., & Alonso, J. M. (2001). Construction of complex receptive fields in cat primary visual cortex. *Neuron, 32,* 515–525. [PubMed] [Article]

Martinez, L. M., Wang, Q., Reid, R. C., Pillai, C., Alonso, J. M., Sommer, et al. (2005). Receptive field structure varies with layer in the primary visual cortex. *Nature Neuroscience, 8,* 372–379. [PubMed] [Article]

Mata, M. L., & Ringach, D. L. (2005). Spatial overlap of ON and OFF subregions and its relation to response modulation ratio in macaque primary visual cortex. *Journal of Neurophysiology, 93,* 919–928. [PubMed] [Article]

May, K. A., & Georgeson, M. A. (2007). Blurred edges look faint, and faint edges look sharp: The effect of a gradient threshold in a multi-scale edge coding model. *Vision Research, 47,* 1705–1720. [PubMed]

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B: Biological Sciences, 235,* 221–245. [PubMed]

Morrone, M. C., Navangione, A., & Burr, D. (1995). An adaptive approach to scale selection for line and edge-detection. *Pattern Recognition Letters, 16,* 667–677.

Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters, 6,* 303–313.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in receptive fields of simple cells in cat's striate cortex. *The Journal of Physiology, 283,* 53–77. [PubMed] [Article]

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607–609. [PubMed]

Peli, E. (2002). Feature detection algorithm based on a visual system model. *Proceedings of the IEEE, 90,* 78–93.

Ratliff, F. (1965). *Mach bands* (pp. 1–365). San Francisco: Holden-Day.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88,* 455–463. [PubMed] [Article]

Schiller, P. H. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences, 15,* 86–92. [PubMed]

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research, 39,* 2697–2716. [PubMed]

ter Haar Romeny, B. M. (2003). *Front-end vision and multi-scale image analysis* (pp. 1–464). Dordrecht: Kluwer.

van Deemter, J. H., & du Buf, J. M. H. (2000). Simultaneous detection of lines and edges using compound Gabor filters. *International Journal of Pattern Recognition and Artificial Intelligence, 14,* 757–777.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences, 265,* 359–366. [PubMed] [Article]

van Warmerdam, W. L. G., & Algazi, V. R. (1989). Describing 1-D intensity transitions with Gaussian derivatives at the resolutions matching the transition widths. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 973–977.

Watt, R. J. (1988). *Visual processing* (pp. 1–168). Hove and London: Erlbaum.

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23,* 1465–1477. [PubMed]

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 1661–1674. [PubMed]

Wilson, H. R. (1983). Psychophysical evidence for spatial channels. In O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 88–99). New York: Springer-Verlag.

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9,* 79–97. [PubMed]

Wilson, H. R., & Kim, J. (1994). A model for motion coherence and transparency. *Visual Neuroscience, 11,* 1205–1220. [PubMed]

Witkin, A. P. (1983). Scale–space filtering. *Proceedings of the 8th International Joint Conference on Artificial Intelligence,* 1019–1022.

Young, R. A., & Lesperance, R. M. (2001). The gaussian derivative model for spatial–temporal vision: II. Cortical data. *Spatial Vision, 14,* 321–389. [PubMed]

Zhang, W., & Bergholm, F. (1997). Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision, 24,* 219–250.