Optimal edge filters explain human blur detection

William H. McIlhagga

Bradford School of Optometry and Vision Science, University of Bradford, Bradford, United Kingdom ✉

Keith A. May

Department of Optometry and Visual Science, City University London, London, United Kingdom ✉

Edges are important visual features, providing many cues to the three-dimensional structure of the world. One of these cues is edge blur. Sharp edges tend to be caused by object boundaries, while blurred edges indicate shadows, surface curvature, or defocus due to relative depth. Edge blur also drives accommodation and may be implicated in the correct development of the eye's optical power. Here we use classification image techniques to reveal the mechanisms underlying blur detection in human vision. Observers were shown a sharp and a blurred edge in white noise and had to identify the blurred edge. The resultant smoothed classification image derived from these experiments was similar to a derivative of a Gaussian filter. We also fitted a number of edge detection models (MIRAGE, N_1, and N_3^+) and the ideal observer to observer responses, but none performed as well as the classification image. However, observer responses were well fitted by a recently developed optimal edge detector model, coupled with a Bayesian prior on the expected blurs in the stimulus. This model outperformed the classification image when performance was measured by the Akaike Information Criterion. This result strongly suggests that humans use optimal edge detection filters to detect edges and encode their blur.

Introduction

It is widely accepted that edges play an important role in visual perception. Edges are usually defined as sudden changes in image intensity, but many edges have a gradual change in intensity. When this occurs the edge is perceived as blurred. Blur may be intrinsic to the physical edge formation process, such as an attached shadow edge on a curved surface or a penumbral shadow boundary. Blur may also be caused by the optics of the human eye. Outside of a small, foveated volume of space, the edges in the retinal image are blurred because of the limited depth of field of the eye. Both intrinsic and optical blur can be exploited to inform us about the world. The intrinsic blur of shadow edges and the sharpness of boundary edges provides information for shape-from-shading, and optical blur is a cue to depth (Marshall, Burbeck, Ariely, Rolland, & Martin, 1996; Mather, 1996, 1997; Mather & Smith, 2000). Optical blur is also the primary driver of accommodation (Kruger & Pola, 1986; Phillips & Stark, 1977), and it may be implicated in emmetropization (Flitcroft, 1998), which is the developmental process that matches the eye's length to its optical power.

Humans are good at detecting blur (Hamerly & Dvorak, 1981; Watt & Morgan, 1983), but it is unclear how they do it. One possibility is that edge blur is extracted as a byproduct of edge detection (Elder & Zucker, 1998; Georgeson, May, Freeman, & Hesse, 2007; Lindeberg, 1994, 1998; May & Georgeson, 2007a, 2007b; Watt & Morgan, 1985). Generally, it is assumed that edges are detected by oriented filters, much like the receptive fields of V1 simple cells, and these filters have a range of sizes, or scales, in order to capture the range of edge blurs that can occur. For example, in Marr's Primal Sketch (Marr & Hildreth, 1980) the image is initially analyzed with Laplacian-of-Gaussian filters of different sizes to yield zero crossings. These are then integrated into edges in V1, in which each edge is characterized by various properties, including its scale, or blur. The MIRAGE model (Watt & Morgan, 1983, 1985) also assumes that an image is analyzed by a bank of filters of different widths. Uniquely, however, the MIRAGE model combines these filter outputs into a single representation of the image in which scale is not explicitly represented; edge location and blur are then extracted by further simple calculations. The N_1 and N_3^+ models (Georgeson et al., 2007), based upon the theoretical work of Lindeberg (1994, 1998), also assume that the image is analyzed by a bank of filters of different scales. In the N_1 and N_3^+ models, the location and blur of an edge are found by looking for peaks in a scale-space representation of the image (Lindeberg, 1998; Witkin, 1983). Other models of edge

detection (Elder & Zucker, 1998) also embody the idea that to detect edges, the image must be analyzed at different scales.

Here we examine the mechanisms of human blur detection using classification images (Beard & Ahumada, 1998; Murray, 2011). To carry out a classification image experiment, one begins with a psychophysical task, such as detecting the difference between a sharp edge and a slightly blurred edge. To each stimulus, white noise is then added. The observer's behavior on each trial, which is influenced by the specific noise pattern used on that trial, can be used to infer how the observer actually detects blur. Classification images have previously been used to characterize human edge detection filters (Morgenstern, Elder, & Hou, 2004), but not in detail.

Our results show that the classification images for blur detection are similar to a difference of Gaussian derivative filters. We also fitted three models of blur detection, the MIRAGE model (Watt & Morgan, 1985), the N_1 model (Georgeson et al., 2007), and the N_3^+ model (Georgeson et al., 2007), to our data by defining appropriate decision variables for those models. We find that none of them fits our blur detection data very well. This is surprising because these models have received support from many experiments—blur thresholds (Watt & Morgan, 1983, 1985), blur matching tasks (Georgeson, 1994; Georgeson et al., 2007; May & Georgeson, 2007a, 2007b), and the reported perception of edge location (Georgeson & Freeman, 1997; Hesse & Georgeson, 2005).

We find however that a new blur detection model, using optimal edge detection filters (McIlhagga, 2011) combined with a Bayesian decision rule, is able to account for the blur detection data better than the classification image.

Methods

Stimuli

In each trial, observers were shown two horizontal edges, one of which was sharp and the other blurred, and had to select the blurred edge by pressing a mouse button. Both edges were embedded in high contrast noise. An example of the stimulus is shown in Figure 1. Each edge image was 1.67° wide and 1.67° tall (400 by 400 pixels), separated by a gap of 0.4°. The stimuli were presented for 250 milliseconds on a gamma-corrected CRT monitor (Mitsubishi 2070SB 22 inches) in a dark room. The monitor was driven by a Bits++ processor (Cambridge Research Systems, Ltd) that generated 14-bit gray-scale images. The background was grey with a luminance of 50.6 cd/m^2.
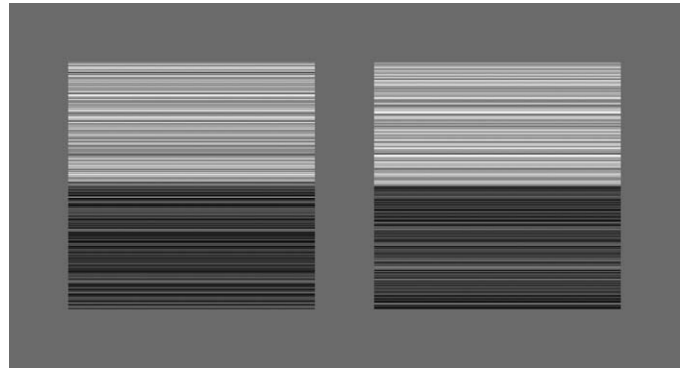

Figure 1. An example of the stimuli, consisting of two horizontal edges embedded in horizontal white noise. The observer had to indicate which edge was blurred, here the left one.

The contrast pattern of the sharp edge on the ith trial was $s_i(x, y) = S(x) + n_i(x)$, where $S(x)$ is a sharp step edge profile and $n_i(x)$ is a Gaussian white noise sample. (Here we use x to refer to the vertical dimension on the stimulus, and y to refer to the horizontal dimension.) The contrast pattern of the blurred edge on the ith trial was $b_i(x, y) = B(x) + m_i(x)$, where $B(x)$ is a blurred edge formed by convolving a step edge profile with a Gaussian filter having scale σ, and $m_i(x)$ is another noise sample. Contrast is defined as the luminance at a point divided by the mean luminance, minus 1. Both sharp and blurred edges had a contrast difference across the edge of 0.4 (i.e., a Michelson contrast of 0.2). Noise was created by adding an independent pseudo-random noise value to each scan line of each edge image. The noise values were drawn from a Gaussian distribution with a standard deviation of either 0.16 (low-noise condition) or 0.32 (high noise condition) in contrast units. The one-dimensional spectral power densities were 0.8×10^{-4} deg^{-1} and 3.2×10^{-4} deg^{-1}, respectively. Before adding the noise signal to the edge, we truncated it to fall between ± 0.8, so that the combined signal fell within the range of physically realizable values, $[-1, 1]$. Each noise sample was stored for classification image analysis as described below. These images are constant in the y direction, so the y coordinate is ignored from here on.

Three observers (the authors KAM and WHM and a student TS) each did 5,000 trials, in 250 trial blocks, at both low and high noise levels. The low-noise experiments were done first, then the high-noise experiments. The scale of the Gaussian blurring filter was chosen for each observer so that their probability correct in the low-noise condition was close to 75%. Experimental procedures conformed to the Ethical Guidelines of the Bradford School of Optometry and Vision Science, as approved by the University of Bradford Life Sciences Ethics Committee. All of the data from these experiments are available in the supplementary material.

Classification images

In a single experimental trial, the observer sees two images, which we will label I_1 and I_2, and must decide which contains the blurred edge. One way they could do this is to compute a weighted sum of the contrasts in each image, and select the image which maximizes this sum as being the most blurred. The vector of weights is called a template. The weighted sum of image j is

$$\sum_x \theta(x)I_j(x) = \theta \cdot I_j \qquad (1)$$

where θ is the template vector, indexed by position x. The observer decides image I_1 contains the blurred edge if $\theta \cdot I_1 - \theta \cdot I_2 > 0$; otherwise they decide image I_2 was blurred. The difference $\theta \cdot (I_1 - I_2)$ is called a decision variable.

The observer will be correct on trial i when $\theta \cdot (b_i - s_i) > 0$, where the actual blurred and sharp images on that trial have been substituted into the decision variable. However, human observers often make a different choice when shown the same stimulus again, which must be caused by some internal randomness, or noise, unrelated to the stimuli. If we assume, for convenience, that the internal noise is a standard logistic variable, the observer's probability correct on the ith trial is a logistic function of the decision variable:

$$p(b_i, s_i, \theta) = \{1 - \exp[-\theta \cdot (b_i - s_i)]\}^{-1} \qquad (2)$$

Now, let c_i be 1 if the human observer actually was correct on the ith trial of the experiment, and 0 otherwise. The log-likelihood of the observer's responses, given the template θ, is then

$$L(\theta) = \sum_i \{c_i \log[p(b_i, s_i, \theta)]$$
$$+ (1 - c_i)\log[1 - p(b_i, s_i, \theta)]\} \qquad (3)$$

An estimate of the template θ is a called a classification image. Classification images are commonly computed from the difference between the mean noise pattern when the observer is correct and the mean noise pattern when the observer is incorrect (Ahumada, 2002; Beard & Ahumada, 1998; Murray, 2011; Murray, Bennett, & Sekuler, 2002). However, we used the maximum likelihood estimate for θ, which can be computed by logistic regression (Knoblauch & Maloney, 2008; Nelder & Wedderburn, 1972). The covariate matrix used in the logistic regression is $X_{i,j} = b_i(j) - s_i(j)$, where the ith row of X contains the difference between the blurred and sharp stimuli on the ith trial. This was regressed against the observation vector c_i.

Classification images are usually very noisy because of overfitting. This occurs because there are so many parameters in the classification image (400 in our case)

that it is possible to exploit wholly accidental correlations between stimulus noise and observer responses to improve the likelihood. Classification images are, for this reason, often improved in appearance by some ad hoc smoothing applied after the image has been estimated. However, smoothing is better accomplished by adding a penalty to the likelihood which is proportional to the sum of squared second differences of the template (Hastie & Tibshirani, 1986; Knoblauch & Maloney, 2008) yielding a penalized likelihood:

$$L_p(\theta) = L(\theta) - \lambda \sum_x [\theta(x-1) - 2\theta(x) + \theta(x+1)]^2 \qquad (4)$$

Smooth classification images can also be estimated by logistic regression using an augmented covariate matrix.

We used the Akaike Information Criterion (AIC) (Akaike, 1974) to choose the best smoothing parameter λ. The AIC is a model selection measure that takes into account both the likelihood of a model and its complexity. It is defined as $-2L(\theta) + 2N(\theta)$, where $N(\theta)$ is the effective number of parameters. The effective number of parameters is the trace of the projection matrix of the logistic regression on the final convergent iteration (Hastie & Tibshirani, 1986) and is reduced as the smoothing increases. The magnitude of the AIC is not meaningful, but differences between AICs are (Burnham & Anderson, 2004). When selecting amongst models, the one with the lowest AIC is to be preferred, so for the smoothing parameter, we chose the value of λ which minimized the AIC.

Alternatives to classification images

While most of the emphasis in classification image experiments is placed on the resulting estimate of the template, the critical aspect of classification image analysis is its attempt to explain human responses on a trial-by-trial basis, using unique stimuli on each trial. The classification image itself is simply a convenient approximation to the human observer, which uses a linear decision variable. Instead of this, however, we could use any other model for blur detection from which we can derive a suitable decision variable.

In general, a decision variable is a function $d(I_1, I_2, \varphi)$ of the two stimuli and a set of parameters φ. The observer will choose stimulus I_1 as being the blurred edge if $d(I_1, I_2, \varphi) > 0$; otherwise they will choose stimulus I_2. Given a decision variable for a particular model, the probability of a correct response is simply

$$p(b_i, s_i, \varphi) = 1 - \exp[-k \cdot d(b_i, s_i, \varphi)]^{-1} \qquad (5)$$

with the model decision variable replacing the linear decision variable in Equation 2. The proportionality factor k is needed because of the assumption that the

internal noise is a standard logistic variable. This probability correct is then substituted into Equation 3, and the model parameters φ can be estimated by maximum likelihood. This approach is an extension of that used by Solomon (2002), who fitted parameterized templates by maximum likelihood. Here, however, we fit entire models.

We use this approach to fit the MIRAGE model (Watt & Morgan, 1985), the N_1 and N_3^+ models (Georgeson et al., 2007), and an optimal edge detector (McIlhagga, 2011) to our blur detection data. Some models of blur detection that focus on predicting blur thresholds (e.g., Watson & Ahumada, 2011) give the magnitude of a decision variable, but not its sign. These kinds of models are not intended to be used for trial-by-trial modeling and so we did not attempt to fit them.

Results

Classification images

Figure 2 shows the unsmoothed and the best smoothed classification images for all three observers in the low noise condition. The AICs for the best smoothed classification images are given in the first row of Table 1. The second row of the table gives the difference ΔAIC between the AIC of the unsmoothed classification image and the AIC of the best smoothed classification image. In all cases, the smoothed classi-

fication image has a better AIC than the unsmoothed one. The classification images in Figure 2 sometimes resemble a difference of two Gaussian first derivative filters, which would be consistent with it being a difference of two simple cell receptive fields (Hawken & Parker, 1987; Parker & Hawken, 1988; Ringach, 2002, 2004), tuned to the sharp and the blurred edges. Our classification images differ from previous classification image estimates of edge detectors (Morgenstern et al., 2004), which appeared more nearly like simple Gaussian first derivative filters. However, our task is different: our observers had to compare two different edges and report which was the most blurred, whereas Morgenstern and colleagues' observers had to compare an edge with a blank and report which was the edge.

The human visual system is almost certainly not linear, unlike the classification image model. However, the classification image is related to the true human decision variable in a straightforward way. Whatever the structure of the true human decision variable $d_{human}(I_1, I_2)$ it can be expanded as a Taylor series in the stimulus contrasts I_1 and I_2. The first order term of this Taylor series is a linear combination of stimulus contrasts I_1 and I_2, like the classification image. Thus the classification image can be thought of as an estimate of the first-order term of the true decision variable. This means that the AIC of the smoothed classification image can be used as a benchmark for accepting or rejecting alternative models for human blur detection. If some alternative model does not have a better AIC than the classification image, then it is worse than a first
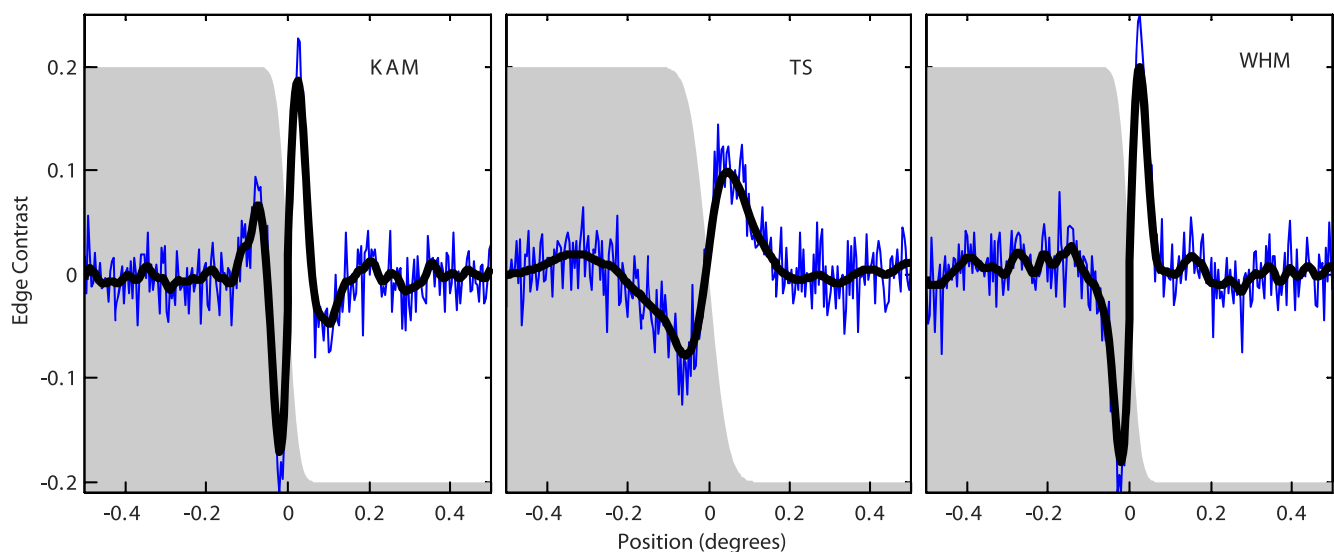

Figure 2. Stimulus profiles and classification images for all observers in the low-noise condition. The gray-shaded area is the luminance profile of the blurred edge, $B(x)$, used for each observer. The Gaussian blur σ used to form the blurred edge was 0.0215°, 0.0367°, and 0.0216°, respectively, for KAM, TS, and WHM. Subject TS needed more blur than the other two to obtain about 75% correct. The x axis is in degrees of visual angle. The y axis is the contrast for the edges. The maximum likelihood classification image is shown as a thin blue line, and the best smoothed classification image by the thick black line. The height of the classification image is arbitrarily determined by the assumption of unit internal noise. However, the classification images share the same scale across all three panels.

Subject:	KAM		TS		WHM	
Noise contrast	0.16	0.32	0.16	0.32	0.16	0.32
1) Smoothed classification image	4494 ($N = 76$)	5800 ($N = 72$)	5083 ($N = 34$)	5038 ($N = 48$)	4291 ($N = 75$)	5439 ($N = 50$)
ΔAIC for						
2) Unsmoothed classification image ($N = 400$)	349	322	264	363	307	387
3) Ideal observer ($N = 1$)	654	570	630	1053	703	978
4) MIRAGE ($N = 1$)	1440	998	1566	1857	1778	1438
5) N_1 model ($N = 2$)	502	523	560	911	957	902
6) N_3^+ model ($N = 2$)	1744	994	1312	1695	1992	1332
7) Optimal edge detector, Bayesian ($N = 6$)	**−178**	**−274**	94	**−1**	**−72**	**−90**

Table 1. AIC for the smoothed classification image (row 1) and ΔAIC for the models mentioned later in the text. The values in brackets in row 1 are the effective number of parameters N for the best smoothed classification image. They vary because different amounts of smoothing proved best for different observers and noise levels. The ΔAIC values are the AIC for the given model (rows 2–7) minus the AIC for the best smoothed classification image. Positive values indicate the model fits worse than the best smoothed classification image in row 1; negative values (in bold) indicate that it fits better. The effective number of parameters N for each of the models is given in brackets after each model name. A simple rule of thumb is that AIC differences less than 2 suggest both models are more or less equivalent, while differences greater than 10 indicate the worse model has essentially no support from the data (Burnham & Anderson, 2004).

order approximation to the true human decision variable. In that case it is unlikely to be correct. Using this criterion, we can evaluate other possible models for human blur detection. We turn to this next.

The ideal observer

The ideal observer (Geisler, 1989, 2011) is a widely used theoretical observer who is statistically optimal for the task at hand, in terms of the probability of a correct choice. An ideal observer who wishes to maximize the probability of correctly selecting the blurred edge from two stimulus images I_1 and I_2 will do so by computing two log-likelihoods. The first is the log-likelihood that image I_1 contains the blurred edge and I_2 contains the sharp edge. The second is the log likelihood of the alternative possibility that image I_1 contains the sharp edge and I_2 contains the blurred edge. They choose the alternative that has the highest likelihood as being the one most likely to be correct. It can be shown that, in additive Gaussian noise, this is equivalent to computing a linear decision variable $\theta_{ideal} \cdot (I_1 - I_2)$, where the ideal template is proportional to the difference between the blurred and sharp edges, $\theta_{ideal}(x) = k[B(x) - S(x)]$, where k is a scaling factor.

We can work out the log-likelihood for the ideal observer by substituting θ_{ideal} into Equation 3. In doing so, we are implicitly adding internal noise to the ideal observer in order to improve their fit. There is one free parameter here, the scaling factor k. The ΔAIC values for the ideal observer, compared to the smoothed classification image, are shown in row 3 of Table 1. In all cases, the ideal observer is substantially worse than the best classification image and so is highly unlikely to

be a correct account of human performance in this task.

MIRAGE

The MIRAGE model (Watt & Morgan, 1985) offers an alternative model for human blur perception. MIRAGE begins by convolving an image with filters that are second derivative of Gaussians of different scales. It then splits the output of each filter into positive and negative halves, and adds the positive halves of all filter outputs together, and the negative halves of all filter outputs together. These half-signals are further parsed into "zero-bounded response distributions" or RESPs. A RESP is a maximal spatial interval in the signal that is nonzero inside and zero outside. It has been suggested that edge blur is encoded in the distance between the centroids of two adjacent RESPs, one in the positive half-signal and one in the negative half-signal.

When applied to our noisy stimulus images, MIRAGE found many RESPs. We assumed that the adjacent RESPs with the largest mass were those most likely to correspond to the edge rather than noise. We chose the most obvious decision variable for the MIRAGE model, which is

$$d(I_1, I_2) = (\text{centroid distance in image } I_1)$$
$$-(\text{centroid distance in image } I_2) \quad (6)$$

This is used in Equation 5 to calculate the MIRAGE probabilities correct, and from these the MIRAGE log-likelihood and AIC can be calculated. The ΔAIC values of the MIRAGE model are shown in row 4 of Table 1.

In all cases, the MIRAGE model fitted the observer responses very poorly compared to the smoothed classification image and is unlikely to be correct.

The N_1 and $N_3{}^+$ models

Georgeson et al. ([2007](#)) described two models for human edge and blur perception. Both models work by computing derivatives of the image at different scales. Both the N_1 and $N_3{}^+$ models yield a scale-space representation of the input image (Witkin, [1983](#)), which is a representation of the image over a range of scales. The scale filters $F(x,\sigma)$ in the N_1 model are normalized derivatives of Gaussians

$$F(x, \sigma) = \sigma^p \exp(-x^2/[2\sigma^2])/\sqrt{2\pi\sigma^2} \qquad (7)$$

The normalization exponent p affects which filter responds best to an edge with a particular blur. The edges in the image are found by looking for local maxima, or peaks, in the scale space. The spatial coordinates of the local maximum gives the location of the edge, and the scale coordinate of the local maximum is proportional to the blur of the edge. A Gaussian blurred edge with scale σ_e will be detected by a filter with scale $\sigma_e\sqrt{p/(n-p)}$, where $n = 1$ for the N_1 model and $n = 3$ for the $N_3{}^+$ model; $p = n/2$ is a conventional choice here, but anything between 0 and n is valid.

Let $Scale(x, \sigma)$ be the scale space produced from an input image by either the N_1 or $N_3{}^+$ filters. Stimulus noise generates many peaks in the scale space, and we have to select the one that corresponds to the sharp or blurred edge. We choose the peak that has the greatest edge contrast. To do this, we find all peaks in scale space and multiply the height of each by a correction factor to get the contrast of the edge, and then choose the one with the highest contrast. Note that the peaks are found before the correction factor is applied. The correction factor for $N_3{}^+$ is derived in May and Georgeson ([2007a](#)), equation 2; the correction factor for N_1 is derived similarly.

The estimate of edge blur is the scale coordinate of the chosen peak multiplied by $\sqrt{(n-p)/p}$. That is, if the best peak for image I_1 is at position x_1 and scale σ_1, our estimate for edge blur is simply $\sigma_1\sqrt{(n-p)/p}$. An obvious decision variable is then $d(I_1, I_2) = (\sigma_1 - \sigma_2)\sqrt{(n-p)/p}$. We used this decision variable to fit the N_1 and $N_3{}^+$ models to observer responses. The scales ranged from 1 to 60 pixels (0.0042° to 0.252°), logarithmically spaced. The exact choice of scales had only minor influence on the fit. We assumed the observer knew the location of the edge and only had to find the peak in scale. (Relaxing this assumption worsened the fit.) ΔAIC values for the N_1 and $N_3{}^+$ models are given in [Table 1](#), rows 5 and 6. These AIC values were obtained by finding the normalization exponent p which yielded the smallest AIC. Neither N_1 nor $N_3{}^+$ fits the data very well, when compared to the fit of the smoothed classification image. The main reason for the poor fit was that in both models, the scale space was overwhelmed by noise peaks.

Optimal edge detection

We turn now to a model that does fit the data better than the best classification image, so it is a strong candidate for an accurate description of human edge detection. It is likely that, whatever process humans use to detect edges, it has been driven towards an optimum by natural selection (Geisler, [2011](#)). We may therefore gain some insight into human edge detection by studying optimal edge detection. The best known approach to optimal edge detection is that of Canny ([1986](#)), who proposed that edge detectors should optimize the product of signal-to-noise ratio and the precision of edge localization. However, Canny oversimplified the localization measure (Koplowitz & Greco, [1994](#); McIlhagga, [2011](#); Tagare & deFigueiredo, [1990](#)) and neglected the impact of nearby edges on the edge detection process (McIlhagga, [2011](#)). When these problems are fixed and the edge detectors further generalized to detect edges of different scales, the optimal detection filter D_σ for an edge of scale σ (here defined as a step edge convolved with a Gaussian filter of scale σ) can be approximated by a convolution of three filters (McIlhagga, [2011](#))

$$D_\sigma(x) = W(x) \cdot g(x, \sigma_0) \cdot M_\sigma(x) \qquad (8)$$

where $W(x)$ is a whitening filter, $g(x, \sigma_0)$ is an auxiliary Gaussian filter with a fixed scale σ_0 and $M_\sigma(x)$ is a filter matched to the shape of an edge of scale σ after it has been whitened. The matched filter $M_\sigma(x)$ is normalized to have an r.m.s. power of 1. The whitening filter $W(x)$ whitens images having a natural-image power spectrum $C^2/f^2 + n_0^2$ (Burton & Moorhead, [1987](#); Field, [1987](#)), where C^2/f^2 is brown noise and n_0^2 is the squared amplitude of the white noise. The whitening filter acts like a smoothed derivative operator. The optimal detector has two parameters, the ratio C/n_0, which is estimated from the image, and the scale σ_0 for the auxiliary Gaussian filter, which should be small. The optimal edge detector is diagrammed in [Figure 3](#).

The convolution of $D_\sigma(x)$ with an image $I(x)$ represents the image at a single scale. To represent the image at all scales, we must convolve an image $I(x)$ with optimal edge detectors at different scales. The collection of these convolutions is a scale-space representation of the image $R(x, \sigma)$, given by

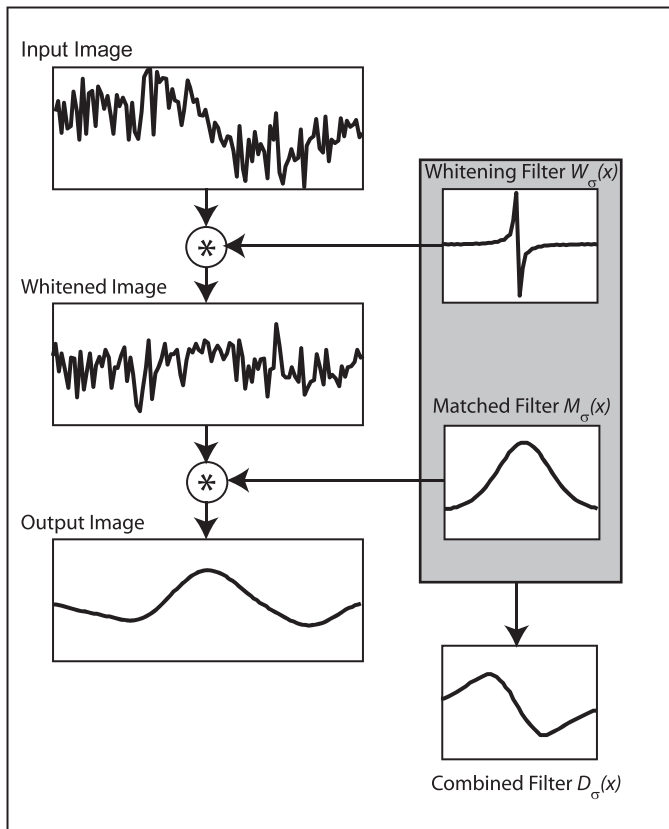$$R(x, \sigma) = I(x) \cdot D_\sigma(x) \qquad (9)$$

Figure 3. Operation of the optimal edge detector at a single scale. A noisy input image is first whitened by a whitening filter to yield a whitened image. The whitening filter is a form of smoothed derivative. The whitened image is then convolved with a filter matched to a whitened edge of a particular scale. This yields a final output image. This is the representation of the input image at a single scale. The collection of all final images at all scales used forms a scale space. An example of the scale space is shown in Figure 4. The process of whitening followed by matching can be collapsed into a single convolution with a combined filter, shown in bottom right. The combined filters are what are shown in Figure 4. (The auxiliary Gaussian filter has been omitted in this diagram.)

The square of the scale space $R(x, \sigma)^2$ is related to the log likelihood of observing the image $I(x)$ given there is an edge at position x with scale σ (McIlhagga, 2011):

$$R(x, \sigma)^2 = \log P(I|\text{edge at } x, \sigma) + \text{constant} \quad (10)$$

If all locations and blurs are equally probable, the maximum of $R(x, \sigma)^2$ gives the location and blur associated with the most probable edge.

Optimal detection with a Bayesian prior

In our experiments, however, the location and scale are not equiprobable; only one location is possible, and only two scales. In this case, the observer should use Bayes' theorem to compute the posterior probability of an edge by combining the likelihood $R(x, \sigma)^2$ with their prior distribution of edge location and scale. For simplicity, we will assume the observer knows the edge position exactly, and so consider only the scale coordinate. A Bayesian observer who views two stimulus images I_1 and I_2 in our experiment may hypothesize that image I_1 contains a blurred edge with scale σ_b, and image I_2 contains a sharp edge with scale σ_s. Letting $\pi(\sigma_b, \sigma_s)$ be the observer's prior probability for this hypothesis, the log posterior probability is

$$\log P(\sigma_b, \sigma_s|I_1, I_2) = \log P(I_1|\sigma_b) + \log P(I_2|\sigma_s)$$
$$+ \log \pi(\sigma_b, \sigma_s) - \log P(I_1, I_2)$$
$$= R_1(\sigma_b)^2 + R_2(\sigma_s)^2$$
$$+ \log \pi(\sigma_b, \sigma_s) + \text{constant}$$
$$(11)$$

where R_1 and R_2 are the scale space representations of images I_1 and I_2 at spatial position $x = 0$. Alternatively, the observer may hypothesise that image I_2 contains a blurred edge with scale σ_b', and image I_1 contains a sharp edge with scale σ_s'. The log-probability of this hypothesis is

$$\log P(\sigma_b', \sigma_s'|I_2, I_1) = R_1(\sigma_s')^2 + R_2(\sigma_b')^2$$
$$+ \log \pi(\sigma_b', \sigma_s') + \text{constant}$$
$$(12)$$

when $\sigma_b' > \sigma_s'$. The constant in this equation is identical to the one in Equation 11. The optimal decision rule is to decide that image I_1 contains the blurred edge and I_2 the sharp edge when

$$\iint\limits_{\sigma_b > \sigma_s} P(\sigma_b, \sigma_s|I_1, I_2) d\sigma_s d\sigma_b >$$
$$\iint\limits_{\sigma_b' > \sigma_s'} P(\sigma_b', \sigma_s'|I_2, I_1) d\sigma_s' d\sigma_b' \quad (13)$$

However, this calculation is computationally expensive, since one would have to take the scale space outputs, exponentiate them, then integrate them. In addition, it does not directly yield an estimate of the edge blur, for which one would have to compute the posterior mean. (It did not fit the data particularly well either.)

An alternative to the optimal decision rule is to prefer the first hypothesis (blurred edge in image I_1) when the maximum posterior probability of the first hypothesis exceeds the maximum posterior probability of the second; that is, when

$$\max_{\sigma_b, \sigma_s}[\log P(\sigma_b, \sigma_s|I_1, I_2)]$$
$$- \max_{\sigma_b', \sigma_s'}[\log P(\sigma_b', \sigma_s'|I_2, I_1)] > 0 \quad (14)$$

The decision variable for this model, $d(I_1, I_2)$, is simply the left hand side of this inequality. It can be easily computed from the output of the optimal edge detector, and estimates of the edge blurs are immediately available as the values of σ_b, σ_s or σ'_b, σ'_s which yielded the maximum.

To fit this model, we need a prior. We assume for simplicity that the prior is an almost separable function

$$\pi(\sigma_b, \sigma_s) = \begin{cases} \pi_b(\sigma_b)\pi_s(\sigma_s) & \text{if } \sigma_b > \sigma_s \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where π_b and π_s are priors for the sharp and blurred edge scales at the known position $x = 0$. This prior is the observer's belief about the distribution of scale, not the true distribution. The sharp and blurred edge priors were modeled as beta distributions because these are flexible distributions with a finite domain.

Fitting the optimal edge detector

The optimal edge detector model has six free parameters: the proportionality factor k, the auxiliary blur in the optimal filters σ_0, and two beta parameters for each of the two priors π_b and π_s. The value of σ_0 was fitted individually by subject and noise level to provide the best fit. No attempt was made to enforce consistency of the parameters within subject. The whitening parameter C/n_0 is not a free parameter and was estimated for each subject and noise level from the collection of all stimuli shown to that subject. The same set of scales adopted for the N_1 and N_3^+ models were used here (up to 60 pixels, or 0.187°), except that the scales for subject TS extended out to 80 pixels (0.25°). The choice of scales affects the AIC only marginally. Fitting of the free parameters for the optimal detector was difficult, and we adopted a semi-Monte Carlo method, in which a Nelder-Mead minimization routine (routine *fmins* in Matlab) was started at many randomly selected initial values, and the best one selected.

We find that this model, based on optimal edge detection, is an overwhelmingly superior to the fit of the best classification image for KAM and WHM (Table 1, last row). Thus it is a strong candidate for the edge detection processes in those observers. The model is illustrated in Figure 4 for observer KAM.

Figure 5 shows a typical psychometric function derived from the optimal edge detection model. This shows the observer probability correct as a function of the optimal detector's decision variable. The psychometric function was constructed as follows. For every possible value d of the decision variable, we selected a subset of trials in an interval around d. We then measured the observer's probability correct over this subset of trials. If the model fits the observer responses,
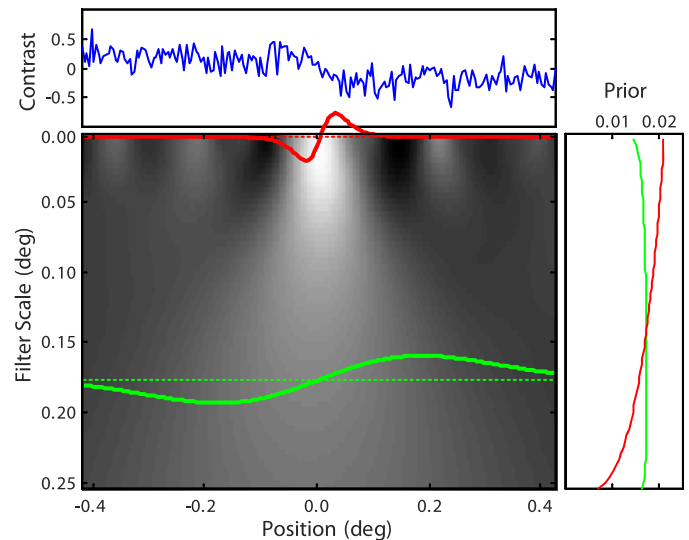

Figure 4. The Bayesian version of the optimal edge detector model. The top panel plots an example luminance profile of a noisy blurred edge $b_i(x)$. Only the central half of the stimulus is displayed. The lower left image shows the output of the optimal edge detectors as a scale space $R(x, \sigma)$. The detector location is along the horizontal axis (in degrees), and the scale σ is plotted along the vertical axis (finer scales at the top, coarser at the bottom). The lower right hand panel shows the priors for the sharp edge (in red) and blurred edge (in green) for each filter scale. The filters drawn on top of the scale space image are the combined filters $D_\sigma(x)$ that correspond to the maxima of the respective priors. This plot is for observer KAM in the low-noise condition.

we would expect that the observer's probability correct should be a smooth logistic function of the model's decision variable.

Although the optimal edge detection model fitted KAM and WHM well, it was rather worse for subject TS; it is rejected for the low noise condition and is essentially the same as the smoothed classification image in the high noise condition. There could be many reasons for this, but one possibility is that, being inexperienced, observer TS did not use a consistent decision rule during the experiment. TS would of course not have a consistent classification image either, but the average classification image might fit all the data better than the average optimal edge detector. To test this, we split TS's data for each noise level into four sequentially recorded parts and fitted each part with both a smoothed classification image and an optimal edge detector model. We found only minor improvement in fit for the low noise data. However, in the high noise data (which was collected last) the optimal edge detector model was better than the best smoothed classification image in all four parts, sometimes overwhelmingly so (the ∆AICs for each part were −19, −19, −1, and −20, all in favor of the optimal edge detector). This suggests that the consistency of observer TS improved over time so that by the time he did the

high noise experiment his decision criterion varied slowly enough for the optimal model to fit his responses well over short runs.

Goodness of fit

The AIC is a model-selection tool. It assesses the relative performance of competing models, but does not indicate whether the best model actually fits the data. Unfortunately, standard goodness-of-fit tests are insensitive in large sparse logistic regressions such as this one (Kuss, 2002), with one data point per cell and many cells. Thus we evaluated goodness of fit by simulation.

Let $p_i = p(b_i, s_i, \theta)$ be the probability correct for the ith trial, as specified by the model. Let c_i be 1 or 0 depending on whether the observer was actually correct. The observed likelihood of the model is $L_{obs} = \Sigma\, c_i \log p_i + (1 + c_i) \log(1 - p_i)$. Now simulate an observer by setting c_i^{sim} equal to 1 if a uniform random variable r_i is less than p_i. The likelihood of the simulated observer is $L_{sim} = \Sigma c_i^{sim} \log p_i + (1 - c_i^{sim}) \log(1 - p_i)$. We can repeat this many times to find the empirical distribution of simulated likelihoods conditional on the model probabilities, i.e., the distribution of observed likelihoods that would occur if the model was precisely correct. If the observed likelihood L_{obs} is consistent with being drawn from this simulated distribution, then the model fits the data.

We found that L_{obs} invariably fell between the 40th and 60th percentiles of the simulated likelihood distribution, so the observer responses are entirely consistent with the optimal edge detector model for all observers and noise levels.

Discussion

The classification image technique is a powerful way of evaluating models for human visual performance. One important insight here is that the fit of the classification image provides a useful benchmark for evaluating the effectiveness of other models in accounting for human perception. We found that the only model that might explain blur perception is one based on a set of optimal edge detectors. In this model, blur detection is a byproduct of edge detection over a range of scales. Our observers combined the outputs of these edge detectors with a Bayesian prior to yield a decision. Both parts of the model, the optimal edge detectors and the Bayesian decision rule, are necessary to explain human performance in this experiment. However, neither the prior identified here nor the specific optimal filters are directly applicable to normal viewing
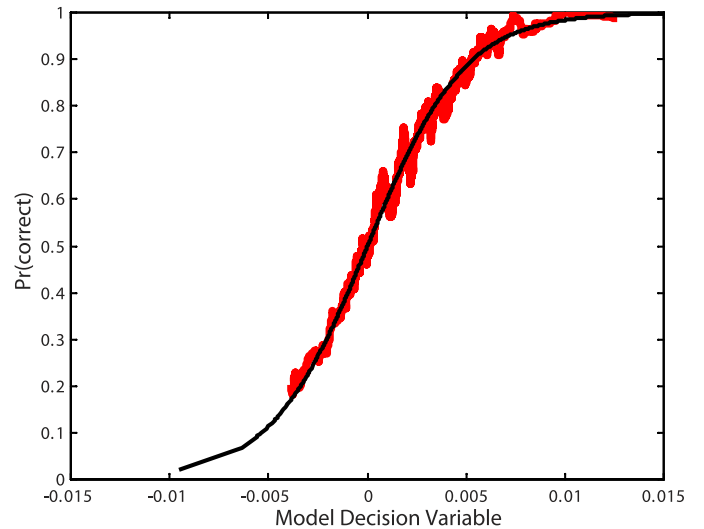

Figure 5. How well the optimal edge detector model accounts for observer KAM's probability correct. The x axis plots the decision variable for the optimal model, and the black curve gives the model's probability correct as a function of the decision variable. The red jagged line shows the human observer's probability correct, as a function of the model decision variable. This was calculated as follows. For each value of x, we selected trials in which the model decision variable was near x, and then calculated the observer's probability correct within that set of trials.

conditions. The prior is obviously task dependent; and under normal viewing the optimal filters would change shape because they depend on image statistics through the parameter C/n_0.

Here we look at some aspects of this model: the role and significance of the whitening stage; relationship to other edge detection models; and the relationship to the ideal observer.

Whitening

One aspect of the optimal edge detector model is the way it adapts to image statistics. The whitening filter $W(x)$ and the matched filter $M_\sigma(x)$ both change in response to changes in image statistics. The main use for this adaptive change in the experiments reported here is to cope with large amounts of noise, but in other circumstances it adjusts the edge detectors to follow the image statistics. In particular, the optimal edge detector will adapt to image blur. Humans also adapt to image blur (Webster, Georgeson, & Webster, 2002). The optimal edge detector model suggests that the adaptive process occurs in order to optimize the edge detection performance, and this is consistent with reports that blur sensitivity improves after adaptation to blur (Cufflin, Mankowska, & Mallen, 2007).

The whitening filter is separate in the optimal edge detector model only as a mathematically convenient way of specifying the optimal filters. However, it may be physiologically separate in the human visual system, too. It has frequently been suggested that the retina performs a whitening-like operation to encode the image for optimal transmission along the optic nerve (Atick & Redlich, 1992; Ruderman, 1994). If so, we could identify the whitening filter in the optimal edge detector with retinal filtering, and the subsequent matched filtering with the computations of neurons in areas V1 and later.

Relationship to other edge detectors

The optimal edge detector model is a generalization of the N_1 model. When the white noise is zero, the whitening filter $W(x)$ becomes a derivative operator, and the matched filter $M_\sigma(x)$ becomes a Gaussian function. Under these conditions, the optimal filter $D_\sigma(x)$ is a derivative of a Gaussian, which is the filter shape suggested by Lindeberg (1998) and the N_1 model (Georgeson et al., 2007), among others. Given that the optimal edge detector is so similar to the N_1 model, perhaps an N_1 model with a Bayesian prior might fit the data better than the simple N_1 model we used. We added the same form of Bayesian prior as used in the optimal model to the N_1 model with normalization exponent $p = 0.5$ (This normalization is needed for the N_1 filter outputs to be interpreted as likelihoods.) This Bayes-N_1 model yielded ΔAIC values of 1012, -20, 630, 57, -42, and 10 (in the same order as the columns of Table 1). While never better than the optimal model, the Bayes-N_1 model does beat the smoothed classification image in two cases.

The N_1 model is good at accounting for human blur perception when noise is absent, but the N_3^+ model is better (Georgeson et al., 2007). The N_3^+ model is nonlinear, so its success implies that human blur perception is not like the linear model proposed here. However, we do not have an optimal theory for nonlinear filters like those used in N_3^+, so we do not yet know if a nonlinear edge detector would account for our data better than the current model. Certainly, the N_3^+ model as it stands is unable to account for our data. It is possible that human edge detection behaves like a set of optimal linear filters for high noise levels, as here, but transitions to a nonlinear detector like N_3^+ at very low noise levels.

Ideal observers and optimal detectors

Although the ideal observer fails to account for human blur perception, there is some similarity in the approach of the ideal observer and the optimal detector theory used here. The ideal observer optimizes the probability of a correct decision. The optimal detector is designed to maximize a numerical criterion of performance. It is possible to frame the optimal edge detector as an ideal observer who maximizes the probability that the edge location and scale is within a particular distance of the true location and scale (Tagare & deFigueiredo, 1990). So to a large extent, the two approaches are very similar.

Where they differ is that the ideal observer is an infinitely flexible observer, who can adapt to the requirements of the particular experiment. The optimal edge detector, however, is pressed into service to carry out a specific psychophysical task for which it may not be well suited. This is because the optimal edge detector is designed to detect edges under general, ecologically reasonable conditions, and where the experiment departs from these conditions, the performance of the optimal edge detector deteriorates. Understanding of human performance in other psychophysical tasks may benefit from this approach: we should assume that the observer uses processes that are optimized to perform a related real-world task. The goal to understanding vision should be to identify this real-world task and derive a mechanism or process—a computational model (Marr, 1982)—that is optimal for it.

Although we find no support for the ideal observer in this study, human observers have often been successfully described as an ideal observer with added noise and a so-called "sampling" inefficiency. However, the sampling inefficiency encompasses a number of ways that the human observer can deviate from the ideal, including—as is the case here—using a completely different template. We believe that, unless the sampling efficiency is quite high, the ideal observer may not be a useful way of modeling how the human observer carries out a task.

Noise

An optimal edge detector will only show a clear advantage over a suboptimal one when the signal-to-noise ratio is low. This doesn't seem to be the case in everyday life, where the retinal image is clear enough, so why would there be strong selection pressure for optimal edge detectors? It is widely believed that the brain needs to minimize the amount of energy expended in computations (Laughlin, 2001; Lennie, 2003; Levy & Baxter, 1996; Vincent, Baddeley, Troscianko, & Gilchrist, 2005), and one way it can achieve this is by minimizing the internal signal level or spiking rate. Under these conditions, internal noise can be significant, and the need for highly optimal edge

detection filters is much greater than might otherwise be expected.

Conclusions

Our experiments suggest that humans detect blur by analyzing the stimuli with a bank of optimal edge detectors, tuned to different scales. In this model, blur detection is a by-product of edge detection over a range of scales. We modeled performance using a recently derived general-purpose edge detector that optimizes itself to the prevailing image statistics to maximize edge detection performance. We found that the optimal edge detector alone could not explain human performance: our model combines the output of the general-purpose edge detector with a task-specific Bayesian prior to yield a decision. Our model predicts human performance in our blur detection task with remarkable accuracy on a trial-by-trial basis. Our model gave an overwhelmingly better fit to the data than other published models of blur perception or even the ideal observer for our task. We argue that the ideal observer may be of limited usefulness in understanding performance in psychophysical tasks: under the assumption that the observer performs the task by recruiting mechanisms that are optimized for a related real-world task, we should try to identify that task and derive a mechanism that is optimal for it. In our case, we find that a mechanism that is optimal for general-purpose edge detection explains human blur detection decisions very well.

Acknowledgments

References

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision, 2*(1):8, 121–131. http://www.journalofvision.org/content/2/1/8, doi:10.1167/2.1.8. [PubMed] [Article]

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6):716–723. doi:10.1109/TAC.1974.1100705.

Atick, J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation, 4*: 196–210.

Beard, B. L., & Ahumada, A. J. (1998). Technique to extract relevant image features for visual tasks. *Proceedings of SPIE*. (pp. 79–85). Presented at the Human Vision and Electronic Imaging III, San Jose, CA, USA. doi:10.1117/12.320099.

Burnham, K. P., & Anderson, D. R. (2004). Multi-model inference. *Sociological Methods & Research, 33*(2):261–304. doi:10.1177/0049124104268644.

Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*(1):157–170. doi:10.1364/AO.26.000157.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis Machine Intelligence, 8*(6):679–698.

Cufflin, M. P., Mankowska, A., & Mallen, E. A. H. (2007). Effect of blur adaptation on blur sensitivity and discrimination in emmetropes and myopes. *Investigative Ophthalmology & Visual Science, 48*(6):2932–2939, http://www.iovs.org/content/48/6/2932, doi:10.1167/iovs.06-0836. [PubMed] [Article]

Elder, J., & Zucker, S. (1998). Local scale control for edge detection and blur estimation. *IEEE Trans. Pattern Analysis Machine Intelligence, 20*(7):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*(12):2379–2394.

Flitcroft, D. I. (1998). A model of the contribution of oculomotor and optical factors to emmetropization and myopia. *Vision Research, 38*(19):2869–2879. doi:10.1016/S0042-6989(98)00087-X.

Geisler, W. S. (1989). Ideal observer theory in psychophysics and physiology. *Physica Scripta, 39*(1):153–160. doi:10.1088/0031-8949/39/1/025.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*(7): 771–781. doi:10.1016/j.visres.2010.09.027.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. *Ciba Foundation Symposium, 184*:147–165. discussion 165–169, 269–271.

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*(1):127–142.

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://www.journalofvision.org/content, doi:10.1167/7.13.7. [PubMed] [Article]

Hamerly, J. R., & Dvorak, C. A. (1981). Detection and discrimination of blur in edges and lines. *Journal of the Optical Society of America, 71*(4):448–452. doi:10.1364/JOSA.71.000448.

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science, 1*(3):297–318.

Hawken, M. J., & Parker, A. J. (1987). Spatial Properties of Neurons in the Monkey Striate Cortex. Proceedings of the Royal Society of London. Series B. *Biological Sciences, 231*(1263): 251–288. doi:10.1098/rspb.1987.0044.

Hesse, G. S., & Georgeson, M. A. (2005). Edges and bars: Where do people see features in 1-D images? *Vision Research, 45*(4):507–525. doi:10.1016/j.visres.2004.09.013.

Knoblauch, K., & Maloney, L. T. (2008). Estimating classification images with generalized linear and additive models. *Journal of Vision, 8*(16):10, 1–19. http://www.journalofvision.org/content/8/16/10, doi:10.1167/8.16.10. [PubMed] [Article]

Koplowitz, J., & Greco, V. (1994). On the edge location error for local maximum and zero crossing edge detectors. *IEEE Trans. Pattern Analysis Machine Intelligence, 16*(12):1207–1212.

Kruger, P. B., & Pola, J. (1986). Stimuli for accommodation: Blur, chromatic aberration and size. *Vision Research, 26*(6):957–971. doi:10.1016/0042-6989(86)90153-7.

Kuss, O. (2002). Global goodness of fit tests in logistic regression with sparse data. *Statistics in Medicine, 21*(24):3789–3801. doi:10.1002/sim.1421.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology, 11*(4):475–480. doi:16/S0959-4388(00)00237-3.

Lennie, P. (2003). The cost of cortical computation. *Current Biology, 13*(6):493–497. doi:16/S0960-9822(03)00135-0.

Levy, W. B., & Baxter, R. A. (1996). Energy efficient neural codes. *Neural Computation, 8*(3):531–543. doi:10.1162/neco.1996.8.3.531.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics, 21*(2):225–270.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision, 30*(2):79–116. doi:10.1023/A:1008045108935.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* Cambridge, MA: MIT Press.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 207*(1167): 187−217. doi:10.1098/rspb.1980.0020.

Marshall, J. A., Burbeck, C. A., Ariely, D., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 13*(4):681–688.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings: Biological Sciences, 263*(1367):169–172.

Mather, G. (1997). The use of image blur as a depth cue. *Perception, 26*(9):1147–1158. doi:10.1068/p261147.

Mather, G., & Smith, D. R. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research, 40*(25):3501–3506. doi:16/S0042-6989(00)00178-4.

May, K. A., & Georgeson, M. A. (2007a). 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.

May, K. A., & Georgeson, M. A. (2007b). Added luminance ramp alters perceived edge blur and contrast: A critical test for derivative-based models of edge coding. *Vision Research, 47*(13):1721–1731. doi:10.1016/j.visres.2007.02.018.

McIlhagga, W. (2011). The canny edge detector revisited. *International Journal of Computer Vision, 91*:251–261. doi:http://dx.doi.org/10.1007/s11263-010-0392-0.

Morgenstern, Y., Elder, J. H., & Hou, Y. (2004). Contrast dependence of spatial summation revealed by classification image analysis. *Journal of Vision, 4*(8):539, http://www.journalofvision.org/content/4/8/539, doi:10.1167/4.8.539. [Abstract]

Murray, R. F. (2011). Classification images: A review. *Journal of Vision, 11*(5):2, 1–15, http://www.journalofvision.org/content/11/5/2, doi:10.1167/11.5.2. [PubMed] [Article]

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision, 2*(1):6,

79–104. http://www.journalofvision.org/content/2/1/6, doi:10.1167/2.1.6. [PubMed] [Article]

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*(3):370–384. doi:10.2307/2344614.

Parker, A. J., & Hawken, M. J. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of the Optical Society of America A, 5*(4):598–605. doi:10.1364/JOSAA.5.000598.

Phillips, S., & Stark, L. (1977). Blur: A sufficient accommodative stimulus. *Documenta Ophthalmologica, 43*(1):65–89. doi:10.1007/BF01569293.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*(1): 455–463.

Ringach, D. L. (2004). Mapping receptive fields in primary visual cortex. *The Journal of Physiology, 558*(3):717–728. doi:10.1113/jphysiol.2004.065771.

Ruderman, D. (1994). Designing receptive fields for highest fidelity. *Network: Computation in Neural Systems, 5*(2):147–155. doi:10.1088/0954-898X/5/2/002.

Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision, 2*(1):7, 105–120. http://www.journalofvision.org/content/2/1/7, doi:10.1167/2.1.7. [PubMed] [Article]

Tagare, H., & deFigueiredo, R. (1990). On the Localization Performance measure and Optimal Edge detection. *IEEE Trans. Pattern Analysis Machine Intelligence, 12*(12):1186–1190.

Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2005). Is the early visual system optimised to be energy efficient? *Network: Computation in Neural Systems, 16*(2–3):175–190. doi:10.1080/09548980500290047.

Watson, A. B., & Ahumada, A. J. (2011). Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision, 11*(5):10, http://www.journalofvision.org/content/11/5/10, doi:10.1167/11.5.10. [PubMed] [Article]

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23*(12): 1465–1477. doi:10.1016/0042-6989(83)90158-X.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*(11):1661–1674. doi:16/0042-6989(85)90138-5.

Webster, M. A., Georgeson, M. A., & Webster, S. M. (2002). Neural adjustments to image blur. *Nature Neuroscience, 5*(9):839–840.

Witkin, A. P. (1983). Scale-space filtering. *Proceedings of the Eighth International Joint Conference on Artificial intelligence*. Volume 2 (pp. 1019–1022). Karlsruhe, West Germany: Morgan Kaufmann Publishers Inc.