Evaluation of a Multiscale Color Model for Visual Difference Prediction

P. GEORGE LOVELL, C. ALEJANDRO PÁRRAGA, and TOM TROSCIANKO
University of Bristol
and
CATERINA RIPAMONTI and DAVID J. TOLHURST
University of Cambridge

How different are two images when viewed by a human observer? There is a class of computational models which attempt to predict perceived differences between subtly different images. These are derived from theoretical considerations of human vision and are mostly validated from psychophysical experiments on stimuli, such as sinusoidal gratings. We are developing a model of visual difference prediction, based on multiscale analysis of local contrast, to be tested with psychophysical discrimination experiments on natural-scene stimuli. Here, we extend our model to account for differences in the chromatic domain by modeling differences in the luminance domain and in two opponent chromatic domains. We describe psychophysical measurements of objective (discrimination thresholds) and subjective (magnitude estimations) perceptual differences between visual stimuli derived from colored photographs of natural scenes. We use one set of psychophysical data to determine the best parameters for the model and then determine the extent to which the model generalizes to other experimental data. In particular, we show that the cues from different spatial scales and from the separate luminance and chromatic channels contribute roughly equally to discrimination and that these several cues are combined in a relatively straightforward manner. In general, the model provides good predictions of both threshold and suprathreshold image differences arising from a wide variety of geometrical and optical manipulations. This implies that models of this class can be generally useful in specifying how different two similar images will look to human observers.

Categories and Subject Descriptors: J.2 [**Physical Sciences and Engineering**]: Engineering; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Psychophysical testing, image difference metrics, color vision

1. INTRODUCTION

There has been much interest in developing computational models to predict how well human observers can discriminate differences between pairs of images. A successful model would have many uses, which include the computation of visibility of targets in natural scenes [Rohaly et al. 1997] and of measures of the perceptual effects of, lossy image compression algorithms [e.g., Lubin, 1995]. One class of model

This research was supported by the EPSRC/Dstl Joint Grant Scheme (UK) (ref RG/S56405/01) and by the BBSRC (UK) (ref S18903).

Authors' addresses: P. G. Lovell, C. A. Párraga, and T. Troscianko, Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK; C. Ripamonti and D.J. Tolhurst, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK.

is based on knowledge of how human psychophysical channels and single neurons in primary visual cortex (V1) respond to simple visual stimuli, such as sinusoidal gratings of different spatial frequency, orientation, and contrast [see Daly 1993; Doll et al. 1998; Lubin 1995; Rohaly et al. 1997; Watson 1987; Watson and Solomon 1997; Watson and Ahumada 2005]. These models recognize that a visual image is processed in parallel (at least in the early stages of visual cortex processing) by channels or neurons with different optimal spatial frequencies, but all with much the same bandwidth of about 1 octave [see Blakemore and Campbell 1969; DeValois et al. 1982; Movshon et al. 1978b; Tolhurst and Thompson 1981; Watson and Robson 1981]. Sometimes, the models are constructed as "pyramids" of increasing sampling density at increasing spatial frequencies.

We, too, have been developing a simple (low-level), physiologically plausible model of achromatic local contrast discrimination to predict human performance for discriminating between pairs of slightly different achromatic natural-scene-based images [Tadmor and Tolhurst 1994; Tolhurst and Tadmor 1997a, 1997b; Párraga and Tolhurst 2000; Párraga et al. 2005]. We compute the band-limited contrast [Peli 1990] at several spatial scales within images. The model then carries out a multiresolution analysis of the two pictures under comparison, detecting differences in local contrast in each spatial frequency "channel". This model (like others) examines several spatial scales in parallel, but unlike those cited above, did not include the orientation tuning so prevalent in visual cortex neurons. However, this type of model has been shown to be very effective in explaining the appearance of natural-scene stimuli under different viewing conditions [Peli 2001; Peli and Geri 2001].

Such models must be validated against real psychophysical experimental data to determine how well they explain human discrimination performance. Generally, such validation has been carried out against psychophysical experiments performed with sinusoidal gratings [e.g., Watson and Solomon 1997; Watson and Ahumada 2005], but there has also been some validation against natural-scene stimuli [Lubin 1995; Rohaly et al. 1997; Párraga et al. 2005]. We have described a variety of psychophysical experiments, measuring thresholds for discriminating small changes in naturalistic images that we control by morphing [Párraga et al. 2005]. We decided to use a morphing technique (as opposed, to a superimposition of two images to different degrees [e.g., Tolhurst and Tadmor 2000], because it produces a set of stimuli where each one of the component pictures is an image of a plausible object (with slightly different shape, color, and texture); each morphed image still shares the natural Fourier statistics of the original ones [see Párraga and Tolhurst 2000].

Here we extend our experiments and modeling to deal with colored images. We describe experiments in which human observers attempt to discriminate small changes in the shape, brightness, texture, and color of images of fruit. We compare the observers' measured thresholds with those predicted by our low-level model of visual cortex processing. We also examine observer ratings of suprathreshold image differences. We are particularly concerned with two issues:

- Often models are developed and tested on a relatively small set of related images. It is important to determine how generalized a model might be by testing it on a great variety of image pairs. Discrimination data (obtained using 2AFC procedures; see Section 3) are "costly" in that many trials are needed to obtain one discrimination threshold. Also, by definition, these techniques measure discrimination thresholds. However, it is important to test the model (a) when there are suprathreshold image differences and (b) for a large image set. The only procedure able to deal with this requirement is the magnitude estimation procedure developed by Stevens [1975] [see also, Lubin 1995]. We describe some experiments using a magnitude estimation technique on a variety of image pairs to show the potential of testing on a large data set.
- Each "channel" within in a multiresolution model, determined by spatial scale, orientation, and opponent channel, may contribute to a greater or lesser extent towards overall discrimination. How

are these many cues combined? Do the many cues contribute equally to discrimination and might there be complicated contingencies between them?

2. A DISCRIMINATION MODEL: BACKGROUND AND IMPLEMENTATION

Several visual discrimination models analyze pairs of images into several spatial scales and then compare the filtered images, spatial scale by spatial scale. The details of the models may differ and steps may be performed in different order. However, the models share very similar features. We will describe the implementation of our model, pointing out some differences from others, and we will show some of the neurophysiological (see review by Lennie and Movshon 2005) and psychophysical observations that lead to steps in such models.

2.1 Calculating Contrast in Several Spatial Frequency Bands

It is a basic tenet of visual psychophysics and neurophysiology that the visual system is sensitive to the *contrast* (*relative* differences in luminance) in a stimulus rather than to absolute differences in luminance or radiance, at least at the higher ranges of mean luminance [e.g. Enroth-Cugell and Robson 1966; Shapley and Enroth-Cugell 1984; Troy and Enroth-Cugell 1993].

Thus, the first stage in any model is to calculate the contrast at each point in an image; we calculate contrast at each of six spatial frequency scales an octave apart [Peli 1990; Tadmor and Tolhurst 1994; Párraga et al. 2005] and at four orientations 45° apart. Contrast at the point (x, y) and in the frequency band F and at orientation ϕ is:

$$C_F(x, y) = \frac{a_F(x, y)}{l_F(x, y)} \tag{1}$$

where $a_F(x, y)$ is a bandpass filtered version of the original image, obtained by convolving the image with a circularly symmetric filter with frequency response given by Eqn 2:

$$A_F(f) = \exp\left[-\frac{(f - F)^2}{2\sigma^2}\right] \tag{2}$$

Spatial-frequency channels are further subdivided into orientation channels by multiplying by a pie-slice filter (Eq. 3):

$$\Gamma(x, y) = \exp\left[-\frac{\phi}{(b/2)}^2\right] \tag{3}$$

where b is the bandwidth of each orientation channel (40°) and ϕ is the orientation (0, 45, 90, 135°). Pointer and Hess (1989, 1990) have demonstrated that contrast sensitivity varies as a function of orientation, among other variables. Currently the models contrast-sensitivity function (CSF) is not varied as a function of orientation. However, while the CSF is constant across orientation in the current implementation, the neural network stage allows for the possibility of reweighting different orientations.

The linear nature of this filtering is justified by the quasilinear summation behavior of simple cells in the visual cortex in response to sinusoidal gratings [Movshon et al. 1978a; Jones and Palmer 1987] and to natural scene stimuli [Smyth et al. 2003]. It is, however, true, that there are a number of nonlinear behaviors that a veridical model of visual cortex should capture [review by Carandini et al. 2005]. $l_F(x, y)$ is the result of convolving the original image with a circularly symmetric low pass operator with frequency response given by:

$$L_F(f) = \exp\left[-\frac{(f)^2}{2\sigma^2}\right] \tag{4}$$

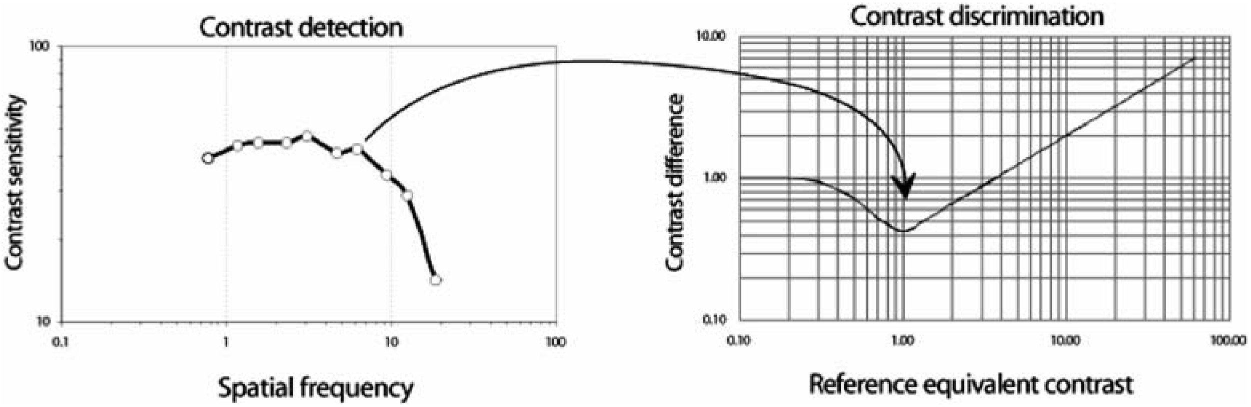


Fig. 1. (*Left*) Typical observer's CSF—measures of the sensitivity for detecting the contrast of gratings. The sensitivity at a given spatial frequency determines the location of the contrast discrimination "dipper" on the x and y axes. The template intercepts the y axis at a contrast difference of 1.0 and the dip is at its minimum for a reference contrast of 1.0. The calculated dipper for a particular frequency band is estimated by multiplying both x and y axis values by the observer's contrast threshold at that frequency.

f is spatial frequency and σ is the spread of the Gaussian frequency-response curves, and is chosen to be $0.3F$ so that the bandpass filters have a frequency bandwidth of about 1 octave and an orientation bandwidth of about 40°, to match estimates of psychophysical channel bandwidth and the most narrowly tuned visual cortex neurons [citations in Introduction]. In fact, the spatial frequency tuning of individual cortical neurons varies widely [Tolhurst and Thompson 1981] and there is a suggestion that higher spatial-frequency channels or neurons have narrower bandwidths [Blakemore and Campbell 1969; Tolhurst and Thompson 1981; DeValois et al. 1982; Baker et al. 1998]. Division of the bandpassed convolution by l_F (the local mean luminance) is a model of the fact that the visual system encodes contrast rather than luminance *per se* [Peli 1990]; the mean luminance is calculated over an area proportional to the period of F. In other modeling contexts [e.g., Field 1994; van Hateren and van der Schaaf 1998; Willmore and Tolhurst 2001] contrast encoding has been modeled by taking the logarithm of the pixel values in an image before applying linear filtering operations.

2.2 Comparing Contrast in Two Images

To model how the visual system discriminates two images, we calculate the $C_F(x, y)$ (Eq.1) for both images at all frequency scales, and then compare the contrasts in the two images, *point-by-point* within each frequency band. We calculate the absolute value of the difference in contrast between the two pictures under comparison at each location and in each frequency band:

$$\Delta C_F(x, y) = |C_{F,J}(x, y) - C_{F,O}(x, y)| \tag{5}$$

where j is the picture number of the test stimulus and $j = 0$ represents the reference picture. We then, must estimate how much each value of ΔC might contribute toward the visibility of the difference between the pictures. We hypothesize that visibility depends not just on ΔC, but that it follows Weber's Law: i.e., we evaluate each ΔC value against the familiar "dipper function" for contrast discrimination for sinusoidal gratings [Campbell and Kulikowski 1966; Nachmias and Sansbury 1974; Tolhurst and Barfield, 1978; Legge 1981; Legge and Foley 1980; Meese, 2004; Chirimuuta and Tolhurst, 2005]. Figure 1 shows such a "dipper function". Note that the linear, "Weber" part of the experimental dipper function for gratings has a slope of only 0.7 on log/log axes rather than unity [Legge 1981]. Each value of $\Delta C_F(x, y)$ is treated as if it is the contrast increment (ΔC) of a test sinusoidal grating of frequency F to be compared with a reference grating, whose contrast is the average of the paired contrast values in the two pictures at that location and frequency band.

$$\bar{C}_{F,J}(x, y) = 0.5|C_{F,J}(x, y) + C_{F,O}(x, y)| \tag{6}$$

We estimated the observer's contrast *discrimination* functions for achromatic gratings indirectly by adjusting the position on the x-axis (contrast reference) and y-axis (contrast difference) of a "dipper function" template for contrast discrimination according to the observer's contrast *detection* thresholds measured for a grating of the same spatial frequency [Párraga and Tolhurst 2000]. In fact, the dipper template used within the model has a different form from the experimental one; the model template is adjusted so that, on solution of the model for discriminating contrast gratings, the model's output will have the same form as the measured experimental data. In fact, the form of the dipper may depend upon stimulus configuration (Meese, 2004) and the present model should be regarded as a simplification. Previously, we determined the model dipper functions from each observer's CSFs separately. Any differences between observer's abilities to discriminate between pictures would hopefully be accounted for by differences in their CSFs [Párraga et al. 2005]. However, in this paper, we develop the model using a *single* averaged CSF, as if there is a single standard observer, since we are partially interested in the question of modeling whether a given image pair might be distinguishable by an average observer.

The strange shape of the psychophysical dipper function has been hypothesized to be because of to a sigmoidal transducer function, relating response magnitude to stimulus contrast [Legge and Foley 1980; but see discussions by Itti et al. 2000, Chirimuuta and Tolhurst 2005]; single neurons in V1 do have a sigmoidal response function (Tolhurst and Thompson 1981; Albrecht and Hamilton 1982). Other models [e.g. Lubin 1995; Watson and Solomon 1997] achieve the same estimate of the likely visibility of a small contrast difference without explicit comparison with a dipper template. Instead, the contrast at each point in each spatial-frequency band is scaled by the observer's threshold for that frequency. It is then transformed through an appropriate sigmoidal transfer function and, finally, the transformed contrasts in the two images are simply subtracted. This latter procedure has the advantage that it allows easy inclusion of *contrast normalization* [Heeger 1992; Foley 1994; Watson and Solomon 1997; Watson and Ahumada 2005], a feature that we have not yet incorporated into our modeling.

2.3 Pooling Discrimination Cues Across Location and Spatial Scale

A measure (V) of how different two pictures might be at a single location and in a single frequency band is given by how far the calculated ΔC is above or below the model's internal dipper template. There will be thousands of minute cues to discrimination, at the many locations, and in the several frequency and orientations bands and opponent channels. To assess the overall discriminability of the two images requires some algorithm for pooling these many cues. Thus, the second stage in the model is to pool the many cues (V) provided at different locations and different frequency and orientation bands to give an overall assessment of whether or not the two pictures differ sufficiently for discrimination to be made. Previously, we have used a weighted average of all the V cues, weighted across all locations and *all* frequency bands, so that there is a single metric for a given pair of pictures rather than one measure per frequency band. We use a Minkowski sum (see Eq.7) with power of four [Rohaly et al. 1997]. The power of 4 derives from an empirical description of the amount of probability summation seen in grating *detection* experiments and relates to the steepness of the psychometric function [Quick 1974; Robson and Graham 1981]. We hypothesize that the same nonlinear weighting would apply to *discrimination* experiments for complex natural scenes. Thus, an overall cue V_4 is given by:

$$V_4 = \sqrt[4]{\sum_F \sum_\pi \sum_y (V_F(x, y))^4} \tag{7}$$

Later in this paper, we will consider whether the cues from the six spatial-frequency scales, from the four orientation bands and from the three luminance/chromatic planes (see below) *should* be so simply summed, or whether they should be summed unequally or combined in more complex ways.

2.4 A Color Version of the Discrimination Model

The above refers essentially to the model we have been developing to try to model human observers' ability to discriminate between *monochrome* images. We now extend this to evaluation of colored images. Since human vision is trichromatic and colored images are represented as three planes (RGB), it seems inevitable that a color model would perform three parallel operations. The Sarnoff model [after Lubin 1995] transforms the RGB images into the CIE L*u*v* space, before performing parallel comparisons on the two images' L planes, on the two u* planes, and on the two v* planes. Jin et al. [1998], enhancing the Daly [1993] model, suggest using the CIELAB transform, which first transforms the RGB planes into LMS planes—representing the calculated responses of human L cones ("red" cones), M cones ("green" cones), and S cones ("blue" cones). The LMS planes are then transformed to one luminance plane and red–green and blue–yellow opponent planes, since it is believed that human vision processes luminance and color information separately and in parallel [Mullen and Losada 1994] and that the color information is processed in red–green and blue–yellow *opponent channels* [Hurvich and Jameson 1957; DeValois 1965; Wiesel and Hubel 1966]. Here we investigate a model explicitly based upon a stylized split of human vision into independent luminance processing and opponent chromatic planes: a luminance plane and red–green and yellow-blue color opponent planes.

The colored images (in a conventional RGB format) are first transformed in order to calculate how the three cone types of human vision (L, M, and S) would respond to the images. This calculation required that we did a spectroradiometric analysis of the wavelength emission of the three (RGB) phosphors of our CRT display used in the experiments and knowledge of the spectral activations of the three cone types [Smith and Pokorny 1975]. The opponent channel values are computed using Macleod and Boynton's [1979] color space, where the channel definitions are as follows: luminance is L + M, red–green is L/Luminance, and blue–yellow is S/Luminance.

Our proposed three channels reflect the psychophysical observations that luminance information in simple grating stimuli seems to be processed independently of isoluminant red–green or blue–yellow information [Mullen and Losada 1994; Mullen and Sankeralli 1999]. The early stages of the monkey's visual system do *not* show such a neat split into three separate channels; individual neurons respond to both luminance and to isoluminant chromatic stimuli to varying degrees [DeMonasterio et al. 1975; Derrington et al. 1984; Lennie et al. 1990; Conway 2001; Johnson et al. 2001].

Thus, we run the image discrimination model *three times* on each pair of colored images. First, we obtain an estimate of the overall discrimination variable V_4 for the luminance plane of the images. We then obtain estimates of V_4 for the red–green and yellow–blue planes. Note that the CSF's for the color-opponent planes are of a different form from that for the luminance plane [Mullen 1985; Mullen and Kingdom 2002]. Measurement of isoluminant thresholds poses many technical challenges including compensation for the effects of chromatic aberrations (Mullen, 1985); however, good estimates of the relative sensitivities of the red–green, blue–yellow and luminance systems can be obtained with stimuli presented on regular CRT monitors (Mullen and Kingdom 2002). As our stimuli were presented on CRT monitors, our model employs the latter estimates of chromatic contrast sensitivity. However, the color

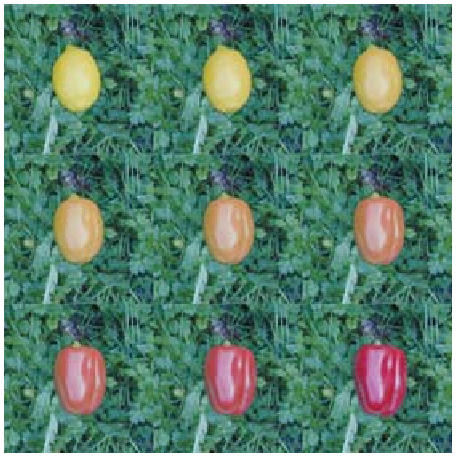


Fig. 2. Examples of morphed images. A lemon (top left) is gradually morphed into a pepper (bottom right).

opponent psychophysical system contains similar 1 octave-wide spatial-frequency channels [Mullen and Losada 1999]. The way in which we should combine the cues from the luminance and chromatic channels is one topic of this paper.

3. DETERMINING OBSERVER SENSITIVITIES TO NEAR-THRESHOLD AND SUPRATHRESHOLD CHANGES IN NATURAL IMAGES

In order to examine how well the current model predicts performance, we conducted two independent studies that examined how well observers could discriminate differences between pairs of natural images. One experiment concentrated upon establishing discrimination thresholds, where differences between images were relatively subtle. Another experiment examined observer ratings of differences, where the changes between image pairs could be relatively large. The acquisition of two disparate datasets allows us to examine the performance of the model under near-threshold and suprathreshold conditions.

3.1 Examining Thresholds for Image Differences in a Morph Sequence

The purpose of this experiment was to obtain a large set of image-discrimination data on which the model could be optimized. In order to achieve this, two sets of images were produced. The first set was of a red pepper morphing gradually into a yellow lemon, all on the same background of leaves with dappled illumination. The morph from one fruit to the other was conducted in 40 steps, so that there were 41 images in a sequence. Figure 2 shows typical basic stimuli (only 9 of the 40 steps are shown). In an experiment, a computer-controlled procedure would determine how much morphing (in %) was needed for an observer to discriminate the initial pepper image from a morphed image.

In fact, the morphed image set was subjected to various filtering operations so that, in all, we obtained 49 different stimulus sequences. The 41 images in the sequence were split into their L, M, and S representations (see above), and then into the three planes of luminance, red–green, and blue–yellow opponent. These three transformed images were Fourier transformed and their amplitude spectra were filtered to either blur or sharpen (edge-enhance or whiten) them. The Fourier spectra were multiplied by a filter of the form:

$$weight(f) = f^{-\alpha} \tag{8}$$

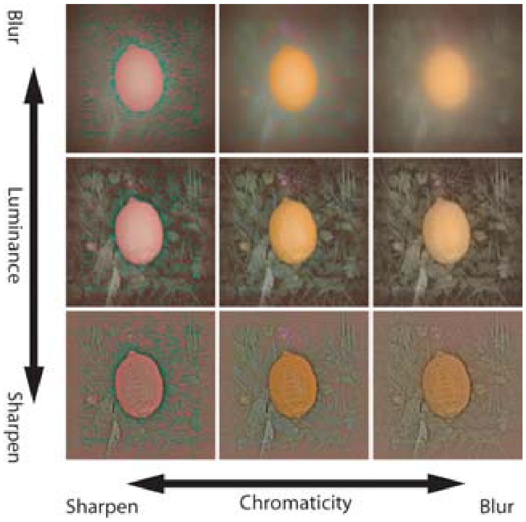


Fig. 3. Examples of a manipulated image. Here a lemon (central image) is either blurred or sharpened separately in the chromatic or luminance channel. The top-right image has been blurred in the luminance and both chromatic planes; the lower-left image has been sharpened in all three planes.

where f is spatial frequency and α is a slope parameter. Positive values of α (+ 0.4, + 0.8, and + 1.2) give different degrees of blurring and a reduction in high spatial frequencies; negative values (−0.4, −0.8, and −1.2) give different degrees of sharpening and a relative increase in the amount of high spatial frequencies; a zero value leaves the images in their original unfiltered forms. The filtered spectra were inverse-transformed back to give modified luminance and modified color-opponent planes. The luminance plane was filtered in seven different ways—from extreme sharpening to extreme blurring. Similarly, the two color-opponent planes were filtered in seven different ways (in fact, we always performed the same operation on the red–green and blue–yellow planes, so that they can be considered as a single "color" plane). These filtering operations were performed in all combinations to give 49 different sets (including one set, which actually had been unfiltered). The modified planes were reverse transformed to calculate the L, M, and S cone values implied; these values were then reverse-transformed to give the RGB values needed to display the desired images on a CRT. Figure 3 shows 9 exemplary images out of our set of 49.

Since these images were derived from pictures of a red pepper and a yellow lemon, the morph sequences resulted in images that changed primarily in luminance or in the red–green opponent plane. To make a sequence with color variations predominantly along the other color axis (blue–yellow opponent channel), we produced a second set of experimental stimuli by exchanging the "R" and "B" planes in the parent image of the red pepper, and morphing this with the uncorrupted yellow lemon, while keeping the rest of the parameters the same (the background was the same as before). Furthermore, the order of the morph sequence was reversed, so observers judged differences from the pepper, rather than from the lemon. The resulting morph sequence shows a "bluish pepper" transforming into a yellow lemon on a "normal" leafy background. Based on these images, we made a second series of image sets (49 combinations of luminance and color filtering, as described before).

3.2 Experimental Methods

Thresholds were measured for several observers for the 49 different filtering combinations of the original morph sequence and for the 49 different combinations of the "bluish pepper" morph sequence. The images were presented on a CRT with overall size 36.5 × 27.4 cm viewed from 200 cm. The individual images were presented one at a time in the centre of the screen; the images measured 11.2 cm square (i.e., 3.2° of visual angle square) and the remaining parts of the screen were held a uniform gray (5.64 cd.m^2). The stimuli were presented with a VSG 2/5 graphics card with pseudo 15-bit output; this allowed compensation for the nonlinear "gamma" of the CRT display. Two alternative forced-choice techniques determined, for each of the 49 conditions, how much a filtered stimulus needed to be morphed in order for reliable discrimination (75% correct) from the parent pepper image [Párraga et al. 2005]. In a single trial, there were three time intervals of 0.5 s each. The observers were free to look at whichever part of the image they wished and were free to make eye movements within the 0.5-s image presentations. The middle interval was always known by the observer to contain a parent image. The first or third interval (chosen randomly by computer) would also contain that same image, but the third or first interval (respectively) would contain the morphed image. The observer's task was to inform the computer whether the first or third interval contained the different image. If the observer chose the wrong interval too frequently, the task was made easier by choosing a morphed image more different than the parent; if the observer chose the correct interval too frequently, the task was made harder. Thus, during an experiment, the "staircase" converged on that percentage morph that the observer could correctly identify approximately 75% of the time. The red–green morph sequences were from lemon to pepper, but the blue–yellow sequences were from bluish pepper to lemon.

After 100 to 200 such trials for each of the 49 conditions of each of the two morph sequences, it was possible to construct psychometric functions for each condition and for each observer separately. A sigmoidal psychophysical function was fitted to the experimental data using *psignifit* [Wichmann and Hill 2001] and threshold or just-noticeable difference (JND) was (arbitrarily) taken by interpolation as the magnitude of morph step that would lead to the observer correctly choosing on 75% of trials. The proportion correct value predicted by the fitted psychometric functions, for each step in the morph sequence were retained; these values were subsequently used in the training or tuning of models discussed in Sections 4.1 and 4.2.

3.3 Discrimination Threshold Results

Figure 4 summarizes the discrimination thresholds for the two morph series. Thresholds for observers KB and CAP were highly correlated. For red–green morphs, the correlation coefficient was 0.74, while for blue–yellow thresholds the correlation was 0.66. Figure 4 left) shows the averaged thresholds for discriminating the various forms of the red pepper to yellow lemon sequence. The thresholds for the 49 conditions are similar, mostly around 11% morph with standard error of about 0.2% morph. Interestingly, the poorest observer performance was for highly sharpened chrominance ($\alpha = -1.2$) and highly blurred luminance pictures ($\alpha = +1.2$) (top-left corner of the plot). Observers were relatively poor at discriminating changes in the image when the color information had been sharpened or edge-enhanced, while the luminance information was blurred. This seems consistent with the finding [Mullen 1985] that the human visual system favors low spatial-frequency color information (i.e., not sharpened) and high spatial-frequency luminance information (i.e., not blurred).

Figure 4 (right) shows a similar plot of the averaged thresholds for discriminating changes in the bluish pepper to yellow-lemon sequences. The threshold surface has a different form, with observer performance worsening for images with their chromatic information blurred (right side of plot). Figure 4 (right) shows a similar plot of the averaged thresholds for discriminating changes in the bluish pepper to yellow lemon sequences. The threshold surface has a different form, with observer performance worsening for images with their chromatic information blurred (right side of plot). It is, at first, surprising that the forms of the R-G and B-Y morph surfaces are different, since the CSFs reported by Mullen (1985) for isoluminant R-G and B-Y gratings are rather similar. Our results indicate

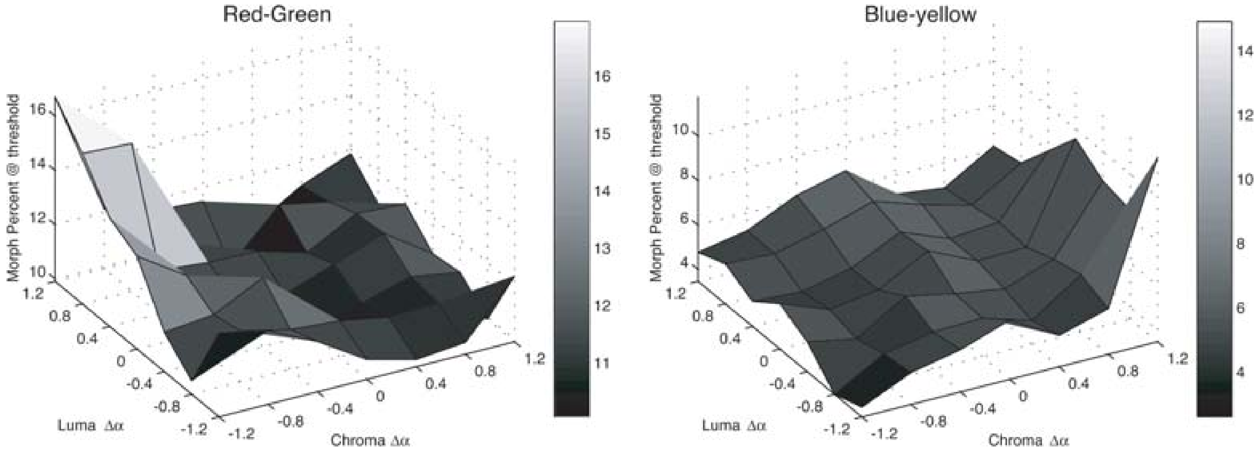


Fig. 4. Surface plots to show the 49 averaged thresholds or JNDs for the variously blurred or sharpened versions of the morph sequences. (Left) Red–green morph series; (right) blue–yellow morph series. Note the difference in the y-axis scales: left is from 10 to 17; right is from 2 to 14.

that perhaps the effective CSFs are not identical. We have not accounted for chromatic aberration in the human eyeball, which is likely to be greater for the BY then the RG images because of the greater wavelength difference in the former compared to the latter. Such imperfections in stimulus display (coupled with the fact that we have only eight-bit resolution on our graphics system) would be likely to generate high spatial-frequency luminance artifacts when isoluminant images are presented (Forte et al. 2006). Therefore, the effective CSFs of the isoluminant channels may not have the assumed shape. Interestingly, the thresholds are generally much lower (approximately 5% morph, standard error equal to 0.2% morph) than for the red pepper sequences.

4. OBTAINING RATINGS OF SUPRATHRESHOLD IMAGE DIFFERENCES

In common with many other studies, we have previously developed and tested our modeling on a relatively small set of psychophysical stimuli. Even seemingly small change in stimulus presentation conditions may require the parameters of such a model to be changed [e.g., Párraga et al. 2005]. It is important, therefore, to validate a model on as wide a variety of different tasks and stimuli as possible [see e.g., Lubin 1995]. The threshold discrimination task that we described above (Section 3) does not lend itself to generating the required large amount of psychophysical data:

1. The 2-AFC staircase procedure is time consuming. We can collect only four to five threshold values in a 30-min session.
2. The nature of the staircase requires that we have to use stimuli which can be graded stepwise between two end points, thus limiting the choice of avoidable natural-scene-based stimuli. The synthesis of such image sequences is time consuming and limited to image pairs in which corresponding points can be identified.
3. The procedure described above deals only with “threshold” perceptual differences between images. We are also interested in how well a model fares in predicting how great a perceptual difference is evoked by an image pair, where the differences are clearly visible and above threshold.

For these reasons, we have been developing a rating procedure [Stevens 1975; Lubin 1995] to obtain measures of the perceptual differences between large numbers of varied image pairs. The observer is presented with a pair of similar natural scene-based images and is asked to provide a numerical rating as to how different the images appear to be. We describe some preliminary observations to show the potential of this technique.

4.1 Rating Scale Methodology

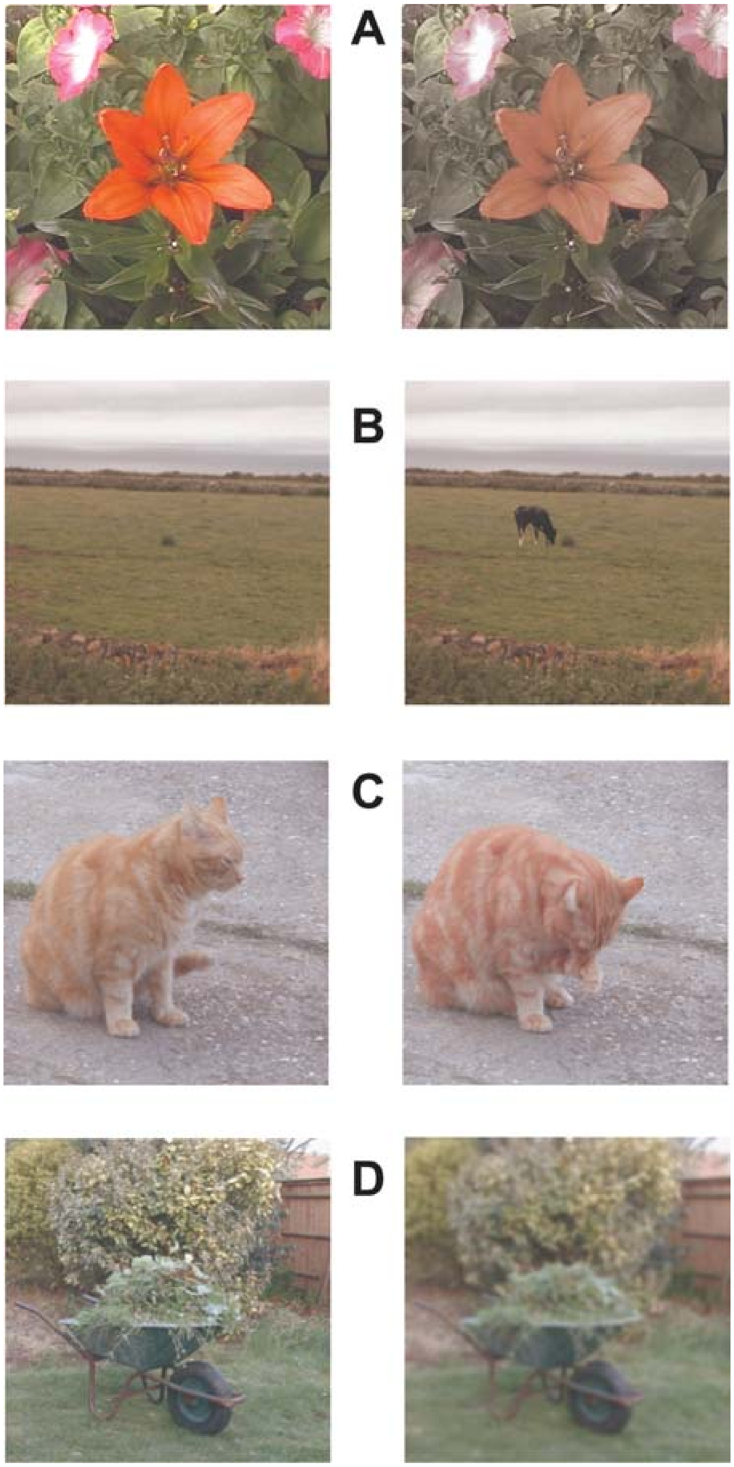
Two observers (naïve to the purpose of the experiments) were presented with 102 pairs of images that were derived from a great variety of digitized photographs of natural scenes—taken with calibrated [Lovell et al. 2005] Nikon 950 or Nikon 5700 still cameras or a JVC digital camcorder. The photographs were loosely grouped into six categories (animals, landscapes, object or "still lifes", people, plants). Within each category, we chose 17 "parent images." For each parent, we chose or synthesized a related image. This might be a second photograph of the scene or a later frame in the JVC video stream after some natural change in the scene (about 1/3 of the image pairs). We could also change the saturation or hue of all or parts of an image or we could blur or brighten all or part of an image. We could also cut and paste features or objects between images. Figures 5A and B show two examples where we made one such change between the images. Figures 5C and D show examples where we made two or more changes. In fact, in 20 of the pairs, the images were identical; these pairs enable us to estimate bias in observer ratings.

In an experimental trial, the observer first viewed the CRT which was held at a uniform mid gray; in the center of the display was a small dark spot and the observer was asked to fixate upon that spot and not to move their gaze during the image presentations (unlike in our earlier experiments). Thus they focused their attention on the center of each image, which measured 3.2° square. One image of a pair was presented for 1 s. The screen then went mid-gray, with the dark spot for 0.25 s. The other image of the pair was presented for 1 s, which was again followed by a 0.25-s interval when the dark spot was presented. Last, the first image was presented again for 1 s. The observer was then asked to make a numerical rating and indicate that number to the control computer.

Before the experiment proper, the observer became acquainted with the technique and learned to make consistent rating judgments using presentations of a different set of images. Frequently, within the training set and within the experiment proper, the observer was shown the image pair of Figure 5A (this was called the "standard") and the observer was instructed to rate all other image pairs with reference to that image. The standard or any other image pair that seemed equally distinct should be given a rating of "20". Any other image pair should be given a higher or lower rating, whose magnitude depended upon how much more or less different the stimuli seemed than the standard. Each observer ran three sessions, so that the rating values we present are the means of three ratings. In the three sessions, the same 102 image pairs were presented, but in a different random order.

4.2 Results

Figure 6 plots the rating results of the two observers against each other. Each point represents the average of the three ratings given by one observer to a given image pair against the average of the three ratings given by the other observer to the same image pair. There is a good overall correlation between the two observers' ratings ($r = 0.864$; $n = 102$). However, it is clear that the two observers have adopted different strategies for rating images, especially for those image pairs where there was, in fact, no difference. Observer 2 has given very low scores to these (mostly zero, rarely above 2) while observer 1 has given a range of values up to about 8. Nevertheless, the good overall correlation shows that the two observers were generally rating in the same way, so that these data can be used to examine how the V1 model copes with suprathreshold data. For the analyses (below) we average together the two observers' ratings.


A
B
C
D

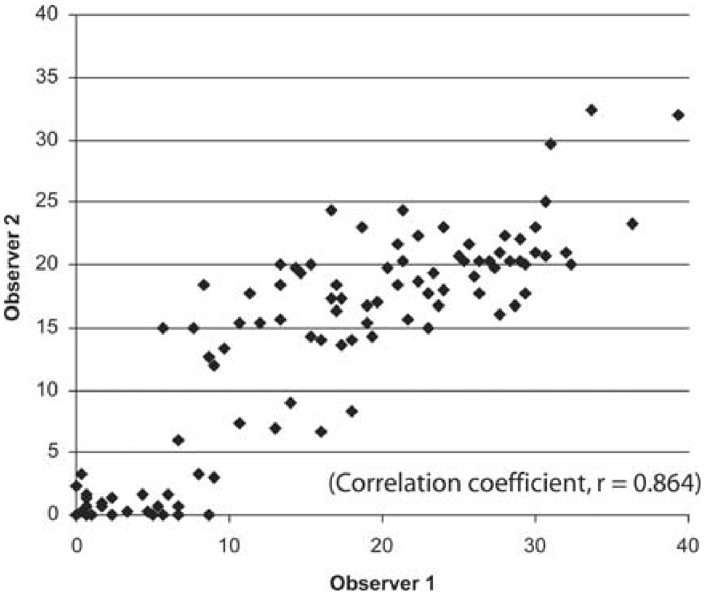


Fig. 6. Correlations between observer 1 and observer 2 for image difference ratings.

5. EXAMINING MODEL PERFORMANCE

Previous implementations of our color model (Tolhurst et al. 2005) combined the weighted fourthe power outputs of the opponent channels. In the current implementation, we investigate whether additional weightings of spatial-frequency and orientation channels might improve the models performance as a predictor of human ratings and thresholds.

In order to test whether there might be a more powerful method for combining the fourth power (V_4) outputs of the model, we examined whether a neural-network stage, that was trained to predict the thresholds and ratings of observers, would give better performance than the current opponent channel weighted output stage. A model-selection procedure was adopted, where all potential weightings of opponent channel, orientation, and spatial frequency where tested. The most complex model may apply a different weight to each combination of opponent channel (3), orientation (4), and spatial frequency (6) differently, resulting in 73 weights (3*4*6 + bias). The simplest model would weight all model outputs equally, i.e., it would have only a single weight (+bias). An *output* of the existing model forms the *input* to the neural-network stage. Each model output was based upon the Minkowski sum (V_4) of all differences in the particular opponent channel, spatial frequency, and orientation (depending upon the model criteria under examination).

We implemented a variety of neural networks, each trained to associate the output of the model with the responses of our observers. Where an individual neural network results in more reliable predictions of observer performance, then this might indicate areas where the basic V1 model may be improved.

←

Fig. 5. Four examples of image pairs used in the rating scale experiment. (A) A pair of images in which the color saturation of the entire image has been changed. This pair was the “reference standard” in the experiment; observers had to learn that this amount of perceptual difference merited a rating of “20.” (B) The image of a cow from one photograph has been removed from another similar image. (C) Two changes have been made. The color of part of the image has been changed and the spatial organization of that object has changed. (D) There two changes were made there are two slightly different images of the same scene, one image is blurred.

The model-selection technique has been used successfully in other areas of vision research, for example, motion perception [Baddeley and Tripathy 1998].

However, it is a relatively straightforward, but fruitless, pursuit to attempt to optimize the performance of the model by adding additional free parameters without ensuring that the improvements achieved generalize to predictions of novel image difference ratings and thresholds. It is easy to have a network over-learn, so that it is excellent at discriminating the training set, but poor at dealing with a novel data set. Thus, two forms of cross-validation are undertaken here; first, model weights are established by training on the morph sequence results and cross-validating on rating stimuli (and vice versa). Furthermore, rating stimuli image-pairs are split into six subgroups (animals, landscapes, garden scenes, object or "still lifes," people and plants), enabling a six-fold cross-validation where training is undertaken with the image-pairs from five groups and generalization is measured with the remaining group. Improved performance on the training sets with poor cross-validation would suggest that the neural network had merely overfitted the characteristics of the training data.

The number of hidden units within the neural network was also varied; networks either featured no hidden layer, or featured one to three hidden units. All networks were fully-connected, i.e., every unit was connected by a weighted value to all other units in the subsequent layer. Bias units were also present and these where connected to both hidden layers (where present) and to the single output unit; Figure 7 illustrates the architecture of network implemented.

Where neural networks were trained to predict morph-sequence performance, training data were the predicted psychophysical function values, but only where the predicted values fell below the asymptote ($<0.95\%$). For rating-trained networks, the average rating value for the two observers were used as training data. Hidden units, where present, utilized the *Matlab* tansig function and output units were linear.

In order to maximize the likelihood that the global minimum was discovered for each network, a two-stage training procedure was followed. For each type of network architecture, 200 networks were created each with randomized initial weights. Initial training using the scaled conjugate gradient back-propagation learning algorithm, limited the number of training epochs to the total number of weights present within a particular network. In a second stage of training, the 25 networks with the smallest mean squared error were selected and trained for a further 200 epochs with the BFGS quasi-Newton back-propagation algorithm. Finally, for each network architecture, the network with the lowest mean squared error was retained.

Where models were trained with rating data, a six-fold cross-validation procedure was followed (as explained above). Consequently, six models were created for each architecture. Figure 8 shows the internal rating cross-validation results. These values represent the mean correlation between the model and observer for each of the six cross-validation groups. Following the six-fold cross-validation process, an average model was generated by calculating the mean of the corresponding weights in the six networks. The results presented in Figure 8 are based upon the performance of the average model, except for the internal cross-validation values.

For human observers, the JNDs for each morph sequence were established using a psychophysical staircase procedure. A single threshold value was defined in order to establish equivalent JNDs for the neural-network outputs. Model outputs for each morph sequence were interpolated using a simple spline (*Matlab*). The point where the interpolated output exceeded the threshold value was taken as the model's JND. The threshold was determined using a least-squares fit for model and averaged observer JNDs. Separate thresholds were established for each neural-network model. Where model outputs are highly nonmonotonic, thresholds could be established incorrectly, because of there being more than one zero-crossing. Consequently, two thresholds were established by locating the first and last zero-crossing in each fitted spline. Where correlations between observer and model performance differed, depending

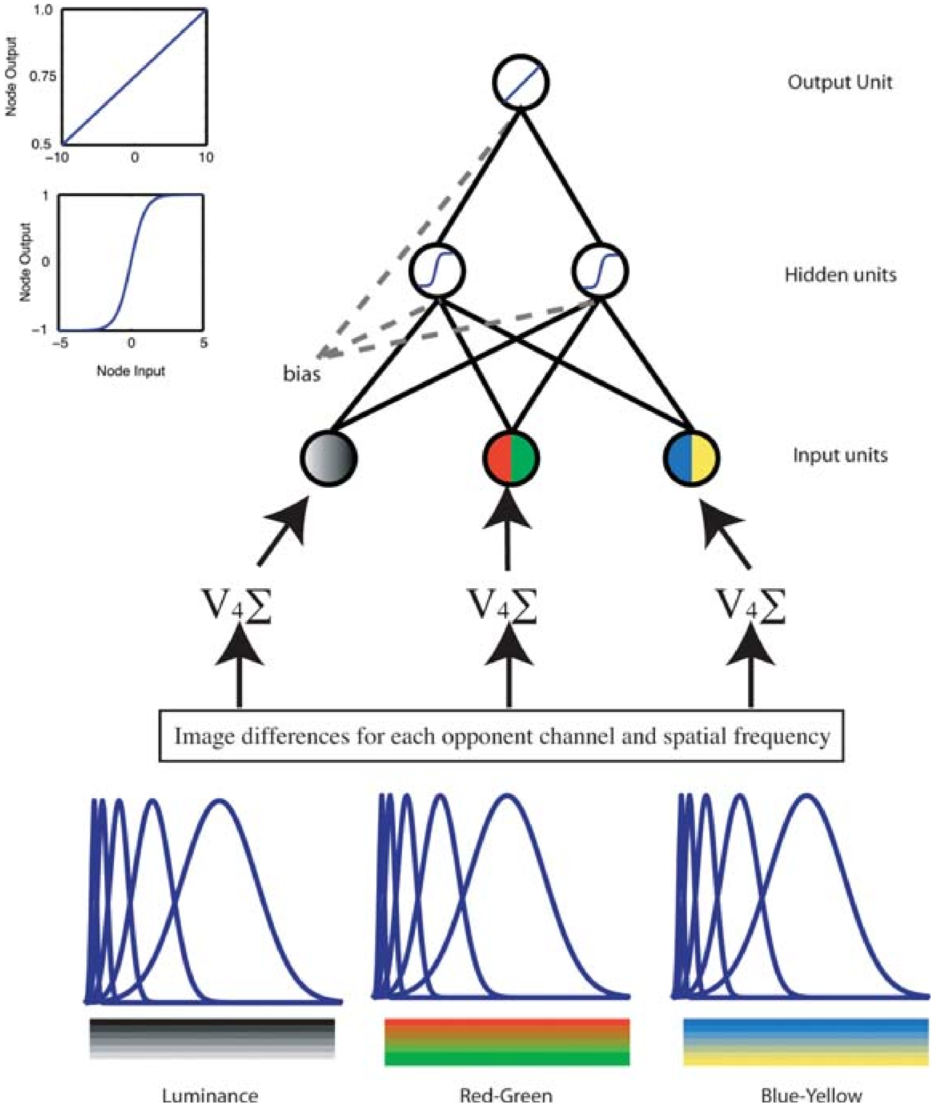


Fig. 7. Schematic diagram of neural network architecture. The diagram illustrates a neural network with three inputs and two hidden units. Where three inputs were present, each input node received the overall activation (Eq. 6) across all spatial frequencies for each opponent channel. The inset plots (top left) depict the transfer functions used within the hidden-layer and output-layer nodes. The bottom row of plots illustrate the spatial-frequency channels used, the peaks of which were at 24, 12, 6, 3, 1.5, and 0.75 cycles per degree.

upon the choice of direction through morph sequence, the lower correlation was reported. This only occurs in the most complex morph-trained models.

5.1 Results

Correlations between the observer responses and the model outputs rose as a function of the number of hidden units present within the neural network. However, cross-validation performance was poor

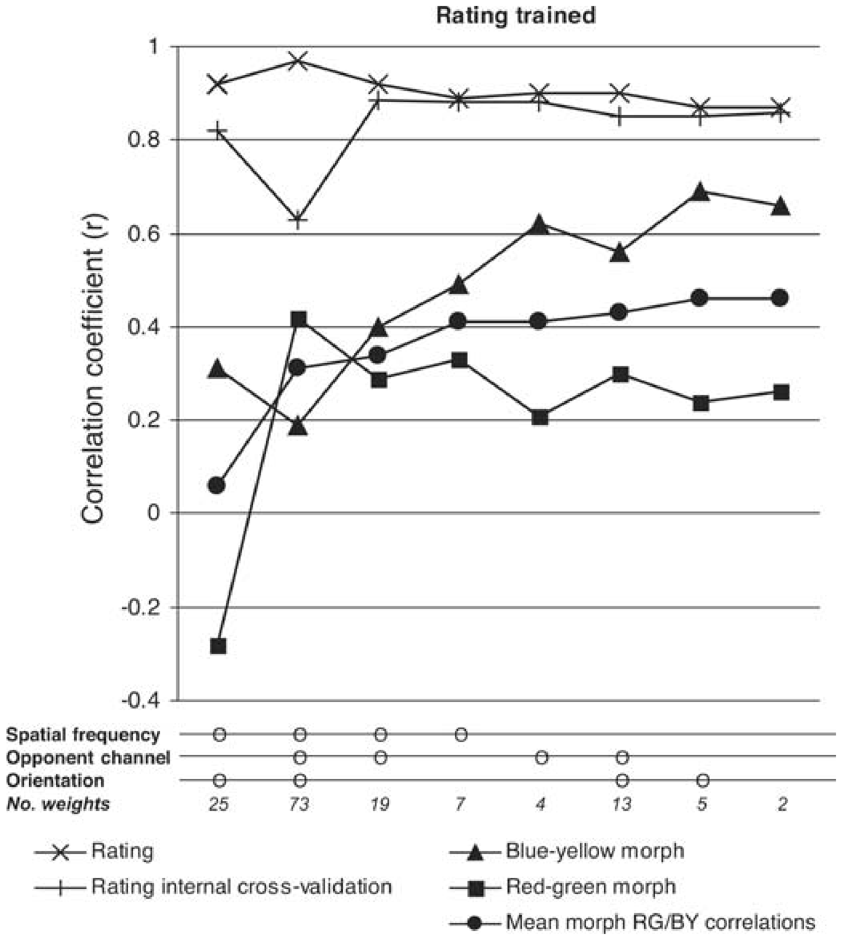


Fig. 8. Correlations between the model predictions and human observer results obtained from morph-thresholds and rating data. Models were ranked in order of cross-validation performance (mean morph RG/BY correlations) and are presented in this order. Cross-validation performance improves as a function of model simplicity, those models featuring the fewest free-weights give rise to the best cross-validation performance. The lower panel of the figure shows how the output of the V1 model was divided into separate neural-network inputs.

in all networks featuring a hidden-layer. As a consequence, only models without a hidden layer are considered below. Figures 8 and 9 show the training and cross-validation performance for models trained with ratings (Figure 8) and with morph thresholds (Figure 9). In both cases, performance for trained datasets is better in the more complex models; conversely, cross-validation performance improves as models become less complex and, thereby, feature fewer weights. The least complex model in both figures is represented as the rightmost set of symbols; the most complex models are represented toward the left of the plots.

For the rating-trained models the cross-validation results for rating images (Rating internal cross-validation) are informative. Correlations for cross-validation image-pair ratings are almost as high as they are for the trained image pairs, apart from the two most complex models; clearly these complex models have overfitted the training data at the cost of performance on the cross-validation data. Cross-validation on morph thresholds improves as models are simplified, i.e., they have fewer free weights.

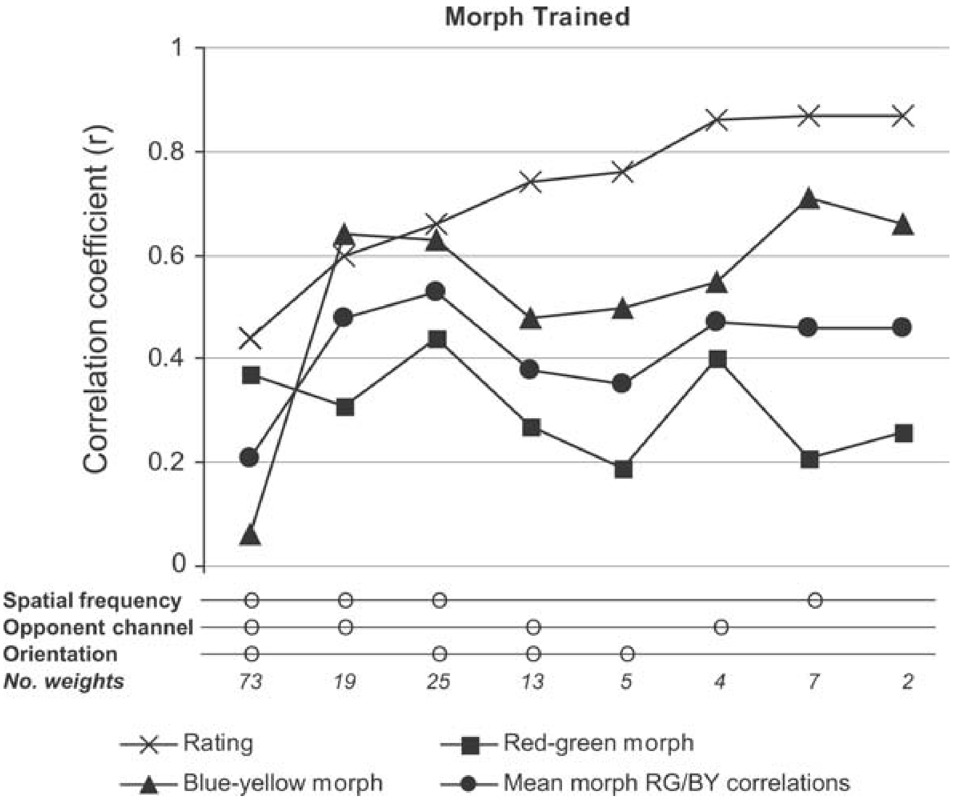


Fig. 9. Correlations of thresholds and image difference ratings for models and human observers. Models were ranked in order of cross-validation performance (image difference ratings) and are presented in this order. As with the rating-trained models, cross-validation performance improves as a function of model simplicity. The lower panel of the figure shows how the output of the V1 model was divided into separate neural-network inputs.

For the rating-trained models, there was a consistent inverse relationship between performance on image difference ratings (training set) and performance predicting morph thresholds. In the case of the morph-trained models, this relationship is less consistent. The more sophisticated models (the three leftmost plots in Figure 9) also show relatively impaired performance in their predictions of morph thresholds and ratings. This is because of a nonlinear relationship between the output of the model and the distance along the morph sequence. Models were only trained with morph imagepairs that were near thresholds. Consequently, the more complex models overfit these images and do not generalize to image pairs further along in the morph sequence. Thresholds identified in these models are unreliable, leading to the lower correlations between observer and model thresholds.

Figure 10 shows observer discrimination thresholds for the morphed fruit stimuli and selected model predictions of these thresholds. Models were selected according to their overall performance on the training and cross-validation sets. For rating-trained models, the most simple model of all provided the best overall performance levels (Figures 10C and F), in this model all channels where equally weighted. For the morph-trained models, some improvement was demonstrated, i.e., some training improvement without significant cross-validation losses when opponent channels where differentially weighted (Figures 10B and E).

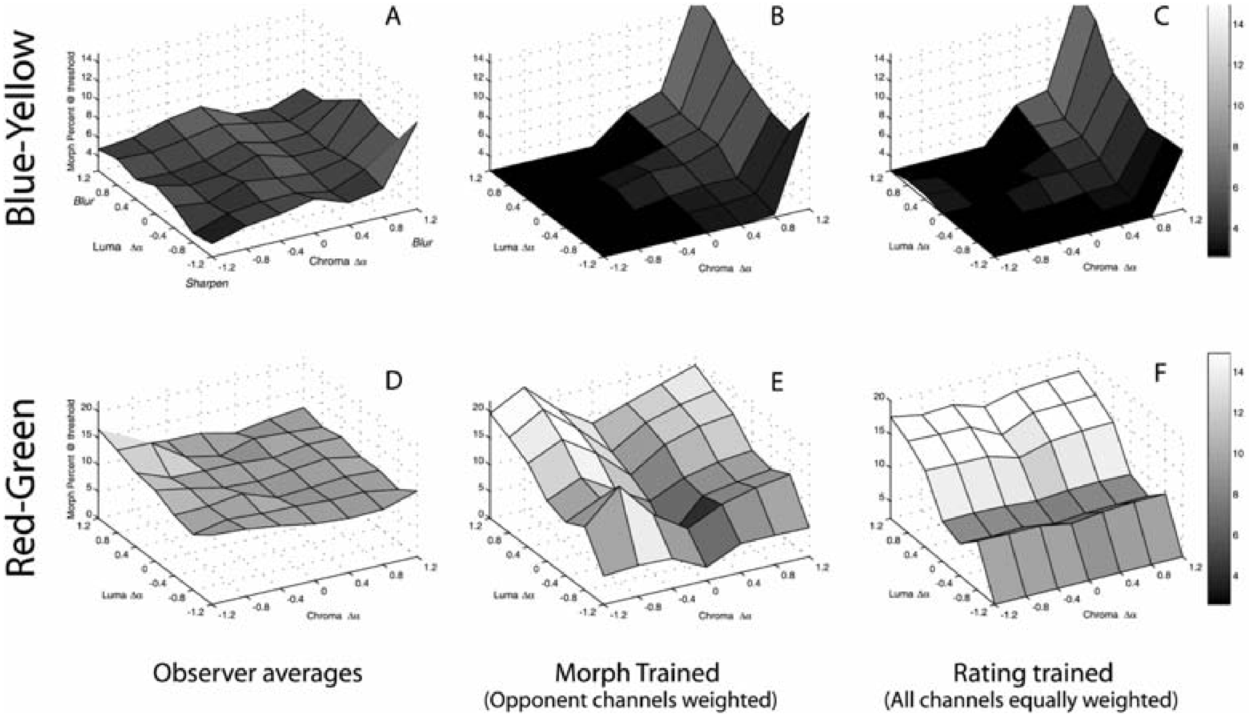


Fig. 10. (A,D) Averaged observer thresholds for morph-sequences with varying blur and sharpening in the luminance and chromatic channels. (B,E) Model predictions of morph thresholds, where the model was allowed to weight the three opponent channels separately. (C,F) Model predictions of morph thresholds where all model outputs are weighted equally.

The model predictions capture something of the gross form of these surfaces; the average correlation is greater than 0.4. However, the forms are somewhat exaggerated. For blue–yellow morphs, the increased threshold, where chromatic image content is blurred is present, but the predicted thresholds are too high. While the increased threshold for red–green morphs that have been sharpened in the chromatic channels and blurred in the luminance channel are present, but, once again, these are exaggerated. We have already speculated about the possible sources of discrepancies between ideal CSF chromatic channels and real data and about the possible differences between RG and YB chromatic planes. However, it should be noted that the models have correctly predicted that the observers' threshold for the blue–yellow stimuli are substantially lower than those for the red–green stimuli.

Figure 11 shows a scatter plot of the observer ratings of image pairs against the least sophisticated model's prediction of these ratings, this model's performance is shown in the right-most data points of Figure 8 (this model was rating-trained, although this factor would have no influence on the rating correlation as all channels within the V1 stage were equally weighted). The overall correlation is good ($r = 0.87$; $n = 102$), although there seem to be some outliers. These seem to fall into a number of categories. Where the model is predicting a higher rating than the one actually measured, the change in the image is often diffuse (circular symbols), e.g., a chromatic or textural change across the whole image. Other changes to which the model is oversensitive include those where an object has moved slightly within a scene (diamond). There are fewer cases where the model predicts a rating that is far lower than those measured in observers. An example is object appearances (triangle "C"). Figure 12 shows examples of these outliers.

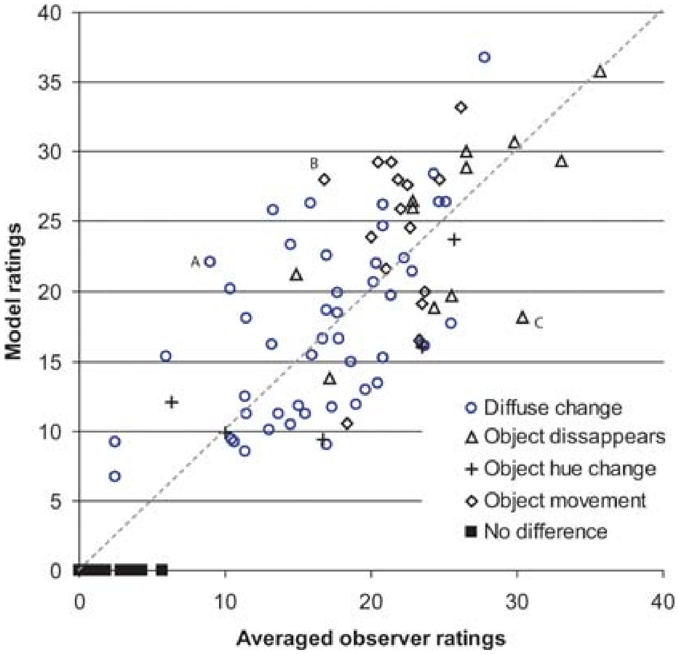


Fig. 11. A plot of the output of the best cross-validated neural-network model against the subjective magnitude ratings of two observers (different symbols) for 82 very different image pairs. “Diffuse change” includes variations in illumination through time-lapse, blurring, hue changes to the whole image, and texture changes through wave movement, etc.

6. CONCLUSIONS

We have examined a V1-inspired model of visual discriminations for colored natural scene stimuli. Figures 8, 9, and 10 summarize this part of the modeling. The left-most column of Figure 10 shows surface plots of the 49 averaged JND values measured in the experiments on the red pepper to yellow lemon morph sequences for two observers. The middle and right-hand plots shows the thresholds predicted by the two of the more successful models.

Figure 10A shows the two observers’ threshold surface plots for the bluish pepper to yellow lemon morphs; Figures 10C and F show the surfaces predicted by the V1 model, with all channels weighted equally. The two model surfaces show many of the features of the actual experimental results: the gross difference in threshold between the red pepper (above) and blue pepper (below) and some hint of threshold elevations in the corners or edges of the plots.

Patterns of cross-validation performance reveal that training on thresholds for a relatively few image-pairs (morphs) may lead to overspecialization in the model, which precludes successful generalization to novel image-pairs. Note that cross-validation to image ratings from morphs is relatively poor where training is undertaken with image-morph thresholds and complex models (Figure 9). It is interesting that the neural-network modeling produced little improvement in this final cross-validation, implying that the rules for combining discrimination cues in such models might be thankfully simple. However, although our version of the V1 family of models explains some of the variance in the rating scale data, it is still certainly true that there is much variance that remains unexplained. Obviously, some variance will be caused by the observers giving slightly inconsistent ratings when they are presented with one image pair more than once, but we should consider other more important reasons why our present model is only partially successful.

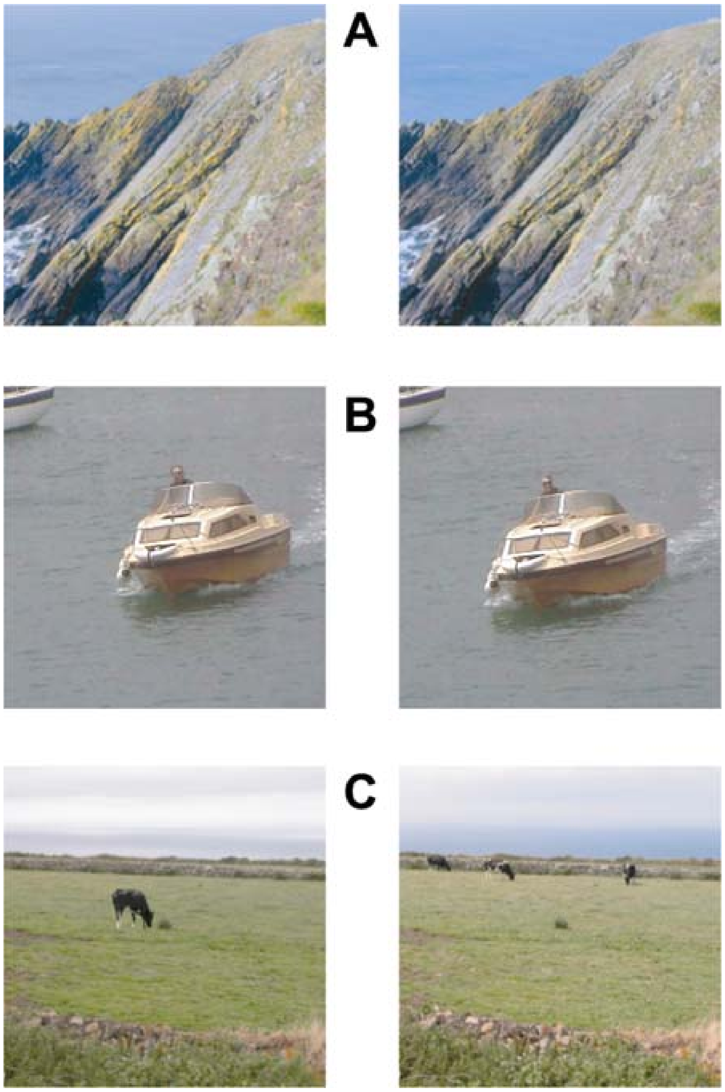


Fig. 12. Examples of outliers in the scatter-plot presented in Figure 11. (A) diffuse change (B) object movement (C) object disappearance.

1. Two observers provided the morph discrimination thresholds from which the model parameters were developed. Their thresholds were different in detail; furthermore, their sensitivities to sinusoidal gratings (one input to the model) were different. Such differences might lead to different image discrimination thresholds and model predictions (see data and modeling for the same two observers in Párraga et al. 2005). Here, we have used an average CSF to develop the model, as if there was a single standard observer. The rating scale data were obtained for two other observers, who quite likely had different CSFs again. If we had to model one single observer's data with that observer's own idiosyncratic CSF, we would expect to obtain better fits. However, if a visual differences predictor model is to have any widespread practical applications, it must probably relate

to a typical observer rather than to specific ones. In the future, it will be necessary to investigate interobserver differences in order to gauge the range of perceptual differences that may apply to a population of people.

2. The model may not be a perfect fit, because it does not follow the true behavior of V1 neurons carefully enough. Unlike some other models [Lubin 1995; Watson and Solomon 1997], ours does not yet include "contrast normalization" or "nonspecific suppression" [Heeger 1992], which is an obvious feature of V1 response behavior. There are, in fact, a variety of other nonlinear behaviors [see Carandini et al. 2005 for review] which ought to be considered, most particularly "surround suppression" [Blakemore and Tobin 1972] where stimuli spatially remote from a neuron's receptive field may nonlinearly suppress its responses to its preferred stimuli. Meese (2004) has shown that this phenomenon too needs to be accounted for in explaining some psychophysical results with grating stimuli.
3. It is well known that a person's sensitivity to simple stimuli falls off steadily the further the stimuli are from the center of gaze (the fovea) [e.g., Robson and Graham 1981]. We have not modeled such visual-field inhomogeneity and, we suspect from other work [Ripamonti et al. 2005], that, for natural-scene stimuli, the decrease in sensitivity is particularly rapid. It is possible that an observer's attention to detecting differences in natural scenes is highly focused on the point where they are presently looking; in the rating experiments, we obliged the observers to look only at the centers of the pictures.
4. There is a more interesting possibility: that V1-based models, however accurately they may represent V1 physiology, may simply not be the complete descriptor of human perceptual differences. There may be some visual discriminations, for instance, that the V1 model might predict are relatively easy, while a human observer finds them especially difficult. For instance, the "neurons" in our V1 model have precisely defined receptive-field locations. For instance, the model will respond to small translations of objects or to changes in the detailed organization of textures, such as pebbles or leaves on the ground. It is likely that a human would find it difficult to detect such changes in a pair of images.

Finally, in a natural-viewing situation rather than the rather contrived psychophysical procedures used here, people scan both images and try to spot the differences, as in a comic-book "spot the difference" task. This is notoriously hard, since it presupposes a pictorial memory that persists across eye movements. Yet, change blindness experiments suggest that there are severe restrictions on encoding and memory in these situations [see Simons and Rensink 2005].

However, given the caveats above, a simple V1 model of pictorial discrimination (now with color) emerges reasonably well from being tested on threshold and suprathreshold image differences, using real, albeit static images, and using both staircase (to determine threshold) and rating procedures (to determine suprathreshold differences). Therefore, one might be cautiously optimistic that further refinements to such models are warranted and that we may be beginning to be able to predict to what extent two images will appear different to a human observer.

ACKNOWLEDGMENTS

This work has been funded by project grants (ref RG/S56405/01) to D. J. Tolhurst and T. Troscianko from the BBSRC and EPSRC/Dstl of the UK, and to T. Troscianko, I. Cuthill, and J. Partridge from the BBSRC (ref S18903). CAP, PGL, and CR were employed on those grants. We thank our observers for giving us their time, especially KB. We also thank Dr Roland Baddeley for very welcome advice on methodologies for training neural networks.

REFERENCES

ALBRECHT, D. G. AND HAMILTON, D. B. 1982S triate cortex of monkey and cat—Contrast response function. *Journal of Neurophysiology 48*, 217–237.

BADDELEY, R. AND TRIPATHY, S. P. 1998. Insights into motion perception by observer modeling. *Journal of the Optical Society of America a-Optics Image Science and Vision 15*, 289–296.

BAKER, G. E., THOMPSON, I. D., KRUG, K., SMYTH, D. AND TOLHURST, D. J. 1998G eniculo-cortical projections and spatial-frequency tuning in area 17 and area 18 of adult ferret visual cortex. *European Journal of Neuroscience 10*. 9357.

BLAKEMORE, C. AND CAMPBELL, F. W. 1969. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology-London 203*, 237–260.

BLAKEMORE, C. AND TOBIN, E. A. 1972L ateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research 15*, 439–440.

CAMPBELL, N. W. AND KULIKOWSKI, J. J. 1966. Orientational selectivity of the human visual system. *Journal of Physiology 187*, 437–445.

CARANDINI, M., DEMB, J. B., MANTE, V., TOLHURST, D. J., DAN, Y., OLSHAUSEN, B. A., GALLANT, J. L., AND RUST, N. C. 2005. Do we know what the early visual system does? *The Journal of Neuroscience 25*, 10577–10597.

CHIRIMUUTA, M. AND TOLHURST, D. J. 2005. Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research 45*, 2943–2959.

CONWAY, B. R. 2001. Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1). *Journal of Neuroscience 21*, 2768–2783.

DALY, S. 1993. The visible differences predictor: an algorithm for the assesment of image fidelity. In *Digital Images and Human Vision*, Watson, A. B. Ed., MIT Press, Cambridge, MA., 179–206.

DE MONASTERIO, F. M., GOURAS, P., AND TOLHURST, D. J. 1975. Concealed colour opponency in ganglion cells of the rhesus monkey retina. *Journal of Physiology 251*, 217–229.

DE VALOIS, R. L. 1965. Analysis and coding of color vision in the primate visual system. *Symposia for Quantitative Biology 30*, (Sensory Receptors), 567–580.

DE VALOIS, R. L., ALBRECHT, D. G., AND THORELL, L. G. 1982. Spatial-frequency selectivity of cells in macaque visual cortex. *Vision Research 22*, 545–559.

DERRINGTON, A. M., KRAUSKOPF, J., AND LENNIE, P. 1984. Chromatic mechanisms in lateral geniculate–nucleus of macaque. *Journal of Physiology-London 357*, 241–265.

DOLL, T. J., MCWORTER, S. W., WASILEWSKI, A. A., AND SCHMIEDER, D. E. 1998. Robust, sensor-independent target detection and recognition based on computational models of human vision. *Optical Engineering 37*, 2006–2021.

ENROTH-CUGELL, C. AND ROBSON, J. G. 1966. The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology 187*, 517–522.

FIELD, D. J. 1994. What is the goal of sensory coding? *Neural Computation 6*, 559–601.

FOLEY, J. M. 1994. Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A 11*, 1710.

FORTE, J. D., BLESSING., E. M., BUZAS, P., AND MARTIN, P. R. 2006C ontribution of chromatic aberrations to color signals in the primate visual system. *Journal of Vision 6*, 97–105.

HEEGER, D. J. 1992N ormalization of cell responses in cat striate cortex. *Visual Neuroscience 9*, 181–197.

HURVICH, L. M. AND JAMESON, D. 1957. An opponent-process theory of colour vision. *Psychological Review 64*, 384–404.

ITTI, L., KOCH, C., AND BRAUN, J. 2000. Revisiting spatial vision: toward a unifying model. *Journal of the Optical Society of America A-Optics Image Science and Vision 17*, 1899–1917.

JIN, E. W., FENG, X. F., AND NEWELL, J. 1998. The development of a color visual difference model (CVDM). *Proceedings of IS&T PICS Conference*, Portland OR. 154–158.

JOHNSON, E. N., HAWKEN, M. J., AND SHAPLEY, R. 2001T he spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience 4*, 409–416.

JONES, J. P. AND PALMER, L. A. 1987. An evaluation of the two-dimensional Gabor Filter model of simple receptive-fields in cat striate cortex. *Journal of Neurophysiology 58*, 1233–1258.

LEGGE, G. E. 1981. A power law for contrast discrimination. *Vision Research 21*, 457–467.

LEGGE, G. E. AND FOLEY, J. M. 1980. Contrast masking in human vision. *Journal of the Optical Society of America 70*, 1456–1471.

LENNIE, P. AND MOVSHON, J. A. 2005. Coding of color and form in the geniculostriate visual pathway. *Journal of the Optical Society of America A-Optics Image Science and Vision 22*, 2013–2033.

LENNIE, P., KRAUSKOPF, J., AND SCLAR, G. 1990. Chromatic mechanisms in the striate cortex of the macaque. *Journal of Neuroscience 10*, 649–669.

LOVELL, P. G., TOLHURST, D. J., PARRAGA, C. A., BADDELEY, R., LEONARDS, U., AND TROSCIANKO, J. 2005. Stability of the color-opponent signals under changes of illuminant in natural scenes. *Journal of the Optical Society of America a-Optics Image Science and Vision 22*, 2060–2071.

LUBIN, J. 1995. A visual discrimination model for imaging system design and evaluation. In *In Vision Models for Target Detection and Recognition* Peli, E. (Ed.), World Scientific, Singapore, 245–283.

MACLEOD, D. I. A. AND BOYNTON, R. M. 1979. Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America A 68*, 1183–1187.

MEESE, T. S. 2004. Area summation and masking. *Journal of Vision 4*, 10, 930–943.

MOVSHON, J. A., THOMPSON, I. D., AND TOLHURST, D. J. 1978a. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology 283*, 53–77.

MOVSHON, J. A., THOMPSON, I. D., AND TOLHURST, D. J. 1978b. Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology 283*, 101–120.

MULLEN, K. T. 1985. The contrast sensitivity of human color vision to red–green and blue–yellow chromatic gratings. *Journal of Physiology-London 359*, 381–400.

MULLEN, K. T. AND KINGDOM, F. A. A. 2002. Differential distributions of red–green and blue–yellow cone opponency across the visual field. *Visual Neuroscience 19*, 109–118.

MULLEN, K. T. AND LOSADA, M. A. 1994. Evidence for separate pathways for color and luminance detection mechanisms. *Journal of Physiology 359*, 381–400.

MULLEN, K. T. AND LOSADA, M. A. 1999T he spatial tuning of color and luminance peripheral vision measured with notch filtered white noise. *Vision Research 39*, 721–731

MULLEN, K. T. AND SANKERALLI, M. J. 1999. Evidence for the stochastic independence of the blue–yellow, red–green and luminance detection mechanisms revealed by subthreshold summation. *Vision Research 39*, 733–745.

NACHMIAS, J. AND SANSBURY, R. V. 1974. Grating contrast discrimination may be better than detection. *Vision Research 14*, 1039–1042.

PÁRRAGA, C. A. AND TOLHURST, D. J. 2000. The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes. *Perception 29*, 1101–1116.

PÁRRAGA, C. A., TROSCIANKO, T., AND TOLHURST, D. J. 2005. The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model. *Vision Research 45*, 3145–3168.

PELI, E. 1990. Contrast in Complex Images. *Journal of the Optical Society of America A 7*, 2032–2040.

PELI, E. 2001. Contrast sensitivity function and image discrimination. *Journal of the Optical Society of America A 18*, 283–293.

PELI, E. AND GERI, G. A. 2001. Discrimination of wide-field images as a test of a peripheral-vision model. *Journal of the Optical Society of America A 18*, 294–301.

POINTER, J. AND HESS, R. 1989. The contrast sensitivity gradient across the human visual-field - with emphasis on the low spatial-frequency range. *Vision Research 29*, 9, 1133–1134.

POINTER, J. AND HESS, R. 1990. The contrast sensitivity gradient across the major oblique meridians of the human visual-field. *Vision Research 30*, 3, 497–501.

QUICK, R. F. 1974. A vector magnitude model of contrast detection. *Kybernetik 16*, 65–67.

RIPAMONTI, C., TOLHURST, D. J., LOVELL, P. G., AND TROSCIANKO, T. 2005. Magnification factors in a V1 model of natural-image discrimination. *Journal of Vision 5*, 595a.

ROBSON, J. G. AND GRAHAM, N. 1981. Probability summation and regional vairation in contrast sensitivity across the visual field. *Vision Research 21*, 409–418.

ROHALY, A. M., AHUMADA, A. J., AND WATSON, A. B. 1997. Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research 37*, 3225–3235.

SHAPLEY, R. M. AND ENROTH-CUGELL, C. 1984V isual adaptation and retinal gain control. *Progress in Retinal Research 3*, 263–346.

SIMONS, D. J. AND RENSINK, R. A. 2005. Change blindness: past, present, and future. *Trends in Cognitive Sciences 9*, 16–20.

SMITH, V. C. AND POKORNY, J. 1975. Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research 15*, 161–171.

SMYTH, D., WILLMORE, B., BAKER, G. E., THOMPSON, I. D., AND TOLHURST, D. J. 2003. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *Journal of Neuroscience 23*, 4746–4759.

STEVENS, S. S. 1975. *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects.* Wiley, New York.

TADMOR, Y. AND TOLHURST, D. J. 1994. Discrimination of changes in the second-order statistics of natural and synthetic-images. *Vision Research 34*, 541–554.

TOLHURST, D. J. AND BARFIELD, L. P. 1978. Interactions between spatial frequency channels. *Vision Research 18*, 951–958.

TOLHURST, D. J., PÁRRAGA C. A., LOVELL P. G., RIPAMONTI C., AND TROSCIANKO T. 2005. A multiresolution color model for visual difference prediction. Presented at the *2nd Symposium on Applied Perception in Graphics and Visualization*. Corona, Spain.

TOLHURST, D. J. AND TADMOR, Y. 1997a. Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra. *Vision Research 37*, 3203–3215.

TOLHURST, D. J. AND TADMOR, Y. 1997b. Discrimination of changes in the slopes of the amplitude spectra of natural images: band-limited contrast and psychometric functions. *Perception 26*, 1011–1025.

TOLHURST, D. J. AND TADMOR, Y. 2000. Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images. *Perception 29*, 1087–1100.

TOLHURST, D. J. AND THOMPSON, I. D. 1981. On the variety of spatial-frequency selectivities shown by neurons in area-17 of the cat. *Proceedings of the Royal Society of London Series B-Biological Sciences 213*, 183–199.

TROY, J. B. AND ENROTH-CUGELL, C. 1993. X-ganglion and Y-ganglion cells inform the cat's brain about contrast in the retinal image. *Experimental Brain Research 93*, 383–390.

VAN HATEREN, J. H. AND VAN DER SCHAAF, A. 1998I ndependent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London Series B-Biological Sciences 265*, 359–366.

WATSON, A. B. 1987. Efficiency of a Model Human Image Code. *Journal of the Optical Society of America A 4*, 2401–2417.

WATSON, A. B. AND AHUMADA, A. J. 2005A standard model for foveal detection of spatial contrast. *Journal of Vision 5*, 717–740.

WATSON, A. B. AND ROBSON, J. G. 1981. Discrimination at threshold: labelled detectors in human vision. *Vision Research 21*, 1115–1122.

WATSON, A. B. AND SOLOMON, J. A. 1997. Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A 14*, 2379–2391.

WICHMANN, F. A. AND HILL, N. J. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics 63*, 1293–1313.

WIESEL, T. N. AND HUBEL, D. H. 1966. Spatial and chromatic interactions in the lateral geniculate nucleous of the rhesus monkey. *Journal of Neurophysiology 29*, 1115–1156.

WILLMORE, B. AND TOLHURST, D. J. 2001. Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems 12*, 255–270.

Received December 2005; revised April 2006; accepted July 2006