A multiresolution color model for visual difference prediction

David J Tolhurst[1], Caterina Ripamonti[4]
Department of Physiology
University of Cambridge

C Alejandro Párraga[2], P George Lovell[3], Tom Troscianko[5]
Department of Experimental Psychology
University of Bristol

Abstract

How different are two images when viewed by a human observer? Such knowledge is needed in many situations including when one has to judge the degree to which a graphics representation may be similar to a high-quality photograph of the original scene. There is a class of computational models which attempt to predict such perceived differences. These are derived from theoretical considerations of human vision and are mostly validated from experiments on stimuli such as sinusoidal gratings. We are developing a model of visual difference prediction based on multi-scale 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. We describe the model, how it has been derived and how we attempt to validate it psychophysically for monochrome and chromatic images.

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

Keywords: psychophysical testing, image difference metrics, color vision

1 Introduction

We have previously shown how a simple (low-level), physiologically-plausible model of achromatic local contrast discrimination predicts human performance for discriminating between pairs of slightly different achromatic morphed pictures [Párraga et al 2000]. The model carries out a multiresolution analysis of the two pictures and detects differences in local contrast in each spatial frequency "channel".

Our model is based on knowledge of primary visual cortex and has much similarity with other models [see Daly 1993; Doll et al. 1998; Lubin, 1995; Rohaly et al. 1997; Watson 1987; Watson 2000]. 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 & Campbell 1969; DeValois et al. 1982; Movshon et al. 1978; Tolhurst & Thompson 1981; Watson & Robson 1981]. Such a model has many uses, which include the computation of visibility of targets in natural scenes [Rohaly et al, 1997] and also the derivation of image quality where quality is defined as the lack of perceptible difference from a given standard, such as in a compressed photograph of a scene. Clearly, there are applications of such a scheme in computer graphics; but it does need to be known that a model has been psychophysically validated. Thus, we have been doing a variety of psychophysical experiments, measuring thresholds for discriminating small changes in naturalistic images that we control by morphing. We decided to use a morphing technique (as opposed, for example to a superimposition of two images to different degrees) because it produces a set of stimuli where each one of the component pictures in an image of a plausible object (with slightly different shape, color and texture), which still shares the Fourier natural statistics of the original ones (see Párraga et al, 2000). Here 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' real thresholds with those predicted by our low-level model of visual cortex processing.

2 A discrimination model

Contrast estimation. We first calculate the contrast at each point in an image, at each of 5 spatial frequency scales [Peli 1990; Tadmor & Tolhurst 1994]. We define contrast at the point *(x,y)* and in the frequency band *F* as:

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

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:

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

while $l_F(x,y)$ is the result of convolving the original image with a circularly-symmetric low pass operator with frequency response given by:

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

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 bandwidth of about 1 octave. 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*; the mean luminance is calculated over an area proportional to the period of F. To model how the visual system compares two images, we calculate the $C_F(x,y)$ for both images at all frequency scales, and then we 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:

$$\Delta C_{F,j}(x,y) = \left| C_{F,j}(x,y) - C_{F,0}(x,y) \right|$$

[1]email: djt12@cam.ac.uk
[2]email: alej.parraga@bris.ac.uk
[3]email: p.g.lovell@bris.ac.uk
[4]email: cr324@cam.ac.uk
[5]email: tom.Troscianko@bris.ac.uk

where j is the picture number of the test stimulus and j=0 represents the reference picture. Then, we must estimate how much each value of ΔC might contribute towards the visibility of the difference between the pictures. We hypothesize that visibility depends not just on ΔC, but that it follows Weber's Law such that the visibility of an increment is affected by the baseline contrast level too (C): i.e. we evaluate each ΔC value against the familiar "dipper function" for contrast discrimination for sinusoidal gratings [Legge 1981; Legge & Foley 1980; Nachmias & Sansbury 1974]. Figure 1 shows such a dipper function. 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 Michelson contrast is the average of the paired contrast values in the two pictures at that location and frequency band.

$$\overline{C}_{F,j}(x,y) = 0.5\left|C_{F,j}(x,y) + C_{F,0}(x,y)\right|$$

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 & Tolhurst 2000]. Thus, the model dipper functions were determined from each observer's contrast sensitivity functions (CSFs). Any differences between observer's abilities to discriminate between pictures should hopefully be accounted for by differences in their CSFs. Note that the linear, "Weber" part of the dipper function for gratings has a slope of only 0.7 on log/log axes rather than unity [Legge 1981].

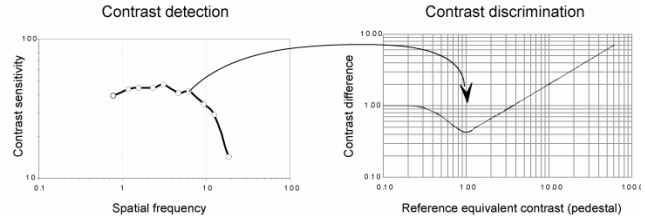


Figure 1. on the left is one observer's CSF – measures of the sensitivity for detecting the contrast of gratings. The sensitivity at a given spatial frequency determines the location of the contrast discrimination "dipper" on x and y axes.

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 dipper. There will be thousands of minute cues to discrimination, at the many locations and in the several frequency bands. To assess the overall discriminability of the two images requires some algorithm for pooling these many cues.

Pooling receptors and channels together. Thus, the second stage in the model is to pool the many cues (V) provided at different locations and different frequency bands to give an overall assessment of whether or not the two pictures differ sufficiently for discrimination to be made. Here we use 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 with power of 4 [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 & 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:

$$V_4 = \sqrt[4]{\sum_F \sum_x \sum_y (V_F(x,y))^4}$$

We have no preconception of the value that V_4 should take. In fact, in our modeling for achromatic stimuli, the criterion value of V_4 is a single parameter that we adjust in order to optimize the fit of the model to many experimental threshold data.

2.1 A chromatic model

The above refers essentially to an *achromatic* version of the model. We now extend this to evaluation of colored images. We suppose that human vision processes luminance (brightness) information and color information separately and in parallel [Mullen & Losada 1994], and that the color information is processed in red-green and blue-yellow *opponent channels* [Hurvich & Jameson 1957]. Here we investigate a model with three 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, and knowledge of the spectral activations of the three cone types [Smith & Pokorny 1975]. The luminance signal is then taken as the sum of L+M, whereas the red-green opponent signal [Párraga et al. 1998; Olmos & Kingdom 2004] is taken as (L-M)/(L+M) which is similar to one direction in Macleod and Boynton's [1979] color space. By analogy, we calculate a blue-yellow opponent signal as (S-0.5(L+M))/(S+0.5(L+M)). Given the lack of S cones in the centre of the human fovea, a dichromatic (luminance plus red-green) version of the model could be used to predict visibility of very small, centrally-fixated, targets. Our visual stimuli were 3.2 degrees square, but the fruit themselves in the centre of the stimuli were only 1.1 degrees. This is probably bigger than the size of the S-cone free area of vision. We then run the image discrimination model *three times* on each pair of colored images. First, we get an estimate of the overall discrimination variable V_4 for the luminance plane of the images. Then we 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]. Sensitivity for color signals is biased towards low spatial frequencies compared to luminance signals. The criterion values of V_4 for luminance and the two color channels are allowed to vary separately in the optimization to fit an observer's experimental data. The observer's ability to discriminate two images is set by the highest value of the three V_4 estimates (after accounting for their different criterion values). In fact, we shall show that the three values of V_4 to fit a set of experimental results are rather similar.

3 A psychophysical experiment

The purpose of this experiment was to obtain a large set of image-discrimination data on which the model could be optimized. 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. Plate A 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, these two morphed image sets were subjected to various filtering operations so that, in all, we obtained 49 different stimulus sequences for each. The 41 images in a sequence were split into their L, M, and S representations (see above), and thence into the three planes of

luminance, red-green opponent 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:

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

where f is spatial frequency and α is a slope parameter. Positive values of α give different degrees of blurring, a reduction in high spatial-frequencies; negative values give different degrees of sharpening, 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 could be filtered in seven different ways – from extreme sharpening to extreme blurring. Similarly, the two color-opponent planes could be filtered in 7 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; and then these values were reverse transformed to give the RGB values needed to display the desired images on a CRT. Plate B shows how filtering affected the appearance of the red pepper image for 9 exemplary image sets out the total of 49.

Thresholds were measured for several observers for the 49 different filtering combinations of each original morph sequence. "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. 2000]. The target images were presented for 0.5 s at a time. Figure 2 shows the thresholds for one observer (2 other observers gave similar results).

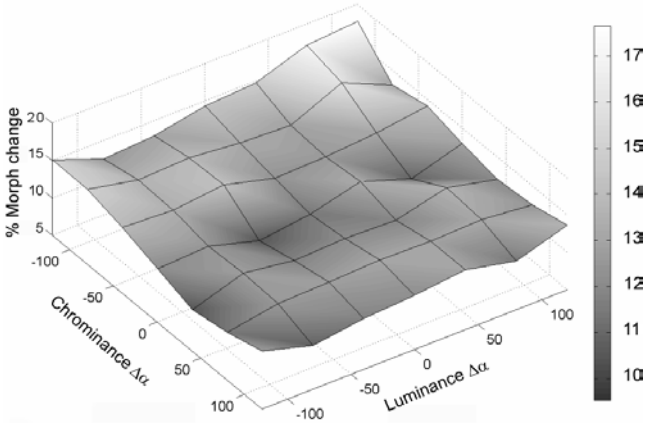


Figure 2. Examples of experimental results for observer CAP (red-pepper/yellow-lemon sequence). The pseudo-3D plot shows the amount of morph needed for discrimination for each of 49 conditions – 7 manipulations of the chromatic plane (left axis) times 7 manipulations of the luminance plane (right axis).

The thresholds for the 49 conditions are mostly similar, and crucially, the poorest observer performance (for highly whitened chrominance and highly blurred luminance pictures –top corner of the plot) was captured correctly by the model. Model performance was worst (higher thresholds) in the right corner of the plot surface where either the luminance or the color opponent planes or both had been subject to extreme filtering. This was especially so for stimuli with negative filtering of color and positive filtering of luminance, i.e., 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). Plates C and E show contour-plots of the experimental results for observers KB and CAP respectively. Since these images were derived from pictures of a red pepper and a yellow lemon, the morph sequence 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 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). 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). Psychophysical experiments were conducted on this sequence in a similar manner (see above). Plate G shows a contour plots of the results for observer KB.

4 Application of the model

The model was applied to the psychophysical data, according to the optimization processes described in Section 2.1 above. Figure 3 shows the results of the modeling applied to the experimental results in Figure 2, presented in the same format. In general, the model has matched the shape of the 49 data-point surface representing the experimental results. The model predicts that human thresholds *should* rise in the corners of the surface, where either the luminance or the color planes or both are highly filtered. However, the model predicted that observers should discriminate better morph sequences that were highly "whitened" in the luminance domain which suggests that there might have been other mechanisms in operation, such as observers not looking at the right places and missing small changes in these sequences , which were captured by the model, etc. These results are also shown in plates D and F as contour plots, where each data point has been colored to represent the channel responsible for the model discrimination (black for luminance, red for red-green opponent channel and blue for the blue-yellow opponent channel).

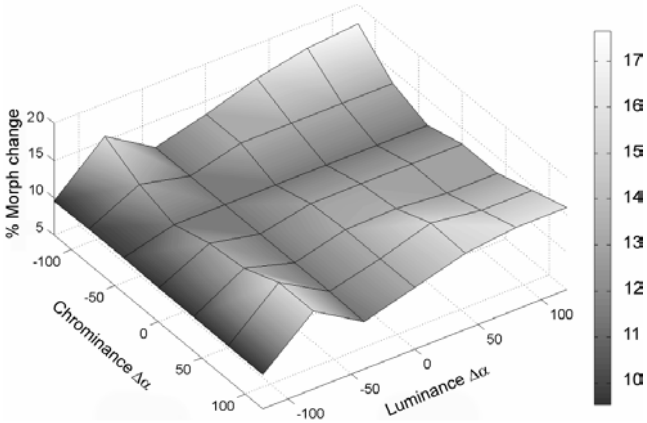


Figure 3. The results of modeling the psychophysical results of Fig.2. Note that the modeled surface here has the same overall shape as the experimental surfaces in Fig.2

The criterion values of V_4 were remarkably similar for the three channels and two observers. (75.07, 68.31 and 68.58 for the luminance, red-green and blue-yellow channels for observer KB and 82.37, 78,24 and 64.08 for observer CAP respectively). These

values are similar to those obtained for other observers using achromatic images (Párraga et al, 2005). Plates G and H show contour plots of experimental results and model predictions for the second experimental set (the "bluish-pepper to lemon sequence"). Here the model again captures the overall shape of the data (higher thresholds for pictures with "blurred" chromatic channels, right side of the plots) despite the fact that the experimental results have a completely different shape than those in Plates E and C. V_4 values were also similar to those obtained before (72.55, 75.61, 68.20) The fact that all the V_4 values are so similar implies that changes in luminance and in color are equally salient in visual discriminations.

5 Conclusions

We have developed a model of how much difference there has to be between two serially-presented visual images, for a human observer to be able to detect the difference at a threshold level. Our model has been validated against human psychophysical data in a natural-image discrimination task. The fit, although not perfect, is good enough to account for the main properties of the psychophysical results. In particular the model captures the differences in the form of the results surface for the two kinds of stimuli that we employed. Additionally, the model needs to be tested against detection performance in a greater variety of naturalistic tasks. A model may be very successful at predicting relative threshold performance in rather similar psychophysical tasks [Rohaly et al. 1997; Párraga et al. 2000]. It is only when one considers a wide range of tasks that discrepancies become evident and demand attention [Párraga et al. 2005]. Our current work in this domain involves predicting the detection distance for traffic signs, and future work will investigate the degree to which our model can predict differences between two images, even when these differences are large and clearly visible. In the field of computer graphics, the model offers the possibility of assessing perceivable differences in rendering quality.

Acknowledgements

This work has been funded by project grants 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. CAP, PGL and CR were employed on those grants.

References

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.

DALY, S. 1993. The visible differences predictor: an algorithm for the assesment of image fidelity. In *Digital images and human vision*. ed. WATSON AB, pp. 179-206. MIT Press, Cambridge, Mass.

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.

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.

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

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.

LUBIN, J. 1995. A visual discrimination model for imaging system design and evaluation. Pp 245-283 In *Vision Models for Target Detection and Recognition* ed E. Peli, World Scientific: Singapore.

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.

MOVSHON, J.A., THOMPSON, I.D. AND TOLHURST, D.J. 1978. 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 LOSADA, M.A. 1994. Evidence for separate pathways for color and luminance detection mechanisms. *Journal of the Optical Society of America A*, 11, 3136-3151.

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

OLMOS, A. AND KINGDOM, F.A.A. 2004. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33, 1463-1473.

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. 2000. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology* 10, 35-38.

PÁRRAGA, C.A., BRELSTAFF, G., TROSCIANKO, T. and Moorhead, I.R. 1998. Color and luminance information in natural scenes. *Journal of the Optical Society of America A 15*, 3, 563-569.

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* submitted.

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

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

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.

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

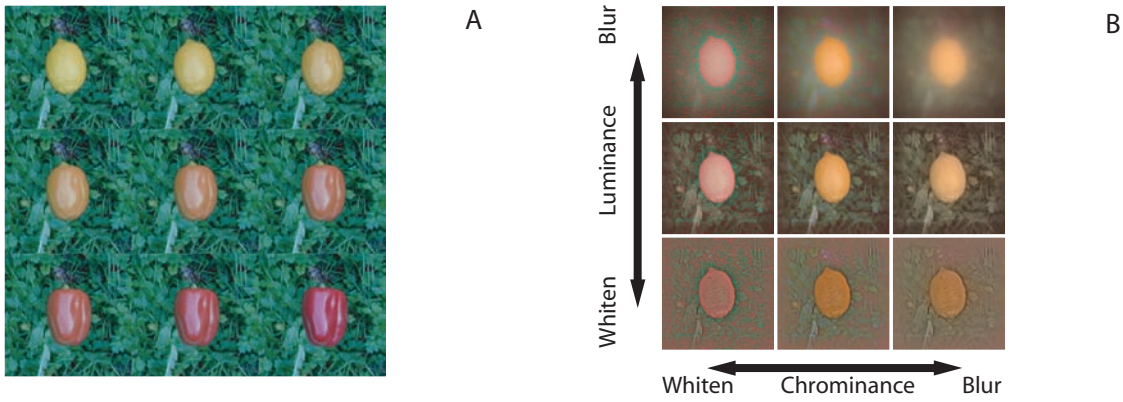
TADMOR, Y. AND TOLHURST, D.J. 1994. Discrimination of changes in the second-order statistics of natural and synthetic-images. *Vision Research* 34, 541-554.

TOLHURST, D.J. AND 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.

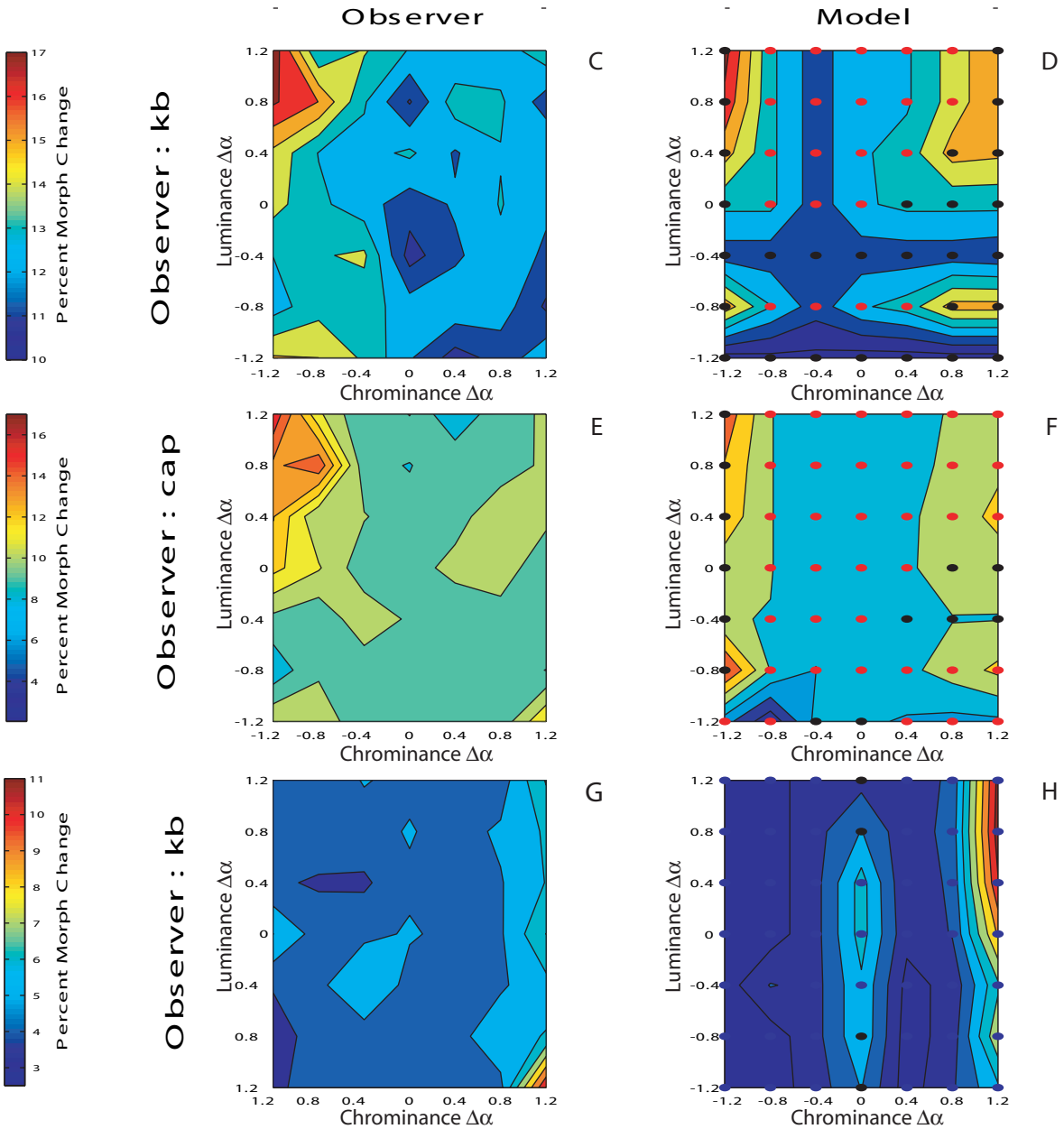
WATSON, A.B. 1987. Efficiency of a Model Human Image Code. *Journal of the Optical Society of America A* 4, 2401-2417.

WATSON, A.B. 2000. Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optical Express* 6, 12-33.

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



Examples of morphed stimuli (A) and filtered versions (B)



Experimental results (left column) and model predictions (right column)