

Papers that [reference](#) this paper.

Simplified Vision Models for Image Quality Assessment

A. J. Ahumada, Jr.
NASA Ames Research Center, Moffett Field CA

Abstract

Recently developed vision models for image discrimination are becoming increasingly complex, so that their use as image quality metrics is often precluded by computational limitations. This paper compares the performance of (1) a single channel model with a simple masking correction to (2) a multiple spatial frequency channel model with "within-channel" masking.

Introduction

Vision discrimination models that have been used as image quality metrics range in complexity from single filter models to multiple channel models with channels that are selective in spatial frequency and orientation (see for a review [1]). The simple filter models can be thought of as representing the visual information at precortical levels of the visual system. These models can predict the variations in the visibility of targets that occur as the target spatial frequency changes. The multiple channel models simulate the orientation and spatial frequency selectivity of cortical cells. Because the channels outputs are nonlinear, the channel models predict masking of targets in high contrast image regions. To facilitate the computation of such models, Watson developed the Cortex Transform and based a model on it [2]. Daly and Lubin have developed similar models [3, 4]. Because they involve computing channel outputs varying in spatial position, spatial frequency, and orientation, these models are computationally complex, even when they use frequency and orientation selective transforms that are simpler to compute [5] or are already being computed [6]. We have reported that a Cortex Transform model with within-channel masking outperformed a simple contrast sensitivity filter model at predicting the detectability of targets in a natural background [7].

Some recent vision models include between-channel interactions that allow the models to predict masking from image components exciting different channels from those responding to the target [8, 9]. These models have even greater computational complexity. Foley developed a formula for predicting the masking of Gabor stimuli by gratings of various spatial frequencies and orientations [10]. Simplifying the formula leads to a simple contrast masking correction. Here we show that applying this correction to a filter model can make its performance comparable to that of the channel model.

The vision models

The models that I will compare in this paper are image discrimination models. They take as input a pair of images and give as output the number of just-noticeable-differences (d') between them. The models are linear in the image difference, so that the background image and the background image plus the image difference multiplied by $1/d'$ are one just-noticeable-difference apart.

Both models have the same initial steps. The images are converted to contrast images by subtracting and then dividing by the mean luminance of the background image. Next they are filtered by a contrast sensitivity filter whose shape and gain are calibrated for each model to predict the detection of 1.33 deg square grating patches according to the formula developed by Barten [11]. The filters have a difference of Gaussian form,

$$S(f) = a_c \exp(-f/f_c)^2 - a_s \exp(-f/f_s)^2, \quad (1)$$

where a_c and a_s are the center and surround amplitude parameters and f_c and f_s are the center and surround frequency cutoff parameters. Table 1 gives the contrast sensitivity filter parameters for the models. The amplitude parameters have the dimensions of JND's per unit contrast and the cutoff parameters have the

dimensions of cycles per degree of visual angle.

Table 1 - Contrast sensitivity filter parameters

model	a_c	f_c	a_s/a_c	f_c/f_s
filter	15.5	20.8	0.77	5.6
channel	18.5	16.4	0.68	7.9

1. The filter models

Let $I_0(k)$ and $I_1(k)$ be the k^{th} pixels of the background and background plus target filtered images. Form the differences of the filtered images,

$$D(k) = I_1(k) - I_0(k) . (2)$$

The predicted difference between the images for the simple filter model is the vector length of the filtered difference image,

$$d'_s = (S_k D(k)^2)^{1/2} . (3)$$

Also compute the RMS contrast of the filtered background image, where the filter is normalized to have unity maximum gain,

$$c = (1/g) ((1/n) S_k I_0(k)^2)^{1/2} . (4)$$

where n is the number of pixels, and g is the maximum gain of the contrast sensitivity filter. The model prediction is given by

$$d'_m = d'_s / (1 + (c/c_2)^2)^{1/2} . (5)$$

The value of c_2 will be kept at a contrast of 0.04 for all the predictions here. Notice that when there is no background image contrast, $c=0$ and $d'_m = d'_s$. The correction therefore does not require any recalibration for the grating thresholds.

2. The channel model

Let $C_0(k)$ and $C_1(k)$ be the k^{th} coefficients of the Cortex Transforms [2] of the background and background plus target filtered images. In all examples, there will be five spatial frequency channels and 4 orientation channels in the transform. Compute the masked transform differences,

$$D(k) = | C_1(k) - C_0(k) | / (1 + | C_0(k) |^2)^{1/2} . (6)$$

The model prediction for the discriminability between the images for the channel model is given by a Minkowski sum of the masked differences with the summation exponent set to 4 to approximate probability summation of the detectability of the differences,

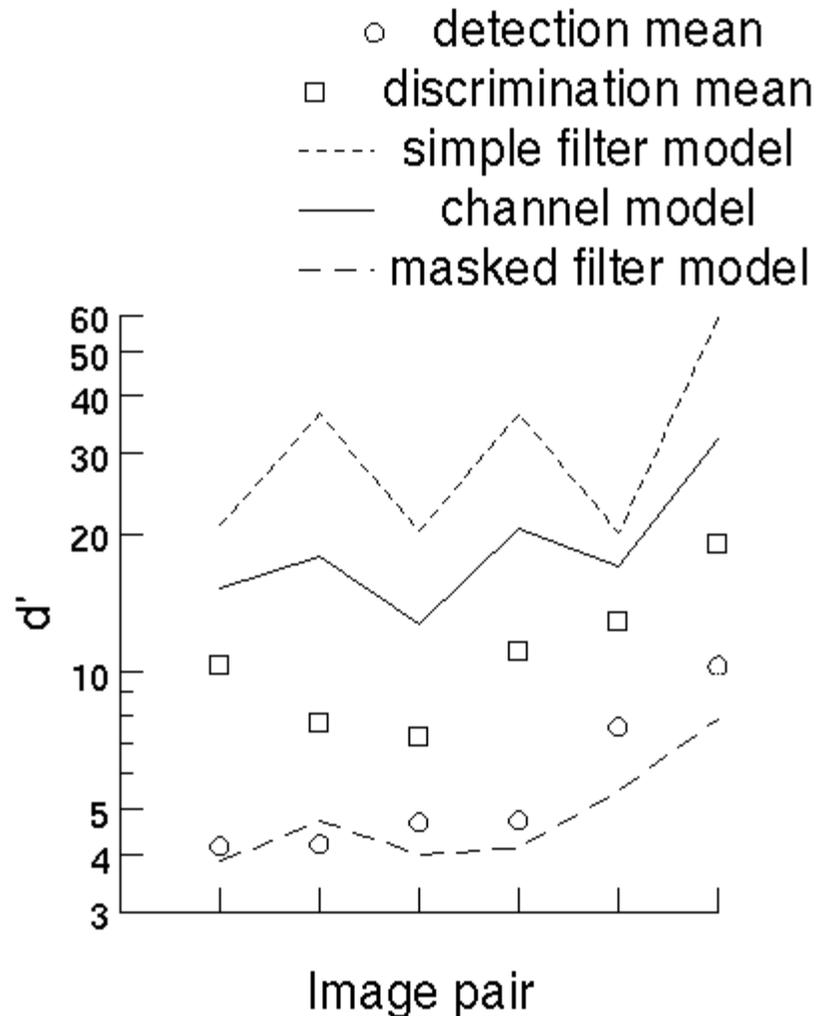
$$d'_c = (S_k D(k)^4)^{1/4} . (7)$$

Object detection in a natural background

Last year, we reported data on the detectability of vehicles by soldiers in natural background color images and showed that the channel model performed better than the filter model with no masking correction [7].

When the filter model is given the masking correction, it outperforms the channel model in predicting those results. The circles in Figure 1 show the mean performance of 2 groups of soldiers for 6 image pairs. Only one number, the difference between the two groups of subjects is available for estimating the accuracy of the average performance over the 6 image pairs. It was 1.7 dB (20 dB = 1 log unit). The estimated error of measurement for the pattern of responses was 0.5 dB, based on the group by image-pair interaction with 5

Figure 1. Object detection in a natural background.



degrees of freedom.

The simple filter model predictions (dotted line in Figure 1) average 14.4 dB higher than the soldier observers. The standard deviation of the prediction deviations when the mean difference is removed is 3.4 dB. As reported previously, the channel model predictions (solid line in Figure 1) are better. The predictions average 8.5 dB too high, with a standard deviation of only 1.9 dB. However, the masked filter model predictions (dashed line in Figure 1) are just 1.3 dB too low and fit the pattern of responses with an error of only 1.2 dB.

We have now measured the discriminability of these six image pairs for three observers, whose average responses are given by the squares in Figure 1. They saw the same gray scale images that were presented to the models, and one image pair was presented per block of trials so that the observer would know the target and its location. The standard error of the average level of performance is 1.1 dB, with 2 degrees of freedom. The estimated error of measurement for the response pattern was 0.7 dB with 10 degrees of freedom. The average detection level of the 3 discrimination observers was 5.7 dB higher than that of the soldiers. Thus, the simple filter model overpredicted the average level by 8.7 dB, the channel model overpredicted by 2.8 dB, and the masked filter model underpredicted by 6.9 dB. The errors in the pattern of predictions were 3.1

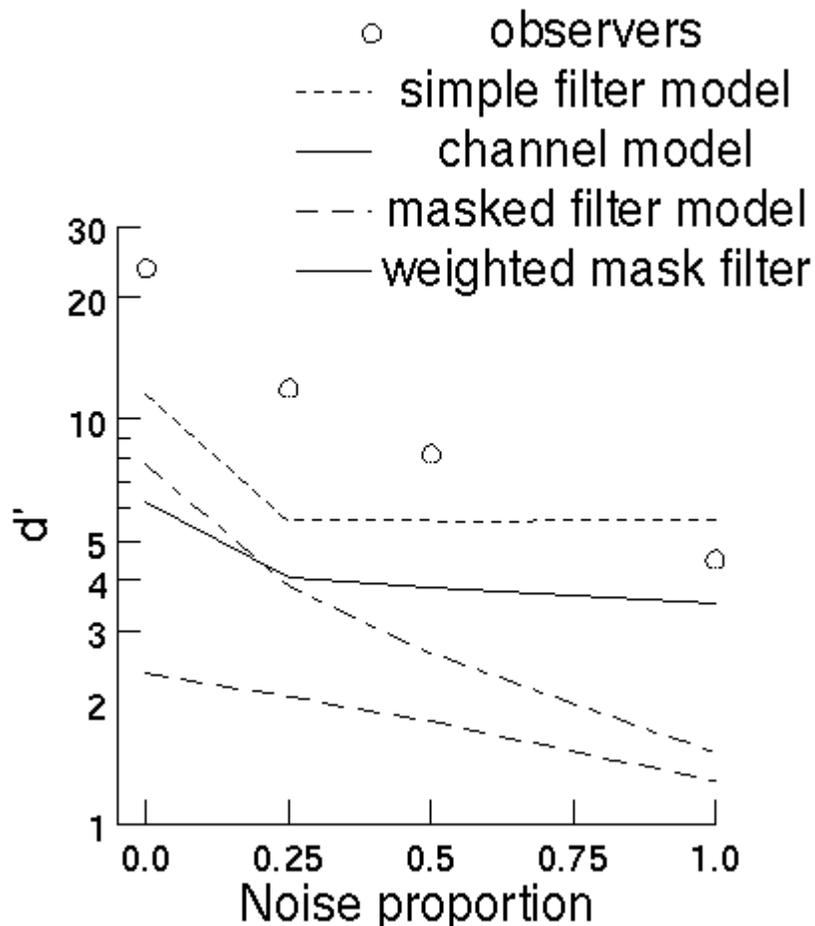
dB for the simple filter model, 1.4 dB for the channel model, and 1.7 dB for the masked filter model.

Overall, the masked filter model and the channel model performed better than the simple filter model both at predicting the overall level of performance and the pattern of performance, but there is no consistent difference between those two.

Target detection in a noisy airport scene

In another study, we measured the discriminability of a target masked by the scene image contrast and additional fixed-pattern white noise [12]. The target was an obstacle airplane on a runway and thresholds were measured as in the previous study for 4 observers. There were 4 image pairs, the airport scene with the airplane present and absent, and those scenes at half contrast with three levels of added white noise. Figure 2 shows the average results for the observers and the predictions of the models. The estimated standard error of the observers average responses was 1.3 dB with 3 degrees of freedom. The estimated standard error for the response pattern was 0.6 dB with 9 degrees of freedom. All models underpredicted the average performance level. The simple filter model (dotted line in Figure 2) by 3.6 dB, the channel model (solid line in Figure 2) by 7.4 dB, and the masked filter model (dashed line in Figure 2) by 14.9 dB. The errors in the overall pattern of predictions were the same for the three models, 3.4 dB for the simple filter model, 3.5 dB for the channel model, and 3.4 dB for the masked filter model. Although there is clearly masking by the noise, which is best predicted by the masked filter model, the model does not do better than the others because it predicts too much masking by the airport scene. When the contrast of the scene is cut in half (0.48) before being combined with the noise contrast, the error in the pattern of predictions is reduced to 0.1 dB (the dot-dash line in Figure 2).

Figure 2. Target detection in a noisy airport scene.



Discussion

The results for the above experiments showed little if any advantage for using the channel model rather than the filter model with the masking correction. The two models performed similarly when predicting the relative pattern of responses in the above experiments. There are two problems with the masked filter model that should be pointed out before one concludes that its computational simplicity makes it clearly the correct choice.

One problem is the summation rule used with the filter models. The probability summation rule is known to be much more accurate for large targets. The good fit of the average prediction for the simple filter model in the second experiment is the result of two problems cancelling out. The model predicts no masking, which raises its sensitivity, but it also underpredicts the sensitivity to targets like the airplane that are much smaller than the calibration targets. If the experiments had used targets that varied greatly in size, this weakness would have been reflected in poorer pattern responses for both filter models. The Minkowski summation rule with an exponent of 4 does better than the exponent of 2 for the filter model as well as the channel model [7]. The exponent of 2 was used here for the filter models because of the many possible applications in which the image is represented in a length preserving transform domain. In that case, it would probably be better to use an exponent of 4 in summing differences in that domain, but the present results show that one can do well with a vector length difference measure.

A second problem is the range of the background image that should be used in computing the masking correction. Here we have used the entire image. In experiments with masking noise that surrounds a target, no masking is found when the mask does not actually cover the target [13, 14]. We recommend, therefore, using an estimate of the image contrast in the region of the target.

In conclusion, the relations among the three models are nicely illustrated by their predictions of the results of Foley on the masking of a Gabor patch target by background gratings of the same frequency, but varying in orientation [10]. The simple filter model predicts no effect of the backgrounds. However, for observer KMF, the performance at the highest contrast level (0.31) of the background grating with the target orientation was 18 dB worse than with no masker. The channel model predicts no masking by an orthogonal grating, but the orthogonal grating masked performance was 16 dB worse than with no masker. The masked filter model errs least by predicting that all orientations have the same masking.

In general masking is greater, the more the masker and the target have the same location in space and spatial frequency (including orientation). The most complex models should allow masking to vary appropriately in these dimensions. In applications where the masking background is spatially homogeneous and wide band in spatial frequency, masking may be predictable from simple models.

Acknowledgments

Assistance was received from B. Beard, R. Horng, A. Watson, A. M. Rohaly, and C. Null. Supported by NASA Grant 199-06-39 and NASA RTOP #505-64-53.

References

- 1) [A. Ahumada, Jr., Computational Image Quality Metrics: A Review](#), SID Digest, 24, 305-308, 1993.
- 2) A.B. Watson, Efficiency of an image code based on human vision, J. Opt. Soc. Am. A, 4, 2401-2417, 1987.
- 3) S. Daly, The visible differences predictor: an algorithm for the assessment of image fidelity, A.B. Watson, ed., Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993.
- 4) J. Lubin, The use of psychophysical data and models in the analysis of display system performance, A.B. Watson, ed., Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993.

- 5) E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multi-scale transforms, IEEE Trans., IT-38, 587-607, 1992.
- 6) A.B. Watson, DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images, SID Digest, 24, 946-949, 1993.
- 7) [A.M. Rohaly, A.J. Ahumada, Jr., A.B. Watson, A Comparison of Image Quality Models and Metrics Predicting Object Detection](#), SID Digest, 26, 45-48, 1995.
- 8) A.B. Watson, J.A. Solomon, Contrast gain control model fits masking data, Inv. Ophth. Vis. Sci., 36 (4 ARVO Suppl.), 438, 1995.
- 9) P.C. Teo, D.J. Heeger, Perceptual image distortion, Proc. ICIP-94, vol. II, IEEE Computer Society Press, 982-986, 1994.
- 10) J.M. Foley, Human luminance pattern-vision mechanisms: masking experiments require a new model, J. Opt. Soc. Am. A, 11 (6) 1710-1719, 1994.
- 11) P. Barten, Physical model for the contrast sensitivity of the human eye SPIE Proc. 1666, 57-72, 1992.
- 12) [A.J. Ahumada, Jr., B.L. Beard, Object detection in a noisy scene](#), SPIE Proc. 2657, 190-199, 1996.
- 13) R.J. Snowden, S.T. Hammett, The effect of contrast surrounds on contrast centres, Inv. Ophth. Vis. Sci., 36 (4 ARVO Suppl.), 438, 1995.
- 14) J.A. Solomon, A.B. Watson, Spatial and spatial frequency spreads of masking: measurements and a contrast gain-control model, Perception, 24 (ECPV Suppl.), 37, 1995.