40.1 A Simple Vision Model for Inhomogeneous Image Quality Assessment

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

Abstract

The single filter model for image quality assessment previously presented at this meeting breaks down when the image luminance or masking contrast energy varies spatially within the image. A slightly more complex version of the single filter model is presented which allows for these inhomogeneities.

Introduction

A simplified vision model for image quality assessment was presented at the 1996 Meeting and elsewhere since [1, 2, 3]. Examples were given demonstrating situations in which this model's performance exceed that of more complex models with channels that represent the spatial frequency selectivity of the simple cells of the visual cortex, but no masking beyond within-cell nonlinearity. The better performance was attributed to the general masking term of the simple model. The general masking term accounts for cross-channel masking that has been added to the cortex type models at the expense of even more computational complexity.

The simplified model has two weaknesses that can be remedied without much trouble. One weakness is that the conversion from luminance to contrast was done at a global level rather than a local level, as is known to be appropriate [4]. This means that in the previous model, contrast in a light part of the image was overestimated, while it was underestimated in dark parts of the image. The second weakness of the simplified model is that the conversion from filtered contrast to masked filtered contrast was done globally, rather than locally [5, 6]. Contrast energy in the target background affects perceived contrast over a large region, but it only masks very locally. Here local luminance and contrast energy images are used in place of the single numbers in the simplified model [1].

The Model Steps

The input to the model consists of two images. The output is a perceptual distance d', representing the number of just-noticeable-differences between the images. Each of the following steps is applied to both images.

Blur. The image I is convolved with a low pass Gaussian filter F_B,

$$B[x,y] = I[x,y] * F_B[x,y].$$

Local luminance. The blurred image B is convolved with a low pass Gaussian filter F_L,

$$L[x,y] = B[x,y] * F_L[x,y].$$

Local contrast. The contrast image is computed from the local luminance,

$$C[x,y] = \frac{B[x,y]}{L[x,y]} - 1.$$

Local contrast energy. Squared contrast image values are convolved with a Gaussian low pass filter F_E,

$$E[x,y] = (C[x,y]^2) * F_E[x,y].$$

Local contrast gain adjustment. The masked visible contrast image is computed using a divisive inhibition formula,

$$V[x,y] = \frac{C[x,y]}{(1 + (g_E\, E[x,y]))^{0.5}}.$$

Summation of image differences. The distance between the masked visibility images is based on a Minkowski metric with an exponent of 4, corresponding to probability summation over space,

$$d' = g_C \left(\sum_{x,y} (V_1[x,y] - V_2[x,y])^4\right)^{0.25}.$$

Parameters. Parameters for the model were estimated by least squares fit to the target detection data for several backgrounds in noise [7]. Table 1 shows the parameter estimates.

Table 1 - Model Parameter Estimates

s_B	blur spread	1.0	arc min
s_L	local luminance spread	9	arc min
s_E	local contrast spread	25	arc min
g_E	contrast gain gain	7	
g_C	contrast sensitivity gain	10.5	

Discussion

When the image is uniform in average luminance, one can simplify the model by dividing by that average luminance value rather than the average local luminance image value. Similarly, if the average contrast energy is constant one can further simplify by using that value. The previous simplified model used both these simplifications, but it also was a linearized model with the property that the d' for a target added to the background is proportional to the target amplitude. This convenient property was obtained by using the luminance and contrast energy values from the background image alone. This could be done here for the contrast energy, but if it is done for the luminance, the bandpass nature of contrast sensitivity is lost because it depends on the local luminance.

In complex models with spatial frequency channels both the luminance normalization and the contrast energy masking have been done after the spatial frequency analysis. Duvall-Destin [8] and Peli [9] have proposed calculating contrast at each spatial scale separately, based on the DC value at the neighboring larger scale. A similar result obtains from Zetzsche and Hauske's ratio of Gaussians model [10]. These models ignore the large amount of local adaptation occurring in the retina before the spatial frequency analysis of the cortex, but perhaps partly correct for the effects of eye movements on retinal adaptation, which would tend to make the adaptation spread to the size of fixated objects.

Models that simulate cortical units now include the lateral interactions that provide scale dependent divisive inhibition from neighboring units stimulated by contrast energy to which they are tuned. For Teo and Heeger [11], the neighbors differed only in orientation. For Cannon [12] the neighborhood was extended in space, but the extension was to explain the drop in perceived contrast, which has a much wider spread than actual masking. In the Watson and Solomon model [6], masking is generated in units tuned to different positions, spatial frequencies, and orientations, but the estimated spatial spread was very small.

Although luminance adaptation and contrast masking are local, the above arguments suggest that fixed spreads cannot always work. Because luminance adaptation at the retinal level is a function of eye movements, and has possible contributions from scale variant processing, no fixed size of spread could correctly predict all possible signal and background situations. The situation is similar for contrast masking. The point of the present model is an engineering approximation, whose usefulness should diminish as more complex models become developed to the point where they can make accurate predictions in complex scenes.

When detectability is the main criterion, it may be better to devote computing resources to modeling masking rather than filtering properties of the visual system. 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 allow masking to vary appropriately in these dimensions. In applications where the artifacts, targets, and/or masking images are not well localized in spatial frequency, the masking may be predictable from this simplified model.

Acknowledgements

This work was supported by NASA RTOP 548-51-12.

References

[1] A. J. Ahumada, Jr., "Simplified Vision Models for Image Quality Assessment," SID International Digest of Technical Papers, Volume XXVII, May 1996, pp. 397-400.

[2] A. J. Ahumada, Jr., B. L. Beard, "Image Discrimination Models Predict Detection in Fixed but not Random Noise," Journal of the Optical Society of America A, Volume 14, pp. 2471-2476, 1997.

[3] A. M. Rohaly, A. J. Ahumada, Jr., A. B. Watson, Object detection in natural backgrounds predicted by discrimination performance and models," Vision Research, Volume 37, pp. 3225-3235, 1997.

[4] H. Wallach, "The perception of neutral colors," Scientific American, Volume 208, January 1963, pp. 107-116.

[5] R. J. Snowden, S. T. Hammett, "The effect of contrast surrounds on contrast centres," Investigative Ophthalmology and Visual Science, Volume 36, Number 4 (ARVO Suppl.), p. S438, 1995.

[6] A. B. Watson, J. A. Solomon, "A model of visual contrast gain control and pattern masking," Journal of the Optical Society of America A, Volume 14, pp. 2379-2391, 1997.

[7] A. J. Ahumada, Jr., B. L. Beard, "Parafoveal target detectability predicted by local luminance and contrast gain control," Investigative Ophthalmology and Visual Science, Volume 38, Number 4 (ARVO Suppl.), p. S380, 1997.

[8] M. Duval-Destin, "A spatio-temporal complete description of contrast", SID International Digest of Technical Papers, Volume XXII, May 1991, pp. 615-618.

[9] E. Peli, "Contrast in complex images," Journal of the Optical Society of America A, Volume 7, pp. 2032-2040, 1990.

[10] C. Zetzsche, G. Hauske, "Multiple channel model for the prediction of subjective image quality," in B. Rogowitz, ed., Human Vision, Visual Processing, and Digital Display, SPIE Volume 1077, pp. 209-216, 1989.

[11] P. C. Teo, D. J. Heeger, "Perceptual image distortion," Proceedings of ICIP-94, Volume II, IEEE Computer Society Press, Los Alamitos, California, pp. 982-986, 1994.

[12] M. W. Cannon, "A multiple spatial filter model for suprathreshold contrast perception," in E. Peli, ed., Vision Models for Target Detection and Recognition, World Scientific Publishing, New Jersey, 1995.