PERCEPTUAL IMAGE QUALITY BASED ON A MULTIPLE CHANNEL HVS MODEL

S.J.P. Westen, R.L. Lagendijk, and J. Biemond
Information Theory Group,
Department of Electrical Engineering,
Delft University of Technology
P.O. Box 5031, NL-2600 GA, Delft, The Netherlands
e-mail: stefan@it.et.tudelft.nl

ABSTRACT

We propose a new measure of perceptual image quality based on a multiple channel human visual system (HVS) model for use in digital image compression. The model incorporates the HVS light sensitivity, spatial frequency and orientation sensitivity, and masking effects. The model is based on the concept of local band-limited contrast (LBC) in oriented spatial frequency bands. This concept leads to a simple masking function. The model has the flexibility to account for the changes in frequency sensitivity as a function of local luminance and is consistent with masking experiments using gratings and edges. Numerical scaling experiments with a test panel and a set a test images that were coded using different coding algorithms showed that the proposed measure correlates better with perceptual image quality than the conventional SNR measure.

1. INTRODUCTION

In optimization and evaluation of digital image compression algorithms, the signal to noise ratio (SNR) is generally used as a measure of image quality. However, the use of a measure of image quality that is based on properties of the human visual system (HVS) will in general lead to a better visual image quality of the reconstructed image [1].

Properties that are usually incorporated into HVS models for perceptual image quality are the sensitivity to light, the spatial frequency sensitivity and masking effects. The sensitivity to light is the dependence of the sensitivity on the local luminance. In general, the HVS is more sensitive in dark areas than in light areas of the image. The spatial frequency sensitivity of the HVS decreases for high frequencies. The frequency sensitivity is usually described with a fixed low-pass or a band-pass filter [2]. However, the frequency sensitivity is dependent on the background luminance. The frequency sensitivity increases and the peak shifts to higher frequencies as luminance increases. The masking effects describe the influence of the image contents on the visibility of distortions. Masking effects occur mainly in the vicinity of edges and in textured regions of images.

Recently, several models of the HVS have been proposed that are based on a hierarchy of spatially oriented band-pass filters [3,4,5,6]. The existence of this hierarchy is suggested by a large number of masking experiments with gratings [7]. Masking effects in these models are explained by threshold elevation in the spatial frequency bands. Distortions in a spatial frequency band are masked by the image contents in that spatial frequency band. Experiments show that at edges maximum masking occurs at the exact position of the edge [8]. However, when linear phase filters are used in the HVS model, the outputs of the band-pass filters will have zero-crossings at the exact positions of edges. When a masking function is used that is based on the exact outputs of these filters, no masking is predicted at the exact positions of edges.

We propose a new model for perceptual image quality that is based on a multiple channel HVS model. We propose a new method to calculate the masking effects that is consistent with masking experiments with gratings as well as edges. The model is centered around the concept of local band-limited contrast (LBC). The model also enables the modeling of the change of spatial frequency sensitivity as function of local luminance, as opposed to the fixed sensitivity in most current models.

2. HVS MODEL

Figure 1 shows the structure of the proposed HVS model. The inputs of the model are the original and the distorted image in the luminance domain (in Cd/m^2). Usually the images will be specified in terms of gray levels, which means that the transfer from gray levels to screen luminance will have to be calculated before the HVS model can be applied. This transfer function depends on the display device that is used and hence we have chosen not to incorporate it into the model.

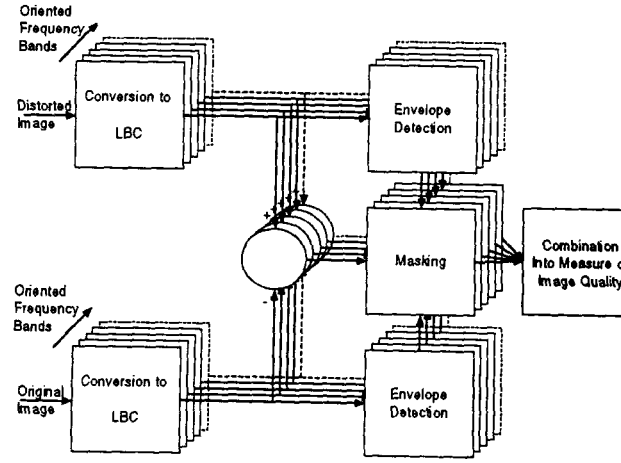

Figure 1: The proposed HVS model.

The original and the distorted image are first transformed into Local Band-Limited Contrast (LBC) for different frequency bands and orientations. The LBC incorporates the effects of light and frequency sensitivity of the HVS. After the LBC calculation, the masking for the different frequency bands and orientations is calculated as a function of the envelopes of the LBCs of the original and the distorted image. It turns out that

0-7803-2431-5/95 $4.00

the concept of LBC leads to a very simple masking function. The last step of the model is the combination of the responses into a measure of perceptual image quality.

2.1. Local band-limited contrast

Figure 2 shows the calculation of the local band-limited contrast. The input image is filtered by a set of low-pass filters and fan filters, after combination this leads to 30 filtered versions with five spatial frequency bands and six orientations. The next step in the model is the calculation of local contrast.

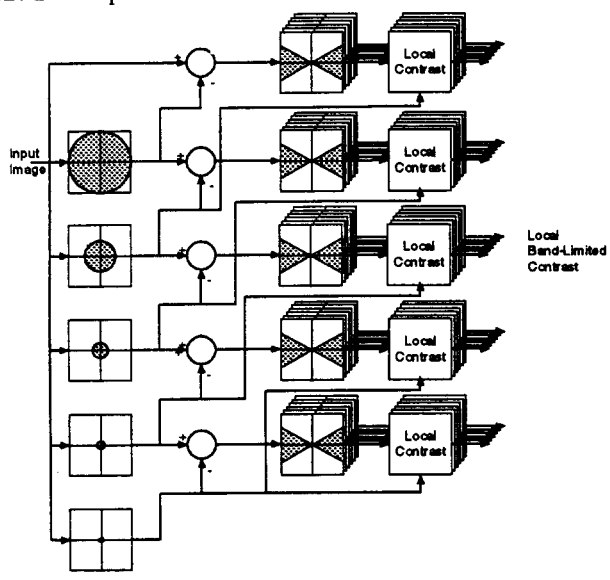

Figure 2: The calculation of LBC.

2.1.1. Low-pass filtering

Evidence from grating and other experiments suggests that the HVS contains band-pass filters with a bandwidth of 1 octave [7]. The original and the distorted image are filtered by a set of low-pass filters with a bandwidth that decreases with a factor 2 for each filter:

$$L_k(x,y)=f_{\mathrm{LOW},k}(x,y)*I(x,y) \qquad k=1..K, \tag{1}$$

where $I(x,y)$ is the luminance distribution of the image, $f_{\mathrm{LOW},k}(x,y)$ is the low-pass filter with index k and $L_k(x,y)$ is the low-pass filtered version of the image with index k. The low-pass filters were designed using the method of Speake and Mersereau [10]. The band-pass filtered versions of the image that result from taking differences of low-pass filtered versions of the image have a bandwidth of 1 octave.

2.1.2. Fan filtering

The next step in the model is the application of orientation sensitive fan filters to band-pass filtered versions of the image:

$$B_{k,l}(x,y)=\begin{cases} f_{FAN,l}(x,y)*\{L_{k+1}(x,y)-L_k(x,y)\} & \forall\ k=1..K-1, l=1..L \\ f_{FAN,l}(x,y)*\{I(x,y)-L_k(x,y)\} & \forall k=K, l=1..L \end{cases} \tag{2}$$

where $B_{k,l}(x,y)$ is the oriented band-filtered version of the image and $f_{FAN,l}(x,y)$ is the fan filter. Experiments give different estimates for the orientation bandwidth depending on the type of experiment that is performed. We choose an orientation bandwidth of 30 degrees for the fan filter and a set of six fan filters for each band-filtered version of the image. The fan filters were designed using the method of Antoniou and Lu [11].

2.1.3. Local contrast calculation

Parameters of visual stimuli in most visual experiments are expressed in terms of contrast. As we wish to incorporate the results from these experiments into our model, we have to convert the frequency bands $B_{k,l}(x,y)$ into some measure of contrast.

For simple stimuli that are symmetric relative to the background luminance, contrast is usually defined as Weber contrast [7]:

$$C_w=\frac{\Delta L}{L}, \tag{3}$$

where ΔL is the luminance difference relative to the background and L is the background luminance. An alternative definition that is often used is the Michelson contrast [7]:

$$C_m=\frac{L_{\max}-L_{\min}}{L_{\max}+L_{\min}}, \tag{4}$$

where $L_{\max}$ and L_{min} are the maximum and minimum luminance, respectively. In the case of our model these definitions cannot be used because a real image is not symmetric and these definitions are global quantities that depend on the average luminance of the entire image. We used a modified version of the definition by Peli [9], namely the ratio between the frequency band under consideration and a lowpass version of the image. Basically this means that the lowpass version is seen as the local image average for the frequency band considered. In this sense this definition is similar to the Weber definition:

$$LBC_{k,l}(x,y)=\begin{cases} A_{k,l}\cdot\dfrac{B_{k,l}(x,y)}{K_{k,l}+L_{k-1}(x,y)} & \forall\ k=2..K, l=1..L \\ A_{k,l}\cdot\dfrac{B_{k,l}(x,y)}{K_{k,l}+\bar{L}_1} & \forall\ k=1, l=1..L \end{cases} \tag{5}$$

Here $LBC_{k,l}(x,y)$ is the local band-limited contrast for frequency band k and orientation l, $\bar{L}_1$ is the average of the image and $A_{k,l}$ and $K_{k,l}$ are constants that can be used to model the frequency and orientation sensitivity of the HVS. By calculating the LBC in this way, we have arrived at a measure of local contrast that is normalized to visibility threshold and that depends on local image properties. The constants A and K give the flexibility to incorporate the luminance dependence of the frequency sensitivity into the model. For use of the model in a certain application, the constants depend on the expected luminance range of the display device, the number of vertical pixels in the image, and the viewing distance. In the experiments described below, we fitted the constants to the expression for the

frequency sensitivity by Barten [2] combined with an orientation sensitivity.
We have now arrived at the local band-limited contrast that is normalized to threshold. When the LBC is equal to one, the frequency component in that frequency band and orientation will be precisely at visibility threshold, taking into account light and frequency sensitivity.

2.2. Envelope detection

For the calculation of the masking at a certain position in the image in a certain frequency band, we use the envelope of the LBC. This leads to the same predictions for the experiments with gratings as the use of the LBC itself, but the advantage is that in the vicinity of edges maximum masking will be predicted at the exact position of the edge. The envelope of the LBC will have a local maximum at the exact position of edges and a masking function based on the envelope of the LBC will lead to a correct prediction of edge masking, as visual experiments indicate that maximum masking occurs at the exact positions of edges [8].

2.3. Threshold Elevation

The masking in the frequency bands is modeled by the threshold elevation function. The threshold elevation function describes how much the threshold for a test stimulus is increased by a masking stimulus. In the case of our model the test stimulus is the distortion that is normalized to visibility threshold by taking the LBC difference between the original and the distorted image, and the masking stimulus is the image itself. The threshold elevation function is based on the minimum of the envelope of the LBC of the original and the distorted image. The question could be asked whether it would be sufficient to derive the masking function from the LBC of either the original or the distorted image, but Daly showed that such an asymmetric model leads to incorrect predictions [6].
Our model is entirely symmetric, and the interchange of the original and the distorted image leads to the same perceptual image quality. The threshold elevation function turns out to be very simple, because it is approximately the same for all frequency bands, orientations and luminance levels if we use the LBC for the calculation of masking [12]. The threshold elevation function is shown in figure 3 and was taken from Daly [6]. When the envelope of the LBC is below 1 (below threshold) no masking will occur (the threshold elevation is equal to 1), and when the envelope of the LBC increases the amount of masking increases. The masked LBC difference $\Delta MLBC_{k,l}(x,y)$ is calculated as a function of the LBC of the original image $LBC_{k,l}(x,y)$, the LBC of the distorted image $LBC^{*}_{k,l}(x,y)$ and the threshold elevation function $TE_{k,l}(x,y)$:

$$\Delta MLBC_{k,l}(x,y) = \frac{LBC_{k,l}(x,y) - LBC^{*}_{k,l}(x,y)}{TE_{k,l}(x,y)} \qquad (6)$$

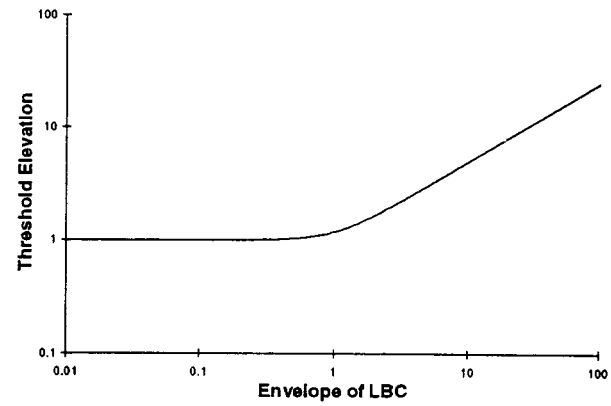

Figure 3: The threshold elevation as a function of the envelope of the LBC.

2.4. Response combination

The last part of the model is the combination of the masked LBC differences in the different frequency bands, orientations, and positions into a measure of perceptual image quality. This is the least understood part of the visual system because it involves more abstract processes. We propose a simple calculation that is based upon a vector norm over frequency bands and positions. The resulting perceptual error measure (PEM) is equal to:

$$PEM = \left(\sum_{x,y} \left| \sum_{k,l} \left| \Delta MLBC_{k,l}(x,y) \right|^{\alpha} \right|^{\beta} \right)^{\gamma} \qquad (7)$$

where α, β, and γ are constants that influence how the responses in different frequency bands, orientations and positions combine into the perceptual error measure. The exponents were determined by means of a number of experiments that are described in the next section.
Another possibility is to combine the responses at each position, which leads to an image with values that represent a local visibility of distortions. In coding applications such a local measure of image quality is probably more useful than a global one.

3. EXPERIMENTS

In order to estimate the parameters of the response combination as described in the last section and to evaluate the performance of the model in comparison with conventional measures of image quality (SNR) we performed experiments with a test panel.

3.1. Experimental conditions

The images in the experiments were the Build, Kiel, Collet, Clown, Karn and Mobile images. The images had a size of 512x512 pixels and were coded using PCM, DPCM, DCT and SBC coding schemes at several different bit rates. This resulted in 108 different images. The test panel consisted of 5 experts and 2 non-experts in image coding. The subjects were asked to give a number between 1 and 10 for the perceptual image quality of the images that were shown, a 1 for a very poor

quality and a 10 for a very good quality. Each image was shown 4 times to each subject. The images Karn and Mobile were only shown to 4 subjects. The viewing distance was 6 times the screen height.

3.2. Experimental results

The parameters A and K of the HVS model were fit to the experimental conditions (viewing distance and display luminance range). For every image we calculated the perceptual error measure with different settings for α, β, and γ. The experiments showed that the largest correlation between the perceptual error measure and the judgments could be obtained by using α=1.5, β=1.0, and γ=0.33. The results for this combination of parameters is shown in figure 5. The results for the conventional Peak to Peak Signal to Noise Ratio (PSNR) are shown in figure 4. The use of the new measure leads to an improvement in the correlation coefficient from 0.78 to -0.84 between the measure of image quality and the judgments when compared with the PSNR measure.

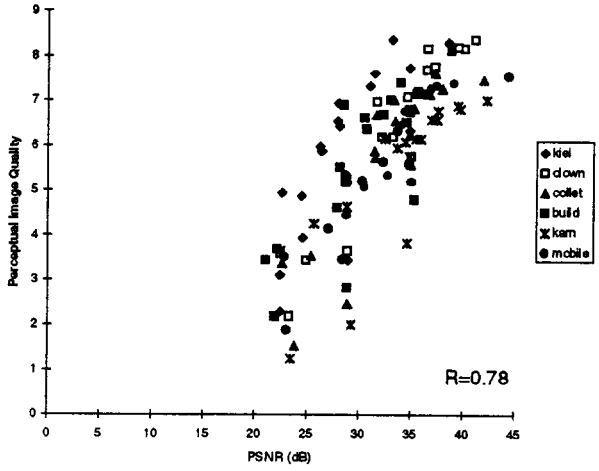

Figure 4: The measured perceptual image quality plotted against the PSNR.

4. Conclusion

The tests that were performed to evaluate the quality of the model only give a first indication of the performance of the model. The real test for the model will be the incorporation of the model into a coding scheme and measuring the number of bits that can be saved to achieve perceptual transparency (no visible differences with the original) compared to the same coding scheme optimized with a conventional measure. Although the response combination in the model is too simple to describe the response combination in the HVS, in our opinion the model is still very useful for image compression, even without response combination. Further research is needed for a better modeling of the response combination and the incorporation of color and motion into the model for use in video compression.

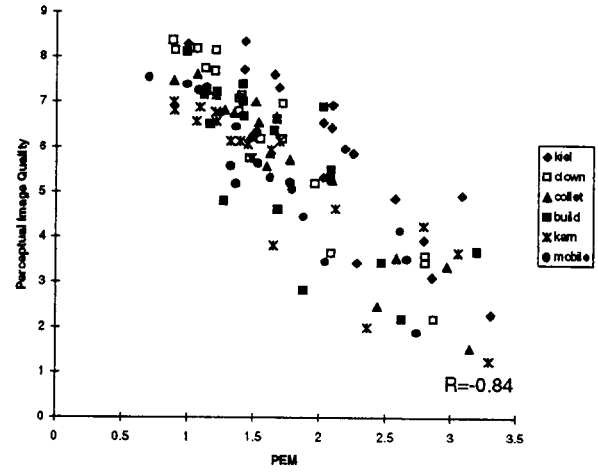

Figure 5: The measured perceptual image quality plotted against the new perceptual image quality measure.

References

[1] N.J. Jayant, J. Johnston and R. Safranek, "Signal Compression Based on Models of Human Perception", *Proc. of the IEEE*, vol. 81, pp. 1385-1422, 1993.

[2] P.G.J. Barten, "Evaluation of subjective image quality with the square-root integral method", *J.O.S.A. A*, Vol. 7, pp. 2024-2031, 1990.

[3] A.B. Watson, "Efficiency of a model human image code", *J. O.S.A. A*, vol. 4, pp. 2401-2417, 1987.

[4] C. Zetzsche and G. Hauske, "Multiple channel model for the prediction of subjective image quality", *SPIE conference on Human Vision, Visual Processing, and Digital Display*, vol. 1077, pp. 209-216, 1989.

[5] J. Lubin, "The Use of Psychophysical Data and Models in the Analysis of Display System Performance", In: *Digital Images and Human Vision*, A.B. Watson, editor, MIT Press, Cambridge, Massachusetts, 1993.

[6] S. Daly, "The Visible Differences Predictor: An algorithm for the Assessment of Image Fidelity", In: *Digital Images and Human Vision*, A.B. Watson, editor, MIT Press, Cambridge, Massachusetts, 1993.

[7] L.A. Olzak and J.P. Thomas, "Seeing spatial patterns", In: *Handbook of perception and Human Performance*, K. Boff, L. Kaufman and J. Thomas, editors, Wiley, New York, 1986.

[8] A. Vassilev, "Contrast sensitivity near borders: Significance of test stimulus form, size, and duration", *Vision Research*, vol. 13, pp. 719-730, 1973.

[9] E. Peli, "Contrast in Complex Images", *J.O.S.A. A*, Vol 7, pp. 2032-2040, 1990.

[10] T.C. Speake and R.M. Mersereau, "A Note on the Use of Windows for Two-Dimensional FIR Filter Design", *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 125-127, 1981.

[11] A. Antoniou and W.S. Lu, "Design of 2-D Nonrecursive Filters Using Window Method", *IEE Proc.*, vol. 137, pp. 247-250, 1990.

[12] A. Bradley and I. Ohzawa, "A Comparison of Contrast Detection and Discrimination", *Vision Research*, Vol. 26, pp. 991-997, 1986.