A Simplified Human Vision Model Applied to a Blocking Artifact Metric

Hantao Liu[1] and Ingrid Heynderickx[1,2]

[1] Department of Mediamatics, Delft University of Technology, P.O. Box 5031, 2628 CD, Delft, The Netherlands
Hantao.Liu@tudelft.nl
[2] Group Visual Experiences, Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands
Ingrid.Heynderickx@philips.com

Abstract. A novel approach towards a simplified, though still reliable human vision model based on the spatial masking properties of the human visual system (HVS) is presented. The model contains two basic characteristics of the HVS, namely texture masking and luminance masking. These masking effects are implemented as simple spatial filtering followed by a weighting function, and are efficiently combined into a single visibility coefficient. This HVS model is applied to a blockiness metric by using its output to scale the block-edge strength. To validate the proposed model, its performance in the blockiness metric is determined by comparing it to the same blockiness metric having different HVS-based models embedded. The results show that the proposed model is indeed simple, without compromising its accuracy.

Keywords: Human vision model, image quality assessment, luminance masking, texture masking, blockiness metric.

1 Introduction

During the last decades a lot of research effort was devoted to the development of objective image quality metrics, which nowadays are widely used in a broad range of image rendering applications, such as for the optimization of video coding or for real-time quality monitoring in displays. In the video chain of a current TV-set e.g., various objective quality metrics, which determine the quality of the incoming video signal in terms of blockiness, noise, blur, etc. and adapt the parameters in the video processing algorithms accordingly, are implemented to enable an improved overall perceived quality for the viewer. To assure that they predict *perceived* quality, objective metrics based on models of the human visual system (HVS) are potentially more reliable for accurate quality prediction [1]. Indeed, including in an objective metric stimulus aspects important to the human eye, while removing perceptual redundancies inherent in metrics purely signal based has been proved to enhance the performance of a metric [1-2].

Advances in human vision research provided crucial information on the structural and functional mechanisms of the HVS [2], which has been primarily adopted to

W.G. Kropatsch, M. Kampel, and A. Hanbury (Eds.): CAIP 2007, LNCS 4673, pp. 334 – 341, 2007.

design a variety of computational vision models in the literature [1-4]. The essential task of modeling the HVS is to quantitatively simulate its operations, which generally involves some lower level processing (e.g. sensitivity and masking) and some higher level processing (e.g. attention) in the visual system [2], as well as to restrictively incorporate them in a vision model [1]. However, as the HVS is extremely complex, HVS based objective metrics often are computationally intensive. Hence, from a practical point of view, it is desirable to reduce the computational complexity of the HVS model without significantly compromising its performance.

Much work has been done trying to incorporate HVS properties into quality metrics [4-7]. In some research parametric vision models including certain HVS aspects, were constructed. The parameters in these models were defined based on the results of a number of psychovisual experiments [6-7]. As a consequence, the accuracy of these models largely depends on the parameter selection, and their robustness cannot be fully ensured. In other research just-noticeable-distortion (JND) profiles, which provide each stimulus being tested with a visibility threshold of the distortion [8-9], are used. In these models, the thresholds for various masking effects are different, which potentially introduces difficulties in combining different masking effects. Instead of only estimating the threshold, the HVS model used in [5] is formulated as a weighting function. The main drawback in this approach, however, is that only one weighting function, intrinsically combining luminance and texture masking, is taken into account, and that efficient integration of different masking effects is not considered. In our paper, we further rely on the approach taken in [5] by using a HVS model as a weighting function for visibility, but extend the idea by including both luminance and texture masking, and by combining both masking effects in a simple way into a single visibility coefficient.

To evaluate this approach, we used the model in a blockiness metric comparable to [5]. The blocking artifact, which manifests itself as an artificial discontinuity between adjacent blocks, is known as the major type of distortion in block-based DCT coding. It is checked whether the simple HVS model helps to quantify the visibility of blocking artifacts in grey-scale images.

2 The Simplified Human Vision Model

The human vision model described in this paper adopts two fundamental properties of the HVS, which affect the visibility of a stimulus in the spatial domain: (1) the averaged background luminance surrounding the stimulus, and (2) the spatial non-uniformity in the background luminance [9]. They are well known as luminance masking and texture masking, respectively. Masking is the reduction in the visibility of one image component (the target) due to the presence of another (the masker), and it is strongest when both components have the same or similar frequency, orientation, and location [10]. In our proposed HVS model both spatial masking effects are estimated in a simple way, and efficiently combined into a single visibility coefficient. The approach is illustrated in Figure 1. A window, representing the local surrounding of a stimulus (i.e. in our case a – possibly deviating – pixel value, e.g. a blocking edge), is scanned over all stimuli (i.e. in our case over all pixels in an image). Both

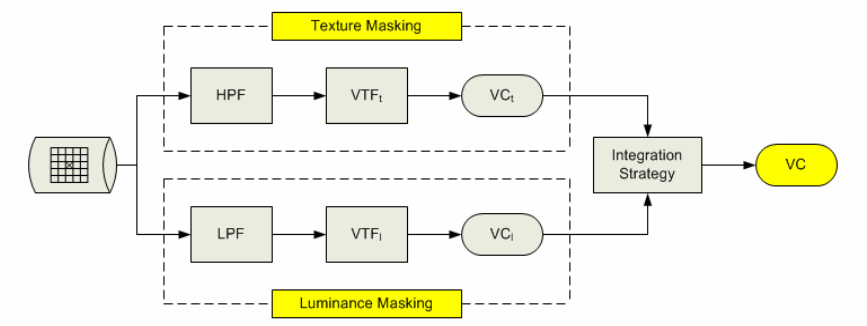

Fig. 1. Schematic overview of the proposed human vision model

masking effects are estimated by analyzing the local signal properties within this window. Based on the results a visibility coefficient (VC), which reflects the perceptual significance of the stimulus, is defined.

2.1 Local Visibility Due to Texture Masking

Texture masking is modeled calculating a visibility coefficient (VC_t). The higher the value of this coefficient, the smaller the masking effect, and hence, the stronger the visibility of the stimulus is. The procedure of modeling texture masking comprises three steps:

- Texture Detection: calculate the local background activity (non-uniformity).
- Thresholding: a classification scheme to capture the active background regions.
- Visibility Transform Function (VTF): obtain a visibility coefficient (VC_t) based on the HVS characteristics for texture masking.

1	2	0	-2	-1
4	8	0	-8	-4
6	12	0	-12	-6
4	8	0	-8	-4
1	2	0	-2	-1

T1

1	4	6	4	1
2	8	12	8	2
0	0	0	0	0
-2	-8	-12	-8	-2
-1	-4	-6	-4	-1

T2

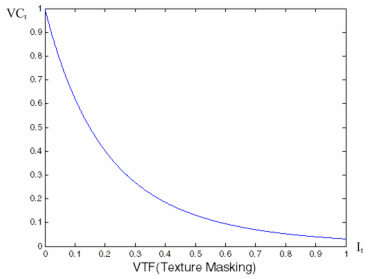

Fig. 2. The high-pass filters for texture detection, and Visibility Transform Function used

Texture detection can be performed convolving the signal with some form of high-pass filter. One of the Laws' texture energy filters [11] is employed here in a slightly modified form. As shown in Figure 2, $T1$ and $T2$ are used to measure the background activity in horizontal and vertical direction, respectively. A pre-defined

threshold Thr ($Thr = 0.15$ in our experiments) is applied to classify the background into 'flat' or 'texture', resulting in an activity value $I_t(i, j)$, which is given by

$$I_t(i,j) = \begin{cases} 0 & if\ t(i,j) < Thr \\ t(i,j) & otherwise \end{cases} \tag{1}$$

$$t(i,j) = \frac{1}{48} \sum_{x=1}^{5} \sum_{y=1}^{5} I(i-3+x, j-3+y) \cdot T(x,y) \tag{2}$$

where $I(i, j)$ denotes the pixel intensity at location (i, j), and T is chosen as $T1$ for texture calculation in horizontal direction, and $T2$ for vertical direction. It should be noted that splitting up the calculation in horizontal and vertical direction, and using a modified version of the texture energy filter, in which some template coefficients are removed, is done with the application of a blockiness metric in mind.

A visibility transform function (VTF) is proposed in accordance to human perceptual properties, which means that the visibility coefficient $VC_t(i, j)$ is inversely proportional (nonlinear) to the activity value $I_t(i, j)$. Figure 2 shows an example of such a transform function, which can be defined as

$$VC_t(i,j) = \frac{1}{(1 + I_t(i,j))^{\alpha}} \tag{3}$$

where $VC_t(i, j) = 1$, when the stimulus is in a 'flat' background, and $\alpha > 1$ ($\alpha = 5$ in our experiments) is used to adjust the nonlinearity. This shape of the VTF is an approximation, considered to be good enough.

2.2 Local Visibility Due to Luminance Masking

In psychovisual experiments it was found that the human visual system's sensitivity to variations in luminance depends on (is a nonlinear function of) the local mean luminance [10]. In this paper, modeling the luminance masking is based on two empirically driven properties of HVS: (1) a distortion in a dark surrounding tends to be less visible than one in a bright surrounding [9], and (2) a distortion is most visible for a surrounding with an averaged luminance value between 70 and 90 (centered approximately at 81) in 8bits gray-scale images [5]. The procedure of modeling luminance masking consists of two steps:

- Local Luminance Detection: calculate the local averaged background luminance.
- Visibility Transform Function (VTF): obtain a visibility coefficient (VC_l) based on the HVS characteristics for luminance masking.

The local luminance of a certain stimulus is calculated using a weighted low-pass filter as shown in Figure 3, in which some template coefficients are set to '0'. The local luminance $I_l(i, j)$ is given by

$$I_l(i,j) = \frac{1}{26}\sum_{x=1}^{5}\sum_{y=1}^{5} I(i-3+x, j-3+y)\cdot L(x,y) \tag{4}$$

where L is chosen as $L1$ for calculating the background luminance in horizontal direction, and $L2$ for the vertical direction. Again, splitting up the calculation in horizontal and vertical direction, and using a modified low-pass filter, in which some template coefficients are set to 0, is done with the application of a blockiness metric in mind.

1	1	0	1	1
1	2	0	2	1
1	2	0	2	1
1	2	0	2	1
1	1	0	1	1

L1

1	1	1	1	1
1	2	2	2	1
0	0	0	0	0
1	2	2	2	1
1	1	1	1	1

L2

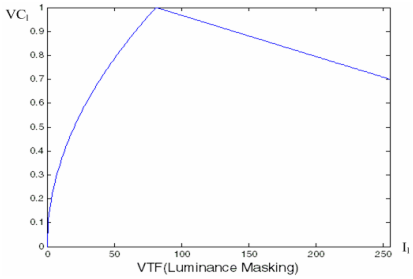

Fig. 3. The low-pass filters for local luminance detection, and Visibility Transform Function used

For simplicity, the relationship between the visibility coefficient $VC_l(i,j)$ and the local luminance $I_l(i,j)$ is modeled by a power law for low background luminance (i.e. below 81), and is approximated by a linear function at higher background luminance (i.e. above 81). This functional behavior is shown in Figure 3, and mathematically described as

$$VC_l(i,j) = \begin{cases} \left(\dfrac{I_l(i,j)}{81}\right)^{1/2} & \text{if } 0 \le I_l(i,j) \le 81 \\ \left(\dfrac{1-\beta}{174}\right)\cdot(81 - I_l(i,j)) + 1 & \text{otherwise} \end{cases} \tag{5}$$

where $VC_l(i,j)$ achieves the highest value of 1 when $I_l(i,j) = 81$, and $0 < \beta < 1$ ($\beta = 0.7$ in our experiments) is used to adjust the slope of the linear part of this function.

2.3 Integration Strategy

The visibility of a stimulus depends on various masking effects co-existing in the HVS, and how to efficiently integrate them is an important issue in obtaining an accurate perceptual model [8]. Since spatial masking intrinsically is a local phenomenon, the locality in the visibility of a distortion due to masking is maintained in the integration strategy of both masking effects. The resulting approach is schematically given in Figure 4. Based on the local image content surrounding a stimulus first the texture

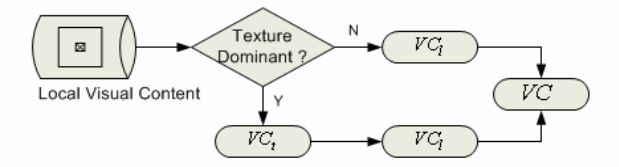

Fig. 4. Integration strategy of the texture and luminance masking effects

masking is calculated. In case the local activity in the area is larger than a given threshold (see equation (1)), a visibility coefficient VC_t is applied, followed by the application of a luminance masking coefficient VC_l. In case the local activity in the area is low, only VC_l is applied.

3 Blockiness Metric Using Proposed Model

Given a DCT-coded image, the block-edge strength (*BS*) can be defined as the inter-pixel difference across block boundaries (e.g. $BS_h(i,j) = |I(i,j) - I(i,j+1)|$ is defined as the inter-pixel difference across horizontal block boundaries, where (i,j) denotes the pixel location) [5]. The output of the proposed human vision model *VC* can be used to locally weight the *BS* to produce a visual blocking strength (*VBS*), which is given by

$$VBS(i,j) = VC(i,j) \times BS(i,j) \tag{6}$$

The *VBS* can be easily implemented in a generalized block-edge impairment metric [5], which is formulated as

$$Metric = \frac{\| \; MO(i,j) \times BS(i,j) \; \|}{\| \; MO(i,j) \times NBS(i,j) \; \|} \tag{7}$$

where $\| \, . \, \|$ is the L2-Norm, and *NBS* is defined as the inter-pixel difference between pixels, which are not at block boundaries [5]. *MO* is used to indicate the output of any HVS model (in our case *VC*). The horizontal and vertical blocking artifacts can be calculated separately using the appropriate filters for *VC*, and then added together to give the resultant blockiness score, i.e. Metric = Metric(h) + Metric(v).

4 Performance Evaluation

The proposed human vision model is validated by its application to an objective blockiness metric. In order to analyze the model contribution rather than the performance of various blockiness metrics, a comparative evaluation is necessarily conducted by embedding different human vision models to the same blockiness metric. Based on the generalized blockiness metric defined in (7), three options are implemented for *MO:* (1) our proposed model, (2) the model used in [5], and (3) the JND profile used in [9]. This results in three blockiness metrics, which we refer to as VBSM, GBIM

and JNDM, respectively. The LIVE database [13], which consists of 233 JPEG images with their subjective Mean Opinion Score (MOS), is used to test the performance of these blockiness metrics. According to the Video Quality Expert Group (VQEG) [12], the performance of the objective metrics can be quantitatively measured by the Pearson linear correlation coefficient and the Spearman rank order correlation coefficient between subjective MOS and objective ratings after nonlinear regression.

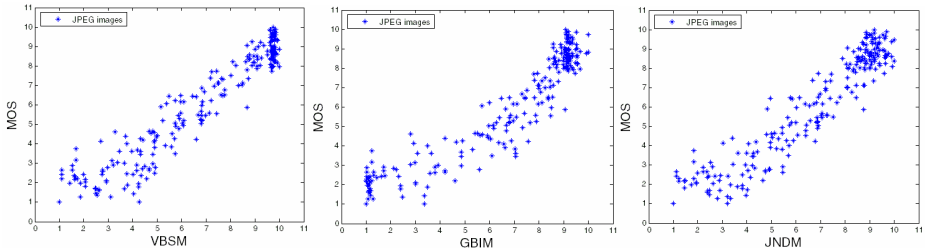

Fig. 5. Scatter plots of MOS vs. VBSM, GBIM and JNDM

Table 1. Performance comparison of three objective metrics for image quality assessment

Metric	Pearson Linear Correlation	Spearman Rank Order Correlation
VBSM	0.9517	0.9251
GBIM	0.9280	0.9116
JNDM	0.9401	0.9176

Figure 5 shows the scatter plots of the MOS vs. VBSM, GBIM, and JNDM, respectively. The corresponding correlation coefficients are listed in Table 1. It is verified that a promising performance is achieved by applying our HVS in a blockiness assessment. In contrast to our model, the vision model used in GBIM [5] intrinsically combines luminance and texture masking into a single weighting function. Although this is statistically acceptable, it might degrade the model's performance in some demanding circumstances, for example when assessing highly textured images. This problem is solved in our model by separating the two masking effects, and by adaptively combining them based on local signal features. Therefore, our model is more reliable in terms of content independency. This is confirmed by repeating the correlation analysis on a limited set of 50 (out of 233) highly textured LIVE database images only. For these images the VBSM gives a Pearson correlation of 0.9391, whereas the GBIM results in a poorer correlation of 0.7695. Our model is comparable to the approach chosen in the JND profile [9] with the exception that the JND profile only considers a threshold, while our model also estimates supra-threshold visibility. This makes our model slightly more accurate and robust (for the limited dataset mentioned above the Pearson correlation for the JNDM is 0.9038). Our model also has the intrinsic advantage that knowledge on the nature of the artifact can simply be taken into account (e.g. by evaluating horizontal and vertical masking separately for blocking artifacts), which makes the model simple and efficient. This simplification is less

obvious in the more generally applicable JND profile model. Nonetheless, we expect also our model to be more generally applicable to the visibility of other artifacts (mainly by changing size and coefficients in the filters).

5 Conclusion

We have presented a simplified and more efficient human vision model based on estimating spatial masking effects of the HVS, such as luminance and texture masking. These masking effects were estimated using spatial filtering followed by a weighting function, and were efficiently combined into a single visibility coefficient. The application of this model in a blockiness assessment resulted in a strong correlation with subjective ratings. The proposed model is unsupervised and does not need to be trained with subjective data. It can be easily integrated into either full-reference or no-reference approaches for measuring blocking artifacts.

References

1 Winkler, S.: Issues in Vision Modeling for Perceptual Video Quality Assessment. Signal Processing 78(2), 231–252 (1999)
2 Osberger, W., Maeder, A.J., McLean, D.: A Computational Model of the Human Visual System for Image Quality Assessment. In: Proc. DICTA-97, pp. 337–342 (December 1997)
3 Yu, Z., Wu, H.R.: Human Visual System Based Objective Digital Video Quality Metrics. In: Proc. Int. Conf. Signal Processing, vol. II, pp.1088–1095 (August 2000)
4 Yu, Z., Wu, H.R., Winkler, S., Chen, T.: Vision Model Based Impairment Metric to Evaluate Blocking Artifacts in Digital Video. In: Proc. of the IEEE, pp. 154–169 (January 2002)
5 Wu, H.R, Yuen, M.: A Generalized Block-edge Impairment Metric for Video Coding. IEEE Signal Processing Letters 70(3), 247–278 (1998)
6 Yeh, E.M., Kokaram, A.C., Kingsburg, N.G.: A Perceptual Distortion Measure for Edge-Like Artifacts in Image Sequences. Human Vision and Electronic Imaging III, pp. 160-172, SPIE (1998)
7 Karunasekera, S.A., Kingsbury, N.G.: A Distortion Measure for Blocking Artifacts in Images Based on Human Visual Sensitivity. IEEE Trans. Image Processing (1995)
8 Yang, X., Lin, W., Lu, Z., Ong, E., Yao, S.: Motion-Compensated Residue Preprocessing in Video Coding Based on Just-Noticeable-Distortion Profile. IEEE Trans. on Circuits and Systems for Video Technology 15(6), 742–751 (2005)
9 Chou, C.H., Li, Y.C.: A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion profile. IEEE Trans. on Circuits and Systems for Video Technology (December 1995)
10 Pappas, T.N., Safranek, R.J.: Perceptual criteria for image quality evaluation. In: Bovik, A.C. (ed.) Handbook of Image and Video Processing, Academic Press, San Diego (May 2000)
11 Laws, K.I.: Texture Energy Measures. In: Proc. DARPA Image Understanding Workshop, Los Angeles, pp. 47–51 (1979)
12 VQEG: Final report from the video quality experts group on the validation of objective models of video quality assessment (2003) http://www.vqeg.org/, Aug
13 Sheikh, H. R., Wang, Z., Cormack, L., Bovik, A. C.: LIVE image quality assessment database. http://live.ece.utexas.edu/research/quality