PERCEPTUAL QUALITY METRICS: EVALUATION OF INDIVIDUAL COMPONENTS

B. Fontaine, A. Saadane, A. Thomas

Institut de Recherche en Communications et Cybernétique de Nantes
Ecole polytechnique de l'université de Nantes
rue Christian Pauc BP 50609, 44300 Nantes France
Email : abdelhakim.saadane@polytech.univ-nantes.fr

ABSTRACT

Based on the numerous quality metrics proposed in the literature, an evaluation of the individual visual model components have been performed. Effort has been focused on the perceptual decomposition and spatial pooling. Four perceptual decompositions have been considered. Three of them include both radial and orientation selectivity while for the fourth one, only a radial decomposition is performed. Five spatial pooling models have been associated to each of these decompositions. The first model is the most used one and consists in a simple Minkowski summation. In the second one, the Minkowski summation is weighted by the occurrence probability of the perceived errors contrast. The last three models are new and include higher level perceptual processes.

1. INTRODUCTION

Visual quality metrics are useful to evaluate image processing algorithms, particularly image compression schemes designed to be perceptually lossless. Currently, their development represents an area of intensive research and several distortion metrics have been proposed. The first metrics were based on simple error measures like Peak signal to noise ratio or mean square error. Compared to the subjective methods, these metrics are neither expensive nor time consuming. However, such measures correlate poorly with perceived quality. To improve performances, most of the recent and numerous metrics are based on models of low level processing of the human visual system and use the same general layout [1, 2, 3]. However, due to the large variety of the vision models, these metrics vary in complexity and present various performances which it is difficult to compare. Moreover, the conducted comparative studies [4] have focused on global performances and thus, are inappropriate to identify new promising research directions.

According to [5], one way to achieve further improvements in the design of reliable visual quality metrics is to evaluate individual visual model components. We think that a second way is to take into account higher level perceptual processes. The purpose of this paper is to combine these two ways by evaluating four perceptual decompositions and five spatial pooling models. The first three decompositions include both radial and orientation selectivity while the fourth one analyses the visual input by only a set of radial channels. The first two spatial pooling models are based on the most used Minkowski summation while the set of the three remaining ones include attentional mechanisms of the human visual system.

The paper is structured as follows: section 2 presents the developed metric which has been used for the components evaluation. Sections 3 and 4 describe the perceptual decompositions and the spatial pooling models respectively. Results and discussions are given section 5. Finally , conclusion and further work under development are outlined.

2. PROPOSED METRIC

Figure 1 gives the general scheme of the developed image quality metric. Original and impaired images are preprocessed by the first block. This latter includes first the gamma correction, in order to express grey levels into luminance signal, and secondly a non linear transform which converts luminance into perceived luminance as visual sensitivity and perception of lightness are logarithmic functions of luminance [6]. The second block models the spatial frequency selectivity of the human visual system and performs a perceptual decomposition on both the original and impaired images. The cortex filters [7] have been used to achieve the four decompositions discussed in section 3. In the third block, for each filtered channel (i,j) the local band limited contrast $c_{i,j}(m,n)$ is computed at each location (m,n) according to the definition of Peli [8]:

0-7803-8554-3/04/$20.00 ©2004 IEEE.

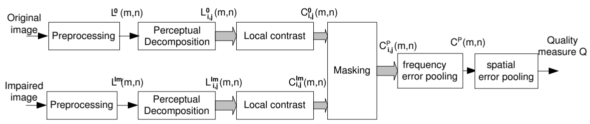

Figure 1: Block diagram of the developed metric

$$c_{i,j}(m,n) = \frac{L_{i,j}(m,n)}{\sum_{k=0}^{i-1} \sum_{l=0}^{card(l)-1} L_{i,j}(m,n)}$$

where $L_{i,j}(m,n)$ represents the luminance. The indexes i and k are relative to radial frequency domain while j and l specify the orientation of the channel in the radial band. card(l) is the number of angular channels in the k^{th} radial band. In the masking block, a masking model is applied to remove all errors which are below their visibility threshold. The accepted model of contrast masking [9] estimates the degree to which a sinusoidal target is masked by the presence of sinusoidal maskers. The complexity of the real images requires an adaptation of this model and constraints the authors to approximations. To avoid such adaptations, the model considered here is derived from psychophysical experiments conducted on natural textured images. For this model, the perceived contrast values $c^P_{i,j}(m,n)$ of output errors are computed according to

$$c^P_{i,j}(m,n) = \begin{cases} \dfrac{c^{err}_{i,j}(m,n) - \Delta c^m_{i,j}}{\Delta c^0_{i,j}} & if \quad c^{err}_{i,j} \geq \Delta c^m_{i,j} \\ 0 & elswhere \end{cases}$$

where $\Delta c^0_{i,j}$ is a normalisation coefficient and $\Delta c^m_{i,j}$ represents the perception threshold when masking is considered. More details are given in [10].

Once perceptual errors are defined for each channel and each location, errors have to be pooled. The frequency error pooling block sums errors across frequency bands to obtain a 2D perceptual error map. Two steps are needed for this pooling. In the first one, the pooling over angular channels is computed by the MAX operator:

$$c^P_i(m,n) = \max_j \left(c^P_{i,j}(m,n) \right)$$

In the second step, the pooling over radial channel is performed by a weighted linear summation:

$$c^P(m,n) = \sum_{i=1}^{4} \alpha_i \, c^P_i(m,n)$$

Finally, the spatial error pooling block sums contrasts across space to produce the quality measure Q of the impaired image. The models used are presented section 4.

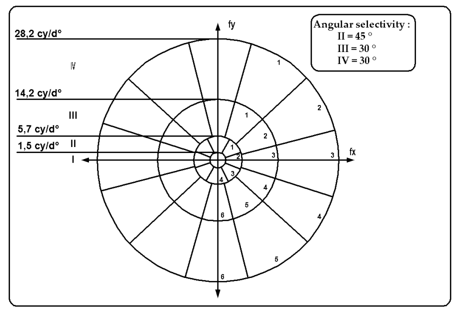

Figure 2: Decomposition of the visual spatial frequency domain into perceptual channels.

3. PERCEPTUAL DECOMPOSITIONS

Four perceptual decompositions have been analysed and compared. The first one, called D1 and given figure 2, is based on a large synthesis of the literature and has been completed by several experimental studies [11]. For this decomposition, three band-pass radial frequency channels are needed (corona II, III and IV), each being decomposed into angular sectors with an orientation selectivity of 45°, 30° and 30° respectively. Corona I is a non directional low-pass radial frequency channel. While bandwidths of this decomposition have been used as they have been measured, the bandwidths of the second and third decompositions, called D2 and D3, have been intentionally constrained by their authors for computational reasons and coding results optimisation. These two last decompositions are well known and have been proposed by Watson [7] and Daly [6] respectively. Both of them use a radial frequency selectivity that is symmetric on a log frequency axis with bandwidths nearly constant at one octave. They consist of one isotropic low-pass and three band-pass channels. The angular selectivity is constant and is equal to 45 degrees for Watson and 30 degrees for Daly. The fourth and last decomposition, D4, uses four radial bands and ignores the orientation sensitivity of the HVS. Results given in [12] show that ignoring angular selectivity has very little effect on visual quality. Hence, computational cost is significantly reduced by avoiding angular filtering and angular frequency pooling.

4. SPATIAL POOLING MODELS

Several approaches have been proposed to reduce a 2D quality map to a single number that reflects the overall image quality. The most well used ones perform a non linear summation of errors over an entire image and across all frequency channels. These latter provide accurate predictions of quality as long as the errors are near their visibility thresholds. When errors are above the visibility thresholds, simple early vision models are not powerful enough and should be completed by the higher level processes. Based on these observations we have compared five spatial pooling models. The first one , called M1, is the most used Minkowski summation given by:

$$Q = \left[\frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(c^P(m,n) \right)^{\beta} \right]^{\frac{1}{\beta}}$$

In the second model, M2, the Minkowski summation is weighted by the occurrence probability of the contrast errors:

$$Q = \left[\sum_{m=1}^{M} \sum_{n=1}^{N} \left(c^P(m,n) \right)^{\beta} \left(\Pr\left(c^P(m,n) \right) \right)^{\gamma} \right]^{\frac{1}{\beta\gamma}}$$

The last three models take into account human visual attention to identify visually salient regions within a scene. According to [13], this can be done through the use of importance maps. These maps identify regions of perceptual interest (ROPI) by using 5 factors known to influence visual attention: contrast, size, shape, location, and foreground/background. The factors are weighted equally and are summed to produce the importance map for the image. Each region in this map is assigned an importance value in the range [0,1], with 1 representing highest importance.

The third spatial pooling model, called M3, uses the ROPI to weight the errors contrast and is given by:

$$Q = \left[\frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} c^P(m,n) ROPI(m,n) \right]$$

In the fourth model, called M4, the Minkowski summation is weighted by the ROPI as follows:

$$Q = \left[\frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(c^P(m,n) ROPI(m,n) \right)^{\beta} \right]^{\frac{1}{\beta}}$$

The last model, M5, is the same as M4 except that different exponents have been used for the ROPI and the errors contrast:

$$Q = \left[\frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} \left(c^P(m,n) \right)^{\beta} \left(ROPI(m,n) \right)^{\gamma} \right]^{\frac{1}{\beta\gamma}}$$

5. RESULTS

5.1. Database

A set of 6 well known images has been used. For each original image of size 512*512 pixels, two kinds of degradations have been generated: blocking errors and blurring due to quantization noise. In the first case, 5 different JPEG quality factors have been used to cover the whole quality scale (MOS varying between 1 and 5) defined by the ITU-R. Thus 30 JPEG images are obtained. In the second case, 15 quantized images, also covering the whole quality scale, have been generated leading to a total of 90 images. A description of the quantization process can be found in [11]. The obtained database has been divided into two subsets. The data of the first one have served as training data while the second subset has been used as the test database.

5.2. Parameters determination

Quality assessment tests have been conducted and the correlation between the resulting scores and the computed quality Q has been used to optimize the parameters of the different spatial pooling models. The results obtained with the training data are given table 1 for each perceptual decomposition.

	M1	M2	M4	M5
D1	β=1.3	β=1.1 γ=2.2	β=1.8	β=1.9 γ=1.6
D2	β=1.5	β=1.3 γ=2.4	β=2.1	β=2.1 γ=1.8
D3	β=1.2	β=1 γ=2.3	β=1.7	β=2 γ=1.5
D4	β=2.3	β=1.6 γ=2.1	β=2.8	β=1.7 γ=2.5

Table 1: Best values for each model and each decomposition.

Compared to the model M1, the parameter β of the model M4 is larger. Knowing that both models use the Minkowski summation, this means that suprathreshold errors, when weighted by their importance, determine the overall image quality. The same observation can be made by comparing the models M1 and M5. On the other hand, the model M2 for which β is smaller, argues that the errors having low levels, if they are more frequent, could more contribute to the quality when they are weighted by their occurrence probability.

5.3. Performances comparison

The parameters of table 1 have been used for each image of the test database. The correlation coefficients of the quantized and jpeg coded images are given tables 2 and 3 respectively. Let us start by comparing the spatial pooling models.

On one hand, by looking at the models M1 and M2, table 2 shows that the occurrence probability improves significantly the performances of model M2 for all the decompositions. On the other hand, by comparing the three last models including the regions of perceptual interest, one can see that the weighted Minkowski summation using different exponents also increases the correlation coefficients of the model M5. Now, if one compares the models M2 and M5, the model M5 remains slightly better since the correlation coefficients increase 3 times out of 4. So, one can highlights the role of the attentional mechanisms in the improved results. This observation however can be discussed if the decompositions are also compared. Indeed the best tradeoff between performance and complexity is given by the association of the model M2 and the decomposition D4. Recall that the latter, by performing only a radial decomposition, reduces the computational cost and avoids the angular pooling.

The results of table 3, in agreement with the made observations, show that the performances of the quality metric remain independent of the type of degradations.

	M1	M2	M3	M4	M5
D1	0.832	0.896	0.878	0.919	0.983
D2	0.854	0.906	0.868	0.906	0.909
D3	0.815	0.883	0.883	0.915	0.939
D4	0.885	0.969	0.824	0.939	0.955

Table 2: Correlation coefficients of the quantized images.

	M1	M2	M3	M4	M5
D1	0.701	0.727	0.779	0.798	0.717
D2	0.704	0.725	0.784	0.812	0.737
D3	0.664	0.71	0.777	0.795	0.775
D4	0.683	0.782	0.767	0.804	0.809

Table 3: Correlation coefficients of the jpeg coded images.

6. CONCLUSION

A perceptual metric for image quality assessment has been presented and investigations have been focused on the perceptual decomposition and spatial pooling. Four decompositions and five spatial pooling models have been evaluated by considering two kinds of distortions. The best correlation with subjective data are obtained with a spatial pooling model including higher level perceptual factors. However, the best tradeoff between performances and complexity is given by a model including the occurrence probability of the errors associated with a simple radial decomposition.

7. REFERENCES

[1] M.P.Eckert, A.P.Bradley, "Perceptual quality metrics applied to still image compression" *Signal Processing*, Vol. 70, pp. 177-200, 1998.

[2] P.C.Teo et D.J.Heeger. "Perceptual image distorsion », Proc. SPIE, Vol. 2179, pp. 127-141, 1994.

[3] S.Daly. *Digital images and human vision*, A.B. Watson editor, MIT Press, pp.179-206, 1993.

[4] VQEG: "final report from the Video Quality Expert Group on the validation of objective models of video quality assessment" see http://www.vqeg.org/

[5] S. Winkler: "Quality metric design: A closer look." in Proc. SPIE , vol. 3959, pp. 37--44, San Jose, CA, 2000.

[6] S.Daly, "A visual model of optimizing the design of image processing algorithm", Proc. ICIP, Vol. II of III, pp.16-20, 1994.

[7] A.B.Watson. "The cortex transform : rapid computation of simulated neural images", *Computer Vision, Graphics and image Processing*, N°.39, pp. 311-327, 1987.

[8] E.Peli, "Contrast in complex images", *JOSA*, A7, N°10, pp. 2032-2040, 1990.

[9] G.E. Legge et J.M. Foley, "Contrast masking in human vision", *JOSA*, Vol.70, pp. 1458-1471, 1980.

[10] A.Saadane, N.Bekkat and D.Barba, *Imaging and Vision Systems: Theory, Assessment and applications*, Nova science Publishers, Inc, Vol.9, 2001.

[11] A.Saadane, D.Barba, H.Senane, "An entirely psychovisual based subband image coding scheme" Proc. VCIP, Taiwan, pp. 702-712, 1995.

[12] T. Mitsa, K.L. Varkur, J.R. Alford, "Frequency channel based visual models as quantitative quality measures in halftoning" Proc. SPIE, Human vision Visual Processing and digital Display, Vol. 1913, pp.390-401, 1993.

[13] W. Osberger and A. Maeder, "Automatic identification of perceptually important regions in an image using a model of the human visual system." 14th International Conference on Pattern Recognition, Brisbane, Australia, Aug 1998.