ON THE CONTRIBUTION OF THE PERCEPTUAL DECOMPOSITION AND SPATIAL POOLING TO THE PERFORMANCES OF FIDELITY METRICS

B. Fontaine, A. Saadane

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: asaadane@polytech.univ-nantes.fr

ABSTRACT

Based on the several visual fidelity metrics proposed in the literature, we sought to determine a metric having low complexity and high performances. A full reference quality metric has been designed and effort has been focused on the perceptual decomposition and spatial pooling. Four perceptual decompositions have been considered. Three of them include both radial and angular sensitivities while the fourth one analyzes the visual input by only a set of radial channels. For each of these decompositions, the performances of two new spatial pooling models have been compared with the most used Minkowski summation.

1. INTRODUCTION

Image distortion measures are useful to evaluate image processing algorithms, particularly image compression schemes which are designed to be perceptually lossless. Two main ways can be used to obtain such measures. First, subjective assessment tests can be conducted in order to obtain subjective judgments or Mean Opinion Scores (MOS) which represent an accurate estimate of the image quality. These tests require appropriate material and accurate experimental protocols. The ITU recommendation 500 [1] gives specifications about observers, viewing distances, quality scales to be used, appropriate methods for a given task and so on.

The second way to assess the image quality is to define computational metrics to get objective measures. Many researchers frequently use Peak Signal to Noise Ratio (PSNR), or Mean Square Error (MSE). Compared to subjective methods, these metrics are neither expensive nor time consuming. However they do not fit well the MOS. To overcome such difficulties, several perceptual image quality metrics that incorporate the Human Visual System (HVS) properties have been proposed [2, 3, 4, 5, 6, 7]. For many of these metrics, common blocks can be identified. These are luminance conversion, decomposition into perceptual channels, contrast conversion, contrast masking and error pooling. For each block, however, the models used differ significantly from one author to another. Several investigations have already tried to compare different metrics. See, for example, the Video Quality Experts Group (VQEG) activities [8] which have evaluated the performances of different quality assessment systems. While these investigations have focused on the performances of entire systems, the improvements in the design of simple and efficient quality metrics need single blocks evaluation. This paper deals with such evaluations and is structured as follows: section 2 describes the proposed quality metric. In sections 3 and 4, four perceptual decompositions and three spatial pooling models are presented. Evaluation results and discussions are given in section 5. Finally, conclusion and future works are outlined.

2. PROPOSED METRIC

Figure 1 gives the general block diagram of the developed image quality metric.

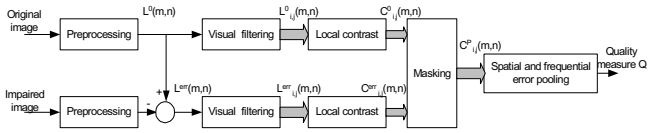

Figure 1: General block diagram of the proposed metric

- **Preprocessing**: this block includes a double non linear transformation. In the first one, grey levels are converted into luminance by considering the screen gamma function. The second non linear transform converts luminance into perceived luminance as the visual sensitivity and perception of lightness are logarithmic functions of luminance [9]. This second conversion is done by raising each objective luminance value normalized by the maximum screen luminance to a power of 0.33.
- **Visual filtering**: This block includes a number of operations that model the frequency selectivity of the HVS by a framework of sub-bands or channels. The decomposition used here is given figure 2. For this decomposition, three band-pass radial frequency channels are needed (numbers II, III and IV), each of them being decomposed into angular sectors associated with angular selectivity of 45, 30 and 30 degrees respectively. Channel I is a non directional low-pass channel. The cortex filters [10] have been used to achieve this decomposition.

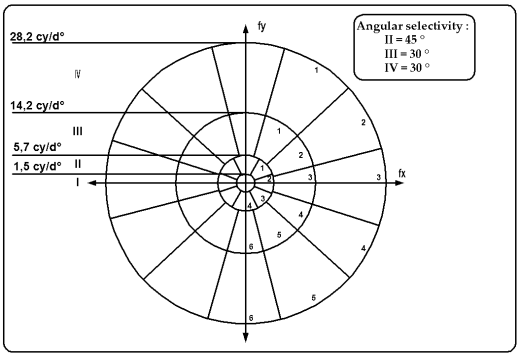

Figure 2: Decomposition of the visual spatial frequency domain into perceptual channels.

- **Local contrast**: 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 [11].
- **Masking**: The masking model is applied on contrast values $c_{i,j}^{err}(m,n)$ of the error images in order to remove all errors which are not perceptible. The perceived contrast values $c_{i,j}^{P}(m,n)$ of output errors are computed according to

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

where $\Delta c_{i,j}^{0}$ is a normalisation coefficient and $\Delta c_{i,j}^{m}$ represents the perception threshold when masking is considered. More details are given in [12].

- **Poolings**: Frequency pooling is performed first. It sums errors across frequency bands to obtain a 2D perceptual error map. This pooling needs two steps. The first one corresponds to the angular frequency pooling and is computed by

$$c_{i}^{P}(m,n)=\max_{j}\left(c_{i,j}^{P}(m,n)\right)$$

Then, the radial frequency pooling is performed as follows:

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

From the obtained 2D perceptual error map, spatial pooling sums contrasts across space to produce the quality measure for the image under test. The Minkowski summation is widely used

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

3. PERCEPTUAL DECOMPOSITIONS

Four perceptual decompositions have been analysed and compared. The first one, used in the metric and given in figure 2, is based on a large synthesis of the literature [13, 14, 15] and has been completed by several experimental studies [16]. While bandwidths of this decomposition have been used as they have been measured, the bandwidths of the second and third decompositions (figures 3 and 4) have been intentionally constrained for computational reasons and coding results optimisation. These two last decompositions are well known and have been proposed by Watson [10] and Daly [17] 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 chan-

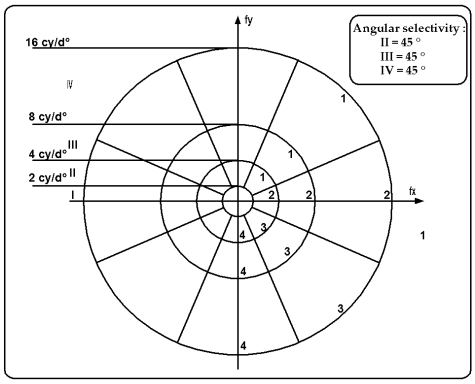

Figure 3: Watson's decomposition

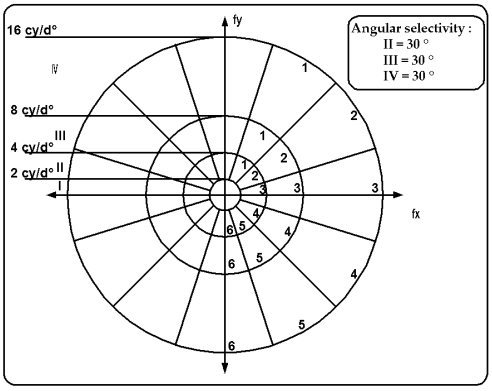

Figure 4: Daly's decomposition

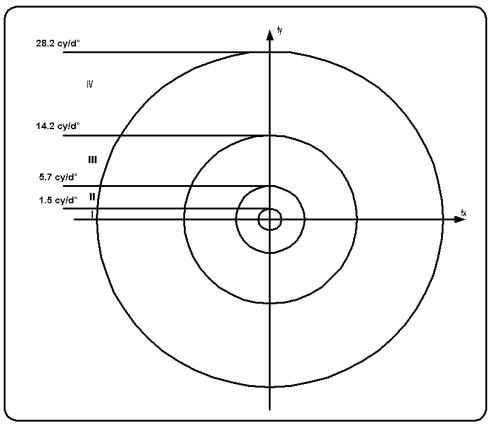

Figure 5: Only radial decomposition is performed

nels. The angular selectivity is constant and is equal to 45 degrees for Watson and 30 degrees for Daly. The fourth and last decomposition (figure 5) uses four radial bands and ignores the orientation sensitivity of the HVS. Results given in [18] 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 POOLINGS

Three spatial pooling models have been compared. The first one, called M1, is the Minkowski summation. It is considered as the most accepted model for the summation of errors across space when distortions are near detection threshold (eq. 1). The second model, called M2, is based on the contrast occurrence probability $\Pr\left(c^P(m,n)\right)$ and is given by

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

The third and last model, M3, is a combination of the two previous ones and consists of a Minkowski summation weighted by the contrast occurrence probability:

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

5. RESULTS AND DISCUSSION

Seven well known original images have been considered in this study.

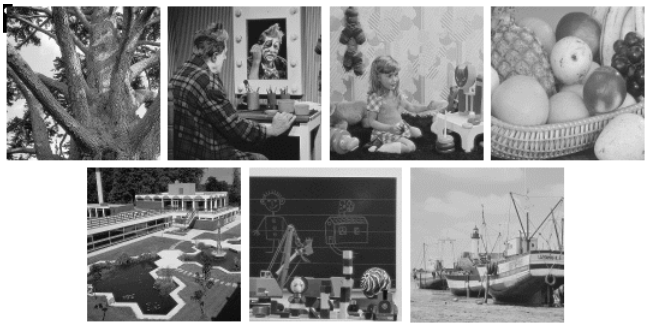

Fig. 2: The original images of the database

For each of these images, fifteen impaired images have been generated to cover the whole quality scale (MOS varying between 1 and 5) recommended by the ITU-R. This 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. 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. Results of such optimization are given table 1 for each perceptual decomposition. PD1 corresponds to the decomposition of figure 1. PD2, PD3 are the decompositions used by

	M1	M2	M3
PD1	β=1.3 Cc=0.9277	β=0.8 γ=0.5 Cc=0.9246	β=1.1 γ=2.2 Cc=0.9689
PD2	β=1.5 Cc=0.9264	β=0.9 γ=0.5 Cc=0.9036	β=1.3 γ=2.4 Cc=0.9717
PD3	β=1.2 Cc=0.9241	β=0.8 γ=0.5 Cc=0.9235	β=1 γ=2.3 Cc=0.9689
PD4	β=2.3 Cc=0.9401	β=1.5 γ=0.5 Cc=0.8514	β=1.6 γ=2.1 Cc=0.9698

Table 1: Best correlation coefficients and the corresponding exponents.

	M1	M2	M3
PD1	0.9521	0.9651	0.9838
PD2	0.9395	0.9621	0.9746
PD3	0.9400	0.9652	0.9811
PD4	0.9701	0.9218	0.9806

Table 2: Correlation coefficients of the first image in the test database.

	M1	M2	M3
PD1	0.9604	0.9676	0.9797
PD2	0.9668	0.9674	0.9801
PD3	0.9660	0.9681	0.9780
PD4	0.9742	0.9542	0.9782

Table 3: Correlation coefficients of the second image in the test database.

	M1	M2	M3
PD1	0.8325	0.9185	0.8965
PD2	0.8548	0.9085	0.9063
PD3	0.8156	0.9207	0.8836
PD4	0.8855	0.8473	0.9696

Table 4: Correlation coefficients of the whole database.

Watson and Daly respectively and PD4 is the radial decomposition. The correlation coefficients Cc obtained with the training data are also reported. Then, these parameters have been used for each image of the test database. The results of the two first images are given tables 2 and 3. Table 4 summarizes the results of the whole database.

For model M1, if one looks at the results obtained image by image, it appears that a simple Minkowski summation associated with a simple radial decomposition is the best choice as well from the point of view of performances as that of complexity. This remark remains true when looking at the table 4 which gathers the results obtained with all the images.

For model M2, the behavior observed image by image also remains in agreement with that of the whole database and the best correlation coefficients are obtained with the decomposition of Daly (PD3). The difference in performances between the decomposition of Daly and that of Watson is mainly due to the angular selectivity as this latter represents the only one

difference. Thus, one can expect that a finer angular selectivity improves the performances of quality metrics. This observation is also confirmed by the first decomposition (PD1) which has the second best results and presents in the medium and high spatial frequencies, the same angular selectivity as PD3.

Finally, the last model (M3) seems to be more sensitive to the image content as the results vary from one image to another and are also different for the whole database. In this last case, and as for the first model M1, the radial decomposition presents the best performances.

To synthesize, we can argue, based on the made comparisons that best correlation with subjective data are obtained with a simple radial decomposition associated to a simple Minkowski summation. Moreover, the performances are significantly improved if the Minkowski summation is weighted by the occurrence probability of perceptual error contrast.

Obviously, this important observation can be discussed. As the radial decomposition avoids the angular frequency pooling, one can argue that the corresponding model used by the three other decompositions is not appropriate. Probably it is the case. However, any effort to make it more effective would undoubtedly increase the metric complexity.

6. CONCLUSION

A perceptual metric for image quality assessment has been presented. This metric includes a perceptual decomposition, a local contrast, a masking model and a frequency and spatial pooling. Investigations have been focused on the perceptual decomposition and spatial pooling. Thus, four decompositions and three pooling models have been evaluated. Based on the made comparisons, it seems that best correlation with subjective data are obtained with a simple radial decomposition associated to a simple Minkowski summation.

To define a simple and efficient measurement tool for image quality, future works, which are under development, concern the evaluation of the remaining components: the contrast and the masking models.

REFERENCES

[1] ITU-R, “Methodology for the subjective assessment of the quality of television pictures”. ITU-R Rec. BT.500-11 (http://www.itu.ch) 2002.

[2] M.P.Eckert, A.P.Bradley, “Signal Processing,” Vol. 70, pp. 177-200, 1998.

[3] S. Comes, O. Bruyndonckx, et B.Macq,” ICASSP”, Vol. 4, 9-12 May 1995, pp. 2635-2638.

[4] C. Zetzsche et G. Hauske, “Human vision, Visual Processing, and digital Display,” in Proc. SPIE, San Jose, Vol. 1077,. 1989, pp.209-216.

[5] P.C. Teo and D.J. Heeger, ” Human vision, Visual Processing, and digital Display,” in Proc. SPIE, San Jose, Vol. 2179, Feb. 1994, pp. 127-141.

[6] M.P.Eckert, A.P.Bradley, “Signal Processing,” Journal, Vol. 70, pp. 177-200, 1998.

[7] T.N.Pappas, R.J.Safranek, *Handbook of Image and video Processing*, AL Bovik ed., Academic press, pp. 669-684, 2000.

[8] http://www.vqeg.org/

[9] S. Daly, “Int. Conf. On Image Proc,” Austin, Texas, Vol. II of III, 13-16 Nov., 1994, pp. 16-20.

[10] A.B.Watson, " Computer Vision, Graphics and Image Processing," Journal, No.39, pp.311-327, 1987.

[11] E.Peli, “J.O.SA,” Journal, Vol. 7, N°. 10, pp. 2032-2040, Oct.1990.

[12] N.Bekkat, A.Saadane, D. Barba, " Human Vision and Electronic Imaging," Proc. of SPIE, San Jose, California, January 23 – 28, 2000.

[13] S.J. Anderson, D.C. Burr and M.C. Morrone," J.O.S.A," Journal, vol. 8, pp1340-1351, 1991.

[14] D.C.Burr, S.A.Wijesundra, “Vis. Research,” Journal, Vol. 31, N°. 7/8, pp. 1449-1452, 1991.

[15] S.J.Anderson, D.C. Burr, “J.O.S.A,” Journal, Vol. A8, N°. 8, pp. 1330-1339, 1991.

[16] A.Saadane, D.Barba, H. Senane, "Visual Communications and Image processing," Proc. Of SPIE, Taipei, Taiwan China, 23-26 May 1995.

[17] S. Daly, *Digital images and human vision*, Chap. 14, 1993.

[18] T. Mitsa, K.L. Varkur, J.R. Alford, “Human vision, Visual Processing, and digital Display,” in Proc. SPIE, Vol. 1913,. Feb.1993, pp.390-401.