Pooling Individual Errors in an Objective Psycho-visually Based Quality criterion: a Critical Stage and New Results

Patrick LE CALLET
IRCCyN UMR n°6597 CNRS
Ecole Polytechnique de l'Université de Nantes, rue Christian Pauc, La Chantrerie, BP 50609
44306 NANTES Cedex - France

and

Dominique BARBA
IRCCyN UMR n°6597 CNRS
Ecole Polytechnique de l'Université de Nantes, rue Christian Pauc, La Chantrerie, BP 50609
44306 NANTES Cedex - FRANCE

ABSTRACT

There has recently been a great research interest in the design of objective quality criteria which can replace the use of visual evaluation of the quality of images by human observers. Many methods have been proposed to do that, all of them use at different levels parts of the human visual system properties in the perception of distortion and the way to merge them into a global and unique value rating the visual quality. This paper points out the way to fuse the distortion of degraded images in the spatial domain in the context of a visual sub-band representation of images. On one hand, we show that the ways of pooling perceived errors across the sub-bands and along the image space are quite different On the other hand , we show that the structure of errors might be used to weight error annoyance. The incorporation of these factors in an objective quality criterion allows to achieve for the correlation coefficient (cc) and rank correlation coefficient (rcc) respectively, values higher than 0.97 between subjective and objective quality measurements on two quite different databases of distorted images.
Keywords: image quality, visual modeling, spatial summation, error pooling

1. INTRODUCTION

The knowledge of image or video quality is a fundamental information in image and video processing, not only for performance evaluation but also for optimizing the processing.. Basically and very often, people takes a pure PSNR or a derived criterion (ie weighted PSNR) as a measure of the relative quality of the distorted image compare to the reference one. This quality measurement is not convenient since PSNR is known as rather poorly correlated with human judgement in quality assessment.

A quite another way to measure image quality is to make subjective experiments where humans have to give their opinion using a quality scale. This kind of methods is very time consuming and complex to run, as the measurements required specialized rooms equipped for visual testing in normalized conditions, so not often used. Therefore the interest to dispose of an objective and realistic criterion is well recognized. In the literature, several quality metrics have been proposed and use models of the human visual system (so-called HVS hereafter) in order to estimate the visual quality. In this paper, we are concerned with the design of a bi-variant objective visual criterion showing good concordance with human judgements. The structure of an objective visual criterion is classically divided into two main blocks. The first one computes a visual representation of the two images (original image and distorted image under evaluation). This kind of block is present in most of criteria already proposed in literature [1][2][3] but different perception models more or less sophisticated are used at this level, derived from different interpretations of experimental psychophysics results. The second block produces usually a single value called the global visual quality from the visual representation obtained at the output of the first block. The second block has to realize at least the merge in one single value of elementary visual distortions spatially distributed along the image plane This transformation or merging is classically called the pooling stage. A correct modeling of this global transformation according to what the human visual system do is quite difficult to get since it involves complex human vision mechanisms at present not very well known. Most authors use a Minkowski summation in this second block for merging, but this kind of approach can be considered rather simple comparing to the refinement of the visual representation used in the first block. Winkler [4] have related the difficulty of modeling the pooling step. In our opinion, this is a crucial step in the design of a realistic criterion and the overall performance of a quality metric in terms of good correspondence between an objective quality measurement and a subjective one depends on the relevance of this modeling. In this paper, we propose new more realistic models for this second block using different perceptual models for the first block and show the impact of different ways of merging locally perceptual errors. Section 2 describes the three perceptual models which have been tested and some information concerning the two data bases of images used in the experiments are given in section 3.

2. PERCEPTUAL MODELS

In order to generate the visual representation of images, we have used and tested three different models, two of them include a sub-band decomposition of the images.

Model I : mono-channel model

The first model is quite basic since it is based on a well known property of the human visual system (HVS) called the contrast sensitivity function (CSF). The CSF exhibits the sensibility of the HVS with the spatial frequency of a signal. Here we used the BARTEN CSF. We apply this CSF on both signals (original and distorted images), then compute the difference between them. This gives one map of perceptual errors (signed errors).

Model II : Visual Difference Predictor

The visual difference predictor (VDP) of S. Daly [5] is a model which integrates several properties of the HVS that one finds in most sophisticated models. It includes luminance adaptation, CSF computation, a sub-band decomposition (with orientation selectivity) into 25 channels, a masking function. The masking function models the variation of the differential visibility threshold of a signal (stimuli) with the signal (masking signal) in its surrounding. In the VDP, masking function have been designed using experimental results of simple signals detection. The entire model is applied on both images (original and distorted), then the difference is computed within every channel and normalized by the visibility threshold value. So, we get 25 maps of perceptual errors, one for each channel.

Model III : Quantization noise modeling

This model have been developed in our laboratory : each of its components has been completely derived from experiments we have conducted. As in model II, we used a perceptual sub-band decomposition but into only 17 channels. The main properties of this decomposition are a non dyadic radial selectivity and an orientation selectivity that increases with the radial frequency. Outputs of the sub-band decomposition are converted into band-limited local contrasts according to a modified version of contrast PELI definition [6]. This definition has allowed us to explain inter-channel masking. Overall masking function is based on experimental results concerning the detection of quantization noise added to mono or multi band-limited textures. This kind of experiments are quite different of classical ones based on measurement of differential visibility threshold of elementary signals. Moreover, compare to model II, inter-channel masking is included in this model. As outputs of the masking function, we get directly 17 maps of perceptual errors, one for each channel.

3. SUBJECTIVE QUALITY EVALUATION

To compare the criterion performance with human judgements, we built two databases of subjective quality evaluations. We used six original images from natural scenes and 2 different types of degradations :

- JPEG image compression for database I (DBI),
- visual quantizations for database II (DBII).

In each database, distorted images were evaluated twice by 5 observers.

In data base DBI, all distorted images are compressed images at five different rates that leads to a total of 30 different images in DBI. Subjective evaluations were made in normalized test conditions at a viewing distance of six times the height of the TV monitor, using a DSCQS (Double Stimulus Continuous Scale) method.

In database DBII, all images were distorted using 15 different levels of quantization steps that leads to a total of 105 images in DBII. Subjective evaluations were made also in normalized conditions using a DSIS (Double Stimulus Impairment Scale)method with five categories.

Different methods have been used for subjective quality evaluation since it is known that results of subjective evaluations are sensitive to the protocol used for the tests.

To compare objective measures (OM) given by the objective criteria and subjective scores (SS), we compute correlation coefficient (cc) and rank correlation coefficient (rcc) between OM and SS. We use these two correlation coefficients to compare the different objective quality criteria.

4. GLOBAL MERGING STRATEGY

From maps of perceptual errors, we have to merge them into a single quality value. Therefore this merging consists to combine all the errors spatially and sub-band distributed in order to get the overall quality measurement. We have already proposed a global merging strategy [8] illustrated in figure 1.

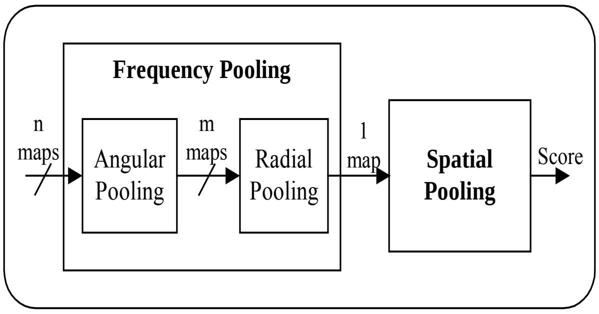

Figure 1 : global merging strategy

In the case of a sub-band decomposition, error merging can be divided in two steps: sub-band pooling and spatial pooling. Concerning the order of them, we definitely decides to perform the sub-band merging before the spatial one. Actually, errors are first perceived locally, taking account the spatial frequency content (frequency pooling) then all the spatially distributed errors are combined (spatial pooling).

Sub-band pooling can be subdivided itself into 2 sub-steps : angular and radial frequency pooling. Angular pooling refers to the fuse of all the errors in a radial frequency band coming from different orientations and radial pooling refers to the fuse of errors of all radial frequency bands, in both cases at the same position. For the radial pooling, we think that errors importance depends on the radial spatial frequency band. Spatial pooling involves errors balance depending on their intensity, their location and their structure. In the two following sections, we propose to analyze thoroughly different models for frequency and spatial pooling.

5. FREQUENCY POOLING

Dependence with spatial frequency band

In order to determine the importance of the different frequency sub-bands in the final quality measurement, we have first combined errors coming from one frequency band. We fixed angular pooling and spatial pooling using a linear summation or a Minkovski summation.

Correlation coefficients with subjective quality measurements have been computed. Results are provided in Table 1 and Table 2.

We can see the radial frequency sub-bands do not show the same importance in the global appreciation of the quality. For both models, middle radial frequency band (5.7 to 14.2 cy/d° for model III and 7 to 14 cy/d° for model II) presents the best correlation (cc and rcc) in the two databases. One can argue it could be a model default : if it is biased for a radial frequency sub-band, it is impossible to get correct results for this particular band. As both models respond in the same way, we can reasonably eliminate this kind of model default.

Another restriction to generalize the importance of middle frequencies is that errors can be limited to frequency domain. It is impossible that impairments used for generate databases I and II concentrate errors only in middle spatial frequencies. Anyway, even if radial frequency sub-band importance is degradation dependent, this should be study carefully since cc and rcc have higher values when taking errors from well-selected sub-bands than taking them from all sub-bands (see Table 1 and Table 2). On an other way, model III gives higher correlation value than model II.

MODEL II	database I		database II	
	cc	rcc	cc	rcc
Base-band	0,192	0,297	0,606	0,623
Sub-band 5	0,202	0,185	0,738	0,755
Sub-band 4	0,888	0,869	0,799	0,823
Sub-band 3	0,872	0,884	0,727	0,749
Sub-band 2	0,489	0,511	0,504	0,525
all sub-bands	0,867	0,87	0,716	0,745

Table 1 : correlation coefficient (cc) and rank correlation coefficient (rcc) between SS and OM (Model II)

MODEL III	database I		database II	
	cc	rcc	cc	rcc
Base-band	0,12	0,08	0,18	0,21
Sub-band 5	0,789	0,834	0,745	0,773
Sub-band 4	0,926	0,959	0,864	0,884
Sub-band 3	0,012	0,11	0,648	0,745
all sub-bands	0,925	0,957	0,87	0,887

Table 2 : correlation coefficient (cc) and rank correlation coefficient (rcc) between SS and OM (Model III)

Angular and radial merging

Angular merging consists here to combine the errors coming from all oriented sub-bands (orientation selectivity) of the same radial sub-band. In this case, we guess that error balance depends more on intensity than on orientation, so we use a Minkowski summation. We have optimized the value of the exponent parameter in the summation in the most important radial sub-band (middle radial frequency) for both models that maximize cc and rcc. Spatial merging is set as a Minkowski summation with an exponent parameter equal to 2. Most best results for cc and rcc between SS and OM (shown in Table 3) are got using a different value of the exponent parameter in the angular and spatial merging respectively. This fact is interesting since most authors in literature use the same exponent value for both of them. These results confirm the need to separate different merging since they probably involve different mechanisms.

	DBI		DBII	
	Model II	Model III	Model II	Model III
cc	0,8867	0,9394	0,8003	0,8709
rcc	0,8751	0,9579	0,8283	0,9057
Exponent	0,8	2	0,8	0,6

Table 3 : Best values of exponent parameter in Minkowski summation for angular merging (in middle frequencies radial band) and maximum values of cc and rcc between OM and SS in DBI and DBII for models II and III.

Radial merging consists to balance the different radial sub-bands according to what we have pointed just before. We have chosen a linear combination , the weights are optimized for both models in order to get best cc and rcc between SS and OM.

6. SPATIAL POOLING

Methods

We have tested four different methods for the spatial merging from a very classical one to some with a special merging effect.

Minkowski summation: The most popular way of merging errors, this summation balances errors according to their magnitude :

$$\text{measure} = \frac{1}{\text{L.H}}\left(\sum_{m=1}^{W}\sum_{n=1}^{H}\left|\text{erreur}(m,n)\right|^{\beta}\right)^{1/\beta}$$

(W and H are the number of pixels along the horizontal and vertical direction respectively).

Probabilistic approach: This approach weights errors according to the probability of the magnitude of the error :

$$\text{measure} = \sum_{a=1}^{a\max} a^{\alpha}.\Pr(\vec{\text{erreur}} = \text{a})^{\gamma}$$

Density error approach: Several errors scarcely distributed along the image are less visually annoying than the same set of errors concentrated in a small region of the image. In order to take this effect into account, we computed the normalized error density (ned) in a window of one degree size (size of the fovea field) for every site and we weighted errors as follows :

$$\text{balanced\_error}(m,n) = \text{error}(m,n) \times \text{ned}^{\alpha}(m,n)$$

then, we computed a Minkowski summation of balanced errors,

Structural approach: Structure of errors is an important information in spatial merging as structured errors are more annoying than non structured ones. We propose to take into account this feature. In perceptual domain, we detect edges in the original image and in the distorted image too. Then we classified the errors in 4 categories:

- errors corresponding to the destruction of edge,
- errors corresponding to the insertion of a new edge,
- errors located on edge,
- others errors,

then, we weighted linearly all the error types and applied a Minkowski summation to get the final quality measurement.

Results

We have fixed the frequency merging parameters according to the results obtained in section 5 for models II and III. For the spatial merging methods described above, we have optimized their parameters in order to maximize cc and rcc between OM for models I, II and III and SS of database I (Table 4) and II (Table 5) respectively.

		DBI		
		Model I	Model II	Model III
Minkowski	cc max.	0,8699	0,9032	0,9416
	rcc max	0,8809	0,8953	0,9747
Probality	cc max	0,8754	0,9102	0,9595
	rccmax	0,8791	0,8862	0,9704
Density	c max	0,8842	0,9197	0,9626
	rccmax	0,8965	0,9291	0,9771
Structure	cc max	0,9026	0,9403	0,9731
	rcc max	0,9011	0,9702	0,9792

Table 4 : Best values of cc and rcc between SS of database I and OM given by the 4 different spatial merging methods with Models I, II and III respectively.

		DBII		
		Model I	Model II	Model III
Minkowski	cc max	0,7051	0,7977	0,8754
	rcc max	0,7108	0,8251	0,9012
Probality	cc max	0,7282	0,8103	0,9035
	rcc max	0,7565	0,8302	0,9105
Density	cc max	0,7682	0,8852	0,9181
	rcc max	0,7348	0,9012	0,9271
Structure	cc max	0,7952	0,9408	0,9485
	rcc max	0,7832	0,9125	0,9317

Table 5 : Best values of cc and rcc between SS of database II and OM given by the 4 different spatial merging methods with Models I, II and III respectively.

For both databases I and II, the best spatial merging model is the structural model, then the density model, the probabilistic model in position 3 and, lastly the Minkowski model.

The rank between these 4 models doesn't change with spatial pooling method : model III provides better results than model II that provides better results than model I.
A very simple model such as model I can produce better results when associated with sophisticated spatial error merging compared to the performances of model II (more sophisticated) associated with a rather simple spatial pooling.

7. CONCLUSION AND PERSPECTIVES

In the design of an image quality criterion, two important points are crucial, each one corresponds to one of the two stages of a quality metric system. The first one, which makes the distinction between perceived and not perceived distortions, gives a quantitative measure of the perceptual distortion spatially localized, in a single or in a multi-band description. It is now rather well known but some refinements can enhance the pertinence of the model. In this paper, we show how to incorporate a more realistic model of the inter sub-band masking effects in the perception of distortions. The second stage has to merge the perceived impairments frequency and spatially distributed into a single value rating the overall quality. This stage, called the merging stage (or the pooling stage), perhaps the most important, is usually performed in different quality models by a simple and uniform Minkovski summation. We shown by testing different way and order of non-linear summations that:
- sub-band pooling has to weight differently radial mid-band spatial frequency with the other,
- spatial pooling which takes into account the structure of the impairments exhibits higher performances that 3 others methods of merging (pure Minkovski summation, probability weighted summation, spatial density weighted summation).

The incorporation of these factors in an objective quality criterion allows to achieve values higher than 0.97 for the correlation coefficient (cc) and rank correlation coefficient (rcc) respectively, between subjective and objective quality measurements on 2 quite different databases of distorted images. In each case, no particular knowledge of the image distortion has been taken in the objective quality criterion.

8. REFERENCES

[1] C. J. van den Branden Lambrecht, "Perceptual models and architectures for video coding applications", Phd thesis, Lausanne, 1996.
[2] C. Zetzsche and, G. Hauske, "Multiple channel for the prediction of subjective image quality", in SPIE Vol. 1077, Human vision, visual processing, and digital display, pp. 209-216,1989.
[3] S. Comes and, B. Macq, "Human visual quality criterion", ICIP, Lausanne, Vol I of III, pp. 869-872, 1996.
[4] Stefan Winkler: "Visual fidelity and perceived quality: towards comprehensive metrics." to appear in Proc. SPIE Human Vision and Electronic Imaging Conference, vol. 4299, San Jose, California, 2001.
[5] S. Daly, "The visible different predictor : an algorithm for the assessment of image fidelity", in Proc. of SPIE, Vol. 1666, Human vision, visual processing and digital display III, pp 2-15, 1992.
[6] E. Peli, "Contrast in complex images", JOSA Vol.7, n° 10, pp 2032-2040, October 1990.
[7] N. Bekkat, A. Saadane, D. Barba "An objective and perceptually based quality criterion", ACIVS 1999,Baden-Baden, 6-7 August 1999
[8] P. Le Callet, and D. Barba, "Frequency and spatial pooling of visual differences for still image quality assessment" in Proc. SPIE Human Vision and Electronic Imaging Conference, vol. 3959, San Jose, California, January 22-28, 2000.