Robust approach for color image quality assessment

Patrick LE CALLET, Dominique BARBA

IRCCyN-IVC (CNRS UMR 6597)
Ecole polytechnique de l'université de Nantes
Rue Christian Pauc, BP 50609, 44306 Nantes Cedex 3 France
{patrick.lecallet, dominique.barba}@polytech.univ-nantes.fr

ABSTRACT

This paper presents a visual color image quality metric assessment with full reference image. The metric is highly based on human visual system properties in order to get the best correspondence with human judgements. Contrary to some others objective criteria, it doesn't use any a priori knowledge on the type of introduced degradations. So the main interest of the metric is on its ability to produce robust results independently of the distortions. The metric can be decomposed into two steps. The first one projects each images, the reference one and the distorted one, in a perceptual space. The second step achieves the pooling of errors between perceptual representation of two images in order to get a score for the overall quality. Since we have shown that these two steps have equivalent importance regarding metric performance, we have particularly paid attention in correct balancing when designing the two steps. Especially, for the second one, that is generally limited to poor consideration in literature, we have developed some new original approaches . We compare results of the metric with human judgments on images distorted with different compression schemes. High performances are obtained leading to assure that the metric is robust, so this approach constitutes an alternative useful tool to PSNR for image quality assessment.

Keywords: Image quality assessment, perceptual image representation, human visual system model

1. INTRODUCTION

Image quality assessment is an active domain of investigation in the image community. The main motivation is to find automatic methods providing computed quality scores which are correlated with quality human judgment. These methods can supply precise control of image compression schemes and more generally a relevant human-related Quality of Service (QoS) for images transmission. Metrics can be classified in three categories. The first category gathers full reference metrics (FR) in which the original image and the distorted image are required to compute the quality score. In the second category, reduced reference (RR) metrics need a description of the original image and the distorted image. The reduced reference is designed to contain relevant information with respect to the image quality to compute. In the third and last category, no reference (NR) metrics only use the distorted image to compute the quality score.

NR or RR metrics development have recently focus most of interest of image quality assessment community. Effectively, in a broadcasting purpose, only RR and NR metrics are suitable in particular for QoS management. For example, RR metrics allow to code the reduced reference to embed it in the bitstream in order to constitute a practical approach to quality evaluation. But, since NR and RR metrics focus on artifacts due to a given transmission scheme, they depend on the service (compression type, transmission channel, ...). So, there is still interest in developing FR metrics for such applications as coding schemes performance comparison. Coding schemes introduce visual distortions between the original image and the decoded one (in fact coded with lost and then decoded) that can be more or less annoying depending on bit rate, image and coding scheme used. Ones uses the PSNR as a measure of the relative differences between original and distorted images. This is not convenient since it is well known that PSNR is poorly correlated with human judgement. Another way to appreciate quality is to make subjective experiments where humans have to give their opinion using a quality scale, this method is very time consuming and complex to run, so not very used. In such context, a FR quality metric, highly correlated with human judgements, provides a precious tool for image coders designers in order to assess performance of their methods. Since FR should be useful for comparison between different image coding schemes, we have to pay attention into not incorporate knowledge on how the images are coded. For instance, several quality criteria in literature[1, 2, 3] use measurement of block effects (typical in DCT-block based methods of compression), postulating that this kind of degradation is introduced, this can be useful for quality regulation of some particular coding schemes, but this is not pertinent for performance evaluation of all type of coding schemes (especially those which are not block-based coding). So, the real challenge for FR metric is to be robust to wide panel of distortions. This problem is up to us not yet solved as the need a second VQEG test plan for FR metric seems to prove it. This paper proposes a metric that constitutes a robust approach for color image quality assessment.

Classically, FR metrics can be divided into two main functions : one to construct the errors between original and distorted images, leading to distortion maps, and an another function to pool the errors providing a global quality score. Two categories of this type of image quality metrics can be found in literature. Metrics from the first category use a human visual system model for low level perception, such as subband decomposition and masking effect, in order to compute distortion maps, but often propose poor error pooling methods, such as Minkowski summation. On the other hand, a second category of metrics use little information about the human visual system for error representation, and push the effort on the pooling stage integrating an a priori knowledge on introduced distortions. This last point makes these metrics specialized so they fail to be robust. Contrary to specialized metrics, the purposed metric doesn't use any a priori knowledge on the type of degradations introduced by any image processing. By the way, we have constructed a generic approach in sense that quality assessment is robust to a wide variety of distortions. For instance, the metric is well adapted to performance comparison between image coding schemes. Moreover, the method can be qualified perceptual since it is based on human visual system (HVS) properties in order to get the best correspondence with human judgments. Our approach is therefore a well-balanced combination between human vision model to compute distortion maps and a pooling stage that tries to extract sense from errors. It includes the definition of a new HVS model to get a perceptual image representation (entirely based on results from psychophysics experiments conducted in the laboratory) and a new approach for the errors pooling stage from perceptual distortion maps based on error density and error structure to make the metric comprehensive.

2. METRIC OVERVIEW

The structure of the criteria is illustrated in figure 1. It includes two main stages. A first stage computes the visual representation of each image, the original one and the distorted one. This stage is highly based on results provided by psychophysics experiments on color perception, and on masking effect. It provides several maps (one map by frequency domain and color component) for each image determining how the information is perceived by the early stages of human visual system. This stage is present in most of criteria in literature[4, 5] but different perception models more or less sophisticated are used derived from different psychophysics experiments results and interpretation. Maps computed from representation of the distorted image are then subtracted from maps computed from the representation of the original image, results are called the distortion maps.

The second stage computes one score called the overall quality, from all the distortion maps. A correct model of this transformation according to human visual system is quite hard to get since it involves complex human vision mechanisms not very well identified. Most authors use Minkowski summation for this stage, this kind of approach can be considered very simple comparing to the refinement of the visual representation stage. Winkler[6] have related the difficulty of modeling the pooling stage. In our opinion, this is a crucial stage of the criterion and the overall performance of a quality metric in terms of good correspondence between objective and subjective quality depends on the pertinence of this stage. Previously[7], we have conducted a study in order to determine the impact on image quality assessment performance of both stages, visual representation and error pooling. We have elected different methods more or less sophisticated for visual representation and error pooling to construct these metrics. Then, we have appreciated performance measuring the correlation coefficient (cc) between human judgments and predictions made by the metric. It was quite amazing to notice that a poor visual representation associated to a sophisticated error pooling method leads to the same performance than a sophisticated visual representation associated with a poor error pooling method. We have clearly seen that effort should be push on both stages.

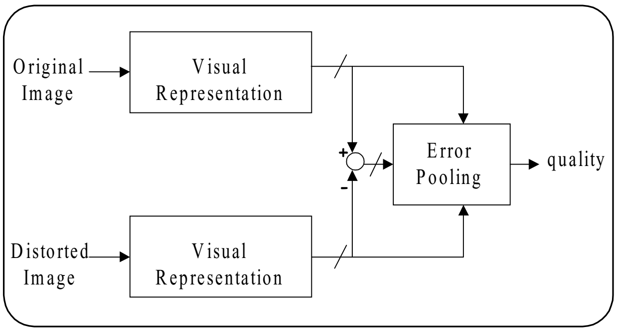

Figure 1 : structure of the criteria

3. VISUAL DISTORTION MAPS

3.1. Overview

Our model of the early vision stages is largely inspired from the VDP of Daly[8] but extended to color. However, several modifications have been made based on psychophysics experiments made in our laboratory. Figure 2 illustrates how we construct the visual representation of images (and so the perceptual maps). This construction is made in four steps. First, we use a color space transformation with adaptation, then a CSF (contrast sensitivity function) for each component of the color space, followed by a PSD (Perceptual Subband Decomposition) and finally a masking stage. Each step of this model is explained in the following sections.

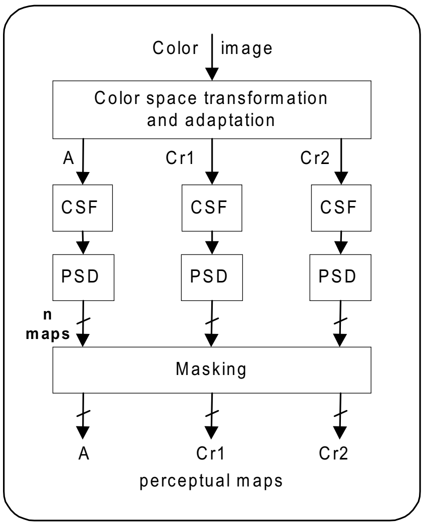

Figure 2 : construction of perceptual maps

3.2. Color space

To take into account perceptual properties some specific color spaces as L*a*b*, L*u*v* have been introduced. We are concerned here with psychophysically based color spaces. Perceptual color spaces are based on the fact that the peripheral parts of the human color vision include two different stages. In the first stage, the light information is transformed into neuro-electrical signals by the three types of cone receptors (L, M, S). The interactive nature of the second stage is generally admitted. It has been shown that the neuro-electrical signals are combined in an opponent manner. However there is no agreement on the receptor weightings needed to describe this opponent interaction.

Several such color vision model have been proposed in the literature, among this variety we have validated in our lab from masking experiments the color space determined by Krauskopf[9]. In this space the achromatic Ach and chromatic Cr1 and Cr2 directions are defined as :

$$\mathrm{Ach} = \mathrm{L} + \mathrm{M}$$

$$\mathrm{Cr1} = \mathrm{L} - \mathrm{M}$$

$$\mathrm{Cr2} = \mathrm{S} - (\mathrm{L} + \mathrm{M}).$$

3.3. Sensitivity function

The CSF describes the variations in visual sensitivity as a function of spatial frequency. For the component A, we use the CSF defined by Daly (band pass filter). For components Cr1 and Cr2, we propose two models issued from experiments made in the laboratory. We use two low pass filters with cut-off frequency of about 5.5 cpd (cycle per degree) and 4,1 cpd for Cr1 and Cr2 component respectively (see figure 3 for Cr1 component). These 2D filters are anisotropic, so they include variations with orientation. Explicit formulations of these filters are respetively for Cr1 and Cr2 component :

$$Scr1(\omega,\theta) = \frac{33}{1+(\frac{\omega}{5.52})^{1.72}}.(1-0.27.\sin(2.\theta)),$$

$$Scr2(\omega,\theta) = \frac{5}{1+(\frac{\omega}{4.12})^{1.64}}.(1-0.24.\sin(2.\theta)).$$

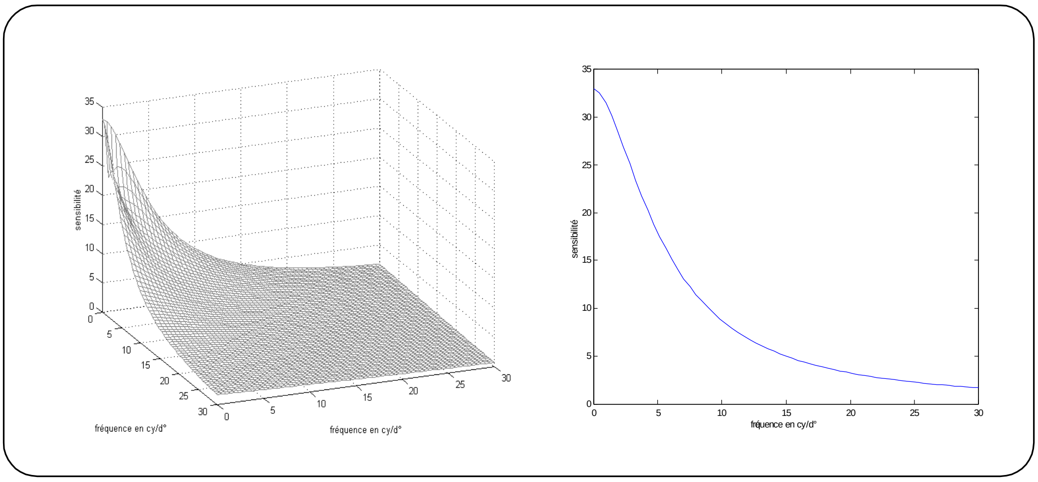

Figure 3 : CSF on component Cr1: CSF2D on the left, CSF in horizontal direction on the right

3.4. Perceptual subband decomposition

As in most approaches, we use a subband decomposition defined by analytic filters for the three components supposed to describe the different channels in the human vision system. Previous study[10] has characterized this decomposition for the A component. The obtained results led us to propose the spatial frequency patch shown in figure 4. We have conducted the same type of experiments for Cr1 and Cr2. Results[11] suggest the same type of decomposition than for component A but limited to subbands I and II.

The main properties of these subband decopomsitions are their spatial frequency and orientation selectivity that vary radial spatial frequency according to what we have measured. While some authors usually use dyadic decomposition with constant orientation selectivity, most of the time under complexity considerations, we have elected to confirm psychophysic observations. The main reason is that our decomposition have led us to well understand some masking experiments results, constituing a good a posteriori validation.

3.5. Masking model

Masking is a well know effect that refers to the changes of visibility (increase or decrease) of a signal due to the presence of a background. The masking stage models this effect. We can characterize two types of masking : intra-component masking and inter-component masking.

3.5.1. Intra-component masking

Intra-component masking is implemented using Daly's model[8] adapted to our perceptual subband decomposition. That means that we only consider there a intra-subband masking.

From experiments, we have identified parameters of the masking function for Cr1 and Cr2 components.

$$T^{i,j}(m,n) = \frac{1+a^{i,j}.\left|f^{i,j}(m,n)\right| + b^{i,j}.\left|f^{i,j}(m,n)\right|^2}{1+c^{i,j}.\left|f^{i,j}(m,n)\right|} ,$$

a, b and c are parameters defined for each subband (i,j).

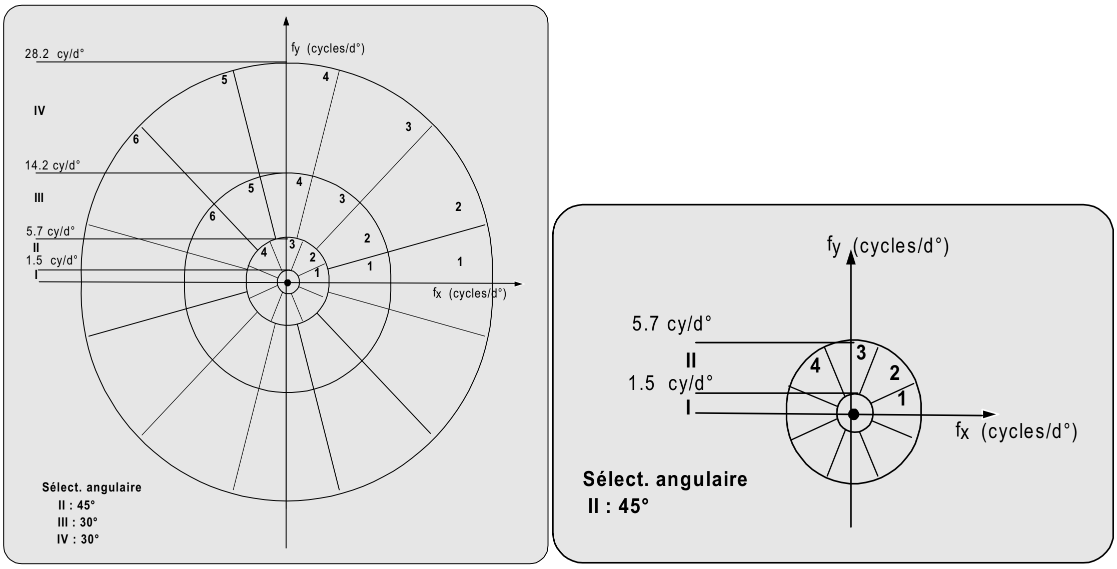

Figure 4 : perceptual subband decomposition for A component

3.5.2. Inter-component masking

Cross-masking between components experiments have shown that the linear transform from LMS signal to opponent-color space cannot efficiently decorrelate the cone responses. This leads to take into account interactions that modify the perception of each components depending on the others in subbands. Therefore, we have to consider the inter-pathway masking effect between the luminance and chromatic components. We conducted several experiments in the laboratory (some results can be found in [12]) in order to characterize these interactions.

Results show that significant interactions are action of Cr1 and Cr2 on A component, and action of Cr2 on Cr1 component. On tab 1, we can see which interactions have been used in our model according to their importance.

	Masqued component		
	A	Cr1	Cr2
A	Intra	I on I et II,j	None
Cr1	I on I, II,j and III,j II,j on I,j and II,j II on III,j	Intra	I on I and II,j
Cr2	none	I on I and II,j	Intra

Tab 1 : main interactions between components

As for Intra-component masking, we have identified parameters of the masking function for each configuration.

4. ERROR POOLING

4.1. Pooling overview

The set of subbands of every components of spatial error perception has to be merged into a single scalar value : the objective quality measurement. We get 27 distortion maps as outputs of the visual representation stage: 17 for the A component and 5 for each chromatic component. We can divided error pooling into three operations : frequency, component and spatial pooling. We have chosen to separate this three pooling functions. Concerning the order of these functions, we have previously expressed that it is more convenient to carry out component pooling then, frequency pooling and finally spatial pooling[7].

4.2. Pooling methods

Component pooling means to merge errors on the component dimension, so coming from the three component of the color space. This pooling is implemented as a linear combination to balance error between component.

Frequency pooling means to merge errors on the spatial frequency dimension, so coming from the different subbands. This pooling is divided in 2 successive stages : angular pooling and radial pooling[7]. Angular pooling is made with a Minkowski summation allowing to balance errors with their levels. For radial pooling, we have showed that a good compromise consists in balancing the perceptual errors of different sub-band. Indeed error appreciation varies with frequency subband, we propose a linear combination for this pooling.

Spatial pooling means merge on the spatial dimension. This polling is the most significant and sensitive. It implies a weighting of the errors according to their location in the image calling upon still ignored mechanisms of the SVH. We have proposed previously several ways to proceed. Here, we use a simple implementation taking into account the following characteristics:

- weighting by density of errors as as several errors distributed in all the image should be less annoying than the same errors concentrated in a particular region. The idea of this approach is to take into account of the density of the errors in areas of size comparable the fovea field. Indeed, several errors distributed in all the image can be less annoying than the same errors concentrated in a particular region. At each site (m,n), we calculate the density of the errors in a window of one degree width then we normalize by the maximum density observed in the image field.

- weighting according to the error structure according to the error structure defining 4 error classes. This weighting techniques has be elected because we think that annoyance appears when semantic content is corrupted. While we cannot access directly to the semantic, we consider that structure errors can be a good marker of troubleshooting of semantic. So, Wwe detect error contours to give more importance to this type of errors.

When we have applied these two weightings, we use a Minkowski summation in order to generate the quality score. All the parameters of the method have been fixed using subjective data in a previous study.

5. RESULTS

5.1. Subjective evaluation database

We have constructed a database in order to compare the criterion performance with human judgements. We several original images (natural scenes) and 3 different coding schemes : JPEG, JPEG 2000 and LAR which is a ROI-based method[13]. These algorithms have the advantage to generate very different type of degradations so we can take benefit to evaluate how the criterion can be useful for coding scheme comparison. All images were compressed with 5 different rates that lead to 165 images in the database. Subjective evaluations were made in normalized conditions at viewing distance of 6 times the screen height using a DSIS (Double Stimulus Impairment Scale) method with 5 categories. Distorted images were evaluated twice by 14 observers, aged from 22 to 43 years.

5.2. Performance evaluation

In order to define parameters values of the errors pooling, we have separated our database into two bases : one for the learning process and one for the evaluation process.

Learning process means that we have optimized parameter values of error pooling in order to get the best correspondence between subjective evaluation (MOS for Mean Opinion Score) and objective evaluation given by the criterion (MOSp for Mean Opinion Score predicted). Then, we use the values defined with the learning process to evaluatte the performance of our criterion to provide a good correspondence with human judgements on the evaluation base. Evaluation process have been conducted using the metrics recommended by VQEG[14]. These metrics allows to qualify criterion relating to accuracy, monotonicity and consistency.

Metrics relating to accuracy : two metrics are provided to qualify accuracy of a criterion. The first one is the simple

$$M1=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(MOS(i)-MOS_p(i)\right)^2}$$

root-mean square error between MOS and MOSp :

For our criterion, we get a value of 0.5302.

The other metric relating to accuracy is the 95% inverse-confidence interval weighted root mean square error :

$$M2=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{MOS(i)-MOS_p(i)}{IC_{95\%}(i)+0.025}\right)^2}$$

With $IC_{95\%}$(i) confidence interval for the ith point (of N points).

This second metric allow to appreciate accuracy of the prediction regarding the precision on subjective evaluation. A value of 1 indicates that the error between MOS and MOSp is comparable to the $IC_{95\%}$. With our criterion, we get a value of 0.7231 suggesting that we can consider that our method is fair accurate regarding to the accuracy of subjective evaluation.

Metric relating to monotonicity : For classical data (not quantified), the purposed metric consists in computing the Pearson linear correlation between MOS and MOSp :

$$CC=\frac{\mathrm{cov}ar(MOS,MOS_p)}{\sigma_{MOS}.\sigma_{MOS_p}}$$

In our case, we get a cc of 0.9407 indicating a good correlation between MOS and MOSp.

Metrics relating to consistency : Two metrics are used here. The first one is an outlier ratio of outlier pints to total

$$\left|MOS(i)-MOS_p(i)\right|>2.\sigma_{MOS}(i)$$

points. Where an outlier point is a point for which :

In our case, we get only 3.08% of outlier points.

The second metric relating to agreement of the metric is the Kappa coefficient, computed on quantized data (MOS and

$$Kappa=\frac{\sum_{c=1}^{5}f_{0(c)}-\sum_{c=1}^{5}f_{E(c)}}{N-\sum_{c=1}^{5}f_{E(c)}}$$

MOSp). The Kappa coefficient is given by :

Where fo is the observed number of agreement between MOS and MOSp for each of the 5 MOS class, and Fe is the number of agreement due to coincidence. The Kappa values are between –1 and 1, a value of 0.4 indicates that the method is efficient. In our case, we get a value of 0.5306 suggesting that we can consider that the method is fair efficient.

6. CONCLUSION

In this paper, we have first described a human visual system model for perception of color images. The originality of this model is that it is fully based on psychophysics experiments and adapted to image quality assessment. Most of the experiments have been conducted in our lab assuming coherence between the different steps of the model. This model is quite complete since it includes most of the well known properties of the human visual system, even the inter pathways masking. We have paid a special attention on the design of the error pooling. We have proposed some new methods and strategies to conduct it. Most of authors in literature focus attention only on one stage (error representation or error pooling). Here, we propose a well balanced approach without using any information on the distortions introduced in the images.

Finally, we get an image quality assessment tool with full reference providing good performance, regarding to metrics defined by VQEG. As it is generic, this tool can be useful in order to evaluate image processing methods that introduce distortion in image such as lossy coding schemes. An adaptation of this FR metric to get a RR metric has been already studied and futher works will focus on extension for video quality assessment.

REFERENCE

[1] Y. Horita, M. Katayama, T. Murai, and M. Miyahara, “Objective Picture Quality Scale for Video Coding”, ICIP, Lausanne, Vol. III of III pp. 319-322, 1996.

[2] A. B. Watson, “DCTune: A Technique for visual optimization of DCT quantization matrices for individual images “ Society for Information Display Digest of Technical Papers XXIV, 946-949 ,1993.

[3] Y. Horita, J. Ohnishi and T. Murai “Quality evaluation model for coded J2000 still picture”, in Proc. of Picture Coding Symposium proceedings, Seoul, 2001.

[4] Stefan Winkler: "A perceptual distortion metric for digital color video." in Proc. SPIE Human Vision and Electronic Imaging Conference, vol. 3644, pp. 175-184, San Jose, California, 1999.

[5] M. D’Zmura, T. J. S. Shen, W. Wu, H. Chen and M. Vassiliou : “Contrast gain control for color image quality”, in Proc. SPIE Human Vision and Electronic Imaging Conference, vol. 3299, pp. 194-201, San Jose, California, 1998.

[6] 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.

[7] P. Le Callet, and D. Barba, ”Image Quality Assesment : From Site Errors to A Global Appreciation of Quality”, in Picture Coding Symposium proceedings, Seoul, 2001.

[8] 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.

[9] D. R. Williams, J. Kraukopf, and D. W. Heeley, “Cardinal directions of color space”, Vision Research, Vol. 22, pp. 1123-1131, 1982.

[10] H. Senane, A. Saadane and, D. Barba “The computation of visual bandwiths and their impact in image decomposition and coding”, International Conference and signal Processing Applications and Technology, Santa-Clara, California, pp. 766-770, 1993.

[11] P. Le Callet, A. Saadane, and D. Barba, “Orientation selectivity of opponent-colour channels”, in PERCEPTION ,vol. 28 supplement ECVP'99 abstracts, Trieste, Italy, August 22-26, 1999.

[12] P. Le Callet, A. Saadane, and D. Barba, “ Interactions of chromatic components on the perceptual quantization of the achromatic component”, in Proc. SPIE Human Vision and Electronic Imaging Conference,vol. 3644, San Jose, California, January 23-29, 1999.

[13] O. Déforges and, J. Ronsin, “Locally Adaptive Method for Progressive Still Image Coding", IEEE International Symposium on Signal Processing and its Applications (ISSPA), Brisbane, Australia, 22-25 august 1999.

[14] A. M. Rohaly, P. Corriveau, J. Libert, A.Webster, V. Baroncini, J. Beerends, J. L. Blin, L. Contin, T. Hamade, D. Harrison, A. Hekstra, J. Lubin, Y. Nishida, R. Nishihara, J. Pearson, A. F. Pessoa, N. Pickford, A. Schertz, M. Visca, A. Watson and S. Winkler “videoquality experts group : current results and future directions”, in Proc. of Visual communications and Image Processing, vol. 3, pages 742-753, 2000