PERCEPTUAL BIT ALLOCATION FOR MPEG-2 CBR VIDEO CODING

O. Verscheure, A. Basso, M. El-Maliki and J.P. Hubaux

TCOM Laboratory, Telecommunications Services Group
Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland
E-Mail : verscheure@tcom.epfl.ch
URL : http://tcomwww.epfl.ch/~verscheu/

ABSTRACT

In this paper we propose a novel bit allocation scheme for MPEG-2 CBR video coding using a model of the early stages of human vision. On the basis of this model we derive a measure of the macroblock activity different from the one proposed in the MPEG-2 test model 5 (TM5). Experiments, obtained by substituting the proposed perceptual activity measure to the one proposed by the MPEG-2 TM5, yield better results among which improvement of the perceptual quality for a fixed bitrate and vice-versa.

Keywords: Vision model, perceptual activity measure, adaptive quantization, MPEG

1. INTRODUCTION

This decade will be characterized by the emergence of a very large number of audiovisual services. ATM technology, efficient compression techniques, and other developments of telecommunications make possible the offer of such services. Among those, MPEG-2 based Video-on-Demand services are a fundamental topic of current research in multimedia communications as they should have a competitive price comparing to video rental. One of the major issues is that, objective measures that are coherent with quality as perceived by human observers are only recently beginning to emerge.

This work addresses MPEG-2 CBR video coding in terms of perceptual video quality improvement using a complete spatio-temporal model of the human visual system. The paper is divided as follows : the spatio-temporal vision model is presented in Sec. 2. The MPEG-2 TM5 bit allocation scheme and its main drawback are subjects of Sec. 3. Section 4 introduces the perceptual activity measure and Sec. 5 presents the perceptual adaptive quantization scheme. Experimental results are presented in Sec. 6. Finally, Sec. 7 concludes the paper.

2. A MODEL OF THE HUMAN VISUAL SYSTEM

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of the Human Visual System [1]. A spatio-temporal model of human vision has been developed for the assessment of video co-ding quality [2, 3]. The model is based on the following properties of human vision:

- The human visual system (HVS) decomposes the visual information into "channels". A channel is characterized by a localization in spatial frequency, spatial orientation and temporal frequency. The structure of the channels is known. The visual information is decomposed into 5 spatial frequency bands, 4 spatial orientation bands and 2 temporal frequency bands [4, 5].

- In a first approximation, channels can be considered to be independent.

- Sensitivity to contrast is a function of the frequency and orientation. This defines the contrast sensitivity function (CSF) which has been specifically estimated for video coding noise [2].

- Interactions between two stimuli are quantized by masking which corresponds to a modification of the detection threshold of a stimulus as a function of the local contrast of the background.

The vision model described in [2] has been used to build a computational quality metric for moving pictures [3] which proved to behave consistently with human judgments. Basically, the metric, termed Moving Pictures Quality Metric (MPQM), decomposes an original sequence and a distorted version of it and computes a perceptual error measure accounting for contrast sensitivity and masking.

Recently, an improved metric, termed Normalized Video Fidelity Metric (NVFM), based on a finer modeling of vision has been introduced in [6].

3. MPEG-2 TM5 RATE CONTROL AND ADAPTIVE QUANTIZATION

The TM5 rate control and adaptive quantization are divided into three steps, namely target bit allocation, rate control via buffer monitoring and adaptive quantization.

- **Target Bit Allocation**
  This step estimates the number of bits available to code the next picture. A complexity measure is computed : the global complexity measures assign relative weights to each picture type (I,P,B). I pictures are usually assigned the largest weight since they have the greatest stability factor in an image sequence.
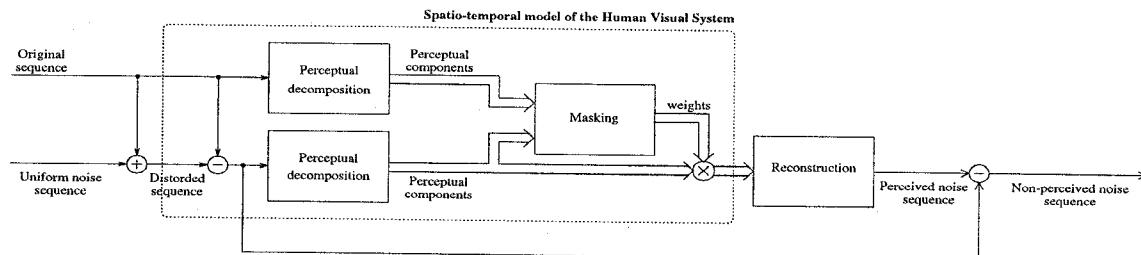
Figure 1: *Generation of the non-perceived noise sequence.*

B pictures are assigned the smallest weight since B energy do not propagate into other pictures and are usually more correlated with neighboring P and I pictures than P pictures are.

The bit target for a frame is based on the frame type, the remaining number of bits left in the Group of Pictures (GOP) allocation, and the history of previously coded pictures.

- **Rate Control via Buffer Monitoring**

Rate control attempts to adjust bit allocation if there is significant difference between the target bits and actual coded bits for a block of data.

Before encoding macroblock j ($j \geq 1$), the fullness of the virtual buffer of picture I, P or B is computed as

$$d_j^I = d_0^I + B_{j-1} - \frac{T_I \cdot (j-1)}{MBcnt} ,$$

$$d_j^P = d_0^P + B_{j-1} - \frac{T_P \cdot (j-1)}{MBcnt} ,$$

$$d_j^B = d_0^B + B_{j-1} - \frac{T_B \cdot (j-1)}{MBcnt} ,$$

where d_0^I, d_0^P, d_0^B are initial occupancy levels of virtual buffers ,B_j is the number of bits generated by encoding all macroblocks in the picture up to and including macroblock j, and $MBcnt$ is the number of macroblocks in the picture.

The reference quantization parameter Q_j for macroblock j is then computed as :

$$Q_j = \frac{31 * d_j}{r} ,$$

with the *reaction parameter r* given by :

$$r = \frac{2 * BitRate}{PictureRate} .$$

- **Adaptive Quantization**

The final step modulates the macroblock quantization step size by a local activity measure.

The activity act_j for a macroblock j is chosen as the minimum among the four 8x8 block luminance variances var^{sblk}_j :

$$act_j = 1 + min(var^{sblk}_j) .$$

The activity measure for the macroblock j is then normalized against the mean activity value (avg) of the most recently coded picture of the same type (I, P, or B) :

$$Nact_j = \frac{2 * act_j + avg}{act_j + 2 * avg} .$$

The quantization scale factor $mquant_j$ for macroblock j is finally obtained by means of the following expression :

$$mquant_j = Q_j * Nact_j ,$$

where Q_j is the reference quantization parameter obtained in the previous step.

Choosing the block of minimum activity means that a macroblock is no better than the block of highest visible distortion (weakest link in the chain). However, in some cases, quantization can produce blocking artifacts indicating that the quantizer parameter is not fine enough. Furthermore, the variance measure as used in MPEG-2 TM5 is poorly correlated with the human perception and can not be used to reliably measure activity in blocks. Therefore, the next section introduces the *perceptual activity measure* computed with a spatio-temporal vision model.

4. PERCEPTUAL ACTIVITY MEASURE

In this section, our objective is to determine a way to classify sequence areas in terms of their relevancy to human perception. In other words, we need to find a **local perceptual activity measure** that can be estimated without knowledge of video encoding process.

Let's consider the block diagram presented in Fig. 1. A uniformly distributed white noise is first added to the original video sequence . For this purpose, a zero-mean uniform noise has been used. We will see in the next section that results are independent of the noise variance chosen [7]. The sequence corrupted by the white noise and the original one are input to the vision model described in Sec. 2. The output of the vision model is reconstructed to compute the perceived noise sequence. The procedure permits to account for pattern sensitivity, accounting for frequency sensitivity and visual masking. The final step consists of computing the *non-perceived noise sequence* $npn(x, y, t)$ which is simply, by definition, the difference between the input noise sequence and the perceived noise sequence.

In this sequence, high activity zones (i.e. zones where the variance is large) correspond to zones of the original sequence in which coarse quantification is possible. Likewise, low activity zones (i.e. zones where the variance is low) correspond to zones in which coarse quantification will introduce visible artifacts (i.e. visual masking does not occur in such areas).

Therefore, we define the perceptual activity measure of a block as the variance of the corresponding block in the non-perceived noise sequence :

$$PAct(a,b,t) \triangleq \frac{1}{X \cdot Y} \sum_{x=a}^{X} \sum_{y=b}^{Y} [npn(x,y,t) - npn_{mean}]^2 .$$

5. PERCEPTUAL ADAPTIVE QUANTIZATION

As previously indicated, the variance measure, as computed in the third step of the MPEG-2 TM5 (Adaptive Quantization), is poorly correlated with human perception. Therefore, we replace the variance measure by the 8×8 block based perceptual activity measure.

We then compute the 8×8 block activity $PAct_j^{sblk}$ on the non-perceived noise sequence and compute the activity $PAct_j^*$ for a macroblock j as :

$$PAct_j^* = \alpha + min(PAct_j^{sblk}) .$$

The value of α is lower than 1 to take into account the fact that in this case, the variance is computed on an error sequence and no more on the original one. In our simulations, we took α equal to 0.001.

As proposed in TM5, we first normalize $PAct_j^*$ against the most recently coded picture of the same type and then we modulate the quantization scale factor $mquant_j^*$ by :

$$mquant_j^* = Q_j * Nact_j^* ,$$

where the normalized parameter $Nact_j^*$ is given by :

$$Nact_j^* = \frac{2 * PAct_j^* + avg^*}{PAct_j^* + 2 * avg^*} .$$

6. EXPERIMENTAL RESULTS

Simulations have been performed on 64 frames (512×512) of the Basket-Ball sequence. This sequence has been compressed at 15 different rates between 1 and 15 Mbits/sec. as interlaced video material, with a constant group of picture (GOP) structure of 12 frames and 2 B-pictures between every reference picture (I or P frames).

In Fig. 2, a frame from the Basket-Ball sequence is shown. On the basis of this frame, Fig. 3 shows the difference between the macroblock activity computed according to the MPEG-2 TM5 and the one computed with the perceptual activity measure. It is interesting to note the better matching between the activity map of our measure and the Basket-Ball image.

Fig. 4 and Fig. 5 show performance comparison of the bit allocation scheme using the perceptual activity measure

Figure 2: *A frame (256×256) from the Basket-Ball sequence.*

and the bit allocation scheme as proposed in MPEG-2 TM5, respectively, for PSNR and NVFM quality rating. It can be seen that the proposed scheme reduces the bitrate for a given quality or improved the quality for a given bitrate.

7. CONCLUSION

This paper presented a new perceptual activity measure. On this basis, a perceptual bit allocation for MPEG-2 CBR video encoding has been proposed. Results have shown interesting perceptual quality improvement for a fixed bitrate and vice-versa. Due to the importance of the subject, further work is planned in this field.

8. REFERENCES

[1] Serge Comes, *Les traitements perceptifs d'images numerisées*, PhD thesis, Université Catholique de Louvain, 1995.

[2] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", *in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2293–2296, Atlanta, GA, May 7-10 1996, available on http://ltswww.epfl.ch/~vdb.

[3] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *in Proceedings of the SPIE*, vol. 2668, pp. 450–461, San Jose, CA, January 28 - February 2 1996, available on http://ltswww.epfl.ch/~vdb.
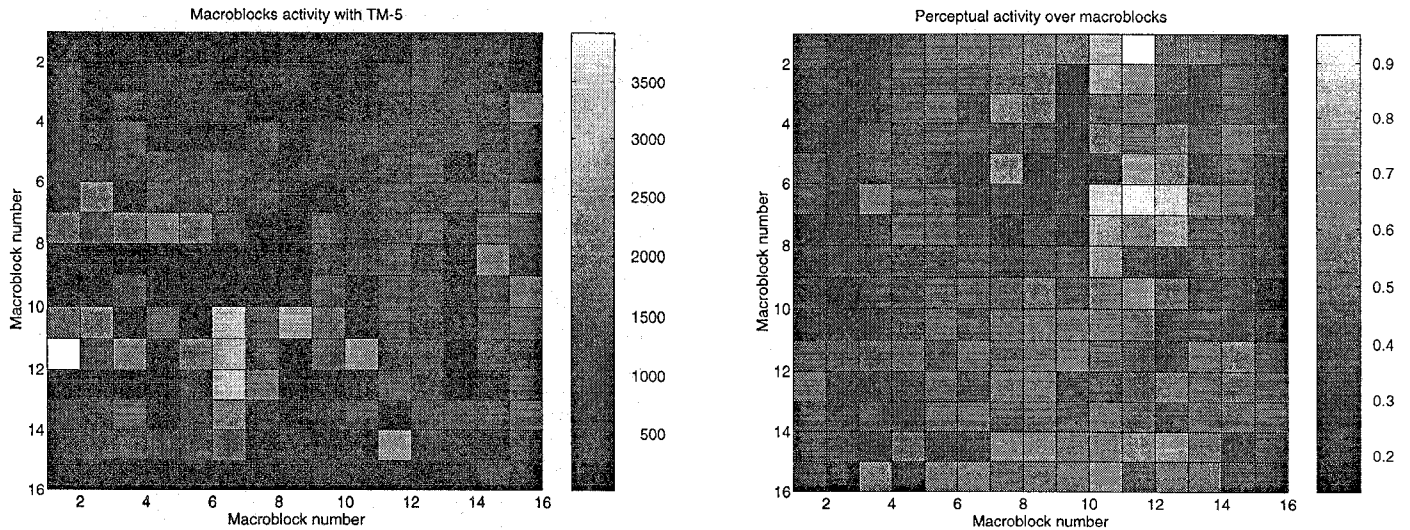
Figure 3: *Macroblock activity maps. Left : Macroblock activity computed according to MPEG-2 TM5. Right : Macroblock activity computed with the perceptual activity measure.*
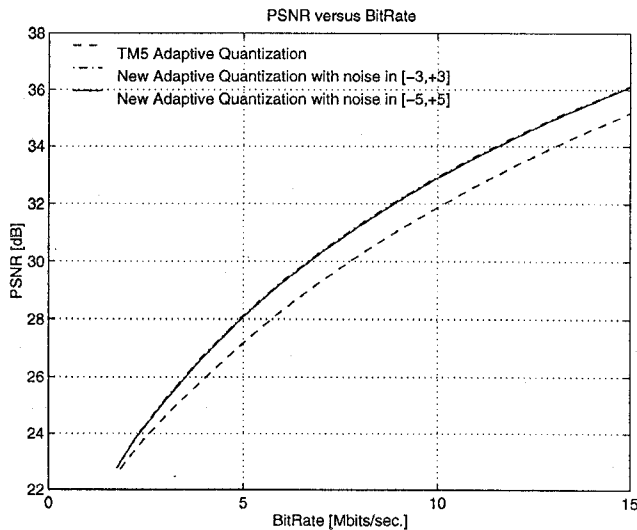

Figure 4: *Mean PSNR versus bitrate : comparison of MPEG-2 TM5 and the new perceptual bit allocation.*
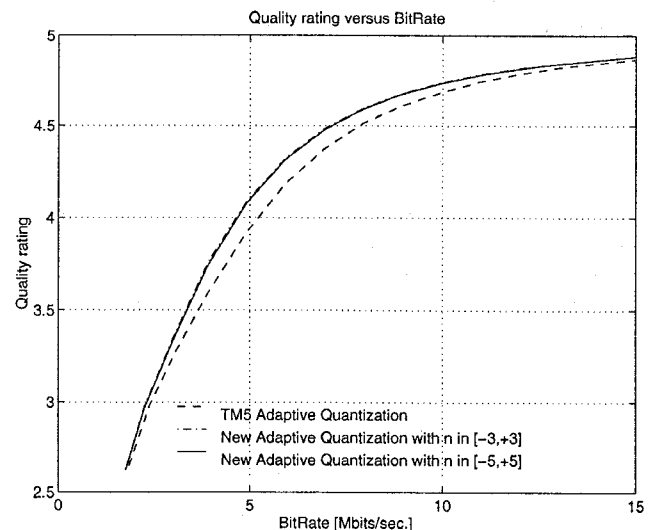

Figure 5: *Mean NVFM quality rating versus bitrate : comparison of MPEG-2 TM5 and the new perceptual bit allocation.*

[4] Andrew B. Watson, *Handbook of Perception and Human Performance*, chapter 6, Temporal Sensitivity, John Wiley, 1986.

[5] L. A. Olzak and J. P. Thomas, *Handbook of Perception and Human Performance*, chapter 7, Seeing Spatial Patterns, John Wiley, 1986.

[6] Pär Lindh and Christian J. van den Branden Lambrecht, "Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences", *in Proceedings of the International Conference on Image Processing*, Lausanne, Switzerland, September 16-19 1996, accepted for publication, available on http://ltswww.epfl.ch/~vdb.

[7] Olivier Verscheure, "Perceptual Video Activity Measure", Technical report, Swiss Federal Institute of Technology, June 1996, in preparation.