

Analysis of a Two Step MPEG Video System

Luís Teixeira (*) (+)

(*) INESC- Largo Mompilher 22, 4000 Porto Portugal

(+) Universidade Católica Portuguesa, Rua Diogo Botelho 1327, 4150 Porto, Portugal

Abstract: The efficiency of a rate control mechanism depends on the knowledge about the video complexity. The high number of different occurrences that can exist in a given video sequence makes it very difficult to encode at a constant bit rate while maintaining constant the picture quality. This paper presents a *'feed forward bit rate control'* to improve the perceptual quality of a video sequence. The algorithm also produces indexing information that allows detecting and classifying scene cuts.

Keywords: MPEG, scene change, scene complexity, bit rate control

1. Introduction

The success of the MPEG family of standards is based on the fact that they were designed to be generic. Therefore, they can be used in a wider range of applications when compared with their predecessors created to meet the requirements of one single application. MPEG1 was designed for multimedia storage applications. MPEG2 standard is expected to cover a larger field of applications such as Digital Terrestrial Television Broadcasting (dTTb), Video-on Demand, pay TV, SNG, ENG.

The MPEG rate control algorithm plays an important role for improving and stabilising the quality of the compressed video sequence. As MPEG does not specify how to control the bit rate several solutions have been presented in the literature [3,4,5,6]. There are two different approaches: *'feed forward bit rate control'* and *'feed backward bit rate control'*. In the last approach we have a limited knowledge of the sequence complexity and the encoder performance. Bits are allocated on a picture basis and spatially uniform distributed throughout the image. Thus too many bits may be overspent in the beginning of the picture while the end of the picture may present a higher degree of complexity. In the first approach, a pre-analysis is performed in order to determine the optimum setting, which will increase the computational complexity. The feed backward bit rate control is suitable for real time applications and feed forward bit rate control for applications where the quality is the main goal and time isn't a constraint. This paper describes a *'feed forward bit rate control'* that allows not only to improve the overall video sequence quality but also to extract additional information that will be useful for operations like indexing and browsing.

2. Bit Rate Control

The bit rate control may be expressed as follows: given a particular video input and a desired bit rate what should be the encoder setting to maintain the picture quality as high and constant as possible. In MPEG encoding, the trade-off between picture quality and the bit rate is controlled by a quantization scale, denoted as q . This parameter is used to compute the step size of the uniform quantizers used for the different AC DCT coefficients [2]. Each macroblock can be quantised with a different q . The scheme for adjusting the value of q between macroblocks within an image frame is called *'adaptive quantization'*. There are several schemes for doing the

adaptive quantization. For example, in MPEG Test Model 5 (TM5) [3], a non-linear mapping based on the block variance is used to adapt the q's. Besides the quantization scale, q, quantization coarseness depends also on the quantization matrix. In MPEG1, the quantization matrix can be altered in each sequence while in MPEG2 on a picture basis. The actual step size is obtained by multiplying the parameter q by a matrix of weights, one for each DCT coefficient.

There are lots of different characteristics in video sequences and the MPEG-2 encoder reacts differently to each of these characteristics. We have implemented a two step coding strategy. First, the video sequence is analysed (pre-coded) using fixed parameters. Statistics are obtained and the video quality is assessed, on a frame basis, in a high quality monitor (in case of assisted encoded). One of the goals is to obtain the *Bit Usage Profile*. The pre-analyser (first step) generates a histogram of the number of bits, the q signal, the PSNR for each macroblock (both for a particular position and regardless of the spatial position). The Bit Usage Profile allows us to determine how, when and where the complexity varies. Thus when the second step coding is performed, the final result presents an improvement of the perceptual picture quality.

2.1 Analysis of Scene Complexity

The first stage analyses the video sequence and extracts characteristics, such as local activity, location of frames that are particularly hard to code, etc., that will enable greater coding efficiency in the second stage. The quality with which a video sequence can be coded at a given data rate is largely dependent on its complexity. As the complexity of a video scene can vary widely we define measures of picture and scene complexity. All interframe processing, such as motion analysis, is also performed in the first stage.

Picture Complexity

The picture complexity is measured by computation of the average value of spatial local activity. Spatial local activity, for the macroblock j, is measured from the four luminance frame-organised sub-blocks and the four luminance field-organised sub-blocks using the original pixel values:

$$act_j = 1 + \min_{sblk=1,8} (var_sblk) \quad (1)$$

where

$$var_sblk = \frac{1}{64} \sum_{k=1}^{64} (P_k - \bar{P})^2 \quad (2)$$

$$\bar{P} = \frac{1}{64} \sum_{k=1}^{64} P_k \quad (3)$$

and P_k are the pixel values in the original 8×8 block.

Hard to code pictures

We defined a *hard to code picture* as a picture for which the prediction fails, resulting in large prediction-error energy and a highly non-smooth vector field. A *hard to code picture* is detected when a mismatch takes place between the coding model (which assumes purely translation motion) and the content of the images (which may contain newly revealed or occluded regions, non-transitional motion or a scene change). To solve this problem, one possible solution would be to encode the *hard to code picture* with intra-coded blocks. As the overhead would still remain high, the *hard to code picture* is associated with the beginning of a new GOP, becoming an Intra frame. Thus there is an alignment between the Intra frames and *hard to code pictures*.

Scene background

In most scenes part of the image does not change within a GOP time frame. The macroblocks belonging to fixed parts of the image, the background, need only to be encoded in the first image of the GOP (as intra). Usually, due to a coarse coding the same spatial macroblock is encoded again in P and B frames (as error information). The presence of noise may affect this decision. Whenever the difference between the sum of pixel values of macroblocks in two consecutive frames, at the same spatial position, is inferior to a fixed threshold, the macroblock is classified as fixed background.

Video Sequences Characteristics

The characteristics studied where: scene cuts, noise characteristics, level of detail and movement. The encoder automatically detects scene cuts and divides the sequence into scenes. Each scene is classified according to the classes identified in MPEG4:

Class	Characteristics	Example
A	Low detail and low movement	Akiyo, Hall Monitor
B	Low detail and medium movement or vice-versa	Coast Guard, Silent
C	medium detail and high movement or vice-versa	Bus, Calendar & Mobile

Table 1) MPEG video classes

After characteristic extraction, each scene goes through a first step encoding process where the input parameters are set according to its class. In images where scene cuts occur all the macroblocks are coded as *Intra blocks* to insure that the block effect won't be visible in that frame.

2.2 Improved rate-control (second step)

In an one-pass scheme, since no information about the video sequence is known in advance, the target number of bits per frame is often estimate in a linear fashion, i.e. allocating to each frame of the same type the same number of bits. This scheme doesn't take advantage of information about local variations of activity and can conduct to inefficient performance. For example, in a GOP where a scene change

occurs within the first frames, a linear distribution of bits as described above is sub-optimal and the perceptual quality of the reconstructed video may be compromised.

As information regarding picture complexity, scene complexity and bit allocation is available at the second encoding stage a different strategy should be used to control rate and buffer. In the first stage, not only characteristics of the video sequence are extracted but also a representation of the data with constant quality is produced: the *Bit Usage Profile*. In the second stage, a rate-control scheme is implemented allocating bits to each frame proportional to the number of bits used for its constant quality representation and complexity. The number of bits necessary to encode each frame in the first stage is a measure of the exigency associated with that frame, and, for constant quality, the target number of bits allocated to each frame in the second stage should be proportional to this exigency.

The target number of bits to encode frame i , for the second stage, is determined by

$$T_i = K_{I,P,B} \times \frac{S_i}{S_{GOP} - S_{i'}} \times R \quad (4)$$

where $K_{I,P,B}$ are weighting factors (introduced to compensated the effect of the distinct quantization matrices applied to distinct picture types), S_{GOP} , the number of bits spent on the first stage encoding the GOP which frame i belongs to, S_i the number of bits spent encoding frame i and $S_{i'}$, the number of bits spent on the first stage encoding up to frame i and R is initially equal to (5) (at the beginning of the GOP) and updated after encoding a picture (6):

$$R = \frac{\text{bit rate}}{\text{frame rate}} \times N^{\circ_of_frame_GOP} \quad (5)$$

$$R = R - S_{I,P,B} \quad (6)$$

where $S_{I,P,B}$ is the number of bits generated to encode the last image.

To determine the weighting factors, $K_{I,P,B}$, several simulations were performed to measure the effect of distinct quantization parameters in the number of allocated bits in variable bit rate, using several SIF625 sequences (Table 2, Figure 1).

Q	2	4	6	8	10	12	14
Frame I	1.6	2.3	2.9	3.7	4.5	5.5	6.6
Frame P	1.5	1.8	2.1	2.5	2.8	3.2	3.6
Frame B	1	1	1	1	1	1	1

Tabel 2.a) Bits Frame Ratio (q)

Sequence	Bit usage	SNR
Flower Garden	$3227.40 m_q p^{-0.6892}$	$47.437 m_q p^{-0.2090}$
Coast Guard	$1798.49 m_q p^{-0.7351}$	$45.030 m_q p^{-0.1850}$
Bus	$2095.52 m_q p^{-0.7042}$	$42.463 m_q p^{-0.2303}$
Akiyo	$812.24 m_q p^{-0.5897}$	$42.581 m_q p^{-0.1122}$

Table 2.b) Modeling bit and SNR in function of quantization step

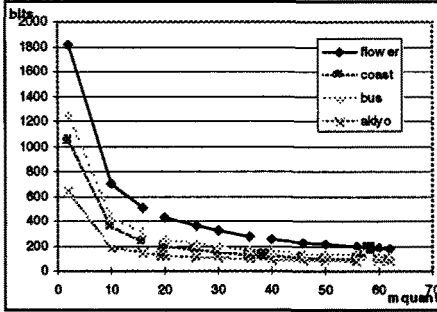


Figure 1.a) MB Bits (q)

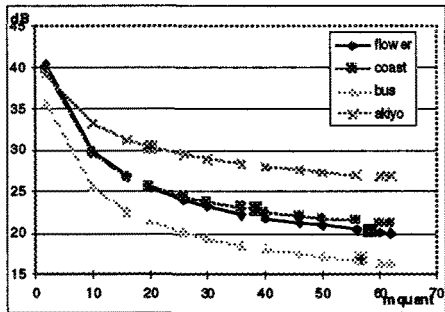


Figure 1.b) MB SNR (q)

The parameter q is modulated by the local and global activity of, respectively, the macroblock and the image.

This bit allocation strategy is susceptible of providing a non-linear bit allocation, as determined by the video sequence characteristics. If, for example, the complexity of the first frame is such that it requires 30% of the total number of the GOP bits in the first stage, the target set for the second stage will also be 30% of the total number of bits available for the GOP. If a linear strategy were applied, this frame would receive roughly 8% of the GOP bits (considering a GOP of 12 frames length). Nevertheless, if a linear allocation is required by the characteristics of the video sequence, it can also be easily achieved.

2.3 Simulation Results

Simulations were performed using SIF625 sequences. On the first stage, we have obtained the complexity of the video sequence and a measure of the exigency associated with each frame: the number of allocated bits in VBR mode (figure 2).

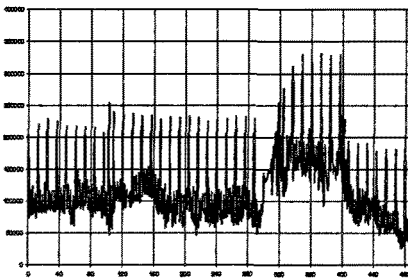


Figure 2.1) Bits per frame (q)

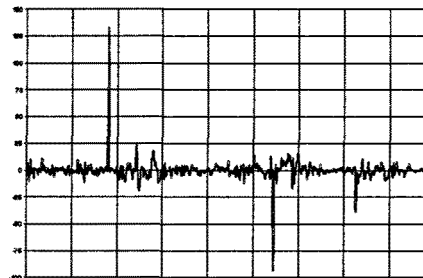


Figure 2.2) Differential Local Activity

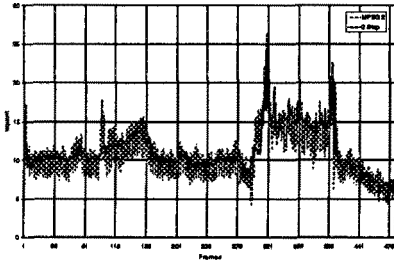


Figure 3.1) Quantization step size

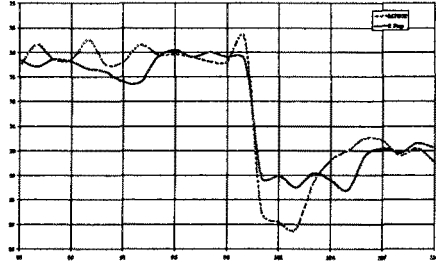


Figure 3.2) PSNR

From the local activity, we have obtained the differential local activity by subtracting adjacent local activity values. By correctly detecting shot scenes (pictures 100, 300 and 400), an improved bit rate algorithm control was developed.

Using the information obtained in the first step the sequence is encoded. In general we obtain a small gain in PSNR (0.3 dB) but the improvement in subjective quality is larger than what might be expected from such gain (mainly in shot scenes and hard to code pictures). The reason is the increased smoothness in the temporal variations of the coding quality and the reduction of the negative peaks associated with scenes changes.

Work with other video sequences is being performed to validate initial results. The way these techniques will interact is also being studied.

3. Discussion

This paper describes a two-stage encoding process. Quality improvement is obtained by using a more global model of the source characteristics. Improvement thus results from two passes with look-ahead bit allocation, special processing of scenes changes and an intelligent rate-control scheme. Future work will also address the problem of detection and classification of scene breaks (including cuts, fades, dissolves and wipes). To achieve this goal we are using techniques to determine real motion such as phase correlation.

4. Acknowledgements

This work was supported by the Junta Nacional Científica e Tecnológica (grant PRAXIS-XXI /4/4.1/BD/2256).

5. References

- [1] ISO/IEC 11172-2, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s-video", Geneva, 1993.
- [2] ISO/IEC IS 13818-2: Generic coding of moving pictures and associated audio, November 1994
- [3] ISO/IEC-JTC1/SC29/WG11 MPEG93/457 MPEG Video Test Model 5 (TM-5), April 1993.

- [4] K. Ramchandran, A. Ortega and M. Vetterli, "Bit allocation for dependent quantization with applications to MPEG video codec", Proc. Internat. Conf. Acoust. Speech Signal Processing. 1993, Minneapolis, March 1993
- [5] Gerjan Keesman, Imran Shah, Rene Klein-Gunnewiek, "Bit-rate control for MPEG encoders", Image Comm., vol. 6 , pp. 545-560, June 1995
- [6] Liang-Jin Lin, A. Ortega, C.-C. Jay Kuo, "Gradient-based buffer control technique for MPEG", VCIP 95, Taipei, Taiwan, May 1995