Statistical Multiplexing of H.264 programs

Luís Teixeira[1,2], Luís Corte-Real[2]

Universidade Católica Portuguesa - CITAR

[1,2] R. Diogo Botelho, 1324, 4169-005 Porto, PORTUGAL

[2] Faculdade de Engenharia da Universidade do Porto / INESC PORTO

Rua Dr. Roberto Frias, nº 378, 4200-465 PORTO Portugal

{lmt, lreal}@inescporto.pt

Abstract- **The advent of H.264/AVC is going to change the way Digital Television programs are broadcast. Each program can be independently encoded or jointly encoded resulting thus in a more efficient way to distribute the available channel bandwidth. This paper presents a combined coding scheme for multi-program video transmission in which the channel capacity is distributed among the programs according to the program complexities. A complexity bit rate control algorithm based on the Structural Similarity Index (SSIM) is proposed. SSIM metric is presented under the hypothesis that the Human Visual System (HSV) is very specialized in extracting structural information from a video sequence but not in extracting the errors. Thus, a measurement on structural distortion should give a better correlation to the subjective impression. Current simulations have demonstrated very promising results showing that the algorithm can effectively control the complexity of the multi-program encoding process whilst improving overall subjective.**

I. INTRODUCTION

The H.264/MPEG-4 Advanced Video Coding standard (H.264/AVC) [1], also referred as ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10), is the latest video coding standard jointly developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). H.264/AVC has accomplished a considerable progress regarding coding efficiency, substantially enhanced error robustness, and increased flexibility and scope of applicability relative to its predecessors [2][3]. It covers all frequent video applications ranging from mobile services and videoconferencing to IPTV, HDTV, and HD video storage [4].

In TV multi-programme broadcast systems, the use of H.264/AVC standards allows economies in transmission bandwidth of programs, while providing a service with a higher quality regarding current systems [5]. TV viewers receive video programs from a number of different video content providers via mixed transmission channels. Consider a simplified broadcasting chain compose by a video coder connected to a video decoder via a multiplexer, digital broadcast channel and demultiplexer. In a fixed multiplexing scheme, each service is assigned to a predetermined part of the total available bandwidth, where the sum of the assign bandwidth for each channel is not greater than total channel bandwidth. One alternative is to allocate different bit rates to each video encoder based on the expected image complexity of the signal to be encoded. Such an allocation scheme should be dynamically adjusted over time, depending on the relative complexity of each channel. This process is called statistical multiplexing (stat-mux). Statistical multiplexing can be defined as:

1. the control required for allocation of bits in proportion to the complexity and importance of each video application within the limits of control allowed by each video encoder, such that:

a the aggregate instantaneous bit rate is less than or equal to the channel capacity;

b. the minimum quality of service (QoS) requirements for all applications are met; and

c. the quality is maximized for applications in the order of their importance; and

2. the control required in cases, where the aggregate instantaneous bit rate is greater than the channel capacity, to minimize the loss in QoS for as minimal a number of applications as possible.

To achieve these levels of control, statistical multiplexing takes into account the variations in bit rate of different video applications when allocating transmission bandwidth. Statistical multiplexing is a highly efficient method to make the best use of a given transponder or cable spectrum bandwidth while maintaining near-constant video quality across all video channels. However, it also presents a significant technical challenge when a custom channel lineup must be created from multiple statistically multiplexed bit streams. Specifically, when multiple channels are extracted from multiple independently statistically multiplexed bit streams, they must be combined to form a new statistically multiplexed bit stream. In this case, the peak bit rates of the various source streams could exceed the total fixed bit rate for the output stream. Because each bit stream may have been generated as part of its own statistical multiplexing process, its individual bandwidth allocation is no longer relevant to the new multiplex.

II. VIDEO QUALITY ASSESSMENT

Television programs are produced for the satisfaction of television viewers so their opinion regarding the video quality is rather important. For broadcasting applications, bandwidth is continuously a valuable resource; therefore we should ideally avoid encoding video information that is outside the human perception or attribute more or equal bits to encode the information of inferior perceptual significance. So incorporate HVS model into broadcasting encoding system is fundamental to additional improve coding efficiency and enhance video quality. We can consider two categories regarding existing video quality evaluation methods [6][7]: subjective testing (human observers provide their opinion regarding video quality) and objective measurements methods (performed with the support a

mathematical algorithm). Although subjective testing is an important part of evaluation system, in practice, subjective evaluation needs to organize the observers to mark the distorted images, which is too inconvenient, time-consuming and expensive. PSNR and MSE are still the most popular employed objective metrics due to their low complexity and clear physical meaning. Nevertheless, both metrics have been target of an high number of critics for not correlating well with HVS [8][9] as they can't signify the exact perceptual quality as they are based on pixel to pixel difference calculation and ignore human perception and the viewing condition

The impact of coding distortion on the subjective quality is still under investigation [10][11][12]. A new scheme for a class of quality metrics, known as structural similarity (SSIM), has been proposed to models perception implicitly by taking into account the fact that the Human Visual System (HVS) is adapted for extracting structural information (relative spatial covariance) from images [8]. Its application in a coding context has just been started to be explored with the work of Brooks modeling typical distortions encountered in video compression/transmission applications and deriving a multi-scale weighted variant of the complex wavelet SSIM (WCWSSIM), with weights based on the human contrast sensitivity function to handle local mean shift distortions [13]. In [14] Zhi-Yi Mai, proposes a R-D optimization using the structural similarity (SSIM) instead of SSD for quality assessment in H.264 I-frame encoder. Improvement of coding efficiency is still not very large, and it is still necessary to study the proposed R-D method regarding motion estimation of inter frame coding.

III. EXPERIMENTAL RESULTS & DISCUSSION

The MPEG rate control algorithm plays an important role for improving and stabilizing the quality of the compressed video sequence. The Rate Control can operate at various levels of video compression, namely, sequence-level, frame level, and macroblock (MB)-level. As MPEG does not specify how to control the bit rate several solutions have been presented in the literature. There are two different approaches: "feed forward bit rate control" and "feed backward bit rate control".

Figure 1. Block Diagram for feed-forward and feed-backward bit-rate control.

Fig. 1 presents diagrams blocks of the two different approaches. In the "feed backward bit rate control" we have limited knowledge of the sequence complexity. Statistical information is gathered by the encoders during the encoding process. This information can be used to determine the video complexity of the program. Bits are allocated on a picture basis and spatially uniform distributed throughout the image. In the "feed forward bit rate control", a pre-analysis is performed in order to determine the optimum setting, which will increase the accuracy of the complexity metrics. In this work we have followed a feed forward bit rate control. A key decision is what statistics should be used to describe the video complexity. In our approach SSIM metric was used.

A. Structural Similarity Index (SSIM)

SSIM is an objective image quality assessment metric which attributes perceptual degradations to structural distortions. The SSIM index has been demonstrated in [8] to be an effective measurement of perceptual global degradations in natural images. It successfully incorporates HVS characteristics without much added complexity. It is also a comparable metric to conventional error-based perceptual quality metrics. In essence, the SSIM index is a measurement of deviations in luminance, contrast and structure between the reference and the distorted images. Luminous, contrastive and structural degradations are represented by the following respectively:

$$l(x, y) = \frac{2\mu_x \mu_y}{\mu_x^2 + \mu_y^2}$$

$$c(x, y) = \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \qquad (1)$$

$$s(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

The SSIM index, shown below, is essentially a product of these three distortions.

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (2)$$

where $C_1 = (K_1 L)^2$ and $C_1 = (K_1 L)^2$ are added constants to ensure stability of the system. K1 and K2 were set to 0.01 and 0.03, respectively [8].

B. Rate Control Algorithm

Regarding statistical multiplexing for digital TV broadcast, sequence-level R-D control and optimization is performed to dynamically allocate the total bandwidth among the TV programs to maximize the statistical multiplexing gain, as well to maximize the objective quality according to the rate-distortion characteristics of the video objects [15][16]. Thus the optimal bit allocation aims to distribute the available bit budget amongst the different programs such that the overall distortion can be minimized:
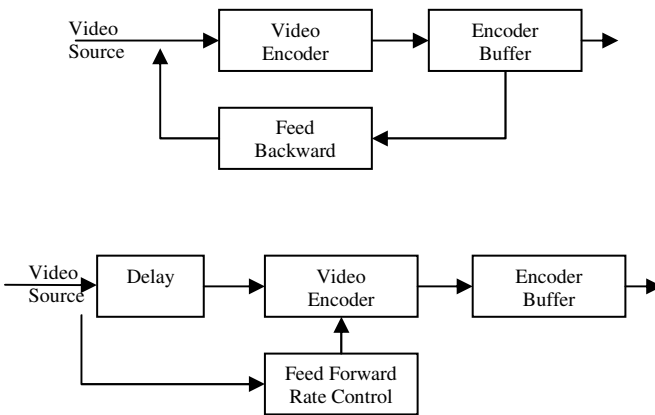
$$\min\left(\sum_{j=1}^{n} D_i\right) \text{ subject to } \sum_{i=1}^{n} R_i \le R_{\text{target}} \qquad (3)$$

where D_i and R_i denote the distortion and bit rate of the ith program respectively, n is the number of video programs. Figure 2 presents an example of a typical employing the proposed approach:
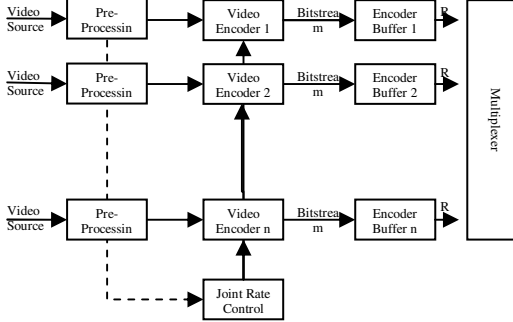

Figure 2. Block Diagram Statistical Multiplexing of H.264 programs.

Each video encoder produces compressed video and corresponding statistics. The joint rate module receives information regarding the relative complexities of each program and the channel buffer fullness. Each encoder changes its bit rate only when a new GOP begins.

In the first step, the reference bandwidth for the next GOP of each video source (BW_{ref}) is determined based on the total available transmission bandwidth, the picture coding complexity and type, GOP structure of each video source and the current state of the total virtual buffer. This is step follows normal H.264 JVT-G030 process.

In the second step, the complexity is measured by encoding at each frame at fixed quantization step size (QP=24) Each image block is first computed within local 4×4 non-overlapped windows and then all the local SSIMs are averaged to a mean SSIM during motion estimation. The SSIM of the whole reconstructed image for each component is computed alike but using a 16×16 slide window instead:

$$MSSIM = 0.7 \times SSIM_Y + 0.15 \times SSIM_U + 0.15 \times SSIM_V \ (4)$$

In the third step, the available bandwidth is allocated to each video source by considering the estimated bandwidth.

$$BW_i = \frac{X_{i_i}}{\sum_{i=1}^{n} X_i} \sum_{i=1}^{n} BW_{ref_i} \qquad (5)$$

where n is the number of video sources, Xi is the complexity of the program i. Finally, in the last step, each video sequence is encoded for a specific bit-rate using a quadratic rate-quantizer (R-Q) model according [21][22]:

$$R_i = C_i \times \left(\frac{\alpha_{1,i}}{Q_i} + \frac{\alpha_{2,i}}{Q_i \times Q_i}\right) \qquad (6)$$

Where R_i is the bits of current sequence I, C is the encoding complexity (sum of absolute difference), Q is the quantization parameter, $\alpha_{1,i}$ and $\alpha_{2,i}$ the model parameter

that are updated by linear regression method from previous encoded parameters [17].

C. Simulations

We have implemented the proposed rate control scheme as using the H.264 JM 10.2 encoder [18]. In this section we present some results from three typical sequences of various encode complexity listed in Table 1.

TABLE I
TEST SEQUENCES (RDO=ON).

Test Sequence	Size	Frame Rate	Frames encoded	Frame type
Akiyo	CIF	30	298	IPPP
Foreman	CIF	30	298	IPPP
Football	CIF	30	298	IPPP

The performance of our proposed scheme is evaluated in comparison with the original encoder JM 10.2 and the existing rate control functionality in the JM 10.2 if the different video sequences were separately encoded. Two scenarios were studied. First each sequence encoded at fixed bit rate of 256kbps and then at fixed 512kbps. Results were evaluated in terms of PSNR and subjective testing was conducted using SAMVIQ methodology [19][20].

The SAMVIQ methodology is currently being standardized within ITU-R. Experimental results using the SAMVIQ methodology may enable subjects to arrive at more appropriate quality ratings for content that they find difficult to judge on a single viewing.

Figure 3. Decoded frames using JM 10.2 (independent coding) for "Akiyo" and "Football" sequences (frame 35) encoded at 256 kbps

Figure 4. Decoded frames using proposed algorithm for "Akiyo" and "Football" sequences (frame 35) encoded at 256 kbps

Several combinations were simulated with jointly combine 3 video streams: Akiyo is represented with letter A, Foreman with letter B and Football with letter C.
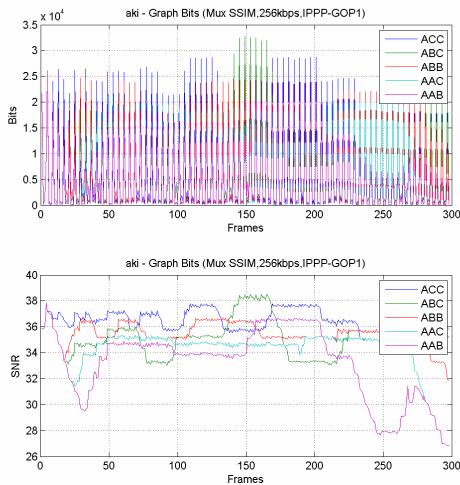
Figure 5. Bits and PSNR variation for Akiyo sequence, 256 kbps with different sequence combination
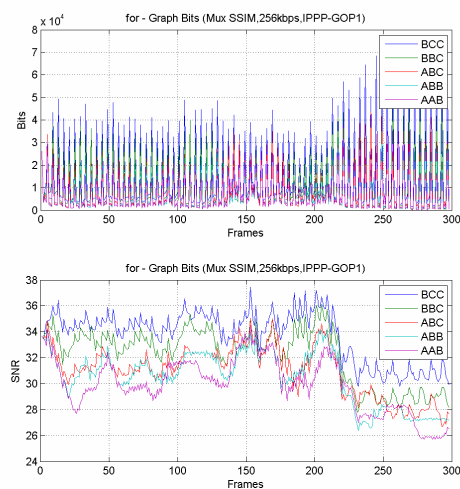

Figure 6. Bits and PSNR variation for Foreman sequence, 256 kbps with different sequence combination
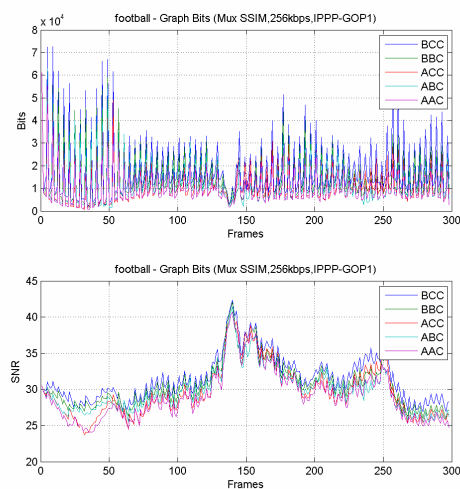

Figure 7. Bits and PSNR variation for Football sequence, 256 kbps with different sequence combination

In this paper we present an algorithm for dynamic bandwidth allocation of H.264 video programs which allots the available bandwidth according to the needs of each video source. The joint rate algorithm is applied at a GOP level resulting in improvement of the picture quality variation.

Preliminary results show that bandwidth gains/quality improvements are more significant when heterogeneous sources are multiplexed together. A good trade-off regarding video quality is observed within the multiple programs as well as within a sequence, compared to independent coding. An increased in subjective quality may be observed in football sequence while the decrease in Akiyo is rather imperceptible. Furthermore, results show that the proposed algorithm results in more uniform picture quality among programs as well as within a program, compared to independent coding. Joint coding can thus improve channel utilization by dynamically distributing the channel bandwidth among video programs according to their respective complexities. Additionally, the advantages of joint coding are obtained even with a reduce number of programs.

REFERENCES

[1] ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Rec. H.264 & ISO/IEC 14496-10, Version 1, May 2003; Version 2, Jan. 2004; Version 3 (with High family of profiles), Sept. 2004; Version 4, July 2005 [Online]. Available: http://www.itu.int/rec/T-REC-H.264

[2] Joern Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi, "Video coding with H.264/AVC: Tools, Performance, and Complexity", IEEE Circuits and Systems Magazine, Vol. 4, No. 1, pp. 7-28, First Quater 2004

[3] A. Luthra, G.J. Sullivan, and T. Wiegand, Eds., IEEE Trans. Circuits Systems Video Technol. (Special Issue on the H.264/AVC Video Coding Standard), vol. 13, no. 7, July 2003.

[4] Thomas Wiegand and Gary J. Sullivan,"The H.264/AVC Video Coding Standard {Standards in a nutshell}", IEEE Signal Processing Magazine, Volume 24, No. 2, March 2007.

[5] Thomas Wiegand, Heiko Schwarz, Anthony Joch, Faouzi Kossentini, and Gary J. Sullivan: Rate-Constrained Coder Control and Comparison of Video Coding Standards, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 688-703, July 2003.

[6] K. Seshadrinathan and A. C. Bovik , "New vistas in image and video quality assessment", SPIE Human Vision and Electronic Imaging, San Jose, California, January 29-01, 2007

[7] H. R. Sheikh, and A. C. Bovik, "Image Information and Visual Quality", IEEE Transactions on Image Processing, Vol: 15 No: 2, February 2006, Page(s): 430 – 444

[8] Wang Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity", IEEE Transactions on Image Processing, Vol. 13, No.4, pp600-613, April 2004.

[9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik , "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms", IEEE Transactions on Image Processing, Vol: 15 No: 11, November 2006, Page(s): 3440-3451

[10] Olivia Nemethova, Michal Ries, Eduard Siffel, Markus Rupp, Subjective Evaluation of Video Quality for H.264 Encoded Sequences, Sympo TIC, Bratislava, Slovakia, October 24-26, 2004

[11] Tobias Wolff, Hsin-Han Ho, John M. Foley, and Sanjit K. Mitra, "H.264 coding artifacts and their relation to perceived annoyance," in European Signal Processing Conference, 2006

[12] Mylène Farias, No-reference and reduced reference video quality metrics: new contributions, Ph.D. thesis, University of California, 2004

[13] Alan C. Brooks and Thrasyvoulos N. Pappas, "Structural similarity quality metrics in a coding context: exploring the space of realistic distortions", IEEE International Conference on Acoustics, Speech, and

Signal Processing, Honolulu, Hawaii, April 15-20, 2007, Pages(s): 869-872

[14] Y. Mai, C. L. Yang, L. M. Po, and S. L. Xie, "A New Rate-Distortion Optimization Using Structural Information in H.264 I-Frame Encoder", Lecture Notes in Computer Science, Vol 3708/2005, Springer-Verlag GmbH, Oct. 2005, pp 435-441

[15] M. Perkins and D. Arnstein, "Statistical multiplexing of multiple MPEG-2 video programs in a single channel," SMPTE Journal, vol. 104, no. 9, pp. 596–599, 1995.

[16] Luís Teixeira, Teresa Andrade, "Exploiting Characteristics of a large number of MPEG video sources for statistically multiplexing video for TV broadcast applications", 4 Bayona Workshop on Intelligent Methods in Signal Processing and Communications, Bayona, June 1996

[17] Z. Chen and K. N. Ngan, "Optimal bit allocation for MPEG-4 multiple video objects," in IEEE Int. Conf. Image Processing, Singapore, Oct. 2004.

[18] Joint Video Team (JVT), "H.264/Advanced Video Coding reference software version 10.2," 2006, http://iphome.hhi.de/suehring/tml/

[19] EBU-UER BNP 056: SAMVIQ – Subjective Assessment Methodology for Video Quality.

[20] Quan Huynh-Thua, Matthew Brotherton, David Hands, Kjell Brunnstr¨om, Mohammed Ghanbari, "EXAMINATION OF THE SAMVIQ SUBJECTIVE ASSESSMENT METHODOLOGY", Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07, Scottsdale, U.S.A., Jan. 25-26, 2007

[21] . Z. G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-G012r1, Mar. 2003.

[22] Keng-Pang Lim, G. Sullivan, and T. Wiegand, "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-W057, San Jose, Apr. 2007