SHOT-BASED RATE CONTROL FOR SINGLE-LAYER AND SCALABLE VIDEO ENCODING

Luk Overmeire[1], Fabio Verdicchio[2], Joeri Barbarien[2], Guy Lateur[1], Lode Nachtergaele[3], Peter Schelkens[2]

[1]E-mail: luk.overmeire@vrt.be Tel: +32-2-648-6142
[1]VRT, Auguste Reyerslaan 52, B-1050 Brussels, Belgium
[2]Vrije Universiteit Brussel, Dept. ETRO, Pleinlaan 2, B-1050 Brussel, Belgium
[3]IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

ABSTRACT

An off-line shot-based rate control approach is proposed for controlling the distortion variation across successive shots of a video sequence when encoding with single-layer (MPEG-4 Baseline, MPEG-4 AVC, Windows Media 9) and scalable (wavelet) video codecs. Consistent quality is achieved by optimally distributing the available bits among the different shots, based on rate-distortion modeling of each shot. The algorithm is formulated for the download and progressive download distribution model. The method can be applied to any codec. To evaluate the performance of the algorithm, a video sequence consisting of several shots is encoded with state-of-the-art video codecs using standard rate control and using our proposed method. The quality consistency in both cases is thereafter compared based on the PSNR of the luminance component. The results indicate that the proposed technique improves the quality consistency significantly.

1. INTRODUCTION

A key challenge in variable bit-rate video compression is to achieve under given constraints a maximal, constant reconstruction quality for varying content. Procedures for rate-distortion optimization continuously make trade-off decisions between bit-rate and overall distortion. Different strategies can be followed. The focus can be on the minimization of the average distortion [1] or the maximum distortion [2] (minimax approach), or on the equalization of the distortion across the coding units [3] (lexicographically optimal approach). In this paper, we will concentrate on avoiding fluctuations of the quality (third strategy), rather than aiming at an optimal average quality. It has been found by visual evaluations that the end-user will judge an entire video sequence based upon the quality variations and the minimum quality [4]. In [3], a two-pass VBR bit allocation method is proposed based on the lexicographical ordering of all the blocks over the entire video sequence. This method is computationally quite intensive, especially for longer sequences and less accurate for dependent coding. In [4], a two-pass encoding system for MPEG-2 is described such that constant visual quality is obtained. Based on experimental results of a training set of different video scenes, a function is derived to predict the optimal bit-rates per frame based on the gathered first pass statistics. The method deals with scene changes. A disadvantage is that a large training set is needed in order to fit a reliable mathematical model to the experimental results. In both cases, the content modeling process is based on individual frames or even macro-blocks. This appears to be inherently inadequate in order to achieve equal quality across different scenes (see section 4).

In this paper, we present a simple, but effective video modeling preprocessing strategy to achieve quality consistency throughout the video sequence based on information delivered by shot segmentation. In [5], an efficient framework is presented that integrates advanced video analysis techniques: shot segmentation, key frame extraction and object segmentation. At present, the structuring metadata extracted by such frameworks plays only a limited role in the enhancement of video encoding. Essentially, in our approach, shot changes will not be treated as exceptional cases to deal with, but as a basic means to structure the input characteristics. Statistics will be gathered on a shot/segment basis and used to distribute the available bits over the different video shots such that constant visual quality is obtained.

In section 2, the optimization problem that needs to be solved is formulated. In section 3, the proposed shot-based rate distortion modeling and quality control mechanism is discussed. An evaluation of the performance of the algorithm and a discussion of the advantages of shot-based rate control are presented in section 4. In section 5, the conclusions of this work are formulated.

2. PROBLEM FORMULATION

The main targeted video distribution models are the file download and progressive download model, in which cases the decoder buffer size can be ignored as an encoding restriction. Progressive download is a pseudo real-time method situated between download and streaming. The whole file is downloaded, but playback starts while the download is still in progress, as soon as enough of the content is available (delay Δ). Frame-level rate control is out of the scope of this paper. We use the PSNR of the luminance component (PSNR-Y) as our primary means of measuring the quality, since it is the most widely accepted objective measure of visual distortion. The challenge we are presented with is to subdivide a video sequence of arbitrary length intelligently and to distribute the available total amount of bits B_{TOT} among the M different shots such that:

$$\begin{cases} \sum_{i=1}^{i=M} B_i = B_{TOT} \\ \forall i, j \in [1, M], i \neq j : PSNR_i = PSNR_j \end{cases} \quad \textbf{(1)}$$

where for each shot i, $i \in [1, M]$, B_i represents the amount of bits and $PSNR_i$ the average PSNR of the luminance component. For progressive download, an additional constraint has to be satisfied:

$$\begin{aligned} B_{cum}(t) &\leq R_{channel} \cdot (t + \Delta) \\ B_{cum}(T) &= B_{TOT} \end{aligned} \quad \textbf{(2)}$$

where $B_{cum}(t)$ is the cumulative bit-rate at time t, $R_{channel}$ the channel rate, Δ the accepted decoding delay and T the duration of the complete video sequence.

3. SHOT-BASED RATE CONTROL

Due to the (increasing) variety of coding options and frame interdependency, practical optimization solutions are only best effort approximations. The modeling of individual frames appears to be inherently inadequate in order to achieve equal quality across different scenes (see section 4). Hence, a more pragmatic approach would be to model more coarse-grained coding units that are not dependent of each other. The solution is to divide the input video into separate shots, using a shot segmentation technique. A shot is defined as a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuing action in time and space. In fact, the idea is to apply the principle of lexicographical optimization on shots instead of on frames or macroblocks. In most cases, the video characteristics (texture, motion) are relatively stable within a shot, which further improves the rate-distortion modeling accuracy and optimization. Thus, we can expect that state-of-the-art codecs are able to control the rate and quality of an individual shot adequately. Assuming the file download model, the proposed rate control algorithm consists of four phases:

- *Segmentation phase*: the entire video is segmented into individual shots based on cut and fade detection [5]. A fade transition is considered as an individual shot.
- *Rate-distortion modeling phase*: each shot is two-pass encoded at a sufficient number of bit rates and the corresponding average PSNR-Y of the shot is measured. This is illustrated in Figure 1. We assume a piecewise linear relationship between the measure points. By carefully selecting five bit rates, this is an acceptable approximation.
- *Bit-allocation phase:* based on the measured rate distortion models for each shot, the total number of available bits is distributed among the different shots, such that the average PSNR-Y is equalized (see equation 1). The target PSNR-Y is the maximum value for which the constraint on the total amount of bits is just satisfied.
- *Optimal encoding phase*: each shot is encoded separately at the calculated optimal bit-rate by the chosen state-of-the-art codec (second pass only suffices by reusing first pass statistics of the R-D phase). Then, the encoded shots are concatenated.

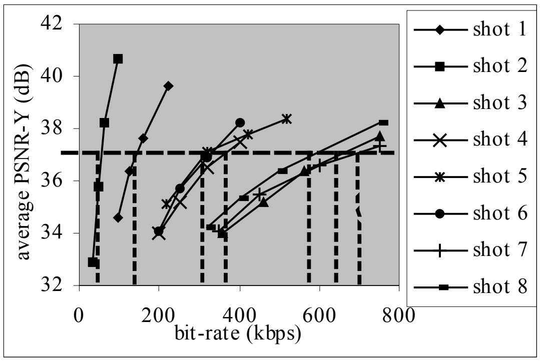

Figure 1: Rate distortion models of the 8 shots of the football news sequence for the wavelet codec. Shot-based rate control is applied by equating the average PSNR-Y of the shots, given a target bit-rate.

For progressive download (equation 2), clearly, if the complex shots are mainly preceding the easy ones, the requirement of the equality of the average PSNR-Y becomes untenable. In this case, the following bit-allocation algorithm is proposed in order to maximize the minimum PSNR-Y and to minimize the PSNR-Y variations between successive shots:

- *Step 1*: find maximum PSNR-Y for which (2) is satisfied for all shots (for simplification, the bit-rate is considered constant within a shot, which will be approximately the case), for a predefined Δ and any *t*. This defines the minimum bit-rate for each shot.
- *Step 2*: find the last shot for which (2) is strictly met, without any margin. Freeze the bit-rate of all

precedent shots. If the end of the input video is not yet reached, go to step 3.

- *Step 3*: increase the bit-rates of the next shots such that the average PSNR-Y increases progressively across the shots, until (2) is strictly met for one of the remaining shots. Go to step 2.

An additional benefit of the algorithm above is that the quality will never decrease.

4. EXPERIMENTAL RESULTS

In this section, we will present experimental evidence showing that our proposed shot-based rate control algorithm provides better quality consistency than the classical rate control algorithms present in state-of-the-art video codecs. We consider rate control in the context of existing video coding standards (MPEG-4, H.264/AVC [6]), a proprietary solution (Windows Media 9) and wavelet-based video coding (motion compensated temporal filtering [7], currently being studied in Part 13 of MPEG-21, focusing on scalable video coding), applied on the download model. Each coding technique has its own optimized rate control scheme, although it is not a normative tool for any coding standard. Rate-constrained coder control for MPEG-4 and AVC is discussed and compared in [8]. The consistency of PSNR-Y between classical and shot-based rate control is compared for MPEG-4, H.264, Windows Media 9 and wavelet video encoding. We used the following codecs, configured in VBR mode:

- **MPEG-4:** DivX Pro, version 5.1 (Simple Profile, Level 3), see http://www.divx.com/divx.
- **Windows Media 9:** Video Compression Manager, http://www.microsoft.com/windows/windowsmedia/9series/codecs/vcm.aspx.
- **H.264:** AHM 2.0, built on JM 6.1 (JVT-F086 added), http://groups.yahoo.com/group/jvt-remd.
- **Wavelet codec**: spatial domain motion compensated temporal filtering (SDMCTF), experimental instantiation used from [9].

The test sequence is a football news sequence (source: VRT). The used color format is YUV 4:2:0. The resolution is 384x224 and the frame rate 25 fps. The sequence consists of the following eight successive shots: female newsreader (frames: 1-155), male newsreader (156-664), players close-up (665-719), free kick (720-886), attack (887-1079), player close-up (1080-1138), goal (1139-1321), cheering (1322-1414). A key frame of each shot is shown in Figure 2. The test sequence is encoded at 256 kbps. The results are shown in Figure 3. The rate-distortion curves show an unstable behavior for each encoder when using classical rate control. Clearly, the measured quality is strongly related to the underlying encoded content.

Figure 2: Key frames from the shots comprising the test sequence.

The PSNR gradually increases and decreases in true harmony with the different shots and their complexity. This undesired effect is partly explained by the fact that the generic models used to predict quality and bit-rate for each frame cannot be perfect. Clearly, the quality variation is much smaller for each codec when using the shot-based rate control approach. Moreover, the PSNR-Y of the high-complexity frames has improved several dBs. In other words, at the same bit-rate, the number of low quality frames has decreased significantly. Our objective is met (while not exceeding B_{TOT}) at the price of a lower overall average PSNR-Y, as expected. We have observed similar effects for PSNR-U and PSNR-V. We also note that, although a stable (in-shot) rate control algorithm is a basic requirement for application of our approach, even the results for the less mature AVC are quite good. Due to the different PSNR-Y range (more than 25dB!) compared to the other codecs, the achieved gains for AVC might wrongly appear less impressive on the figure. For SDMCTF, a fluctuation in PSNR between successive frames is observed. A solution for this problem was presented in [10]. However, this solution was not implemented in the codec we had at our disposal.

5. CONCLUSIONS

We have presented an off-line shot-based rate control approach and algorithm which optimizes the consistency of the quality of the encoded video. The algorithm is codec independent. It perfectly fits into an integrated video analysis framework. The results indicate that the fluctuations in achieved quality over successive shots can be significantly reduced, at the expense of an average quality loss. The method is particularly interesting for scalable encoders, as almost no extra cost is involved due to the fact that a scalable video coder produces a single, embedded bitstream that immediately supports all quality levels. This approach has as an additional benefit that it allows for a parallelized execution of the bit-allocation processes. Finally, shot-based rate control perfectly fits into an integrated metadata extraction framework for a video production environment. The proposed framework can be improved by accelerating the R-D modeling phase and by adding a subjective quality component, which is part of our current research.

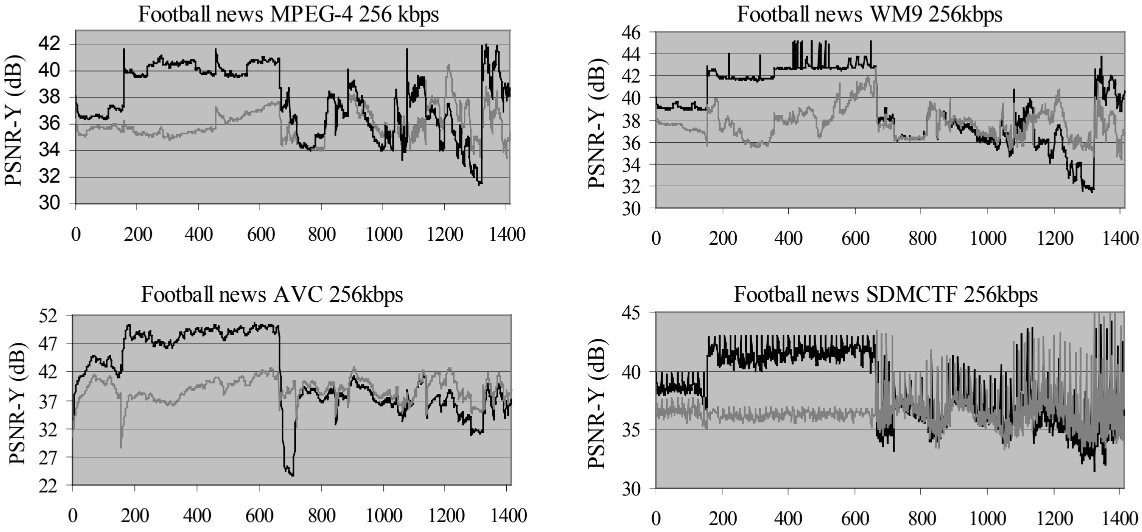

Figure 3: Comparison of quality consistency between classical (black line) and shot-based (gray line) rate control of a football news sequence for MPEG-4, AVC, Windows Media 9 and wavelet encoding. Horizontally, the frame number is displayed.

ACKNOWLEDGEMENTS

This work is a result of a joint collaboration between Vlaamse Radio en Televisie (VRT, public broadcaster of Flanders), Interuniversity MicroElectronics Center (IMEC) and Vrije Universiteit Brussel (VUB). This is one of the VRT projects funded by the Flemish government. P. Schelkens has a post-doctoral fellowship with the Fund for Scientific Research Flanders (FWO).

6. REFERENCES

[1] Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445-1453, September 1988.

[2] G. M. Schuster, G. Melnikov and A. K. Katsaggelos, "A Review of the Minimum Maximum Criterion for Optimal Bit Allocation Among Dependent Quantizers", *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 3-17, March 1999.

[3] D. T. Hoang, E. L. Linzer and J. S. Vitter, "Lexicographic Bit Allocation For MPEG-Video", *Proc. Data Compression Conference*, Snowbird, Utah, pp. 101-110, March 1997.

[4] P. H. Westerink, R. Rajagopalan and C.A. Gonzales, "Two-pass MPEG-2 variable-bit-rate encoding", *IBM Journal of Research and Development,* vol. 43, number 4, July 1999.

[5] T. Caljon, J. Barbarien, S. Rousseaux, L. Overmeire and L. Nachtergaele, "Integrated video shot segmentation, key frame extraction and object segmentation", *Proc. ACIVS*, Ghent, Belgium, pp. 124-131, September 2003.

[6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*, John Wiley & Sons, England, November 2003.

[7] D. Turaga and M. van der Schaer, "Wavelet Coding for Video Streaming Using New Unconstrained Motion Compensated Temporal Filtering", *Proc. Internat. Workshop on Digital Communication*, Capri, Italy, pp. 41-48, September 2002.

[8] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini and G. 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.

[9] Y. Andreopoulos, J. Barbarien, F. Verdicchio, A. Munteanu, M. van der Schaar, J. Cornelis and P. Schelkens, "Response to Call for Evidence on Scalable Video Coding", *ISO/IEC JTC1/SC29/WG11, M9911*, MPEG 65th meeting, Trondheim, Norway, July 2003.

[10] A. Munteanu, Y. Andreopoulos, M. van der Schaar, P. Schelkens and J. Cornelis, "Control of the Distortion Variation in Video Coding Systems Based on Motion Compensated Temporal Filtering", *IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003.