CONSTANT QUALITY RATE-CONTROL FOR VIDEO ENCODING BASED ON ACTIVITY SEGMENTATION

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

[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 enhanced, off-line, segment-based rate control approach is proposed for controlling the distortion variation across successive segments of a video sequence when encoding with single-layer (MPEG-4 Baseline, MPEG-4 AVC) and wavelet video codecs. Consistent quality is achieved by a time-efficient, predictive rate-distortion modeling per segment. The individual segments are either shots or sub-shots, that are defined based on shot segmentation and activity analysis techniques. The rate control method solves the quality stability problem of state-of-the-art codecs, especially for less mature rate control modules (such as present in MPEG-4 AVC). The processing overhead compared to classical two-pass VBR encoding is limited while the distortion variation is significantly reduced.

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. In this paper, the focus will be on the equalization of the distortion across the video segments, rather than aiming at an optimal average quality. It has been found by visual evaluations [1] that the end-user will judge an entire video sequence based upon the quality variations and the minimum quality. In [1], a two-pass encoding system for MPEG-2 is described such that constant visual quality is obtained. Based on experimental results of an heterogeneous training set of video scenes, a function is derived to predict the optimal bit rates per frame based on the gathered first-pass statistics. The main disadvantage of this approach is that a large training set is needed to deliver a reliable mathematical model, as there will always be scenes that do not exactly fit the statistics of the training set. In addition, context-dependent characteristics, such as activity, are ignored in the derived bit allocation model.

In [2], an off-line shot-based rate control approach is presented which optimizes the quality consistency of the encoded video while solving the above shortcomings. Statistics are extracted on a shot basis and used to distribute the available bits over the different video shots. The method perfectly fits into an integrated video analysis framework of [3]. The only drawbacks are the complexity of the rate-distortion modeling phase in case of single-layer encoding and the necessity of a mature rate control module. However, the latter is many times more complex that the video coding process itself, especially for advanced codecs such as MPEG-4 AVC (due to the extensive coding options).

In this paper, a faster, yet effective modeling approach is proposed. Furthermore, shots will be refined into smaller segments if appropriate, based on the measured activity variation within the shot. Finally, we will experimentally show that the proposed method eliminates the rate control stability problem for MPEG-4 AVC.

In section 2, the segment-based rate control is briefly explained. In section 3, several enhancements to the rate-distortion modeling and quality control mechanism are discussed. An evaluation of the performance and a discussion of the improved algorithm are presented in section 4. Finally, the conclusions of this work are formulated in section 5.

2. SHOT-BASED RATE CONTROL

The challenge we are presented with is to intelligently subdivide a video sequence of arbitrary length and to distribute the available total amount of bits B_{TOT} among the M different segments such that:

$$\forall i, j \in [1, M], i \neq j : PSNR_i = PSNR_j \quad \textbf{(1)}$$

where for each segment i, $i \in [1, M]$, $PSNR_i$ represents the average PSNR of the luminance component. In [2], a pragmatic shot-based rate control approach is proposed based on the following four phases:

0-7803-8554-3/04/$20.00 ©2004 IEEE.

- *Segmentation phase*: the entire video is segmented into individual shots based on cut and fade detection.
- *Rate-distortion modeling phase*: each shot is two-pass encoded at a number of bit rates; the corresponding average PSNR-Y of the shot is measured.
- *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.
- *Optimal encoding phase*: each shot is encoded at the calculated optimal bit rate and the shots are concatenated.

The algorithm can be applied to single-layer (MPEG-4 Baseline and AVC [4], Windows Media 9) and scalable (wavelet [5]) state-of-the-art codecs. The main targeted video distribution models are the file download and progressive download model. 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. The results prove that the proposed technique improves the quality consistency significantly. This approach has as an additional benefit that it allows for a parallelized execution of the bit-allocation processes. In the following paragraphs, this approach will be enhanced and a solution to some shortcomings will be presented.

3. SEGMENT-BASED RATE CONTROL

3.1. Activity-based shot refinement

Usually, the video characteristics are relatively stable within a shot, but this is not always the case. Constant quality is more difficult to achieve if the video characteristics, such as activity, vary significantly. The activity of the *m*-th frame in the shot is defined as:

$$act(m) = \frac{100}{N \cdot \Delta} \cdot \sum_{x,y} \left| Y_m(x,y) - Y_{m-1}(x,y) \right| \quad \textbf{(2)}$$

with $Y_m(x,y)$ the luminance value of the pixel (x,y) of the *m*-th frame, *N* the number of pixels in the frame and Δ the maximum luminance difference value. The defined activity measure is resolution independent. The purpose is to subdivide the shot into segments of nearly constant activity. Therefore, an activity change indicator $\alpha(m)$ is calculated, based on the difference of the average activity of two adjacent windows of *L* frames:

$$\alpha(m) = \frac{1}{L} \cdot \left(\sum_{x=m}^{m+L} act(x) - \sum_{y=m-L}^{m-1} act(y) \right) \quad \textbf{(3)}$$

The window length *L* is a parameter and was set to 10 for our experiments.

The local maxima and minima of $\alpha(m)$ are possible split points. Subdivision is performed if a predefined threshold $\alpha_{\max}$ (set to ±1.5% in our experiments) is exceeded. Generally, a substantial change in activity indicates a modification of content and corresponding characteristics. In such a case, rate-distortion modeling of separate sub-segments is preferred.

3.2. Time-efficient rate-distortion modeling

In contrast to quality scalable coding, the rate-distortion modeling phase is rather complex for single-layer coding, although the obtained results can be stored and reused anytime. The proposed enhancements are twofold:

- *One-pass constant quality encoding*

Instead of two-pass encoding, one-pass encoding at a number of well-chosen fixed quantization parameters (QP) can be applied to calculate the R-D models of each segment. Two-pass only outperforms one-pass encoding in case of alternating easy and complex scenes. However, shot segmentation and activity-based shot refinement (section 3.1) guarantees relatively constant characteristics within a separate segment.

- *Predictive R-D modeling*

The R-D model of each shot can be predicted as follows:
Step 1 - Each segment is once (and only once) fully encoded at a specific segment-dependent basic quantization parameter QP_0.
Step 2 - Each segment is subdivided in N_{sb} sub-blocks of *N* frames, the last sub-block contains the remainder of the frames. The first *M* frames of each sub-block are (one-pass) encoded at a number of fixed QP values (QP_0 not included). In our experiments, *N*=50 and *M*=10. The estimated motion vectors of step 1 can be reused. A small value for QP_0 guarantees high-quality motion estimation results. Typically, motion estimation is the most complex part of the video encoding process. Therefore, the extra overhead of step 2 will be limited.
Step 3 - For each QP value, the required bits and obtained PSNR-Y are analysed on a frame basis. Assuming no bi-directional encoding, the total shot rate corresponding to a specific QP value is then estimated as follows (coding assumption: initial I frame followed by P frames only):

$$\begin{cases} r(QP) = \dfrac{c_r}{N_{sb}} \cdot \sum_{i=1}^{N_{sb}} r_i(QP) \\ r_i(QP) = \dfrac{fps}{N_f}\left(b_{i1}(QP) + b_{i2}(QP) + \dfrac{N_f - 2}{M-2} \cdot \sum_{j=3}^{M} b_{ij}(QP)\right) \\ c_r = \dfrac{r_{real}(QP_0)}{r_{est}(QP_0)} \end{cases} \quad \textbf{(4)}$$

where *r* and r_i are the estimated rates of the shot and its *i*-th sub-block respectively, N_{sb} the number of sub-blocks in

the shot, *fps* the frame rate, N_f the total number of frames in the shot, b_{ij} the amount of bits used for the j-th frame in the i-th sub-block, c_r the bit rate correction factor, $r_{real}(QP_0)$ and $r_{est}(QP_0)$ the real and estimated bit rate at the basic quantization parameter QP_0. The amount of bits of the first P-frame is treated separately, because it will typically differ slightly from the other P-frames.
Equivalently, the average PSNR-Y of the shot for a specific QP is estimated by:

$$\begin{cases} PSNR(QP) = \dfrac{c_p}{N_{sb}} \cdot \sum_{i=1}^{N_{sb}} PSNR_i(QP) \\ c_p = \dfrac{PSNR_{real}(QP_0)}{PSNR_{est}(QP_0)} \end{cases} \quad \textbf{(5)}$$

where $PSNR$ and $PSNR_i$ are the average PSNR-Y of the shot and its i-th sub block respectively, c_p the PSNR-Y correction factor, $PSNR_{real}(QP_0)$ *and* $PSNR_{est}(QP_0)$ the real and estimated PSNR-Y value at QP_0.
The smaller the differences between the QP measurement values, the more reliable the above method is. The accuracy of the estimation method is explained by the fact that the constructed segments approximately have constant characteristics and small variations are taken into account by a well-balanced set of representative frames.

3.3. Object of interest based quality metric

Another possible enhancement is to use available metadata about movement of objects acquired by object tracking techniques in order to fine-tune the quality measurement process. Instead of considering the quality of a whole frame, the PSNR-Y value of a (moving) object of interest obviously plays a dominant role in perceived quality of the whole video, especially if the background is pre-processed with for instance a blurring filter. In that case, the PSNR-Y of the whole frame would overestimate the quality of the object of interest. Once more, this example demonstrates the advantages a segment-based rate control approach offers in combination with available metadata from video analysis frameworks.

4. EXPERIMENTAL RESULTS

In this section, we will present experimental evidence showing that the proposed enhancements increase the efficiency and applicability of a segment-based rate control algorithm. 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).

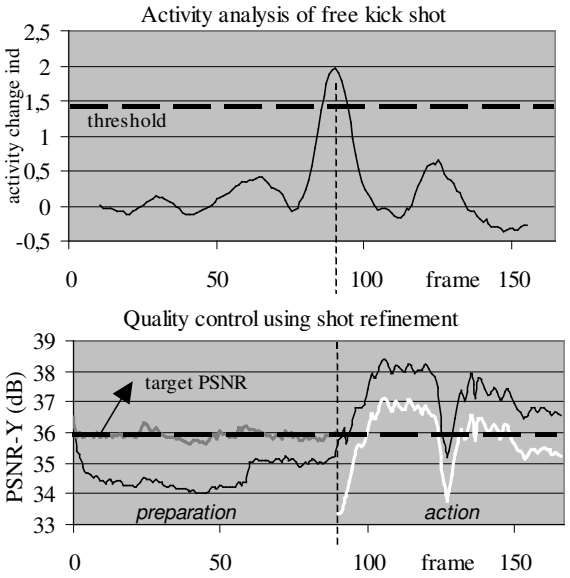

Figure 1 – Refinement of the *free kick* shot based on activity analysis and comparison of the corresponding PSNR-Y variation between classical VBR (black curve) and segment-based rate control (concatenation of gray and white curve).

Applying the activity-based shot refinement of section 3.1, the *free kick* shot will be further split up in a low-activity *free kick preparation* and a high-activity *free kick action* segment as shown in Figure 1. The maximum of the *activity change indicator* coincides with the transition between the two segments. Remark that, in case of classical VBR, the obtained quality in the action segment is surprisingly better. This can be explained by an overestimation of the complexity of the segment by the heuristical model assumed in the rate-control module and/or the tendency of state-of-the-art codecs to aim at constant quality by fixing the quantization parameter. Figure 1 also compares the variation of the PSNR-Y of the MPEG-4 (Baseline) encoded *free kick* shot using standard VBR and segment-based rate control respectively, for a target average PSNR-Y of 36dB. The used MPEG-4 codec is DivX Pro, version 5.1, see http://www.divx.com/divx (no B frames). By applying the enhanced segment-based rate control, the variance of the PSNR-Y is reduced with a factor 5. Unfortunately, the PSNR-Y curves of the two segments doesn't link up closely at the transition point. Also, in busy scenes, isolated low PSNR-Y values may appear. To eliminate these imperfections, frame-level quality control is required, which is out of the scope of this paper.

Figure 2 shows the excellent accuracy of the proposed predictive one-pass (constant quality) based rate-distortion modeling for three shots of the test sequence: *female newsreader* (shot 1), *attack* (shot 5) and *player close-up* (shot 6). The same conclusions can be drawn for the other shots.

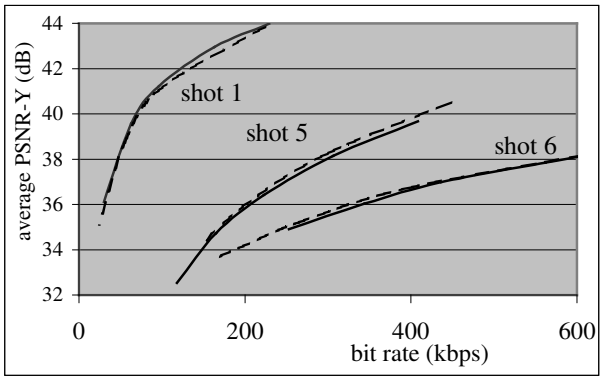

Figure 2 - Accuracy of the predicted rate-distortion models (dashed line) compared to the real R-D models for three shots of the test sequence.

By applying predictive R-D modeling, the overall complexity of segment-based rate control is significantly reduced. If five QP measurement points are used, only a limited amount of extra overhead is needed compared to the total processing time for state-of-the-art VBR encoding, while preserving the benefits of the segment-based rate control technique, i.e. strongly reduced distortion variation.

In Figure 3, the results for our enhanced segment-based rate control algorithm are compared with classical rate control and with two-pass R-D shot-based rate control [2] for MPEG-4 AVC. The used codec is AHM 2.0, built on JM6.1 (JVT-F086 added). The test sequence is encoded at 256 kbps. The measured total variance of PSNR-Y for classical, two-pass R-D and one-pass predictive R-D segment-based rate control respectively are in the proportion of 100 to 10 to 1.

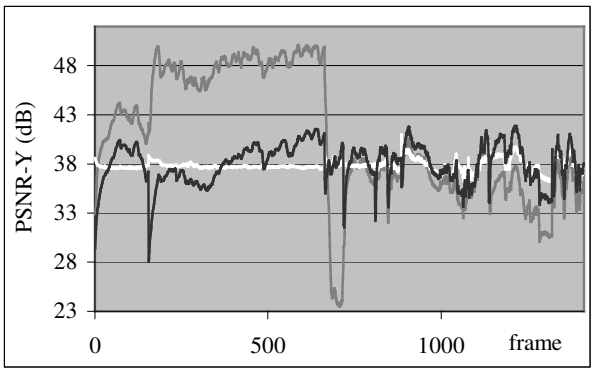

Figure 3 - Comparison of quality consistency between classical (grey line) and segment-based rate control applying one-pass (white line) and two-pass (black line) encoding based R-D modeling for MPEG-4 AVC encoding of the test sequence.

5. CONCLUSIONS

We have presented an enhanced segment-based rate-control algorithm, based on efficient one-pass rate-distortion modeling. The proposed improvements increase the applicability of the method presented in [2] significantly. The reuse of metadata (available via integrated video analysis frameworks) such as shot segmentation and activity information yields nearly constant quality when encoding the constructed segments at fixed quantization parameters, thereby speeding up the R-D modeling and optimal encoding phases. Predictive R-D modeling produces excellent results for an acceptable processing complexity. The advantages of using a object tracking-based quality metric have been discussed.

The optimal threshold value for the *activity change indicator* for different types of segments (segment classification) needs further investigation. Also, the method needs to be extended with frame-level rate control to deal with deficiencies at segment borders and isolated low-quality frames. Finally, the quality metric can be further improved based on shot/segment classification using available metadata in order to come closer to subjective quality.

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. Special thanks to Guy Lateur for his appreciated collaboration.

6. REFERENCES

[1] 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.

[2] L. Overmeire, F. Verdicchio, J. Barbarien, G. Lateur, L. Nachtergaele and P. Schelkens, "Shot-based rate control for single-layer and scalable video encoding", accepted for the *International Workshop on Image Analysis for Multimedia Interactive Services*, April 2004.

[3] 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.

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

[5] 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.