Subjective Assessment of H.264/AVC Video for Low-Bitrate Multimedia Messaging Services

Patrick Brun, Gert Hauske, and Thomas Stockhammer
Institute for Communications Engineering (LNT)
Munich University of Technology (TUM)
Munich, Germany

Abstract— **In this work we investigate the performance of the H.264/AVC video coding standard for low bit rate mobile multimedia messaging services (MMS). We focus on the appropriate selection of the quantization parameter and the temporal resolution. For this purpose, a psycho-visual experiment has been designed. It is revealed that only limited set of parameter values is necessary to span almost the entire range of quality levels. In general quantization parameters of 34 frame rates of 10 fps and bit rates below 64 kbit/s are sufficient to provide good quality. Sports sequences ask for slightly higher frame rate, 15 fps is sufficient. To provide sufficient quality, quantization parameters below 40 and frame rates below 5 fps should definitely not be used.**

I. Introduction

The demand for fast and location-independent access to multimedia services offered on today's Internet is steadily increasing. Hence, most current and future cellular networks, like GSM-GPRS, UMTS, or CDMA-2000, contain a variety of packet-oriented transmission modes allowing transport of practically any type of IP-based traffic to and from mobile terminals, thus providing users with a simple and flexible transport interface. The third generation partnership project (3GPP) has selected several multimedia codecs for the inclusion into its multimedia specifications [1]. Especially applications like multimedia messaging services (MMS) to and from wireless clients will become important features in future mobile systems and may be a key factor to their success. Since the end-user's costs are likely to be proportional to the transmitted data volume, compression efficiency is the main target for wireless video and multimedia applications. Initially, to provide basic video service in the first release of the 3G wireless systems, the well-established MPEG-4 visual simple profile has been integrated. The choice was based on the manageable complexity of the encoding and decoding process as well as on the maturity and simplicity of the design. However, the new joint ISO/ITU standard H.264/MPEG-4 AVC [2] promises significant gains in terms of compression efficiency. Therefore, the optional integration of the baseline profile of H.264 is planed for the next release of 3GPP standards and will likely be mandatory for future releases. The encoder of the standard however provides a huge flexibility in the encoding parameters.

In this work we concentrate on the selection of the appropriate quantization parameter as well as the frame rate for MMS applications. In previously performed tests to evaluate video quality for mobile applications, e.g. [3], only the performance for a fixed encoder configuration is assessed. Since for our purpose no empirical data for the quality of such low-rate and low-resolution video sequences is available, a psycho-visual experiment has been designed. In the following we formulate the problem in more detail; we explain the subjective test environment, and provide and discuss the test results.

II. Video in Mobile Multimedia Messaging Services

In MMS applications encoding, transmission, and decoding, of the video sequence are in general completely separated. These download-and-play applications allow the network to transport the encoded video stream reliably to the distant user. Therefore, error resilience issues - typically important for wireless real-time transmission - can be neglected for MMS applications. Due to the likely business models for MMS as already outlined in the Introduction, the encoding bit-rate is the most critical cost and should be as low as possible while maintaining a certain minimum desired quality. Typically, the end user can select from very few discrete quality categories, such as excellent, good, satisfying, sufficient, and poor. The problem now is that these simple subjective quality definitions cannot easily be mapped to common video standards such as H.264/AVC as the encoding parameters are manifold, and are not specified in the standard. The normative part of a video coding standard only consists of the appropriate definition of the order and the semantics of syntax elements and the decoding of error-free bit-streams.

A typical encoder for H.264/AVC as used for our experiments is shown in Figure 1 with the main encoding options relevant for our work highlighted. In the following we discuss and justify predetermined settings of encoder parameters and extract two variables in the encoding process which allow the control of bit rate and quality for the considered applications. Due to the limited display size for typical wireless devices the spatial resolution is fixed to QCIF. The frame structure applied is IPPP…. Additional Intra frames are not used, since neither random access nor error resilience is of major importance in MMS video. Also, any other error resilience features are completely ignored for the same reasons. We also dispense on B-frames in the encoding as these frame types are not supported in the baseline profile due to the high encoding and decoding complexity. However, we use all block

partition modes as well as five reference frames in the encoding process. For the selection of the macroblock mode, the reference frames, as well as the motion vector rate-distortion optimized mode selection based on Lagrangian formulation is used as already implemented in the applied test model software JM4.0. For more details on rate-distortion optimization we refer the reader to [3] and [5]. For real-time applications which are transmitted over constant bit rate channels usually a rate control is introduced to maintain a constant bit rate and to fulfil the constraints imposed by a Hypothetical Reference Decoder (HRD). However, it is well-known, that variable bit-rate (VBR) encoded video with fixed quantization parameter has superior performance than imposing rate control [6]. As for the considered low bit rate download-and-play applications HRD constraints are of little relevance, VBR encoding with a fixed quantization parameter for the entire sequence is definitely interesting. It is also worth to mention that a fixed bit rate, even with VBR encoding, might result in an excellent quality for one sequence, but in a very poor quality for a different sequence dependent on the complexity of the sequence.

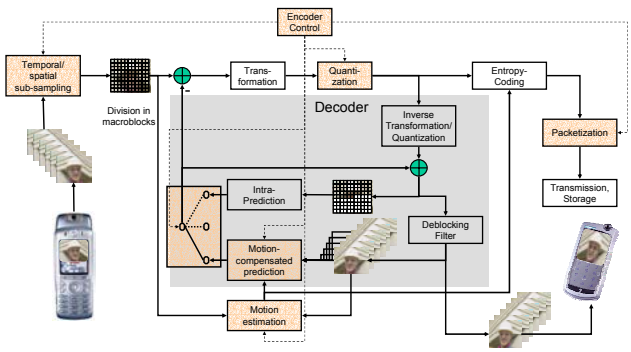

Figure 1. Typical H.264/AVC encoder

In H.264 the quantization parameter can be selected from q=0,…51 where low quantizers result in high quality for each encoded video frame, but also in high bit rates and vice versa. The quantization parameter controls the step size of the scalar quantizer of the transform coefficients in a logarithmic way. Note that reducing the quantization parameter by six approximately doubles the bit-rate.

An additional parameter to control the resulting bit rate is the frame rate f of the encoded video. Cameras integrated in low-complexity mobile devices commonly capture video at a frame rate of f=30 fps, but temporal subsampling can be applied by skipping frames before coding the next. Lower frame rates obviously result in lower bit-rates. However, note that the resulting bit rate does not decrease linearly with decreasing frame rates as the correlation between more distant frames decreases resulting in less coding efficiency when using temporal prediction.

To summarize, for MMS applications it is useful to control the quality and the bit rate by two global parameters, namely

- the quantization parameter q to be selected from the set q=0,…51, and
- the frame rate f, selected from the set f=30/(1+s) fps with s the frame skip being any integer s=0,1,…

The encoder design now obviously asks to select the parameters appropriately based on some quality measures. Whereas the quality of a sequence for fixed frame rates can usually be expressed reasonably well by the mean PSNR computed as the average of the PSNR of all encoded frames, it is difficult to compare video sequences encoded with different frame rates. In addition, it would be interesting what parameter combinations (q,f) match the simple user quality specification, e.g., from excellent to poor. It is not clear if and how this selection depends on the sequence to be encoded. We are also curious on the bit rates necessary to support MMS with a certain quality. Finally, it is interesting to know reasonable parameter ranges for the frame rate f and q for low bit rate MMS applications. For this reason, we have set up a subjective test which should reveal some of the problems and questions discussed.

III. The Subjective Test

As subjective tests usually are costly and time-consuming, a careful pre-selection of test sequences and encoding parameters is essential. In addition, the execution of the test should be as formal and controlled as possible, but also provide the subjects with a typical application environment. Finally, it has to be taken care that the test results provide sufficient significance. Based on these considerations we set up the subjective test as follows. The four test sequences consist of a mix of typical MMS clips, partly professional, partly amateur. All sequences have length of 10 seconds. We will briefly characterize the sequences; the first frame of each sequence is shown in Figure 2. *Architecture* is an amateur movie showing a 180° camera pan of large bright buildings with objects moving in the foreground. *Moving Person* is an amateur movie showing a self-recorded sequence of a walking person with shaky camera movements. *Soccer* is a professionally filmed wide angle soccer scene with the camera following moving players and the ball. Finally, *Cinema Trailer* shows a professionally filmed trailer with fast movements and several scene cuts.

Figure 2. First frames of test sequences.

As already discussed, these sequences now have been encoded using JM4.0 with only two parameters varied, namely quanti-

zation parameter q and frame rate f. Obviously, only a reduced amount of combinations can be tested. As we are mainly interested in selecting the appropriate temporal resolution, we use six frame rates in the encoding process, namely f=3, 5, 7.5, 10, 15, 30 fps. In addition, for each test sequence we have specified six bit rates ranging from 8 kbit/s to 64 kbit/s, where the selection of the set of bit rates depends on the complexity of the sequence. The quantization parameters q are selected such that for each frame rate f we obtain the desired bit rate. The maximum deviation of the real bit rate to the desired bit rate is less than 10%. To summarize, we have specified 144 test cases where each test case is specified by the test sequence, the frame rate f, and the bit rate r. The pre-selection of test cases was based on observations from preliminary tests, partly based on results for H.263 [7]. In addition, reference sequences have been generated at bit rate 128 kbit/s and f=30 fps providing a maximum achievable quality impression.

Ten paid subjects (students of age between 19 and 25 years, 6 of them female and 4 male) were employed in the test procedure consisting of three runs on separate days. Within each run the video sequences were presented in different order. To allow parallel test execution, a WWW-based test environment has been set up. The sequences were streamed over an over-provisioned Intranet to standard PCs with CRT display (EIZO Flexscan T67), the gamma of the monitor was 2, the maximum luminance 92 cd/cm^2. The viewing distance of about 30 cm corresponds to 6 times the picture height. The videos were presented pair-wise with the reference video first followed by video to be tested. The web site was designed such that the sequences are presented with a virtual mobile telephone as surrounding frame. After the presentation of each test case sequence the subjects had to give their estimates on the perceived on a "Absolute Category Rating" scale [1] with six grades reaching from "1, excellent", "2, good", "3, satisfactory", "4, sufficient", "5, imperfect" to "6, poor"[1]. The quality estimates of the user are interactively collected into a data base. In the future, it is planned to provide this test environment also over the public Internet.

IV. Results and Discussion

In the following we will present selected results of the subjective test. Although the test results reveal many additional interesting aspects, we focus in the sequel on the discussion of problems and questions raised at the end of section II.

Since our data shows no significant differences between runs (3) and subjects (10) we use the average of the 30 marks available for each of the 144 test cases. The average marks versus the frame rate, f, for all 36 test cases for test sequence *cinema trailer* are shown in Figure 3. It is worth to note that the selection of test cases was reasonable as marks are distributed over the entire available range from 1 to 6 with higher density around the mean at 3.5. Let us first state some non-surprising results: At a constant frame rate higher bit rates obviously result in higher perceptual quality. It is also obvious that at a constant bit rate varying the frame rate produces different perceptual quality, with an optimal frame rate between 5 and 10 fps depending on the bit rate. In general, the optimum frame rate decreases with decreasing bit rate. Let us now focus on the on the optimum configuration, i.e. the frame rate resulting the maximum perceived quality for a certain constant bit rate. It can be observed that frame rates of 10 fps already provide more than good quality if the bit rate is high enough, in this case 50 or 32 kbit/s. However, frame rates at 5 fps or below cannot provide satisfying quality, and, at 3 fps not even sufficient quality.

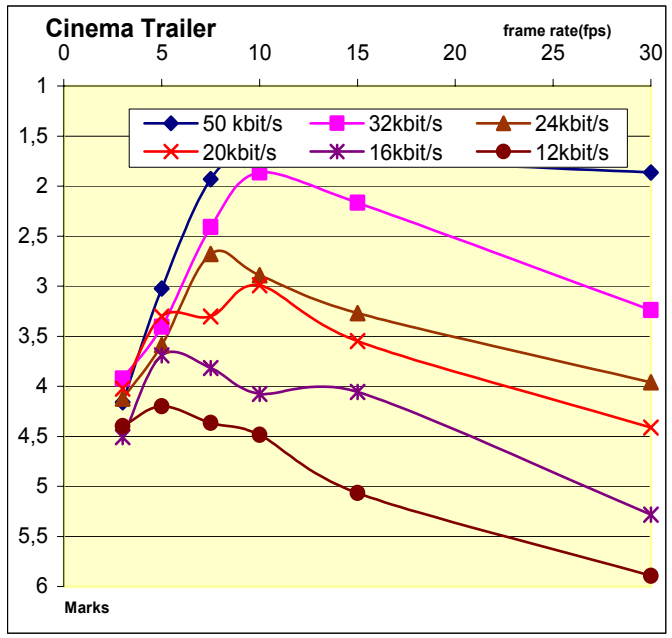

Fig. 3 Average marks versus frame rate for test sequence *cinema trailer*.

The other test sequences show similar results, namely that the selection of the parameters for the test cases was proper, and, that there exists an optimal frame rate for a given bit rate. Therefore, we concentrate in the following only on points providing maximum perceptual quality for a fixed bit rate. For this optimal configuration, Table 1 shows different parameters at different bit rates for all test sequences, namely the optimal frame rate, f, the resulting PSNR for each test, the applied quantization parameter, and the average marks. Since marks are equally distributed, it is evident that the selection of test case parameters for other test sequences appears to be proper.

We will first extract some global characteristics among sequences before investigating more specific details. Depending on the complexity of sequences, the perceived quality at the same bit rate can be significantly different, e.g. for r=32 kbit/s ranging from 1.7 (Architecture) to 3.4 (Moving Person). With negligible exceptions a higher bit rate results in better subjective quality. For all test sequences it is also obvious that quantization parameters at around 34 are sufficient to obtain at least good quality, whereas quantization parameters higher than 40 lead to worse than 4 and should therefore not be used. For all tested sequences a bit rate of 64 kbit/s can provide at least good quality if the appropriate frame rate is selected. Frame rates higher than 15 fps seem not to be necessary,

[1] This corresponds to the German school grading system.

whereas frame rates below 10 fps cannot provide good, below 7.5 fps not even sufficient quality. For all sequences it is also obvious that changing the bit rate, or the desired quality, the optimal configuration generally changes in both dimensions, namely in frame rate and quantization parameter. In general, the quantization parameter increases, and the frame rate decreases, for decreasing bit rate or decreasing desired quality.

TABLE I. PARAMETERS FOR OPTIMAL FRAME RATES *F* AT FIXED BIT RATE *R* FOR DIFFERENT TEST SEQUENCES

r (kbit/s)	*f* (fps)	PSNR(dB)	*q*	Marks
Cinema Trailer				
50kbit/s	10	37.303	30	1.6
32kbit/s	10	35.008	34	1.9
24kbit/s	7.5	34.501	35	2.7
20kbit/s	10	32.67	38	3.0
16kbit/s	5	33.83	36	3.7
12kbit/s	5	32.693	38	4.2
Soccer				
64kbit/s	15	30.896	33	2.0
50kbit/s	15	30.239	34	1.9
32kbit/s	15	28.948	36	2.7
24kbit/s	7.5	28.949	36	3.6
20kbit/s	10	27.808	38	3.6
16kbit/s	10	27.808	38	4.0
Architecture				
32kbit/s	30	30.321	36	1.7
24kbit/s	10	31.142	35	2.2
20kbit/s	15	29.816	37	2.1
16kbit/s	10	29.873	37	2.3
12kbit/s	10	28.547	39	3.0
8kbit/s	5	27.992	40	3.7
Moving Person				
64kbit/s	10	30.279	35	1.8
50kbit/s	10	29.016	37	2.5
32kbit/s	5	29.017	37	3.4
24kbit/s	5	27.675	39	3.9
20kbit/s	3	28.402	38	4.1
16kbit/s	5	26.408	41	4.4

Looking at specific sequences some interesting characteristics for different sequence types are revealed. The sequence *Cinema Trailer* is different from the rest as the average PSNR is much higher than for the other three sequences partly due to additional intra information in scene changes, partly due to the darkness of the sequence. The sequence *Soccer* is different as higher frame rates are selected for good quality. It seems that for sport sequences temporal resolution is more important than the quality of individual frames. The sequence *Architecture* is unique in the sense that the required bit rates for good and sufficient quality are rather small. The panorama view can be encoded easily due to simple global motion in the sequence.

Finally, the sequence *Moving Person* is unique because the perceived quality is rather low for all bit rates. This comes from the shaky recording due to the handheld camera. In this case changing the encoding parameters could not increase the perceived quality significantly. Preprocessing to remove the shaky recording should be considered for these sequences.

V. CONCLUSIONS

In this work we have investigated the performance of the H.264/AVC video coding standard for low bit rate mobile multimedia messaging services (MMS) for bit rates up to 64 kbit/s. For this purpose, a psycho-visual experiment has been designed. We have focused on the appropriate selection of the quantization parameter and the temporal resolution for a fixed specified bit rate. For H.264/AVC encoding a bit rate of 64 kbit/s is sufficient to provide at least good quality. It is revealed that only a limited set of parameter values is necessary to span the entire range of quality levels. Quantization parameters of 34 and frame rates of 10 fps are sufficient to provide good quality. Sports sequences ask for slightly higher frame rate, but 15 fps is sufficient. To provide sufficient quality quantization parameters below 40 and frame rates below 5 fps should definitely not be used, for sports sequences at least 7.5 fps, better 10 fps should be used.

As a consequence of these results a combination of the user perceived quality and the resulting bit rate could be used to support end users with a simple method to select their encoder configuration. It is planned to integrate a simple mapping of user perceptual quality to encoder's quantization parameter and frame rate. Thereby, the sequence type should be taken into account which requires specifying a limited set of sequence types, e.g. sports, panorama, etc. with certain characteristics. The specification should be based on simple accessible measures as processing power for handhelds is limited.

ACKNOWLEDGMENT

We would like to thank test participants for their dedication during the tests as well as Antonios Tsetsos, Günther Liebl, and Johannes Zangl for providing assistance in setting up the web-based test environment.

REFERENCES

[1] 3GPP Technical Specification 3GPP TR 26.140: Multimedia Messaging Service (MMS); Media formats and codecs.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems, vol. 13, no. 7, July 2003.

[3] Stefan Winkler, and Frédéric Dufaux: "Video quality evaluation for mobile applications," in *Proc. SPIE/IS&T Visual Comm. and Image Proc.*, vol. 5150, pp. 593-603, Lugano, Switzerland, July 2003.

[4] G. J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression," IEEE Signal Processing Magazine, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[5] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-Constrained Coder Control and Comparison of Video Coding Standards," IEEE Transactions on Circuits and Systems, vol. 13, no. 7, July 2003.

[6] C.-Y. Hsu, A. Ortega, and A. R. Reibman, "Joint selection of source and channel rate for VBR video transmission under ATM policing constraints," IEEE Journal of Selected Areas in Communications, vol. 15, no. 6, pp. 1016-1028, August 1997.

[7] G. Hauske, R. Hofmeier, and T. Stockhammer, "Subjective Image Quality of Low-Rate and Low-Resolution Video Sequence,", in Proc. International Workshop in Mobile Multimedia Communications, Munich Germany, October 2003.

[8] ITU-T, "Subjective Video Quality Assessment Methods for Multimedia Applications", ITU-T Recommendation P.910, August 1996.