STANDARDS BASED VIDEO COMMUNICATIONS AT VERY LOW BIT-RATES

Bernd Girod, Niko Faerber, Eckehard Steinbach
Lehrstuhl für Nachrichtentechnik, University of Erlangen-Nuremberg
Cauerstrasse 7, D-91058 Erlangen, Germany
Tel: +49 9131 857100; fax: +49 9131 303840
`e-mail: {girod,faerber,steinb}@nt.e-technik.uni-erlangen.de`

Invited Paper

ABSTRACT

Video communication at very low bit-rates has made significant progress recently through the new ITU-T standard H.263. In this paper, we are reviewing the performance advances over the 1990 ITU-T standard H.261, and present a novel extension that allows robust transmission of moving video over highly unreliable channels, such as the mobile channel.

1 THE H.263 VIDEO COMPRESSION STANDARD

The ITU-T draft international standard H.263 [2] is closely related to the well-known and widely used ITU-T recommendation H.261 [3], which has also been devised by Study Group XV. This close relationship helped to arrive at the new standard in a short period of time, including not only the video coding algorithm but also the corresponding audio (G.723), multiplex (H.223), control (H.245) and system (H.324) aspects. H.261 and H.263 share the same basic codec structure consisting of block based motion compensation (MC) and DCT based transform coding of the remaining prediction error. However, there is a significant improvement in performance. Side-by-side comparisons show that the same subjective image quality can be achieved with less than half the bit-rate. This performance gain is due to improved and optimized coding techniques, which are either included in the default mode of H.263 or are part of optional coding-modes ("options").

In the default mode, two major differences in the prediction loop can be observed. Firstly, H.261 is limited to MC with integer-pel accuracy, while H.263 provides half-pel accuracy. The significant improvement due to half-pel MC is well understood [4] and has already been utilized successfully in ITU-T H.262 (MPEG-2) [5]. Secondly, no loop filter is included in H.263. Though spatial lowpass filtering is successfully utilized in H.261, it is not equally important in H.263 because the bilinear interpolation used for half-pel MC introduces spatial lowpass filtering as a side affect. In addition, one of the H.263 options includes overlapped block motion compensation, which has an inherent filtering effect as well.

An H.263-coder may use optional coding techniques ("options") to further improve its performance. Options have to be negotiated with the decoder via external means (for example within ITU-T H.245). Four options are available in H.263: Unrestricted Motion Vector mode (UMV), Advanced Prediction mode (AP), PB-frames mode (PB), and Syntax-based Arithmetic Coding mode (SAC). For more information, the reader is referred to Annex D, E, F and G of the H.263 standard [2].

2 RATE-DISTORTION PERFORMANCE

In this section we compare the rate-distortion performance of H.263 and H.261. We use the averaged peak signal-to-noise ratio (PSNR) of the luminance component as a distortion measure for a whole sequence, i. e., first we calculate the PSNR for each luminance frame n according to

$$PSNR_n = 10 \log \frac{1}{M} \sum_{i=1}^{M} \frac{255^2}{(o_i - c_i)^2}, \qquad (1)$$

where M is the number of samples in a frame, and o_i and c_i are the amplitudes of the original and coded frame, respectively. Then the PSNR values for each frame are averaged for N frames in the sequence according to

$$PSNR = \frac{1}{N} \sum_{n=1}^{N} PSNR_n. \qquad (2)$$

The rate is expressed in kbps (1000 bit/s) and includes the portion of the two chrominance components. All data presented in the following were obtained for the "Foreman" test sequence in QCIF resolution (frames 0-200). Note that a bit-rate control strategy would make a fair comparison more difficult. This is especially true if the bit-rate control results in a variable frame-rate, where different frames are encoded in different simulations. Therefore, all sequences were coded at a fixed frame-rate using the fixed quantizers 31, 24, 20, 16, 12, 10, 8 and 6, respectively. Simulations were carried out using available software codecs [6] [7].

2.1 Performance of H.263 vs. H.261

Fig. 1 shows the comparision of H.261 with H.263. At a bit-rate of 64 kbps, the following observations can be made. H.263 w/o options outperforms H.261 by approximately 2dB. Another dB is gained if we use all of the H.263 options (top curve). Two thirds of the maximum performance gain are apparently due to features not included in the H.263 options. As can be shown by further analysis [8], half-pel MC is the main reason for this performance gain.

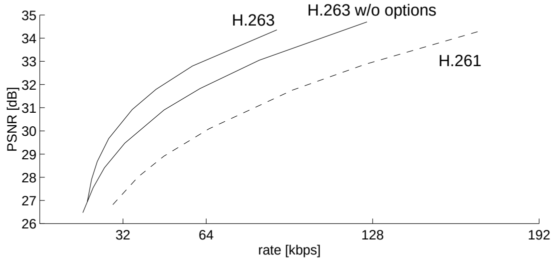


Figure 1: Performance of H.261 and H.263 at a frame-rate of 12.5 fps.

2.2 Performance of H.263 options

The following paragraphs evaluate the effectiveness of the H.263 options compared to the default mode of H.263, thus providing a more differentiated view of the performance gain due to single options.

Advanced Prediction mode: In Fig. 2 the performance gain due to the AP-mode is illustrated. Because the Unrestricted Motion Vector mode is automatically included in the Advanced Prediction mode, this option is not investigated separately. At 64 kbps, the AP-mode results in a performance gain of approximately 1.2 dB. It should be mentioned that four motion vectors per macroblock were not used very frequently during the simulations (less than 15%).

Syntax-based Arithmetic Coding mode: The improvement due to the SAC-mode is very small, approximately 0.2 dB at 64 kbps (Fig. 2). Because SAC is a different (lossless) entropy coding scheme, the PSNR for a given quantizer is uneffected, but fewer bits are produced. In terms of reduced bit-rate, the average gain for inter-coded macroblocks is 3-4%. For intra-coded macroblocks, the gain is higher, on average about 10%.

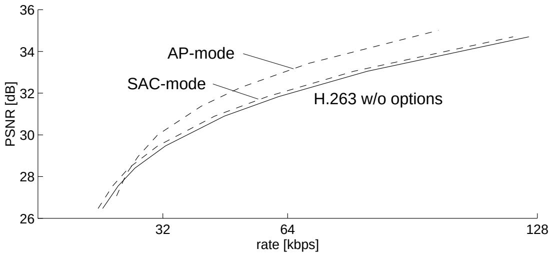


Figure 2: Performance of the H.263 options 'Advanced Prediction mode' and 'Syntax-based Arithmetic Coding'. The frame-rate is 12.5 fps.

PB-frames mode: The main purpose of PB-frames in H.263 is to increase the frame-rate without increasing the bit-rate too much. Typically, the frame-rate is doubled when the PB-mode is used. Consider a sequence coded at 6.25 fps using H.263 w/o options (Fig. 3, top curve). Though the quality of single frames is good (33.5 dB at 64 kbps), the low temporal resolution results in jerky motion. However, increasing the frame-rate to 12.5 fps while maintaining the bit-rate at 64 kbps results in an significant loss of image quality (1.7 dB).

A better compromise between temporal and spatial resolution is possible when the PB-mode of H.263 is used. The dashed curves in Fig. 3 shows the performance of the PB-mode at 12.5 fps. Because the quality of P- and B-frames differs significantly, the averaged PSNR according to (2) is calculated separately. As can be seen, the PSNR for P-frames drops only by 0.6 dB compared to the top curve. Note that the same number of P-frames per second (6.25) are now transmitted with only little loss of quality. With the use of the B-frames, however, the frame-rate is doubled. Though the quality of the B-frames is low, they provide the subjective impression of smooth motion. According to this concept, the fraction of the bit-rate allocated to the B-part of a PB-frame is kept low, on average about 15-20%. In fact, H.263 specifies that a B-macroblock always has to be quantized more coarsely than its corresponding P-macroblock.

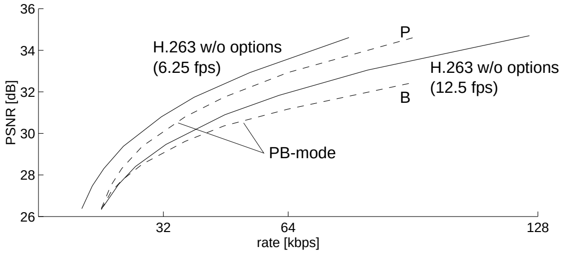


Figure 3: Performance of the H.263 option 'PB-frames mode'.

3 ROBUST VIDEO TRANSMISSION FOR UNRELIABLE CHANNELS

As can be seen from the previous section, significant progress in compression has been made through the new ITU-T standard H.263. However, similar to other video compression standards, H.263 is very sensitive to transmission errors. In this section we present a novel approach for robust video transmission which does not require any modification of the H.263 bitstream syntax.

Many existing networks cannot provide a guaranteed quality of service. This may result from the underlying medium access control, like in 802.3 based Local Area Networks (Ethernet), or from the limitations of the transmission channel, e.g., in mobile environments where remaining errors may not be avoided during fading periods. Delayed packets in Local Area Networks (LANs) have to be considered as lost for real-time conversational services like videoconferencing, if the delay exceeds a maximum value. Transmission errors of a mobile communication channel may range from single bit errors up to burst errors or even a temporal loss of signal. Those varying error conditions limit the effective use of Forward Error Correction (FEC), since a worst case design leads to a prohibitive amount of overhead. Networks with these limitations are characterized as "best effort" networks and require increased robustness for the transmission of video.

Motion compensated prediction in H.263 leads to high coding efficiency, which is essential to cope with limited bandwidth and low delay requirements in mobile networks. However, motion-compensated prediction also causes spatio-temporal error propagation, i.e. visible distortion due to transmission errors generally remains visible for several seconds.

Error propagation can be reduced efficiently sending negative acknowledgments (NAKs) via a feedback channel between transmitter and receiver. The proposed system tolerates errors, but limits their effect by error control techniques in the source codec. Error concealment is employed to hide visible distortion and residual errors are compensated using the acknowledgment information from the receiver.

3.1 Error Concealment

Packet loss or severe burst errors lead to information loss at the decoder. Error concealment is employed in order to minimize the resulting visible distortion. Corrupted Group of Blocks (GOB) are concealed considering all MBs in the GOB as not coded, i.e. the image content of the GOB is copied from the preceding frame. This technique works almost perfectly for non-moving parts of the sequence, e.g. stationary background, but introduces severe distortion for moving image regions. Fig. 4 shows the loss of picture quality (Δ PSNR) after concealment of 3 successive GOBs. The QCIF sequence *Foreman* is coded at 64 kbps and 8.33 fps, resulting in an average PSNR of about 34 dB in the error-free case. 25 simulations are conducted with different temporal and spatial location of the lost GOBs (dotted lines). The solid line shows the averaged result, indicating that a residual loss of approximately 3 dB still remains in the sequence after several seconds. The PSNR is computed over the entire frame. However, the distortion is generally concentrated in certain image regions.

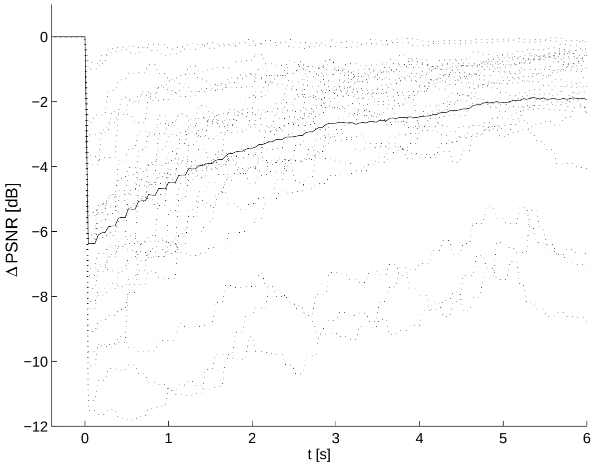


Figure 4: Decrease in PSNR for error concealment of three successive GOBs

3.2 Error Compensation with Feedback Channel

Because the picture quality recovers only slowly, special action should be taken to stop error propagation and shorten recovery time. Our approach utilizes the INTRA mode to stop temporal error propagation but limits its use to severely affected image regions only. Using a feedback channel, the temporal and spatial occurrence of an error is reported to the transmitter. The decoder sends negative acknowledgments (NAK) for GOBs which could not be decoded successfully and had to be concealed. The transmitter evaluates the acknowledgment information and incorporates it into the coding control. Two strategies for rapid error recovery relying on feedback information have been investigated and compared:

Error Tracking strategy: The location and extent of propagated errors is reconstructed at the transmitter when the NAK is received. Only the most severely affected MBs are INTRA coded.

Same-GOB strategy: The entire reported GOB is INTRA refreshed. In other words, only temporal error propagation is taken into account. Please note that this approach must be sub-optimal if the round-trip delay is high and the error has already propagated from its original location due to motion-compensated prediction. However, the evaluation of the NAKs is simple and does not require additional intelligence in the encoder.

Let us assume for the moment that for the Error Tracking strategy the encoder gains complete knowledge about the error distribution. This assumption is useful for simulation purposes only, since re-coding and storing of past frames is involved. For a practical system the error propagation hat to be estimated with a low complexity algorithm [1]. Fig. 5 compares the two strategies. The test sequence *Foreman* is coded under the same simulation conditions as in Fig. 4. Both strategies, Error Tracking and Same-GOB, achieve rapid error recovery as soon as the NAK arrives at the encoder (in Fig. 5 after 700 ms). The Error Tracking strategy outperforms the Same-GOB strategy by about 0.5 to 1 dB. The average decrease in PSNR due to concealment from Fig. 4 is included for comparison. Fig. 6 shows example frames for the two strategies after the loss of two GOBs.

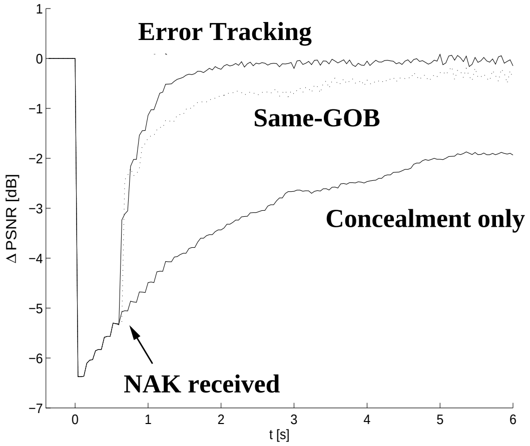
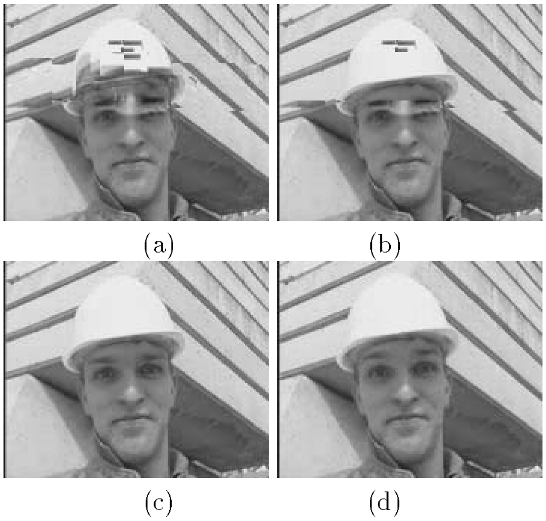


Figure 5: Error recovery with feedback channel

3.3 Experimental Results

400 frames of the sequences *Mother and Daughter*, *Carphone*, and *Foreman* are coded and packets of size 255 bits are transmitted over a simulated DECT (Digital European Cordless Telephony) channel. The corresponding bit error sequence exhibits severe burst errors and provides a maximum bit rate of 32 kbps at an average bit error rate of 2×10^{-2}. We compared different combinations of Forward Error Correction (FEC), Automatic Repeat on reQuest (ARQ) and the two error compensation strategies, Error Tracking and Same-GOB. The total round-trip delay assumed for NAKs is 250 ms. The code rate of the BCH code is 0.875 and for


(a) (b)

(c) (d)

Figure 6: (a) Frame 90 of sequence *Foreman* after two GOBs were lost and concealed in frame 75 (b) Same-GOB strategy (c) Error Tracking (d) Frame 90 without GOB loss in frame 75 for comparison.

ARQ only one retransmission of corrupted packets is allowed. Fig. 7 shows averaged simulation results. No protection of the bit stream leads to a dramatic decrease in PSNR. FEC alone improves the result, but is still far from satisfactory. ARQ combined with FEC leads to a considerable improvement of the resulting image quality. If ARQ is not feasible due to large round-trip delay as it generally is the case for satellite links, the Error Tracking and Same-GOB strategies in combination with FEC achieve about the same level as ARQ-FEC. For *Mother and Daughter* they perform better, for the other two sequences they perform worse. The best result is observed for a combination of Error Tracking, ARQ, and FEC. The PSNR gain over SG-ARQ-FEC is only marginal in the average. However, subjectively Error Tracking still outperforms the Same-GOB strategy in the same way as described in Fig. 6.

4 Conclusions

In this paper we systematically evaluated the rate-distortion performance of H.263 and the various options provided by the standard. Half-pel accuracy of motion compensation yields a typical gain of 2 dB over integer-pel accuracy for our test sequences. The Advanced Prediction Mode that includes overlapped block motion compensation typically yields another 1 dB. The PB-frames mode allows to almost double the frame rate with only little loss of picture quality for the P-frames. Only a very small gain is realized with Syntax-based Arithmetic Coding. Finally, we showed how negative acknowledgments sent over a feedback channel enable a robust transmission for unreliable channels such as LANs or mobile radio channels. The novel scheme does not change the H.263 bit-stream syntax and provides almost instantaneous recovery from transmission errors by an intelligent coder control.

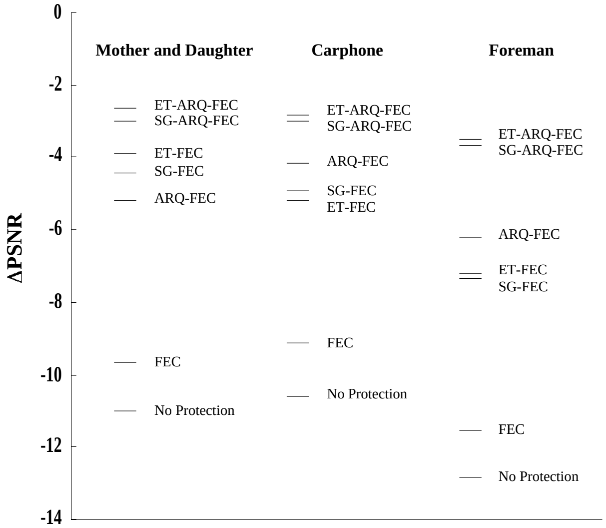


Figure 7: Averaged simulation results for a simulated DECT channel with an bit error rate of 2×10^{-2}. **ET:** Error Tracking strategy; **SG:** Same-GOB strategy; **ARQ:** Automatic Repeat on reQuest (one retransmission only); **FEC:** Forward Error Correction (BCH code with code rate c=0.875).

References

[1] N. Faerber, E. Steinbach, B. Girod, “Robust H.263 Compatible Video Transmission Over Wireless Channels,” International Picture Coding Symposium, Melbourne, March 1996.

[2] “Video Coding for narrow telecommunication channels at < 64 kbit/s”, DRAFT ITU-T Recommendation H.263, April 1995.

[3] “Video Codec for audiovisual services at p x 64 kbits/s”, ITU-T Recommendation H.261, Geneva, 1990.

[4] B. Girod, “Motion-Compensating Prediction with Fractional-Pel Accuracy”, IEEE Transactions on Communications, vol. 41, no. 4, April 1993, pp. 604-612.

[5] ISO/IEC 13818 (ITU-T H.262), “Generic coding of moving pictures and associated audio information, Part 1: Systems, Part 2: Video, Part 3: Audio”, March 1994.

[6] TMN-Test model for H.263, Version 1.4a Telenor Research & Development, 1995 (Internet: http://www.nta.no/brukere/DVC/).

[7] H.261 software codec, Portable Video Research Group, Stanford (Internet: ftp://havefun.stanford.edu/pub/p64/).

[8] B. Girod, E. Steinbach, N. Faerber, “Comparison of the H.263 and H.261 Video Compression Standards”, SPIE Proceedings Vol. CR60, Standards and Common Interfaces for Video Information Systems, October 25-26, 1995, Philadelphia, USA.