First publ. in: Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis : Rome, Italy, September 18 - 20, 2003 ; Pt. 1 / Ed.: Sven Lončarić ... - Zagreb : Fac. of Electrical Engineering and Computing, Univ. of Zagreb, 2003. - S. 370-374. - ISBN 953-184061-X

Packet Loss Protection of Scalable Video Bitstreams Using Forward Error Correction and Feedback

Youssef Charfi, Raouf Hamzaoui
University of Konstanz
Department of Computer and Information Science
D-78457 Konstanz
charfi,hamzaoui@inf.uni-konstanz.de

Abstract

We propose a system for packet loss protection of quality scalable video coders. The compressed bitstream is composed of a non-scalable base layer and an embedded enhancement layer as in MPEG4-Fine Granularity Scalability (FGS) and the emerging H26L-FGS. But the system can also be used with embedded video bitstreams like the 3D-SPIHT bitstream. Our system allocates a given transmission bitrate between the base layer bitstream and the enhancement layer bitstream in an efficient way using feedback. The base layer is protected with equal packet loss protection. For the enhancement layer, unequal packet loss protection is used to provide graceful degradation of video quality in the presence of packet erasure. Preliminary results for a hypothetical operational distortion-rate function and the 3D-SPIHT video coder show promising performance of the proposed system.

1 Introduction

Streaming video over the Internet has received tremendous attention due to the increasing demand for multimedia services and still poses many challenges due to its real-time constraints [1]. One of these challenges concerns improving reconstructed video quality in the presence of packet loss.

Quality-scalability or embeddedness is desirable for transmission over channels with different bandwidth and packet loss conditions. An efficient scalable video coding scheme is the MPEG4 fine granularity scalability [2]. An FGS video encoder compresses the raw video sequence into two substreams, a base layer bitstream and an enhancement layer bitstream. The base layer is a non-scalable bitstream that should be perfectly decoded. The enhancement layer is an embedded bitstream that can be truncated anywhere to achieve the target bitrate. This FGS scheme was recently proposed for the emerging H26L video coding [3].

To provide error protection for embedded bitstreams against packet loss, multiple description (MD) systems are largely used. They are also called unequal packet loss protection (ULP) systems. In an embedded bitstream, the first bits are more important than the following bits. The main idea of the MD systems is to provide ULP using forward error correction (FEC) to the embedded bitstream and to generate independent descriptions or channel packets. Recovering any number of the transmitted channel packets provides a certain reconstruction quality. Several researchers devised efficient end-to-end rate-distortion optimization algorithms for such systems [4, 5, 6, 7] using SPIHT [8] and JPEG2000 [9] for images and 3D-SPIHT [10, 11] and an embedded version of H263 [12] for video sequences.

The robustness of the MPEG4-FGS for Internet video streaming over packet erasure channels was studied by Schaar and Radha [13], who showed the suitability of ULP for MPEG4-FGS. However, no rate-distortion optimization was used. Yang et al. [14] proposed a degressive error protection algorithm for the MPEG4-FGS enhancement layer bitstream. They used the MD system described above with the hill-climbing algorithm proposed in [5] for the end-to-end rate-distortion optimization. However, they assumed that the base layer bitstream is received without loss and did not consider its error protection.

Majumdar et al. [15] used the efficient Hybrid-Automatic Repeat reQuest for packet loss protection of non-scalable codes. The non-scalable code was provided with exactly the amount of protection needed for a lossless recovery using forward error correcting codes with an efficient use of feedback.

In this paper, we propose a new system that provides packet loss protection to both the base layer and the enhancement layer . The base layer is non-scalable and therefore any loss in the base layer makes the enhancement layer useless and results in a very poor reconstructed video quality. In the proposed system, the base layer bitstream is protected using a technique similar to the one given in [15]. The base layer (BL) bitstream is protected using Reed-solomon (RS) codes and sent followed by the protection packets until the receiver acknowledges that the BL is correctly recovered. The remaining bandwidth is spent for the protection of the enhancement layer bitstream using ULP to provide graceful degradation in video quality in the presence of packet loss. For the rate-distortion optimization of the ULP of the enhancement layer, we use the Lagrange-based algorithm of [6]. This algorithm is faster than the hill-climbing algorithm of [5] used previously in [13].

The proposed system is also suitable for embedded video coders. To our knowledge, in all previous works that have considered packet loss protection of embedded video bitstreams, the ULP was applied to the whole bitstream. However, in many applications, the streaming server should adhere to a required minimum quality of service to deliver a video bitstream with a reconstruction quality that would be useful for the receiver. This means that a video bitstream with a bitrate of a minimum number of bits, R_0, should be correctly received to ensure the required minimum quality. Therefore, the embeddedness of the first R_0 bits of the bitstream would be useless and any loss in the base layer would not be tolerated. Thus, our system can be used for embedded video bitstreams by considering the first R_0 bits of the bitstream as the base layer and what remains as the enhancement layer.

The rest of the paper is organized as follows. In Section 2, we describe the proposed packet loss protection system. In Section 3, we derive an average expected distortion measure to evaluate the performance of our system. The efficiency of the proposed system is shown in Section 4 through experimental results. Section 5 gives the conclusions and suggests future work.

2 Packet loss protection system

We assume that the bitstream of a video sequence is divided into messages that can be decoded independently. A message can be the bitstream of a frame or a group of frames (GOFs). In the following, we consider the transmission of a GOF bitstream composed of a non-scalable base layer and an embedded enhancement layer over packet erasure channels.

The packet loss protection scheme is illustrated in Figure 1. Let N be the number of channel packets of length L bytes available for the transmission of a GOF. We split the BL into k packets of length L bytes and we append $(N-k)$ RS protection packets to the k BL packets. This means that L $(N, k, N-k+1)$ RS codes, which can correct up to $N-k$ erasures, are used. Thus, by correctly receiving at least k channel packets, the BL bitstream can be recovered.

In the packet loss protection system, the transmitter starts by sending the k BL packets to the receiver, followed by the RS protection packets until an acknowledgment (ACK) from the receiver arrives meaning that the k BL packets were correctly received, or the time deadline for the transmission of the GOF is reached (see Figure 1). Due to the real-time requirements of video, we limit the transmission deadline to N packets. We assume that x channel packets were sufficient for the correct reconstruction of the BL bitstream (see Figure 1).

Once the ACK is received, the transmitter starts the transmission of the embedded enhancement layer (EL) using the remaining bandwidth for the considered GOF. Let $M = N - x$ be the number of channel packets that remain for the transmission of the EL bitstream. We split the EL bitstream into M sublayers $m = 1, 2, ..., M$ indexed in order of decreasing importance such that each sublayer m can be split into m equal EL blocks $\mathrm{B}(m,b)$, $b = 1, 2, ..., m$ and protected with $M-m$ protection blocks $\mathrm{B}(m,b)$, $b = m+1, m+2, ..., M$ using $(M, m, M-m+1)$ RS codes. Therefore, M channel packets are formed such that a channel packet b, $1 \leq b \leq M$, is formed from the bth blocks $\mathrm{B}(m,b)$, $m = 1, 2, ..., M$. This ULP strategy ensures that all sublayers up to the bth sublayer can be decoded if any b out of the M channel packets are correctly received.

Let R_0 be the bitrate of the BL bitstream. We denote by R_b, $1 \leq b \leq M$ the bitrate of the EL bitstream up to sublayer b. The bitrate partition $\mathbf{R} = (R_1, \ldots, R_M)$ of the M sublayers can be optimized using efficient rate-distortion optimization algorithms [4, 5, 6, 7].

In the next section, we derive an average expected distortion measure to evaluate the end-to-end performance of the proposed system.

3 Expected distortion of the packet loss protection system

We consider a packet erasure channel where packets are either lost or correctly received. Suppose that the video bitstream is protected and sent through a packet erasure channel. We denote by $P_N(n)$ the probability of receiving exactly n packets out of N. Let X denote a random variable whose value is the number of channel packets that should be sent through the packet erasure channel so that k of them would be received. The average expected distortion can be expressed as

$$\mathsf{E}_N = \sum_{n=0}^{k-1} P_N(n) D_0 + \sum_{x=k}^{N} Prob(X = x) E_{N-x}(\mathsf{R}^{N-x}),$$

where D_0 denotes the distortion when the BL could not be recovered, $\mathsf{R}^{N-x} = (R_1, ..., R_{N-x})$ is the bitrate partition that minimizes the average expected distortion

$$E_{N-x}(\mathbf{R}^{N-x}) = \sum_{n=0}^{N-x} P_{N-x}(n) D(R_n)$$

of the ULP of the EL bitstream using $M = N - x$ channel packets, R_0 is the bitrate of the base layer, and $D(R_n)$, $0 \leq n \leq N - x$ denotes the distortion when the first R_n bits of the GOF bitstream are recovered.

If we assume a memoryless packet erasure channel with erasure probability p_e, the probability that the random variable X takes a value x is:

$$Prob(X = x) = \begin{cases} P_x(k) & \text{for} \quad x=k \\ P_{x-1}(k-1)(1-p_e) & \text{for} \quad k<x\leq N \end{cases}$$

where $P_N(n) = \binom{N}{n}(1-p_e)^n p_e^{N-n}$. Note that

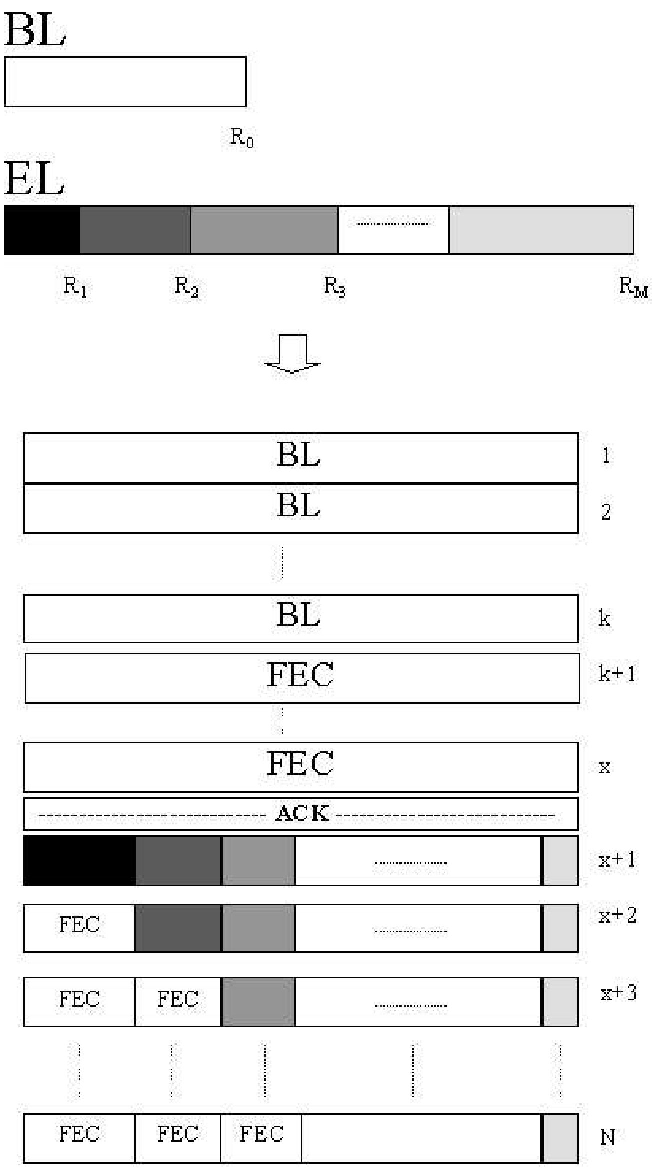

Figure 1. The packet loss protection scheme. Darker colors represent EL sublayers that are more important than the ones represented by lighter colors.

$$\sum_{n=0}^{k-1} P_N(n) + \sum_{x=k}^{N} Prob(X = x) = 1.$$

4 Results

In this section, we provide experimental results using first a hypothetical operational distortion-rate function and then the 3D-SPIHT video coder. We compare the proposed system with an MD system that uses ULP for the whole bitstream.

We assume that receiving a video bitstream with a bitrate of $R < R_0$ would be useless since the quality of the reconstructed video would be unacceptable. Therefore, we assume for $R < R_0$, $D(R) = D_0$, a constant. This assumption is reasonable in many applications and R_0 can be chosen by the streaming server (sender), taking into consideration both the performance of the video coder and the specification of the quality of the reconstructed video required by the receiver.

In all simulations, we model the communication channel as a memoryless packet erasure channel with various erasure probabilities $p_e = 0.05, 0.1, 0.15, 0.2$.

The first simulations were done using the operational distortion-rate function

$$D(R) = \begin{cases} D_0 & \text{for} \quad R<R_0 \\ D_0 2^{-2\alpha R} & \text{otherwise} \end{cases}$$

with $D_0 = 2000$ and $\alpha = 7/R^*$ chosen such that under lossless conditions at the target transmission bitrate of $R^* = NL$ bytes, $SNR = 10\log_{10}(D_0/D(R^*)) = 42$ dB.

We assume a frame rate of 10 frames per second (fps) and a GOF of 10 frames. We chose the bitrate of the base layer $R_0 = 32$ kbits, which corresponds to 32 kbits per second (kbps) and the length of the channel packets $L = 125$ bytes. Table 1 shows the SNR of the average expected distortion for the proposed system and the MD system at various transmission bitrates and an erasure probability of $p_e = 0.1$.

rate (kbps)	48	64	96	128
MD	11.56	16.11	24.62	33.09
Proposed	13.49	17.56	25.93	34.42

Table 1. SNR in dB of the average expected distortion of the proposed system and the MD system for a packet erasure channel with an erasure probability of $p_e = 0.1$ and various transmission bitrates.

Table 2 shows the same results for a transmission rate of 128 kbps at various erasure probabilities. The two tables show that the proposed system has up to about 2 dB better average expected reconstruction quality than the MD system.

p_e	0.05	0.1	0.15	0.2
MD	36.18	33.09	30.57	27.62
Proposed	37.17	34.42	31.95	29.70

Table 2. SNR in dB of the average expected distortion of the proposed system and the MD system for a transmission bitrate $R^* = 128$ kbps and various erasure probabilities.

We now present results for the embedded video coder 3D-SPIHT. We consider the standard 30 fps YUV 4:2:0

QCIF (176 × 144) Foreman video sequence. The original sequence was divided into GOFs of 16 frames each. The simulations were done on the first GOF. Note that in 3D-SPIHT, the GOFs are encoded and decoded independently. We chose a frame rate of 10 fps. The size of the channel packets was 640 bytes: 40 bytes for the packet header and 600 bytes for the payload. We chose the source bitrate of the base layer $R_0 = 38.4$ kbits, which corresponds to 24 kbps. We consider transmission rates up to 128 kbps.

The packet loss protection optimization algorithms require the operational distortion-rate curve of the original video, whose generation is time consuming. Because of the real-time constraint of video, the rate-distortion curve must be quickly computed. To this end, we compute the protection solutions using the four-parameter Weibull model of [18] that models the operational MSE-rate curve of 3D-SPIHT. However, the average expected mean square error of the luminance component (Y-MSE) was computed using the true distortion-rate curve.

Table 3 gives the PSNR of the average expected Y-MSE for the proposed system and the MD system at various transmission bitrates.

rate (kbps)	48	64	96	128
MD	24.47	25.49	29.64	31.10
Proposed	28.90	29.76	31.25	32.50

Table 3. PSNR in dB of the average expected Y-MSE of the proposed system and the MD system for a packet erasure channel with an erasure probability of $p_e = 0.1$ and various transmission bitrates. The video bitstream was generated by encoding the first 16 frames of the QCIF Foreman sequence using 3D-SPIHT

Table 4 shows the same results for a transmission bitrate of 128 kbps at various erasure probabilities. These results confirms the superiority of the proposed system over the MD system.

p_e	0.05	0.1	0.15	0.2
MD	31.83	31.10	30.94	29.99
Proposed	32.92	32.50	32.16	31.81

Table 4. PSNR in dB of the average expected Y-MSE of the proposed system and the MD system for a transmission bitrate $R^* = 128$ kbps and various erasure probabilities. The video bitstream was generated by encoding the first 16 frames of the QCIF Foreman sequence using 3D-SPIHT

5 Conclusions

We propose a packet loss protection system for quality scalable video bitstreams. It can be applied to both fully embedded bitstreams and bitstreams that are composed of a non-scalable base layer and an embedded enhancement layer. The system combines forward error correction and ARQ with an efficient use of protection bits and acknowledgments (one ACK per GOF). The simulation results for a hypothetical operational distortion-rate function and the embedded video coder 3D-SPIHT show the efficiency of the proposed system. Packet erasure channels were considered in this paper. A possible future work would be to incorporate the product codes of [16, 17] into the proposed system for the transmission over channels that suffer from both bit errors and packet loss.

References

[1] Wu, D., Hou, Y. T., Zhu, W., Zhang, Y., Peha, J. M., *Streaming video over the Internet: Approaches and directions, IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 1, Feb. 2001.

[2] Li, W., *Overview of fine granularity scalability in MPEG-4 video standard, IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 3, March 2001.

[3] He, Y., Yan, R., Wu, F., Li, S., *H.26L-based fine granularity scalable video coding, Q.6/SG16, Video Coding Experts Group (VCEG)*, 15th meeting, Pattaya, Thailand, Dec. 2001.

[4] Mohr, A. E., Ladner, R. E., Riskin, E. A., *Approximately optimal assignment for unequal loss protection, Proc. IEEE ICIP-2000*, vol. 1, pp. 367–370, Vancouver, Sept. 2000.

[5] Mohr, A. E., Riskin, E. A., Ladner, R. E., *Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction, IEEE Journal on Selected Areas in Communications*, vol. 18, no 7, 819–828, 2000.

[6] Puri, R., Ramchandran, K., *Multiple description coding using forward error correction codes,* in: *Proc. 33rd Asilomar Conf. on Signals and Systems*, vol. 1, pp. 342–346, Pacific Grove, CA, Oct. 1999.

[7] Stanković, V., Hamzaoui, R., Xiong, Z., *Packet loss protection of embedded data with fast local search, Proc. IEEE ICIP-02*, Rochester, New York, Sept. 2002.

[8] Said, A., Pearlman, W. A., *A new fast and efficient image codec based on set partitioning in hierarchical trees, IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.

[9] Taubman, D., Marcellin, M., *JPEG2000: Image Compression Fundamentals, Standards, and Practice*, Kluwer, 2002.

[10] Kim, B.-J., Xiong, Z., Pearlman, W. A., *Low bitrate scalable video coding with 3D set partitioning in hierarchical trees (3D SPIHT)*, *IEEE Trans. Circuits and Systems for Video Tech.,* vol. 10, pp. 1365–1374, Dec. 2000.

[11] Puri, R., Ramchandran, K., Lee., K. W., Bharghavan, V., *Forward error correction (FEC) codes based multiple description coding for Internet video streaming and multicast, Proc. of the Packet Video 2000 Workshop*, Sardinia, May 2000.

[12] Goshi, J., Ladner, R. E., Mohr, A. E., Riskin, E. A., Lippman, A., *Unequal loss protection for H.263 compressed video, the IEEE Data Compression Conference (DCC'03),* Snowbird, Utah, March 2003.

[13] Schaar, M. V., Radha. H., *Unequal packet loss resilience for fine-granular-scalability video, IEEE Trans. Multimedia*, vol. 3, no. 4, pp. 381-393, Dec. 2001.

[14] Yang, X. K., Zhu, C., Li, Z. G., Feng, G. N., Wu, S., Ling, N., *A degressive error protection algorithm for MPEG-4 FGS video streaming, Proc. IEEE ICIP-02*, Rochester, New York, Sept. 2002.

[15] Majumdar, A., Sachs, D. G., Kozintsev, I. V., Ramchandran, K., Yeung M. M., *Multicast and unicast real-time video streaming over wireless LANs, IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, June 2002.

[16] Sherwood, P. G., Zeger, K., *Error protection for progressive image transmission over memoryless and fading channels*, *IEEE Trans. on Communications,* vol 46, no. 12, pp. 1555–1559, 1998.

[17] Sachs, D. G., Anand, R., Ramchandran, K., *Wireless image transmission using multiple-description based concatenated codes*, *Proc. SPIE'00*, vol. 3974, pp. 300–311, Jan. 2000.

[18] Charfi, Y., Hamzaoui, R., Saupe, D., *Model-based real-time progressive transmission of images over noisy channels*, *Proc. WCNC-03, IEEE Wireless Communications and Networking Conference*, New Orleans, March 2003.