CONCEALMENT DRIVEN SMART SLICE REORDERING FOR ROBUST VIDEO TRANSMISSION

Enrico Baccaglini, Tammam Tillo, Gabriella Olmo

Dipartimento di Elettronica - Politecnico di Torino
Corso Duca degli Abruzzi 24 - 10129 Torino - Italy
Ph.: +39-011-5644094 - FAX: +39-011-5644149 - E-mail: `name.surname@polito.it`

ABSTRACT

In this paper we address a novel scheme to protect video sequences according to slice importance based on slice reordering, ULP and error-concealment techniques. The approach does not require the modification to the video decoder although an application-layer channel coding is required. Simulation results show that the proposed algorithm outperforms state-of-the-art approaches, reducing the gap with the upper-bound error-free performance curve. Moreover, the complexity of the additional stage required to pilot the protection allocation stage is negligible with respect to traditional ULP schemes.

Index Terms**—** Video coding, Image communication

1. INTRODUCTION

On-demand peer-to-peer (P2P) video streaming has recently gained interest, and is becoming popular due to the increasing number of wideband home connections and the recent advances in video coding technologies [1]. In such a scenario, there is no centralized approach to store and stream multimedia data. Instead, a client can receive data from many peers at the same time, without relying on index servers and multicast trees. All peers should provide resources (bandwidth, storage space) to the community, and the distributed nature of the network increases robustness in case of failures by replicating data over multiple peers.

In spite of this, P2P streaming presents many challenging issues that should be addressed to increase efficiency and robustness of the video transmission. A non-negligible packet loss rate must be coped with, mainly due to the unreliability of peers and links. In fact, during a video transmission, a sender may turn off, or a link be inefficient especially if wireless access is considered, or part of the network is congested. This may cause the loss of some macroblocks (MBs) or even of whole frames at the decoder.

Usually, retransmission of the missing portions of the data is not feasible due to the constraints of real-time content delivery, or too expensive in case of multicast transmission; therefore, techniques that insert a proper amount of redundant information to deal with information losses could be of great benefit. These techniques include error resilient coding schemes (such as flexible MB ordering and Intra MB refresh for the H.264/AVC standard [2]). They may mitigate the effect of channel errors, but usually increase computational complexity and reduce compression efficiency. Error concealment algorithms are applied at the decoder side, and can mitigate the effect of lost information. A simple concealment algorithm may recover missing slices by copying the same slice from the previous frame or performing more sophisticated spatio-temporal interpolation of missing slices [3].

Other methods to face packet losses include error-protection algorithms which assign forward error correcting (FEC) codes to the data. The amount of recovered data depends on the strength of the FEC codes. Unequal Loss Protection (ULP) algorithms [4] assign unequal amounts of FEC protection to the data so as to obtain different protection strength according to the importance of various portions of the data. ULP techniques provide graceful degradation of the quality, and this latter is then proportional to the number of successfully received packets. The main objective of these algorithms is to maximize the expected quality of the received video stream, and this is achieved by using different optimization techniques. In [5], ULP is applied to motion-compensated H.263 video over the Internet by reordering the bitstream so as to make it embedded. In [6], unequal amounts of protection are allocated to the different frames of a group of pictures (GOP). Stronger FEC codes are allocated to the former frames of a GOP while less redundancy is added to the latter ones to account for the different impact on the error propagation (drift). In [7], the error resilient features of H.264/AVC codec are exploited together with Reed-Solomon (RS) codes to enhance the protection of the stream. The optimal classification of MBs into slice groups and the optimal channel rate allocation are achieved by iteration.

Here, we propose a novel scheme named *concealment driven smart slice reordering* (CD-SSR) which combines ULP algorithms with concealment tools in order to increase the robustness of a video streaming. Video information is partitioned into slices, which are first sorted by relevance (considering their contribution to the distortion) and then protected with different FEC codes, according to the probability of packet loss. Hence, the most important slices are successfully decoded with high probability, because they have been protected with powerful FEC codes whereas less important slices are concealed leading to low distortion.

This paper is organized as follows. In Sec. 2 we introduce the proposed algorithm based on slice classification and FEC allocation. In Sec. 3 we provide some experimental results and in Sec. 4 we report some conclusions, highlighting the benefits of this encoding procedure and sketching future developments.

2. PROPOSED ALGORITHM

The proposed algorithm is composed of two stages. In the first stage, slices are classified and reordered according to a metric of importance. In the second stage, ULP is applied to the slices so that the most important ones are the most protected and will be successfully decoded with high probability.

This work has been supported in part by the European Union under the SEA project.

2.1. Encoder: Slice classification and reordering

In order to determine the importance of each slice of a video sequence, the encoder emulates the decoder concealment stage. A distortion Δd_i is evaluated for each slice i, as the mean squared error between the quality achieved if the slice is received, and the reconstructed version if the slice is lost and the concealment algorithm is applied. In the following, the distortion caused by the drift effect is not taken into account, and the related analysis is left to future research.

In addition to Δd_i, the proposed algorithm also requires the rate Δr_i devoted to encode each slice. This value is calculated in the video entropy coding stage of the encoder [2] and does not increase the complexity of the system.

After collecting the pairs (Δd_i, Δr_i) for all the slices, the CD-SSR algorithm sorts them from the most to the least important. The metric used to determine the importance of a slice is the normalized residual distortion (NRD) $\Delta d_i/\Delta r_i$. Slices with high NRD are classified as more important, and as a consequence, they will be protected with a stronger FEC code. The normalization term Δr_i quantifies the relative importance of a slice, with respect to the other slices that do not necessarily have the same rate. If slice i has higher $\Delta d_i/\Delta r_i$ than slice k, then the unit value of rate in slice i carries more information than in slice k.

The ordered list of pairs (Δd_i, Δr_i) allows the encoder to generate a particular rate-distortion function (from now on, called *rate-residual-distortion function*, RRD). This is made of discrete points $[\hat{R}_j, D(\hat{R}_j)]$ where $\hat{R}_j = \sum_{i=0}^{j} \Delta r_i$ (cumulative sum of the rate of the j most important slices), $D(\hat{R}_j) = \sigma^2 - \sum_{i=0}^{j} \Delta d_i$, $\sigma^2 = D(R_0) = D(0)$ being the distortion when no information is received. The RRD is used to pilot the redundancy allocation stage as described in Sec. 2.2 and does not represent the source distortion for a given rate; instead, it classifies the slices in terms of their relative importance, and is evaluated on a GOP basis. As an example, in the upper part of Fig. 1, the term R_1 marks the first 2 most important slices ($R_1 = \hat{R}_2$) and $R_2 = \hat{R}_5$, $R_3 = \hat{R}_{10}$. The values R_1, R_2, R_3 are selected by the ULP allocation scheme.

2.2. Encoder: ULP allocation

The level of inserted FEC codes at the encoder depends on the RRD function of each GOP and the probability of packet loss over the network. In [8], a Lagrangian rate-allocation procedure has been proposed to evaluate the suitable level of protection for different quality layers, using RS erasure codes and assuming independent losses. This approach is adopted also in this paper, using the information collected in the concealment-emulation stage at the encoder. An example of RRD function and the corresponding ULP allocation scheme is reported in Fig. 1. From now on, and without loss of generality, RS codes are assumed for ULP protection. Raptor codes or Fountain codes [9] can be applied as well.

To create M packets of R bits each, the Lagrangian optimization partitions the sequence of ordered slices into groups (chunks). A chunk is made of a variable number of slices and is protected with an (M, k) RS code, so that it can be decoded if at least k packets out of M are received. The Lagrangian optimization marks the RRD function at M different rates; chunk c is bounded by rates R_{c-1} and R_c ($1 \leq c \leq M$). As shown in Fig. 1, a chunk protected with the (M, k) code is divided into k parts of equal length, and accommodated into k different packets. The remaining $(M - k)$ packets host the RS parity bits required to protect the information according to an (M, k) RS code. Thus, each packet is composed by data and/or FEC of all the chunks. It contains both slices and redundancy information.

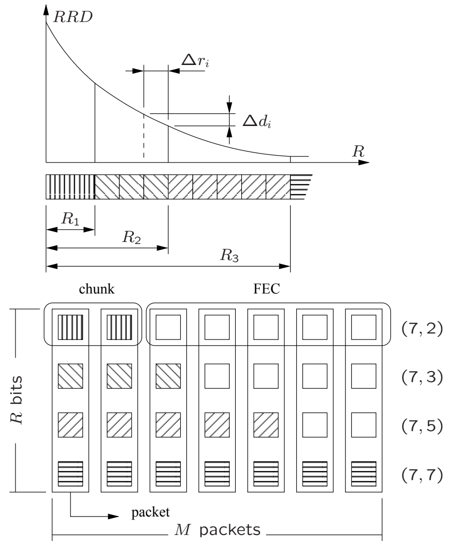

Fig. 1. The RRD function and the ULP coding scheme. $M = 7$ packets, $c = 4$ chunks.

The resulting total output rate [8] is $R_t = \sum_{j=1}^{M} \alpha_j R_j$ with

$$\alpha_j = \frac{M}{j(j+1)} \text{ for } j = 1, \dots M - 1$$

$$\alpha_M = 1$$

and denoting with $R_1, \dots R_M$ the M rates in which the RRD function is divided. This represents the Lagrangian constraint on the ULP system. The allocation procedure may partition the RRD function in $c \leq M$ rates; consequently not all the M rates are different, as some of them may overlap.

Denoting with p the probability of losing a packet in the network, the end-to-end distortion is given by [8]:

$$E_d = q_0 \sigma^2 + \sum_{j=1}^{M} q_j D(R_j) \tag{1}$$

with q_j denoting the probability that j out of M packets are delivered to the destination and $D(R_j)$ being the RRD function evaluated in R_j, for $1 \leq j \leq M$.

To obtain the values of R_j that minimize (1) while satisfying the rate constraint of the ULP system, a Lagrangian multiplier λ is used:

$$\mathcal{L} = q_0 \sigma^2 + \sum_{j=1}^{M} q_j D(R_j) + \lambda \left(\sum_{j=1}^{M} \alpha_j R_j - R_t \right) \tag{2}$$

To get the minimum of this function, the partial derivative of the Lagrangian function with respect to R_j, $j = 1, \dots, M$, should be equal to zero:

$$\frac{\partial \mathcal{L}}{\partial R_j} \triangleq 0 \Rightarrow \frac{q_j}{a_j} \frac{dD(R_j)}{dR_j} + \lambda = 0 \text{ for } j = 1, \dots, M \tag{3}$$

The optimal solution is obtained by locating the rates on the RRD function, where slopes are in proportion to each other according to the ratio q_j/a_j. In a practical implementation of the scheme, the rate R_j is only possible at slice level, and the slopes on the RRD function are quantized and defined at slice level. Then, the algorithm should take into account the approximation $\frac{dD(R_j)}{dR_j} \approx \frac{\Delta d_i}{\Delta r_i}$, with $i = f(j)$ and f being a mapping function that generates the index of the slice that corresponds to the rate R_j in the RRD function.

This scheme is similar to some extent to a multiple description coding scheme (MDC) [10]. Each packet sent over the network is independent of the others and the quality at the receiver (hence, the number of successfully decoded slices) increases with the number of received packets, regardless of the arrival order. However, the main difference with a traditional MDC scheme is that the decoder has to receive at least k_{min} packets in order to begin decoding, with k_{min} being the strongest FEC code (M, k_{min}). For example, in Fig. 1, the strongest code is $(7, 2)$.

2.3. Decoder side

At the receiver side, packets are collected within a time window shorter than or equal to the playout deadline. Packets received after the end of the time window are assumed to be lost. FEC decoding starts when the time window has elapsed. The loss of some packets results in missed slices, which are recovered by the concealment stage. The algorithm decodes the most important slices first, as they are protected with stronger codes which require k_{min} packets to be decoded. As packets are received, more slices are successfully decoded. Since slices are not received according to the corresponding frame ordering, the application at the receiver should take into account a delay of at least one GOP size. In fact, given that the decoder cannot display a frame before receiving all its slices, as a worst case, if the first slice of the first frame of a GOP is marked as the less important at the encoder, it will be available at the decoder only when nearly all packets of the corresponding GOP are received. In general, when all packets belonging to a GOP are received and FEC-decoded, the decoder can reorganize the sequence of slices according to the correct order and move it to the video application for displaying.

3. EXPERIMENTAL RESULTS

A set of simulations has been carried out using the first 90 frames of the standard CIF Foreman sequence @ 30 fps.

For the proposed algorithm, a GOP structure containing only P slices is adopted, and each slice is constrained to contain a maximum number of 66 MBs, leading to 6 slices per frame. Slicing policy yields a certain degree of error resilience; in fact, packet losses and consequently slices loss will appear as partial picture losses, thus making the subsequent concealment procedure more effective. As a consequence, small slice size improves system robustness while increasing header information. The number of reference frames is set to 5 and the entropy coding scheme is CAVLC. The H.264/AVC standard includes many resilience options that may further improve the results reported in the following. Nevertheless, the optimization and the joint benefits of other features, such as Intra MB refresh, flexible MB ordering, B slices, data partitioning, etc. are beyond the scope of the present work and not adopted here.

The value of packet loss probability p used to pilot the ULP allocation scheme is assumed to be known at encoding time, and can be determined by a feedback channel or by estimating the transmission characteristics. Each packet, which contains both data and FEC protection, is assumed to be lost using the Bernoulli model with probability p. The packet size R is set to 400 bytes and is kept unchanged for all simulations so as to enable fair comparisons of the results under the same transmission constraints. In particular, for all the addressed schemes, each slice is smaller than 400 bytes and fits into a packet.

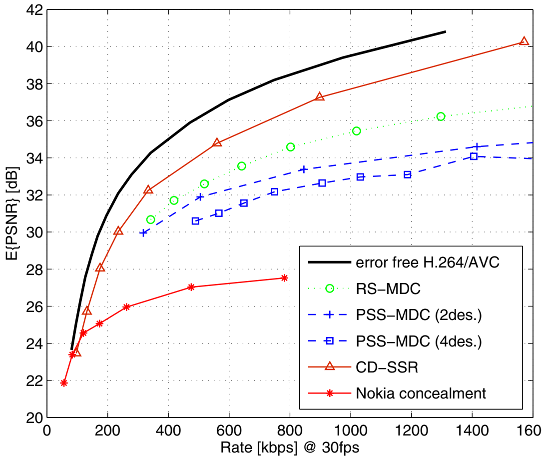

Fig. 2. Foreman CIF sequence, 30 fps, GOP= 11, $p = 10\%$. Average PSNR versus total output rate compared to RS-MDC, PSS-MDC, standard H.264/AVC with Nokia concealment and error-free H.264/AVC.

The performance is measured in terms of average luminance PSNR, obtained with 100 independent transmission trials. This amounts to the transmission of more than $4 \cdot 10^4$ packets, when working at 1000 kbps, which yields significant results from the statistical point of view.

In order to appreciate the robustness of the proposed scheme, we compare CD-SSR with the algorithms proposed in [11] and [12]. These are MDC algorithms based on polyphase spatial sub-sampling (PSS-MDC). In [11], four descriptions are created by H.264/AVC encoding of four QCIF sequences, obtained by sub-sampling the original CIF video, whereas in [12] two sub-sequences are generated by splitting odd and even rows of the frames. The two generated sub-sequences are then encoded using H.264/AVC. The results in [11] and [12] are obtained using both P and B pictures. The results of the redundant slice MDC scheme (RS-MDC) proposed in [13] are also reported. This latter is an MDC technique based on the redundant slice representation option, defined in the H.264/AVC standard. The redundant representation of a slice is used to replace missing portions of the compressed bitstream, thus yielding a certain degree of error resilience. The performance of the Nokia spatio-temporal concealment [14] are shown as well, and represent the quality obtained at the receiver if a concealment algorithm is applied but no slice reordering and no FEC insertion are performed at the encoder side. Finally, we also compare CD-SSR with the error-free H.264/AVC curve as a performance upper-bound.

In Fig. 2, the average PSNR obtained with the proposed allocation, $p = 10\%$ and GOP size 11 is shown versus the output rate. For all the algorithms, the total output rate varies from approximately 100 to 1600 kbps, according to the Quantization Parameter QP that varies from 22 to 51. The proposed concealment driven approach exhibits a noticeable performance improvement with respect to PSS-MDC. The gain range is between 2 and 5 dB. The CD-SSR

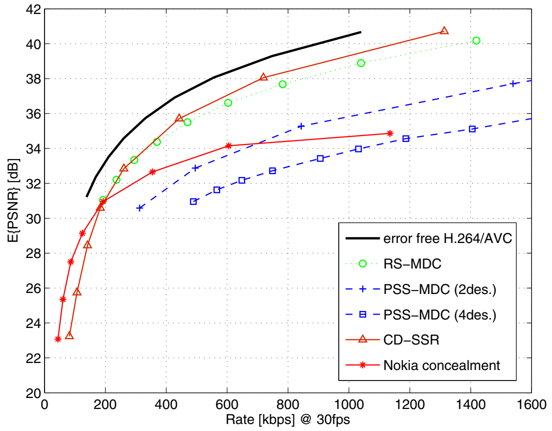

Fig. 3. Foreman CIF sequence, 30 fps, GOP= 45, $p = 1\%$. Average PSNR versus total output rate compared to RS-MDC, PSS-MDC, standard H.264/AVC with Nokia concealment and error-free H.264/AVC.

algorithm significantly outperforms also a system in which a concealment is applied without any knowledge of slice importance (i.e. Nokia). The system gain is up to 8 dB thus making the proposed approach robust for high probability of packet loss. As for the comparison with RS-MDC, the proposed algorithm gains approximately 2.5 dB. It is worth noticing that CABAC is used in the RS-MDC scheme while in CD-SSR, CAVLC is adopted. Preliminary results show that if CABAC is used also in CD-SSR, this allows for an additional 0.5 dB gain.

Similar results for $p = 1\%$, reported in Fig. 3, show that a gain of approximately 3 dB can be obtained with respect to PSS-MDC (2 descriptions), while the gain versus RS-MDC is approximately 1 dB and up to 4 dB versus the Nokia spatio-temporal concealment.

As for complexity, the CD-SSR algorithm requires an additional stage at the encoder to sort slices and insert FEC, and an additional stage at the decoder to collect packets and to decode FEC codes. A standard H.264/AVC encoder running on a Intel Pentium IV 2 GHz, 1 GB RAM requires 2.3 s to encode a frame with the parameters reported above. The generation of the RRD function required an additional time of 0.06 s/frame and the Lagrangian allocation procedure is performed in a negligible time. Concerning the decoding time, the FEC decoding time strictly depends on the code applied to the data. For example, if Raptor codes are applied, then the decoding time is nearly linear with the number M of packets to be decoded.

4. CONCLUSIONS AND FUTURE WORK

In this work we propose a novel scheme to protect a video sequence according to the importance of its slices. The importance of a slice is evaluated at the encoder side using its normalized residual distortion, evaluated by emulating the concealment stage. Simulation results showed that the proposed algorithm outperforms state-of-the-art approaches by approximately 1 dB at low p and up to 2.5 dB at high probability of packet loss. The complexity of the additional stage required to pilot the allocation stage is negligible with respect to traditional ULP schemes. Ongoing work includes the study of the propagated distortion of a slice to the following frames and a detailed study of the complexity of the encoding/decoding stages.

5. REFERENCES

[1] "IEEE Transactions on Multimedia," vol. 9, no. 8, Dec 2007.

[2] T. Wiegand, G.J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul 2003.

[3] Yanling Xu and Yuanhua Zhou, "H.264 video communication based refined error concealment schemes," *IEEE Trans. on Consumer Electronics*, vol. 50, no. 4, pp. 1135 – 1141, Nov 2004.

[4] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE Journal on Selected Areas in Commun.*, vol. 18, no. 6, pp. 819–828, Jun 2000.

[5] J. Goshi, A.E. Mohr, R.E. Ladner, and E.A. Riskin, "Unequal loss protection for H.263 compressed video," *IEEE Trans. on Circuits and Systems for Video technology*, vol. 15, no. 3, pp. 3232 – 3234, Mar 2005.

[6] Tao Fang and Lap-Pui Chau, "A novel unequal error protection approach for error resilient video transmission," *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 4, pp. 4022–4025, May 2005.

[7] N. Thomos, S. Argyropoulos, N.V. Boulgouris, and M.G. Strintzis, "Robust transmission of H.264/AVC video using adaptive slice grouping and unequal error protection," *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 593 – 596, Jul 2006.

[8] R. Puri and K. Ramchandran, "Multiple description source coding using forward error correction codes," in *Proc. of Thirty-Third Asilomar Conference on signals, systems and computers*, Oct 1999, vol. 1, pp. 342–346.

[9] M. Luby, "LT codes," in *Proceedings of The 43rd Annual IEEE Symposium on Foundations of Computer Science*, Nov 2002, vol. 1, pp. 271 – 280.

[10] V.K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep 2001.

[11] R. Bernardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, "Polyphase spatial subsampling multiple description coding of video streams with H.264," in *Proc. of IEEE International Conference on Image Processing*, Oct 2004, pp. 3213–3216.

[12] A. Vitali, F. Rovati, R. Rinaldo, R. Bernardini, and M. Durigon, "Low-complexity standard-compatible robust and scalable video streaming over lossy/variable bandwidth networks," in *Proc. of ICCE. 2005 Digest of Technical Papers. International Conference on Consumer Electronics*, Jan 2005, pp. 7–8.

[13] T. Tillo, M. Grangetto, and G. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding (to appear)," *IEEE Trans. on Circuits and Systems*, Jan 2008.

[14] Ye-Kui Wang, M.M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," in *Proc. of IEEE International Conference on Image Processing*, Sept 2002, vol. 2, pp. 729–732.