

ROBUST TRANSMISSION OF H.264/AVC VIDEO USING ADAPTIVE SLICE GROUPING AND UNEQUAL ERROR PROTECTION

N. Thomos^{1,3}, S. Argyropoulos^{1,3}, N. V. Boulgouris², and M. G. Strintzis^{1,3}

¹Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece

²Department of Electronic Engineering, Division of Engineering, King's College London, UK

³Informatics and Telematics Institute, Thessaloniki, Greece

ABSTRACT

We present a novel scheme for the transmission of H.264/AVC video streams over lossy packet networks. The proposed scheme exploits the error resilient features of H.264/AVC codec and employs Reed-Solomon codes to protect effectively the streams. The optimal classification of macroblocks into slice groups and the optimal channel rate allocation are achieved by iterating two interdependent steps. Simulations clearly demonstrate the superiority of the proposed method over other recent algorithms for transmission of H.264/AVC streams.

1. INTRODUCTION

The demand for multimedia transmission over best effort networks facing packet erasures, like the Internet, motivated most recent research on real-time streaming applications. Considering that the network is unaware of the transmitted content, we realize that packet erasures during transmission could cause significant problems in demanding applications such as video streaming. Error resilient coding schemes like the H.264/AVC standard [1] have been proposed to overcome these problems. Unfortunately, H.264/AVC error resilient tools increase computational complexity and have a negative impact on compression efficiency. Therefore, schemes combining unequal error protection (UEP) algorithms with error resilient tools are often shown to be advantageous for transmission of H.264/AVC coded streams, while maintaining the computational cost at reasonable level.

In a recent work [2], data partitioning of H.264/AVC and high-memory RCPC codes were proposed for video transmission over wireless channels. Data partitions were unequally protected according to their significance. A similar approach was presented in [3], which also uses the data partitioning mode of H.264/AVC. Unequal channel rate allocation was performed using lagrangian techniques. Robust transmission of H.263 streams was examined in [4]. A packetization method of slices and a UEP algorithm for joint optimization of macroblock coding parameters and selection of FEC codes were presented. An algorithm which adaptively classifies the data packets of MPEG-2 encoded video streams into two quality of service (QoS) classes was proposed in [5]. Packet classification into priority groups was also studied in [6]. Intra-frame interleaving and RS codes were used to improve error resilience.

The proposed scheme is based on macroblock classification into three slice groups and UEP of H.264/AVC streams. Prior to transmission, macroblocks are classified into three slice groups by examining their contribution to the overall video quality. The slice groups are unequally protected using Reed-Solomon (RS) codes determined

by a UEP algorithm based on dynamic programming techniques. To the best of our knowledge, the present method is the first jointly utilizing the explicit mode of the H.264/AVC Flexible Macroblock Ordering (FMO) [7] and channel coding techniques. The resulting system is evaluated and is shown to outperform the recently proposed method in [3].

2. ADAPTIVE MACROBLOCK SLICE GROUPING

In this section, we present the macroblock classification policy employed by the proposed scheme. In H.264/AVC, the macroblocks are coded in groups termed slices, exploiting this way the spatial dependencies more effectively by partially sacrificing the error localization capabilities of the decoder. The encoding parameters of macroblocks are declared in a header which includes the encoding parameters of all macroblocks in a slice. Therefore, slices are self-contained in the sense that they can be independently decoded without utilizing data from other slices of the current frame. In the following, each such slice is assumed to be transmitted in a single transmission unit which is termed "packet".

This work assumes that macroblocks are classified in three categories. This is depicted in Fig. 1(a). Due to this classification, if a slice is erased, only the macroblocks at slice boundaries can be concealed effectively using neighboring slices that were received errorlessly at the decoder. Specifically, erroneous frame areas are efficiently concealed using non-normative concealment methods.

The limitation of the above conventional slice formation is partially overcome in H.264/AVC, in which error concealment is improved by means of FMO. Using FMO, groups of macroblocks, known as slice groups, are formed. Slice groups consist of one or more slices; this enables better error localization. As reported in [7], the FMO mode, in conjunction with advanced error concealment methods applied at the decoder, maintains the visual impact of the losses at a low level even at loss rates up to 10%. Apart from predefined patterns, fully flexible macroblock ordering (explicit mode) is also allowed which changes dynamically the macroblock classification throughout the entire video sequence based on the video content.

The provision for dynamic formation of slice groups is exploited by the proposed system. Specifically, the macroblocks are classified into three categories, as in [6], labelled as "high", "medium", and "low" according to the following rules:

- if $D_{MB} < T_l \cdot D_{mean}$, the examined macroblock is classified to the "low" importance slice group.
- if $T_l \cdot D_{mean} \leq D_{MB} < T_h \cdot D_{mean}$, the examined macroblock is classified to the "medium" importance slice group.
- if $D_{MB} \geq T_h \cdot D_{mean}$, the examined macroblock is classified to the "high" importance slice group.

This work was supported by the EC under contract FP6-511568 3DTV.

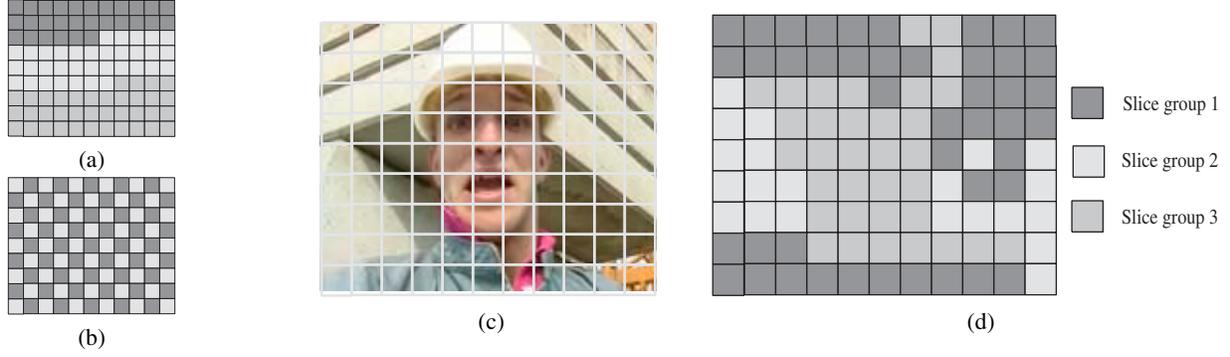


Fig. 1. Macroblock classification (a) without FMO, (b) employing FMO (checker board), (c) original frame of “Foreman”, (d) Classification map using explicit FMO mode.

where T_l, T_h ¹ denote the classification thresholds, D_{MB} is the distortion of the examined macroblock, and D_{mean} is the mean value of the macroblocks distortion. The distortion D_{MB} initially used is determined assuming the frame as a single slice group. After the classification of macroblocks the compression efficiency will degrade and thus, more bits will be needed for the encoding of each macroblock than those initially estimated. This is taken into account by the rate-control algorithm at the encoder. In Figs. 1(c) and (d), a frame of the “Foreman” sequence and its MacroBlock Allocation map (MBAmap) for three classes, according to the above rules, are presented. It is worth noting that other classification techniques could be applied, such as classification between foreground and background. The proposed macroblock classification has been selected, since it boosts the performance of the overall system, in terms of subjective quality, due to the more efficient protection of intense motion and high-texture image areas.

Since the transmission scenario is over packet erasure networks, channel codes should be used for the efficient protection of H.264/AVC streams. To this end, we developed an algorithm for the efficient channel rate allocation. This is presented in the ensuing section.

3. CHANNEL RATE ALLOCATION

In the preceding analysis for an optimal classification, it was assumed that the distortion between the original and reconstructed coefficients is known. In practice, however, the actual distortion depends on the reconstructed coefficients *after* channel decoding. This means that the processes of slice grouping and channel allocation are actually interdependent. For this reason, the formation of slice groups and their unequal error protection are optimized in our system by iterating two interdependent steps which are presented in detail in the following.

The employment of the FMO makes the formation of slice groups of unequal importance possible. In our approach, the unequally important slice groups consist of equally-sized slices (packets). RS codes were chosen for UEP protection due to their excellent error recovery properties for transmission over packet erasure networks. Nevertheless, other channel codes could be used with the proposed scheme without any performance loss. Since different frames have, in general, different classification maps, channel rate allocation is performed at the frame level. The optimization objective is to find:

- the optimal classification of macroblocks into slice groups

¹The average values of T_l and T_h are 0.7 and 1.1 respectively.

- the optimal RS channel protection of slice groups.

The optimization algorithm intends to minimize the average expected distortion \bar{D} subject to a channel rate constraint determined by experimentation. This is necessary to avoid overprotection of the first frames. Specifically, without this constraint, the first frames in the sequence would allocate the maximum allowable RS protection and drift would occur.

The average expected distortion \bar{D} in case of s slice groups is equal to:

$$\bar{D} = \sum_{l=1}^s \left\{ \sum_{i=1}^{N_l} D_{f,l} \cdot P_l(i) + \sum_{i=N_l+1}^{N_l+K_l+1} D_{f,i,l} \cdot P_l(i) + D_{f,PC,l} \cdot P_l(N_l + K_l) \right\} \quad (1)$$

where K_l and N_l are the number of source and RS packets of the l_{th} slice group respectively and $P_l(i)$ is the probability of losing i packets out of the $N_l + K_l$ packets of the l_{th} slice group and is given by:

$$P_l(i) = \binom{N_l + K_l}{i} \cdot p^i \cdot (1-p)^{N_l+K_l-i} \quad (2)$$

where p is the packet erasure probability.

The distortion $D_{f,PC,l}$ in the last term of (1) expresses the distortion when all packets of the l_{th} slice group are erased and concealed by slice group replication. Finally, $D_{f,i,l}$ represents the distortion introduced when the current frame slice group is concealed by slices received intact and $D_{f,l}$ the distortion when the RS protection is sufficient to recover all erased packets. It should be noted that the distortion terms do not consider error propagation. This does not affect seriously the estimated distortion since macroblock updates usually cope effectively with drift.

3.1. Reed-Solomon rate allocation

In this section, we present a solution to the previously formulated optimization problem. The optimization objective is actually two-fold. Specifically, it includes the determination of both the number of slices classified into each slice group and their RS protection. In general, reaching an optimal solution of the joint problem is a difficult task. In this work, we propose a two-step optimization procedure, which iteratively determines the packet classification and the RS protection. Although, this approach to the solution of the optimization problem does not guarantee global optimization, in prac-

tice it yields very satisfactory results. The optimization procedure is summarized as follows:

1. Determine the RS protection of each frame.
2. Determine thresholds T_h and T_l .
3. Classify macroblocks into slice groups according to T_h and T_l .
4. Find the optimal RS protection for the above classification.
5. Calculate the expected distortion of allowable neighboring macroblock classifications with the restriction that a single packet can be exchanged between successive classes.
6. Compare the expected distortion of the ancestor classification with the lowest average distortion of all descendant classifications of step 3. If a classification with lower expected distortion exists, it is considered as optimal and steps 2 to 6 are repeated, else the algorithm is terminated. When the same packet is exchanged, in two successive iterations, between two slice groups the algorithm is again terminated.

Our objective is to optimize the RS allocation by minimizing the expected distortion given by (1). We follow the dynamic programming algorithm presented in [8] to reduce the computational cost.

4. EXPERIMENTAL RESULTS

The proposed scheme for transmission of H.264/AVC streams over IP/UDP/RTP was evaluated for two standard QCIF sequences “Foreman” and “Carphone” coded at 10 *frame/s* (fps) and CIF sequence “Paris” coded at 30 fps. Group of Pictures (GOPs) of *IPPP* . . . structure consisting of 100 and 300 frames were considered for QCIF and CIF sequences respectively. Intra update of 10% of the total number of frame macroblocks according to distortion criteria (the macroblock with the largest prediction error) was used to prevent error propagation. Although the intra updated macroblocks could be classified to the “high” importance slice group, this approach was not followed because this leads very often to an undesirable increment of the rate used for the coding and the protection of this slice group.

Channel simulations were carried out using the NS-2 event simulator, employing a uniform bit error model. Video sequences were encoded using JM 8.3 of the H.264/AVC standard. Non-normative advanced error concealment methods were applied at the encoder for the estimation of the end-to-end distortion. At the decoder the same concealment policy was followed.

The JM 8.3 was modified to support fully flexible MBAmapping for each frame. The picture parameter set (PPS) packets containing the classification maps, were protected using the (3,1) RS codes which are able to correct all possible error patterns occurring in the considered channel conditions. The use of these RS codes is affordable because of the small PPS packet size. The selected error protection policy ensures that high-quality video sequences is decodable even in the case of high packet-error rates.

The packet sizes were 50 and 200 bytes for the QCIF and CIF sequences respectively. The use of relatively small packet sizes endowed our scheme with the ability to achieve better error localization and prevent drift. Fig. 4 illustrates that small packets guarantee the decoding of video sequences of satisfactory quality, whereas in error-free cases schemes with larger packets benefit. Considering the above, our choices of packet sizes achieve a good tradeoff between robustness and compression efficiency. To avoid the increased bandwidth for packet headers, the RoHC was used to reduce

the IP/UDP/RTP header from 40 bytes to approximately 3 bytes. The main disadvantage of RoHC is the increased processing delay at routers, which leads to end-to-end delays. However, as shown in several other techniques (*e.g.*, in [9]) it is possible to use RoHC for real-time communication over multi-hop networks.

Three variants of the proposed scheme were considered for comparison purposes: the full scheme, which classifies macroblocks into three slice groups according to the rules presented in Section 2, a scheme which divides the image into two slice groups according to the checker board pattern and finally a simplified scheme which treats each frame as a single slice group (non-FMO coding). The RS protection for the above schemes was determined using the UEP algorithm of Section 3.

The transmission schemes were evaluated for a variety of channel conditions. All reported results are PSNR averages over 100 simulations. The proposed schemes are compared with an implementation of the method in [3] since it is a joint source/channel coding scheme which is in the spirit of our method. In Figs. 2(a), 2(b), and 2(c), results for transmission over packet networks with 10% packet losses are presented for the “Foreman”, “Carphone”, and “Paris” video sequences. Optimization was performed assuming 10% packet error rate. From Figs 2(a), 2(b), and 2(c), it can be easily seen that the three slice group variant of the proposed method decodes higher quality videos more frequently than the rest of the methods. The performance gap between our best-performing scheme and the method in [3] is significant and grows wider as the transmission bit rate increases. The performance gain is attributed to the synergistic cooperation of adaptive group slicing which allows better error localization and to the UEP algorithm which enables the application of less powerful RS codes, saving rate for the transmission of source rate.

The proposed scheme was also evaluated for transmission in channel mismatch conditions. In Fig. 3, results are presented for the “Foreman” sequence coded at 128 *kbps* for the case where the schemes are optimized for packet error rate equal to 10% and transmitted over channels which exhibit various packet error rates. The results show that the proposed full scheme is superior to the method in [3] and the other variants of the full scheme. When the transmission is error free the proposed full scheme has lower performance due to the application of stronger RS codes and the inferior compression efficiency when FMO is used. The gain achieved by the full scheme over the other methods becomes more impressive when the channel conditions deteriorate. Specifically, for the most of the considered transmission scenarios, the performance gap is roughly 2 *dB*. It is worth noting that our three slice group method provides graceful degradation in image quality when the channel becomes more erroneous, whereas the other methods collapse. This is due to both the UEP algorithm and the adaptive group slicing.

5. CONCLUSIONS

A novel method was proposed for the transmission of H.264/AVC coded sequences over packet erasure channels. The proposed scheme exploits the error resilient features of H.264/AVC codec and employs Reed-Solomon codes to protect effectively the resulting streams. A framework for optimal classification of macroblocks into slice groups and optimal unequal error protection was also proposed. Experimental evaluation showed the superiority of the proposed method in comparison to well-known schemes for transmission of H.264/AVC streams.

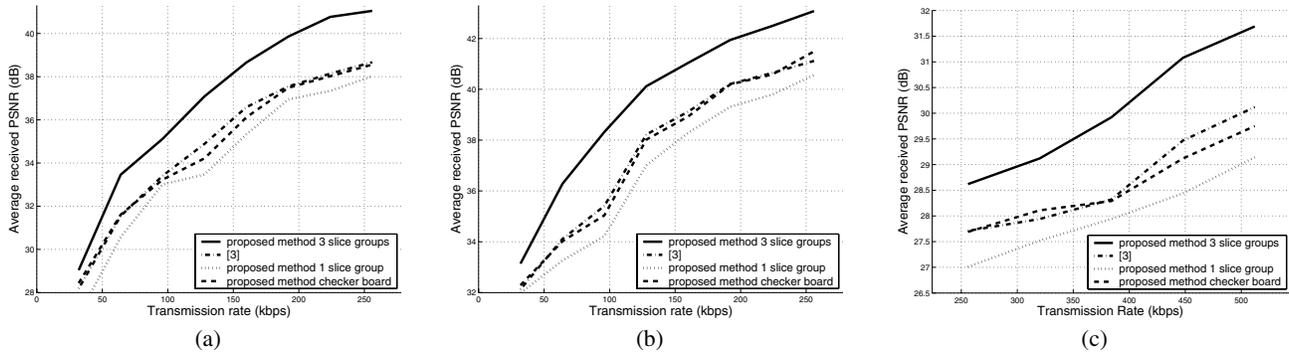


Fig. 2. Comparison of the proposed methods with the method in [3] for the transmission of the QCIF sequences “Foreman” and “Carphone”, and the CIF sequence “Paris”. Reconstruction quality in terms of Mean PSNR is reported. Results for packet error rate equal to 10%: (a) Foreman, (b) Carphone, (c) Paris.

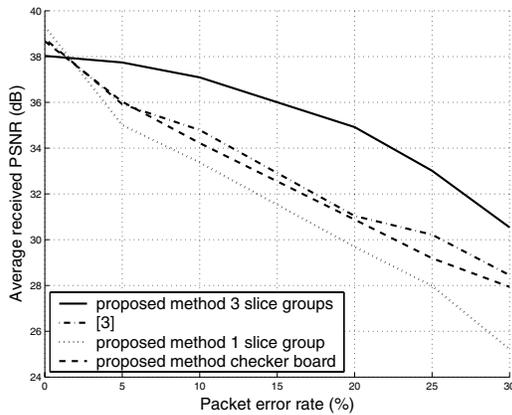


Fig. 3. PSNR comparison for the transmission of the QCIF sequence “Foreman” coded at 128 kbps as a function of the packet error rate. The scheme was optimized for 10% packet error rate and tested for various packet error rates.

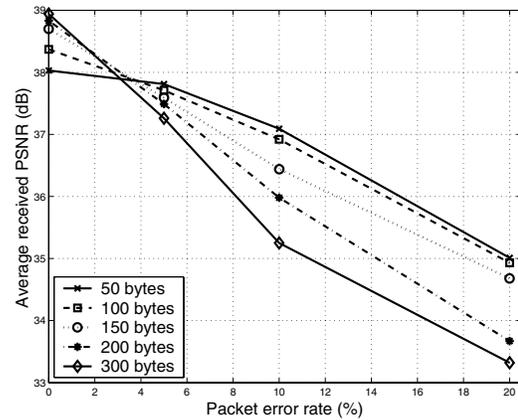


Fig. 4. Average received Mean PSNR for transmission of the “Foreman” sequence coded at 128 kbps over channels facing packet error rates in the range [0,20] for various packet sizes.

6. REFERENCES

- [1] *Information Technology - Coding of Audio-Visual Objects - Part 10: Advanced Video Coding. Final Draft International Standard*, ISO/IEC FDIS 14 496-10, 2003.
- [2] T. Stockhammer and M. Bystrom, “H.264/avc data partitioning for mobile video communication,” in *Proc. IEEE Int. Conference on Image Processing*, Singapore, Oct 2004, pp. 545–548.
- [3] O. Harmanci and A. M. Tekalp, “Optimization of h.264 for low delay video communication over lossy channels,” in *Proc. IEEE Int. Conference on Image Processing*, Singapore, Oct 2004, pp. 3209–3212.
- [4] E. Masala, H. Yang, K. Rose, and J. C. De Martin, “Rate-distortion optimized slicing, packetization and coding for error resilient video transmission,” in *Proc. DCC Data Compression Conference*, Snowbird, UT, USA, Mar 2004, pp. 182–191.
- [5] E. Masala, D. Quaglia, and J. C. De Martin, “Adaptive picture slicing for distortion-based classification of video packets,” in *Proc. IEEE Workshop on Multimedia Signal Processing*, Cannes, France, Oct 2001, pp. 111–116.
- [6] Q. Qu, Y. Pei, J. W. Modestino, and X. Tian, “Error-resilient wireless transmission using motion-based unequal error protection and intra-frame packet interleaving,” in *Proc. IEEE Int. Conference on Image Processing*, Singapore, Oct 2004, pp. 837–840.
- [7] S. Wenger, “H.264/AVC over IP,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645–656, July 2003.
- [8] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, “Wireless image transmission using turbo codes and optimal unequal error protection,” *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1890–1901, 2005.
- [9] F. Fitzek, S. Rein, P. Seiling, and M. Reisslein, “RObust Header Compression (ROHC) Performance for Multimedia Transmission over 3G/4G Wireless Networks,” *Wireless Personal Communications*, vol. 32, no. 1, pp. 23–41, 2005.