ERROR-RESILIENT TRANSMISSION OF H.264/AVC STREAMS USING FLEXIBLE MACROBLOCK ORDERING

N. Thomos[1,3], S. Argyropoulos[1,3], N. V. Boulgouris[2], and M. G. Strintzis[1,3]

[1] Electrical and Computer Engineering
Department
University of Thessaloniki
Thessaloniki, GR-54124, Greece
e-mail: strintzi@eng.auth.gr

[2] Department of Electronic Engineering,
Division of Engineering
King's College London
Strand, London, WC2R 2LS, UK
e-mail: nikolaos.boulgouris@kcl.ac.uk

[3] Informatics and Telematics Institute
Thermi-Panorama Road
GR-57001 Thermi
Thessaloniki, Greece
e-mail: {nthomos, savvas}@iti.gr

Keywords: H.264/AVC, error-resilient coding, FMO, unequal error protection, channel coding.

Abstract

We present a novel scheme for the transmission of H.264/AVC video streams over packet loss networks. The proposed scheme exploits the error resilient features of H.264/AVC codec and employs Reed-Solomon codes to protect effectively the streams. A novel technique for adaptive classification of macroblocks into three slice groups using Flexible Macroblock Ordering is also proposed. The optimal classification of macroblocks and the optimal channel rate allocation are achieved by iterating two interdependent steps by means of dynamic programming techniques. 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, like the Internet, motivated most recent research on real-time streaming applications. Since networks employing the Internet Protocol (IP) face packet erasures and are, in general, unaware of the transmitted content, error resilient coding schemes endowed with valuable error resilient tools, like the H.264/AVC standard [10], have been proposed. However, these tools increase the computational complexity, necessitating the careful selection of the employed error resilient tools and the appropriate amount of channel protection using Forward Error Correction (FEC) codes [4].

In a recent work [8], data partitioning of H.264/AVC and high-memory Rate Compatible Punctured Convolutional codes (RCPC) [2] were proposed for transmission over wireless channels. Data partitions were unequally protected according to their significance. Robust transmission of H.263 streams was examined in [6]. A packetization method of slices and an Unequal Error Protection (UEP) algorithm for joint optimization of macroblock coding parameters and selection of FEC codes was 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]. An approach which also uses the data partitioning mode of H.264/AVC was presented in [3]. The transmitted data were protected by Reed-Solomon (RS) codes and channel rate allocation was performed using lagrangian techniques. Packet classification into priority classes was also studied in [7]. Intra-frame interleaving and RS codes were used to improve error resilience.

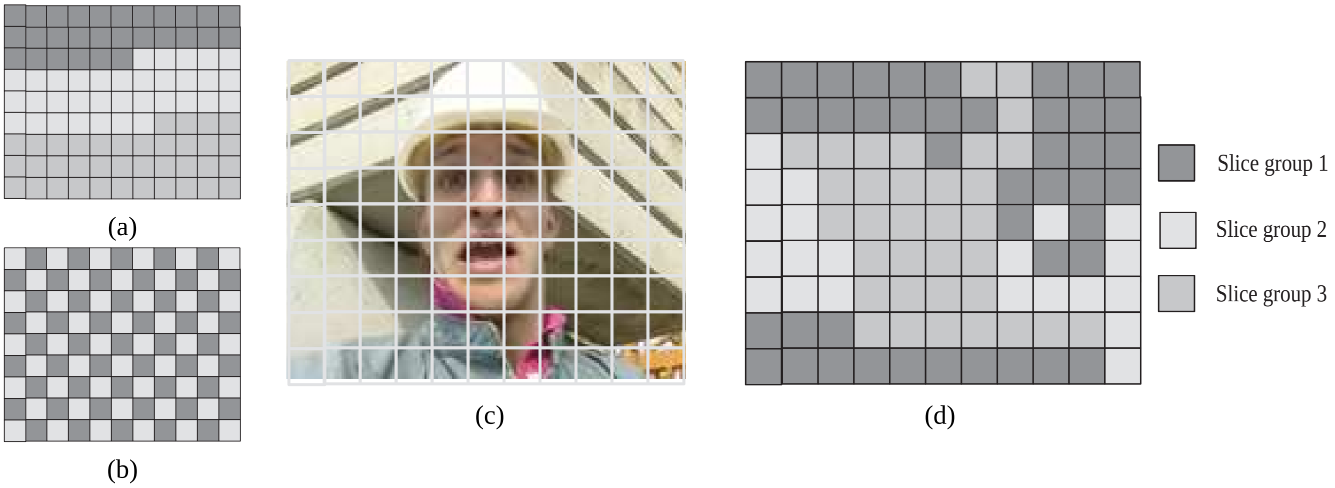
The proposed scheme is based on macroblock classification and UEP of H.264/AVC streams. A novel algorithm is proposed to classify macroblocks into three slice groups using a new error resilient tool in H.264/AVC, termed Flexible Macroblock Ordering (FMO). RS codes are selected for the efficient protection of video streams. The RS protection of each slice group is determined using a channel rate allocation algorithm based on dynamic programming techniques. The overall system is evaluated and is shown to outperform significantly the recently proposed method in [3].

2 Adaptive macroblock slice grouping

In this section, we present the macroblock classification policy employed by the proposed system. In H.264/AVC, macroblocks are coded in groups, termed slices, which are structures of jointly-encoded macroblocks. Slices are self-contained in the sense that their decoding does not require data from other slices of the current frame. A macroblock classification into three slices is depicted in Fig. 1(a). Due to this classification, if a slice is lost during transmission, only the macroblocks located at slice boundaries can be concealed effectively using neighboring slices that were received intact at the decoder. The limitation of the above conventional slice formation is partially overcome in H.264/AVC, in which error concealment is improved by means of H.264/AVC FMO mode. Using FMO, macroblocks can be allocated to slices in non scan order and form slice groups. The H.264/AVC standard provides

---

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



(a) (b) (c) (d)

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

some macroblock classification patterns, like the checker board (Fig. 1(b)). Apart from predefined patterns, fully flexible macroblock ordering (explicit mode) is also allowed. According to this mode, macroblock classification into slice groups may change dynamically based on the video content throughout the entire video sequence.

The provision for dynamic formation of slice groups is exploited by the proposed system. Specifically, macroblocks are classified into three slice groups with respect to their relative importance. As a measure of macroblock importance (based on the mean square error), we use the distortion D_{MB} defined as

$$D_{MB} = \frac{1}{x_{MB} \cdot y_{MB}} \cdot \sum_{i=1}^{x_{MB}} \sum_{j=1}^{y_{MB}} \left(c_{i,j} - \widetilde{c}_{i,j}\right)^2 \qquad (1)$$

where x_{MB}, y_{MB} are macroblock dimensions and $c_{i,j}$, $\widetilde{c}_{i,j}$ are the original and the reconstructed pixel values of a macroblock respectively. Prior to macroblock classification, the mean value D_{mean} of the macroblock distortions of the current frame is computed as

$$D_{mean} = \frac{1}{N_{MB}} \cdot \sum_{i=1}^{N_{MB}} D_{MB_i} \qquad (2)$$

where N_{MB} is the total number of macroblocks in a frame and D_{MB_i} is the distortion (in terms of MSE) associated with the i_{th} macroblock. Subsequently, each macroblock is classified into one out of three slice groups, as in [7], labelled as “low”, “medium”, and “high” according to the following rules:

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

The distortion D_{MB} initially used is determined assuming that the frame consists of only one slice group. The determination of the thresholds T_l and T_h is discussed in the Section 3. In Figs.1(c) and (d) a frame of the “Foreman” sequence and its Macroblock Allocation map (MBAmap) into three classes following the above rules are presented.

3 Channel rate allocation

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

The proposed algorithm takes into account the importance of each slice group and allocates more RS packets (RS slices) to slice groups carrying important information and less to the rest. Each slice will be assumed to be transmitted in a single transmission unit which will be termed “packet”. The terms “packets” and “slices” will be used interchangeably in the analysis below, with “packet” meaning the transmitted stream corresponding to a slice. During the channel rate allocation process, slices are transferred from one slice group to another leading to new slice group formations. The algorithm classifies optimally the macroblocks into slice groups and determines their optimal channel protection, as it will be described below. The packet formation of a slice group after RS encoding is illustrated in Fig. 2.

The optimization objective is two-fold. Specifically we intend to find:

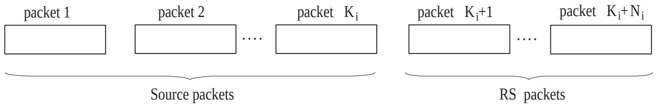


Figure 2: Packet formation of a slice group.

- the optimal classification of macroblocks into slice groups
- the optimal RS channel protection of each slice group.

The optimization algorithm minimizes the average expected distortion $\overline{D}$ subject to a channel constraint, which limits the allocated RS protection, preventing overprotection of the first frames and degradation in image quality. For specifying the maximum number of RS packets N_f (per frame) we calculate $\overline{D}$ for a large set of channel rates r_c. This is given by:

$$r_c = \frac{\sum\limits_{i=1}^{N_{seq}} N_{f,i} \cdot p_l}{r_T} \tag{3}$$

where N_{seq} is the number of frames in a sequence, $N_{f,i}$ the number of RS packets in frame i, p_l the packet length and r_T the transmission rate. From the examined channel rates r_c, the one achieving the lowest distortion is considered as optimal.

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

$$\overline{D} = \sum_{l=1}^{s} \left\{ \sum_{i=1}^{N_l} D_{f,l} \cdot P_l\left(i\right) + \sum_{i=N_l+1}^{N_l+K_l-1} D_{f,i,l} \cdot P_l\left(i\right) + \right. \\ \left. + D_{f,PC,l} \cdot P_l\left(N_l + K_l\right) \right\} \tag{4}$$

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

$$P_l\left(i\right) = \left(\begin{array}{c} N_l + K_l \\ i \end{array} \right) \cdot p^i \cdot \left(1-p\right)^{N_l+K_l-i} \tag{5}$$

where p is the packet erasure probability.

The last term in Equation (4) $D_{f,PC,l}$ expresses the distortion when all packets of the l_{th} slice group are erased and concealed by slice group replication of the previous frame. Finally, $D_{f,i,l}$ denotes the distortion introduced when a slice of a slice group in the current frame is concealed by other slices which have been received intact and $D_{f,l}$ the distortion of the l_{th} slice group when the RS protection is sufficient to recover all erased packets.

In this work, we propose a two-step optimization procedure, which iteratively determines the packet classification and the RS protection. Although the above procedure does not guarantee global optimization, in practice 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 the 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 according to Equation (4) of neighboring macroblock classifications with the restriction that a single packet is exchanged between successive classes.
6. Compare the expected distortion of the parental 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 a packet is exchanged, in two successive iterations, between two slice groups the algorithm is again terminated.

From the above analysis, it is obvious that when the classification thresholds are close to the optimal values, then the channel rate allocation algorithm converges rapidly. Thus the computational cost is significantly reduced. The thresholds used for the first frame of a sequence are initially determined by experimentation. In the sequel, the resulting classification of macroblocks of each frame is used for the refinement of the thresholds values which are used for the initial classification of the next frame.

The dynamic programming algorithm of [9] is used for reduced computational complexity. In [9], each branch in trellis corresponds to the application of a RS code to a slice group. Firstly, the algorithm determines the RS protection of more important slice groups and then the respective protection of less important slice groups. The trellis nodes are intermediate stages where decision is made about the best RS allocation up to the s_{th} slice group protection. From the paths converging at a node (corresponding to allocations with equal source and transmission rate), the path attaining the lower expected distortion is retained (survivor), while the rest are pruned. At the final stage, from the survivor paths, the one with the overall lower expected distortion is considered as the resulting RS allocation. The state number in the trellis depends on the allowable RS protection levels.

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 $IPPP\ldots$ 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 very often leads to an undesirable increment of the rate used for the coding and the protection of this slice group.

The NS-2 simulator employing a uniform bit error model was used for channel simulations. The video sequences were encoded using a modified version of JM 8.3 of the H.264/AVC standard to support fully flexible macroblocks allocation map (MBAmap) for each frame. After the application of the channel rate allocation algorithm, a refinement of the MBAmap could be applied to move single macroblocks to the same slice group as its neighboring. However, this approach is not followed, in general, since it harms the overall compression performance, due to the use of the arithmetic coder. For the estimation of the end-to-end distortion, 30 independent channel-decoder pairs have been used in the encoder and non-normative advanced error concealment methods have been applied [1]. Robust Header Compression (RoHC) was incorporated to reduce IP/UDP/RTP packet overhead from 40 bytes to approximately 3 bytes. The packet size was set to 50 and 200 bytes for QCIF and CIF sequences respectively.

Three variants of the proposed scheme were considered for comparison issues:

- the full scheme which classifies macroblocks into three slice groups as described in section 2.
- a scheme which divides the image into two slice groups following checker board pattern.
- a simplified scheme which treats each frame as a single slice group (non-FMO coding).

All schemes are optimized using the algorithm of section 3. The proposed schemes are compared with an implementation of the method in [3] which uses two data partitions and employs slices of fixed number of macroblocks. This method was selected for comparison purposes since it is a joint source-channel coding scheme which is in the spirit of our method. In Figs. 3 (a), (b) and (c) results are reported for transmission over packet erasure networks experiencing 10% packet error rate, while in Figs. 3 (d), (e) and (f) the channel encounters 20% packet losses. It can be easily concluded from Fig. 3 that the proposed system decodes more frequently higher quality videos. The performance gap between the full scheme and the method in [3] is significant and grows as the transmission rate increases. The performance gain is attributed to the adaptive slice grouping which improves error concealment and to the efficient channel protection. It must be noted that the UEP algorithm boosts the performance of the proposed scheme, since it enables the application of less powerful RS codes, and thus, saves rate which can be used for source coding. Considering the above, the performance gain should not be attributed solely to the adaptive slice grouping itself or the UEP algorithm, but rather to their synergistic cooperation.

The proposed scheme was also evaluated for channel mismatch scenarios. In Fig. 4 results are presented for transmission of "Foreman" sequence coded at 128 *kbps*. All schemes are optimized for 10% packet loss rate and tested for a large variety of channel conditions. The results demonstrate that the proposed full scheme performs better than its other variants and the method in [3]. The worse performance of the full scheme in error free case is attributed to the inferior compression efficiency when FMO is used. The gain of the full scheme over the other methods widens when the channel conditions deteriorate. Specifically, for the majority of transmission scenarios the performance gap is close to 2 dB. It is worth noting that the performance of the full scheme decreases more gracefully unlike the other methods due to the adaptive slice grouping and the efficient channel protection. An instance of our simulation is presented in Fig. 5 which shows that the full scheme has less fluctuation in image quality in comparison to the method in [3].

In Fig. 6, we present a visual comparison of the decoded sequences by the proposed method. From this figure, it can be easily seen that the adaptive slice group scheme outperforms the other schemes in which one or two slice groups are used. It should be also noticed that the proposed method does not induce any annoying artifacts.

5 Conclusions

A novel method was proposed for the transmission of H.264/AVC coded sequences over packet erasure channels. The proposed scheme exploits FMO for adaptive slice grouping to improve error resilience. Moreover, RS codes are employed to protect effectively the resulting streams. Adaptive slice grouping is used for improved error resiliency. 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.

References

[1] *Joint Model Reference Encoding Methods and Decoding Concealment Methods*. JVT-I049d0, San Diego, USA, Sep 2003.

[2] J. Hagenauer, E. Offer, and L. Papke, "Iterative Decoding of Binary Block and Convolutional Codes," *IEEE Trans. Information Theory*, vol. 42, no. 2, pp. 429–445, Mar. 1996.

[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] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, 1983.

[5] E. Masala, D. Quaglia, and J. C. D. 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.

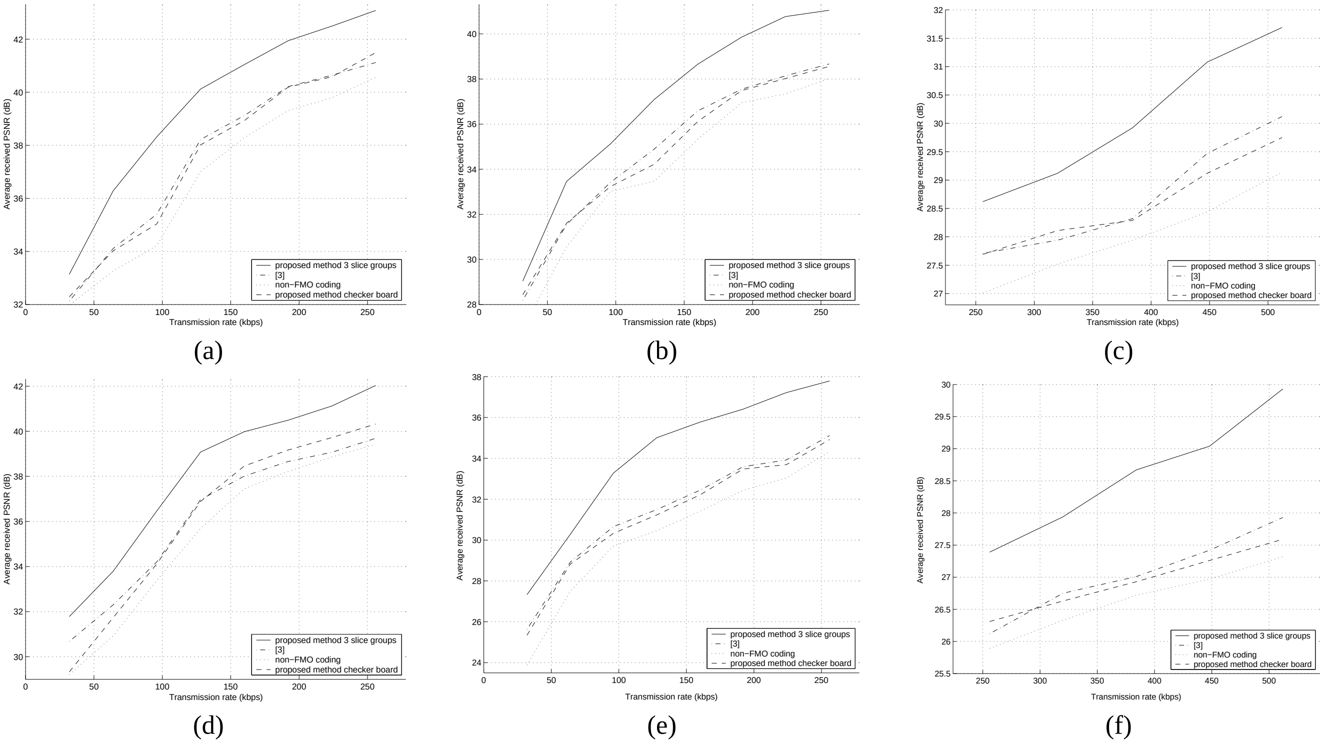
[6] E. Masala, H. Yang, K. Rose, and J. C. D. Martin, “Rate-distortion optimized slicing, packetization and coding for error resilient video transmission,” in *Proc. DCC Data Compression Conference*, Snowbird, UT, USA, Mar 2004, pp. 182–191.

[7] 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.

[8] 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.

[9] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, “Wireless image transmission using turbo codes and optimal unequal error protection,” *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1890–1901, November 2005.

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



(a) (b) (c)

(d) (e) (f)

Figure 3: Comparison of the proposed methods with the method in [4] 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. Comparison for packet error rate equal to 20%: (d) Foreman, (e) Carphone, (f) Paris.

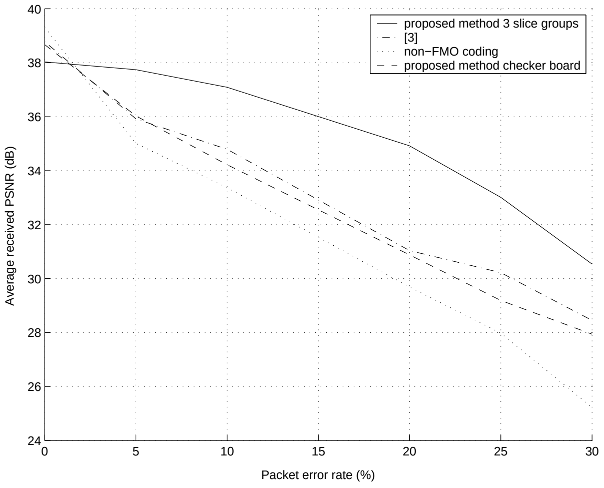


Figure 4: PSNR comparison of the proposed full scheme with the method in [3] for the transmission of the QCIF sequence “Foreman” coded at 128 *kbps* over packet erasure channel with 10% packet losses.

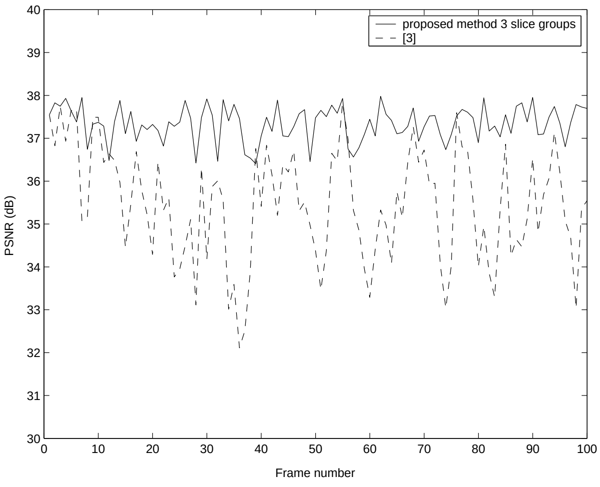


Figure 5: PSNR comparison for the transmission of the QCIF sequence “Foreman” at 128 *kbps* as a function of the packet error rate. The scheme was optimized for 10% packet error rate and tested for various packet error rates.

(a) (b) (c) (d) (e) (f)

Figure 6: Visual comparison of the proposed methods for the frame 68 of the “Foreman” sequence coded at 96 *kbps*. Comparison of visual artifacts induced due to transmission over packet networks encountering 10% packet error rate. Error free transmission of the (a) single slice group variant of the proposed scheme (non-FMO coding) (37.32 *dB*), (c) two-slice groups (checkerboard) variant of the proposed scheme (36.58 *dB*), (e) three-slice groups variant of the proposed scheme (36.06 *dB*). Frames harmed by noise when sequences are encoded using the (b) single slice group variant of the proposed scheme (non-FMO coding) (32.86 *dB*), (d) two-slice groups (checkerboard) variant of the proposed scheme (33.47 *dB*), (f) three-slice groups variant of the proposed scheme (34.93 *dB*).