H.264 MULTIPLE DESCRIPTION CODING BASED ON REDUNDANT PICTURE REPRESENTATION

Tammam Tillo, Marco Grangetto, Gabriella Olmo

Dipartimento di Elettronica - Politecnico di Torino
Corso Duca degli Abruzzi 24 - 10129 Torino - Italy
Ph.: +39-011-2276510 - FAX: +39-011-5644099
`tammam.tillo@polito.it`
`marco.grangetto@polito.it`
`gabriella.olmo@polito.it`

ABSTRACT

In this paper a novel H.264 multiple description technique is proposed. The coding approach is based on the redundant slice representation option, defined in the H.264 standard. In presence of losses, the redundant representation can be used to replace missing portions of the compressed bitstream, thus yielding a certain degree of error resilience. This paper addresses the creation of two balanced descriptions based on the concept of redundant slices, while keeping full compatibility with the H.264 standard syntax and decoding behavior. Moreover, a practical algorithm for redundancy tuning as a function of the packet loss rate is provided. Experimental results demonstrate that the proposed technique favorably compares with other H.264 multiple description approaches.

Index Terms— **Multiple Description Coding, H.264, redundant slices**

I. INTRODUCTION

Nowadays, most video compression techniques are based on hybrid transform coding and predictive motion compensation. However, the presence of the prediction loop makes the compressed data very sensitive to errors. As a consequence of the possible error propagation, video transmission over networks subject to packet losses is very challenging. In order to overcome this vulnerability, one can consider the use of multiple description coding (MDC), where different non hierarchical representations (or *descriptions*) of the same data, yielding mutually refinable information, are generated in order to be transmitted over independent paths.

The most popular methods to generate MD are based on the pioneering MD scalar quantizer (MDSQ) proposed in [1]; this principle has been applied to video coding in [2]. Another class of methods employs correlating transforms [3]; this approach has been applied to motion compensated MD video coding in [4]. These methods, although providing good performance, are conceived as stand alone codecs, and are incompatible with video standards such as H.264 [5]. Some MDC schemes based on pre- and post-processing stages [6], [7], [8] have been proposed in order to make the multiple description generation independent of the actual video codec. In [9], the authors suggest to use the *slice group* coding tool available in H.264 in order to create two balanced descriptions. Although the generated descriptions are indeed H.264 compliant, the use of the slice group modality impairs the compression efficiency. In order to mitigate this effect in [10] the authors suggest using three-loop slice group MDC; nevertheless, this latter solution still exhibits a performance impairment in terms of coding efficiency at the two side encoders. In [11] spatial sub-sampling is used to design a MDC technique and results in the case of H.264 video coding are shown. In this paper we propose a novel MDC coding approach which is fully compliant with the H.264 video coding standard. This goal is achieved by exploiting the redundant slice coding option available in the standard. Finally, the proposed technique permits to optimally allocate the MDC redundancy according to the network status.

II. PROPOSED ALGORITHM

The proposed algorithm aims at creating two balanced[1] descriptions of a video sequence, each of which is an H.264 compliant bitstream. This objective is fulfilled using the concepts of *primary* and *redundant* slices, defined in the H.264 standard [12]. Primary slices are used to code the primary picture, and are associated to a normative decoding procedure. On the other hand, redundant slices represent an alternative representation of a picture; the H.264 recommendation does not specify a normative decoder behavior in presence of redundant slices. Clearly, when some of the samples in the decoded primary picture cannot be correctly decoded due to errors or transmission losses, whereas the redundant slice can be correctly decoded, the decoder shall replace the samples of the decoded primary picture with the corresponding ones of the decoded redundant slice.

The proposed coding approach is summarized in Fig. 1. First of all, a single H.264 bitstream is generated, where a redundant representation of each coded picture is encoded. A redundant picture, which is generally composed of several redundant slices, is obtained employing a quantization parameter $QP_r = QP_p + \Delta QP$, where QP_p is the quantization parameter of the primary picture. The value of ΔQP allows one to tune the amount of coding redundancy inserted in the compressed bitstream. It is important to notice that each redundant P slice is predicted using only the previously encoded primary slices. In other words, the redundant slices cannot be used to predict the following

---

[1] Two descriptions are said to be balanced if they are encoded at the same rate and yield the same distortion when separately decoded

redundant pictures. As a consequence, when a redundant slice is decoded and used to replace a lost primary one, a certain error is introduced in the decoder prediction loop; this error propagates for a certain number of pictures; nevertheless, in the H.264 case the generated decoder drift does not produce strong artifacts, as reported by [13] in the case of bitstream switching.

The single redundant H.264 bitstream can now be reorganized in order to form two balanced descriptions. This task can be easily accomplished by interlacing primary and redundant slices so as to create two H.264 bitstreams which contain alternatively the primary and the redundant representation of each slice, as depicted in Fig. 1. In order to guarantee that the two bitstreams can be independently decoded, the crucial information contained in the *Sequence Paramater Set* (SPS) and *Picture Parameter Set* (PPS) must be duplicated in both descriptions. The output of this process are two balanced descriptions, which are fully H.264 compliant.

At the decoder side, if both descriptions of a slice are received, the decoding of the primary representation is guaranteed. On the other hand, if a whole description is lost, the received one is a compliant H.264 bitstream, and can be decoded yielding inferior, yet acceptable quality. In the more general case, the two H.264 descriptions are transmitted across two independent physical or virtual channels, characterized by a certain packet loss probability. At the receiver side, the best representation of each slice, available within its play-out deadline, can be decoded. Only in the case that both descriptions are lost, the decoder must invoke a concealment algorithm. The experimental results presented in the following section are obtained using the temporal concealment available in JM9.4 reference software for the missing predicted macroblock (MB). When losses occur in an I slice, the simple MB replacement with the co-located MB in the previous frame is used.

It is worth noticing that, as in every MDC scheme, the introduced redundancy is beneficial in the case of single description reception, whereas it impairs the overall performance in case of two descriptions reception. Therefore, the introduced redundancy should match the network condition in order to achieve the best average performance. This objective can be achieved by estimating the quantization parameters for the redundant and primary slices depending on the network status.

III. PRACTICAL REDUNDANCY ALLOCATION PROBLEM

In order to optimize the rate devoted to the redundant representation, we develop an analytical model for the expected distortion at decoder side. To this end, let us assume that the GOP comprises N frames, and that the k-th slice is replaced by its redundant representation. The impact on the total distortion of the current GOP, $d_t[k]$, can be evaluated as:

$$\begin{aligned} d_t[k] &= d_r[k] + d_m[k] + d_m[k]\cdot f[2] + \ldots + d_m[k]\cdot f[N-k] \\ &= d_r[k] + d_m[k]\sum_{n=1}^{N-k} f[n] \end{aligned}$$

where $d_r[k]$ is the distortion due to the k-th redundant slice, which is used to replace the lost primary slice, and $d_m[k]$ is the propagated distortion due to the mismatch generated in frame k, between the primary and redundant representation of the slice. The $f[n]$ function is used to model the distortion attenuation over time experienced at distance n from the mismatch [14].

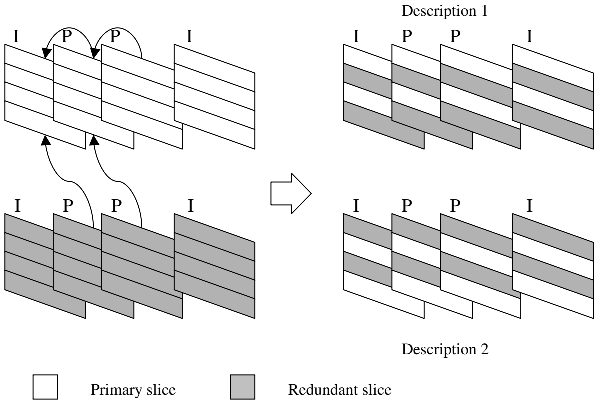

Fig. 1. Multiple description scheme for H.264 using redundant slice.

Assuming that p is the probability of loss, and that each slice fits a single packet, each primary representation is received with probability $(1-p)$, whereas the probability of having a primary representation replaced with the redundant one is $p(1-p)$, and no representation is received with probability p^2. Assuming that this latter term is negligible, the expected distortion experienced in one GOP, can be evaluated as:

$$\overline{d}[k] \approx (1-p)d_p[k] + p(1-p)d_t[k] \tag{1}$$

where $d_p[k]$ is the distortion due to the fine quantization of the slice.

The quantization error is known to be uncorrelated with the sample values, especially at high coding rates [15]; as a consequence, also the mismatch error injected by different samples can be considered uncorrelated [15]. Therefore, we can work out the following expression for the overall expected distortion over the GOP:

$$\overline{D} \;=\; \sum_{k=1}^{N}\overline{d}_k = (1-p)\sum_{k=1}^{N}\left(d_p[k] + p d_t[k]\right) \tag{2}$$

The optimal allocation problem can then be formulated as a constrained minimization represented by:

$$J = \overline{D} + \lambda\sum_{k=1}^{N}\left(r_p[k] + r_r[k]\right) \tag{3}$$

where $r_r[k]$ and $r_p[k]$ are the rates of the redundant and the primary representations of the slice, and λ is the Lagrangian multiplier. Imposing $\nabla J = 0$ we get

$$\frac{\partial d_p[k]}{\partial R_p[k]} = p\frac{\partial d_t[k]}{\partial r_t[k]}\ ,\ \forall k \in [1, N] \tag{4}$$

The optimal condition expressed by (4) allow one to design an optimal redundancy allocation strategy. This can be accomplished using a closed loop rate-control technique, able to optimally partition the rate budget between the primary and redundant slices. This can be performed using the following standard H.264 R-D approximation [16]:

$$\frac{\partial d}{\partial r} = -0.85 \times 2^{(\frac{QP-12}{3})} \tag{5}$$

Assuming that $d_r[k] \gg d_m[k]$, and as a consequence $d_t[k] \approx d_r[k]$ we obtain:

$$QP_r[k] - QP_p[k] = \Delta QP = -\lfloor 3\log(p) \rfloor \tag{6}$$

Equation (6) states that the optimal ΔQP, varies as a function of the packet loss probability p.

IV. EXPERIMENTAL RESULTS

In this section, the proposed technique is validated by means of a number of experimental trials. A demo version of the software is available at: *http://www.telematica.polito.it/sas-ipl/diva/*.

Since the compressed video is going to be transmitted across a packet network, it is important to adopt a proper data partitioning strategy. To this end, each picture in the video sequence is partitioned into a certain number of slices so as to guarantee that each compressed H.264 *Network Access Level* (NAL) unit is smaller than the network maximum transfer unit (MTU). Moreover, the picture slicing yields a certain degree of error resilience; in fact, packet losses will appear as partial picture losses at the decoder side, thus making the concealment procedure more effective. As a consequence, a limit on the maximum number of MBs per slice is enforced as well.

The experimental results reported in this section are obtained with the *Foreman* CIF sequence at 30 frame per second using the following settings. The primary slice quantization parameter QP_p is selected in the interval $(22, 38)$, in order to span a wide range of coding rates. A GOP of 21 pictures containing only one I followed by all P frames is adopted and a frame buffer containing the last 5 coded pictures is employed. MTU of 400 bytes are considered and each slice is constrained to contain a maximum number of 80 macroblocks, i.e. 5 slices per frame in the case of CIF sequences. It is worth pointing out that the H.264 standard includes other resilience options that may further improve the results reported in following. Nevertheless, the optimization and the joint benefits of other features, such as Intra MB refresh, flexible MB ordering, etc. are beyond the scope of the present paper. The performance is measured in terms of average luminance PSNR, obtained with 100 independent transmission trials. This amounts to the transmission of more that $4 \cdot 10^4$ packets, which yields significant results from the statistical point of view.

In Fig. 2 the average PSNR obtained by the proposed technique, when the descriptions are transmitted across channels with packet lossy probability $p = 0.01, 0.05$ and 0.10 are shown versus the total coding rate R. Using $\Delta QP = 8$ yields an average coding redundancy, measured as the ratio between the redundant rate and the total rate, $\rho = 0.27$. The error free curve, i.e. the central decoder performance, is reported in order to evaluate the penalty, in terms of average PSNR, induced by packet losses over the two links. In the same figure the results available in [11] are shown. This latter is a MDC algorithm based on polyphase spatial sub-sampling (PSS-MDC). Four descriptions are created by H.264 encoding of four QCIF sequences, obtained by sub-sampling an original CIF video. In [11] the video is encoded on 4 H.264 compliant bitstreams, whereas the decoder must be properly enhanced to handle the descriptions. This technique has been selected as a benchmark for our algorithm performance. The results in [11] are obtained with the same GOP $N = 21$, but using both P and B pictures; the transmission conditions and the slice dimensions are the same as those used in this paper. From Fig. 2 it turns out that the proposed MDC approach exhibits a noticeable performance improvement with respect to [11]. The average PSNR gain ranges from more than 2 dB in the case $p = 0.10$, up more than 4 dB in the case $p = 0.01$. It is worth pointing out that the performance of the proposed approach is improved by selecting the best value of redundancy for each value of p. This can be easily achieved by tuning the parameter ΔQP according to (6). On the other hand, [11] is constrained to a rigid spatial sub-sampling, which does not allow one to control the amount of redundancy. This turns out to depend on the spatial correlation among the sub-sampled sequences. Moreover, H.264 coding of four QCIF sequences can be highly inefficient with respect to the original full resolution H.264 coding performance, so imposing a high redundancy level. Finally, the proposed scheme, which is based on a single H.264 encoder/decoder (plus a simple multiplexing stage at the decoder side in order to merge the two descriptions into a single H.264 bitstream), favorably compares with [11] in terms of computational complexity as well. In fact, at the encoder side, PSS-MDC requires the operation of four encoders. At the decoder side, four H.264 decoders must synchronously process the received bitstreams and share a common reference frame buffer containing the interpolated full resolution pictures.

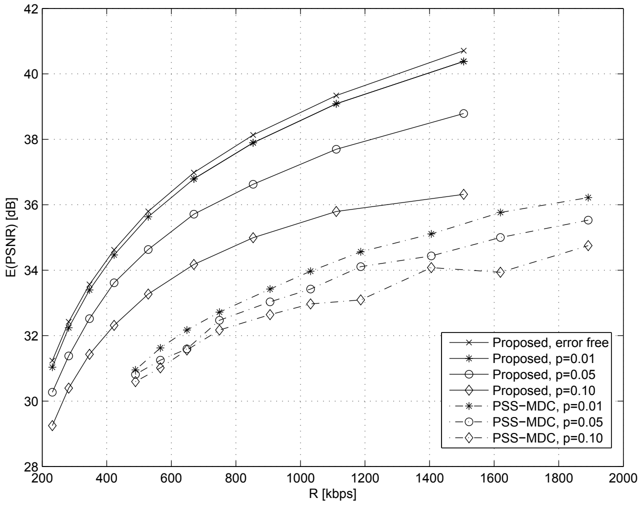

Fig. 2. Average PSNR of the proposed technique, compared to PSS-MDC [11] in the cases $p = 0.01, p = 0.05, p = 0.10$.

The performance of the proposed scheme can also be evaluated in terms of central and side distortion (measured as the average of the two side distortions). In Fig. 3 the central and side performance in the three cases $\Delta QP = 5, 8, 11$, corresponding to $\rho = 0.34, 0.27, 0.23$, are shown versus R. Obviously, ΔQP

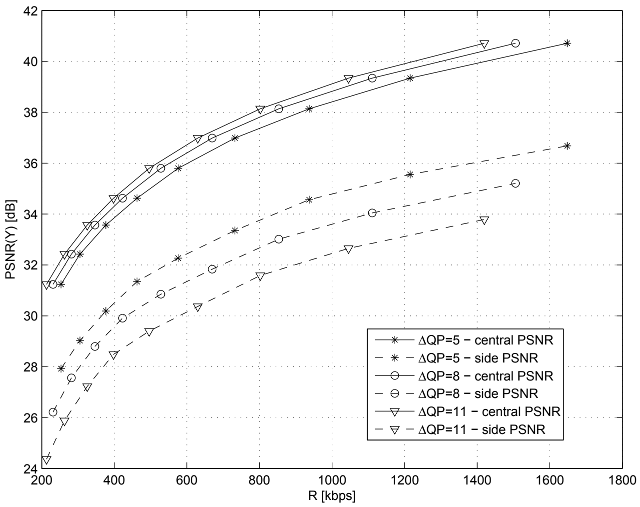

Fig. 3. Luminance PSNR for central and side decoder with $\Delta QP = 5, 8, 11$.

can be used to tune the amount of redundant rate. Increasing the value of ΔQP, corresponds to reducing the redundant rate and therefore, given a certain total rate R, a larger central PSNR is obtained, paid in terms of a lower side performance. Clearly, the optimal operative point depends on the network packet loss probability. It is worth noticing that the side decoder performance also represents the quality obtained by a H.264 decoder, which is unaware that a MDC technique is being used. This latter observation makes the proposed technique backward compatible with standard single description coding.

V. CONCLUSIONS

In this paper the redundant slice concept has been employed in order to design a novel MDC technique, producing two balanced descriptions of a video source. Moreover, a practical algorithm for redundancy tuning as a function of the packet loss rate is provided. The proposed approach yields better performance with respect to state-of-the-art MDC video codecs. Finally, it presents a number of advantages with respect to competing algorithms, namely the capability to flexibly tune the inserted redundancy, the use of standard H.264 encoder and decoder and the limited added computational cost.

VI. REFERENCES

[1] V. Vaishampayan, “Design of multiple description scalar quantizers,” *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[2] V. Vaishampayan and S. John, “Balanced interframe multiple description video compression,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Oct. 1999, vol. 3, pp. 812–816.

[3] Y. Wang, M.T. Orchard, V. Vaishampayan, and Reibman A.R., “Multiple description coding using pairwise correlating transforms,” *IEEE Trans. on Image Processing*, vol. 10, no. 3, pp. 351–367, Mar. 2001.

[4] Reibman A.R., H. Jafarkhani, Y. Wang, M.T. Orchard, and M. Puri, “Multiple description coding for video using motion compensated prediction,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Oct. 1999, vol. 3, pp. 837–841.

[5] 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, July 2003.

[6] S. Shirani, M. Gallant, and F. Kossentini, “Multiple description image coding using pre- and post-processing,” in *Proc. of International Conference on Information Technology: Coding and Computing)*, Apr. 2000.

[7] N. Franchi, M. Fumagalli, and R. Lancini, “Flexible redundancy insertion in a polyphase down sampling multiple description image coding,” in *Proc. of IEEE International Conference on Multimedia and Expo (ICME)*, Aug. 2002.

[8] T. Tillo and G. Olmo, “A low complexity pre-post processing multiple description coding for video streaming,” in *Proc. of the IEEE International Conference on information and communication technologies: From Theory to Applications (ICTTA 04), Damascus, Syria*, Apr. 2004, pp. 519–520.

[9] D. Wang, N. Canagarajah, and D. Bull, “Slice group based multiple description video coding using motion vector estimation,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Sept. 2004.

[10] D. Wang, N. Canagarajah, and D. Bull, “Error concealment for slice group based multiple description video coding,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Sept. 2005.

[11] R. Bemardini, 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] Joint Video Team JVT of ISO/IEC MPEG and ITU-T VCEG, *International Standard of Joint Video Specification (ITU-T Rec. H.264 , ISO/IEC 14496-10 AVC)*, Mar 2003.

[13] T. Schierl, T. Wiegand, and M. Kampmann, “3gpp compliant adaptive wireless video streaming using h.264/avc,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, Sept. 2005.

[14] B. Girod and N. Farber, “Feedback-based error control for mobile video transmission,” in *Proc. of the IEEE*, Oct. 1999, pp. 1707–1723.

[15] N.S. Jayant and P. Noll, *Digital Coding of waveforms*, Prentice Hall, 1984.

[16] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, “Rate-constrained coder control and comparison of video coding standards,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, July 2003.