

# Multiple Description Coding for H.264/AVC with Redundancy Allocation at Macro Block Level

Chunyu Lin, Tammam Tillo, *Member, IEEE*, Yao Zhao, *Member, IEEE*, and Byeungwoo Jeon, *Senior Member, IEEE*

**Abstract**—In this paper, a novel multiple description video coding scheme is proposed to insert and control the redundancy at macro block (MB) level. By analyzing the error propagation paths, the relative importance of each MB is determined. The paths, in practice, depend on both the video content and the adopted video coder. Considering the relative importance of the MB and the network status, an unequal protection for the video data can be realized to exploit the redundancy effectively. In addition, a simple and effective approach is introduced to tune the quantization parameter for the variable rate coding case. The whole scheme is implemented in H.264/AVC by employing its coding options, thus generating descriptions that are compatible with the *baseline profile* and *extended profile* of H.264/AVC. Due to its general property, the proposed approach can be employed for other hybrid video codecs. The results demonstrate the advantage of the proposed approach over other H.264/AVC multiple description schemes.

**Index Terms**—H.264/AVC, multiple description coding, rate allocation.

## I. INTRODUCTION

MULTIPLE description coding (MDC) has emerged as an effective method for video transmission over unreliable and non-prioritized networks. It can effectively combat packet loss without retransmission thus satisfying the demand of real-time services and relieving the network congestion. In the MDC approach, at least two coded streams (or *descriptions*), of the same data, are generated and sent through separate channels. If only one channel works, the source can be reconstructed by the side decoder with certain acceptable distortion, called *side distortion*. When more channels work,

the reconstruction quality can be enhanced up to the smallest *central distortion* upon the reception of all descriptions. In this paper, we will only consider the two-channel balanced MDC scheme, i.e., two descriptions having similar rate and distortion.

The objective of the MDC scheme is to make the distortions as small as possible for any given rate. However, achieving the best central and side performance at the same time is shown to be conflicting [1]. In fact, the central performance can be obtained by the corresponding single description coder (SDC) with a smaller total rate [1]. The extra bits (or redundancy) of MDC are introduced to achieve robustness, and the main task of the redundancy tuning mechanism in any MDC scheme is to control the redundancy according to the network condition.

One of the most popular solutions to MDC problem is multiple description based on scalar quantization (MDSQ) [2], which is applied to video coding in [3]. However, the main drawback of MDSQ is that it yields descriptions that are not standard compliant. In [4], MDC method with correlating transform is proposed, and it is employed in video coding in [5]. However, this method also yields descriptions that cannot be decoded with standard tools.

Another class of multiple description video coding scheme is based on the concept of polyphase subsampling. This class includes polyphase spatial subsampling that is used in [6], [7]. In [8], an approach based on polyphase subsampling in the transformed domain is proposed for intra-frame wavelet-based video coding. This scheme has a good redundancy tuning mechanism and overcomes the limitations posed by the prediction stage in hybrid video codecs. However, it suffers from the low coding efficiency, which is a common problem for the intra-frame-based video codecs. In [9], a new approach is proposed by segmenting the video data in both the spatial and frequency domains. This approach mainly focuses on the four-description scheme. Although these schemes are very simple and can be compatible with the coding standard if used as pre-processing and post-processing stage in the non-transformed domain, the performance is affected by the relevant amount of inherent redundancy that cannot be efficiently controlled.

The unequal loss protection (ULP) [10] methods adopt the concept of priority encoded transmission on MDC, i.e., allocating more powerful codes to more important data layers. These methods differ as for the family of codes employed and the code optimization strategy. However, it is recognized that they are very sensitive to the variations of the estimated packet

Manuscript received November 10, 2009; revised August 29, 2010; accepted November 16, 2010. Date of publication March 17, 2011; date of current version May 4, 2011. This work was supported in part by 973 Program, under Grant 2011CB302204, the National Science Foundation of China for Distinguished Young Scholars, under Grant 61025013, Sino-Singapore JRP, under Grant 2010DFA11010, National Natural Science Foundation of China, under Grants 60776794, 60903066, 61073142, and 90604032, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, under Grant 20100008030. This paper was recommended by Associate Editor J. Ridge.

C. Lin and Y. Zhao are with the Institute of Information Science, Beijing Jiaotong University, Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China (e-mail: yuailian@gmail.com; yzhao@bjtu.edu.cn).

T. Tillo is with Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: tammam.tillo@xjtlu.edu.cn).

B. Jeon is with the Department of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon 440-746, Korea (e-mail: bjeon@skku.edu). Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2011.2129270

loss rate (PLR). In [11], the rates of Reed–Solomon (RS) codes are obtained, using a practical optimization algorithm, so as to maximize the expected performance at the video receiver. In [12], a method is proposed to generate two unbalanced descriptions of video streams, and it is verified that ULP suffers from the presence of a cliff phenomenon that explains the strong dependency of its performance on the correctness of the estimated PLR. Another negative point for the ULP methods is that they require the reception of high percentage of the transmitted packets to generate the first acceptable quality level.

In [13] and [14], the slice group coding tool available in H.264/AVC is exploited to create two balanced descriptions, whereas the dynamic slice group is used to create multiple description based on H.264/AVC codec in [15]. However, the main drawback of using slice group tool is that it degrades the coding efficiency. In [16], the redundant slice concept defined in the baseline and extended profile of H.264/AVC [17] is used to achieve the same objective. In [18], two descriptions are generated by splitting the video pictures into two threads, and then redundant pictures are periodically inserted into the two threads. In the following, we will refer to this latter algorithm as RP-MDC and the one proposed in [16] as RS-MDC. To improve the performance of RS-MDC, another scheme performing rate-distortion optimization at the slice level is proposed [19], which will be denoted as slice-level rate-distortion optimized MDC (SLRD-MDC). The main flaw of these approaches is that the redundancy is introduced at slice/frame level, namely, all macro blocks (MBs) belonging to the same slice/frame will be regarded as equally important. This degrades the performance of these approaches, especially for non-stationarity video content. In fact, MBs have different characteristics, which is the reason for having different coding modes for different MBs in H.264/AVC. Consequently, it would be better to tune the inserted redundancy at MB level.

In this paper, we propose an MDC scheme in which the redundancy is effectively allocated at MB level. This is achieved by practically analyzing the error propagation paths, and then an effective approach that classifies the MBs is presented to evaluate the importance of each MB in terms of its contribution to the overall distortion. The evaluation process considers the error propagation paths, the video contents, and the network status. Based on the relative importance of each MB, the MB quantization parameter (QP) is properly determined to provide corresponding protections, and this permits better exploitation of the redundancy. The whole scheme is implemented by exploiting the redundant slice concept and the possibility to tune the QP at MB level in H.264/AVC. This assures full compatibility with the baseline and extended profile of H.264/AVC standard. Nevertheless, the compatibility with other profiles is still possible with the use of a pre-processing stage. Moreover, we derive a closed-loop formula to determine the QPs for the open-loop rate control scenario, which allows optimally selecting the amount of redundancy and determining how the redundancy should be fractionated among the MBs according to the network status. Finally, we propose a low complexity approach to insert a user-defined amount of redundancy.

It is worth mentioning that although the proposed approach is tailored for H.264/AVC, it is general and can be used with

any other hybrid video codecs, even though, in this case the compatibility with the original standard definition may not be guaranteed.

The remainder of this paper is organized as follows. In Section II, an overview of the mismatch distortion and its propagation paths are presented, and the outline of the proposed scheme is sketched. The propagated distortion is analytically modeled in Section III-A, and a practical approach to evaluate the relative importance of MBs is devised in Section III-B. Section III-C proposes a mechanism to determine the QPs, thus allowing optimally tuning the redundant rate, in the unknown PLR case. Section III-D analyzes the performance of the proposed algorithm over a channel with Bernoulli losses. In addition, an algorithm to insert a given amount of redundancy is presented in Section III-E. In Section IV, experimental results are presented. Finally, the conclusions are drawn in Section V.

## II. PROPOSED ALGORITHM

In this section, the mismatch distortion and its propagation paths will be first analyzed, and then the proposed algorithm will be presented.

### A. Mismatch Distortion and Its Propagation Paths

Efficient compression of video data is mainly achieved by block-based motion prediction and compensation, which allows exploiting the temporal correlation between subsequent frames. However, the introduced dependency among the successive encoded frames, due to the prediction stage in the video codec, makes the compressed sequence highly vulnerable to errors. Although a corrupted or missed slice can be partially recovered by concealment techniques, a *mismatch error* will be introduced in the decoder loop because of the quality difference between the original and recovered version of the slice. The mismatch error will propagate for a certain number of frames, due to the motion compensation and prediction mechanism used in the hybrid video codec. Hence, the mismatch error will affect not only the current lost slice but also some of the subsequent frames, which makes video transmission over networks a very challenging topic.

In [20], it is reported that the generated decoder drift due to *bitstream switching* in H.264/AVC, does not produce strong artifacts. This partly explains the superior performance of RS-MDC [16], in which the redundant slices are used as a low-quality backup version to replace the missed or corrupted primary slices. In fact, the substitution process helps to mitigate the disturbing effect generated by corrupted or missed slices. The second reason for the superior performance of RS-MDC is that the proposed mechanism for redundancy allocation, takes into account the propagated distortion by the power transfer function [21], which assumes that the propagated distortion for all MBs are governed by the same model and have equivalent impact on the total distortion. Consequently, the analytical solution to the redundancy allocation problem for RS-MDC shows that the parameters that determine the redundancy and coding mode of the redundant slice, depend only on the frame index within the group of pictures (GoP) and the PLR.

However, the main problem with RS-MDC is that it does not take into account the propagated distortion that may partially,

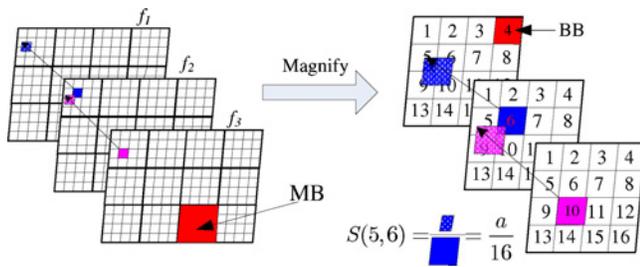


Fig. 1. Error propagation paths due to the eventual loss of some slices.

or completely, be stopped due to intra-mode MBs. Moreover, it does not consider the case that the mismatch distortion may pass to the following frames without any attenuation with skip-mode MBs. Therefore, it becomes important to practically determine the propagation paths of the mismatch error, which allows better understanding of the distortion contribution of each MB, so as to have a good redundancy allocation strategy.

H.264/AVC supports motion compensation with block sizes ranging from  $4 \times 4$  to  $16 \times 16$ , with  $4 \times 4$  being the largest common divisor for all the sizes. Hence, this latter block size is denoted as the basic block (BB), which will be used as the basic unit for analysis. If a block with other size (different from  $4 \times 4$ ) is used for motion compensation, all its BBs will have the same MV. From now on, we will use the notation  $BB(\beta, f)$  to denote a BB in the frame  $f$  with index  $\beta$  (Greek letters will be used to denote the index of the BB within the frame).

Fig. 1 shows the error propagation paths of some BBs. In this figure, it can be seen that  $BB(6, 2)$  uses  $BB(5, 1)$ ,  $BB(6, 1)$ ,  $BB(9, 1)$ , and  $BB(10, 1)$  as references, which means that losing the slice that embodies  $BB(9, 1)$  will also affect  $BB(6, 2)$  due to the direct propagation of the mismatch error generated at  $BB(9, 1)$ . This mismatch error can be caused by concealment or the replacement of the primary slice with its redundant one. Moreover, the reference BB for  $BB(10, 3)$  partially overlaps  $BB(6, 2)$ , which means that the mismatch generated at  $BB(9, 1)$  will indirectly affect  $BB(10, 3)$ . However, since  $BB(13, 1)$  is not used for reference, losing it will not induce any perturbations in the following frames. Hence, different BBs in the same frame have different importance, and the one that will cause larger distortion and longer propagation paths is more important. This can be generalized to MB level and the importance of MB can be estimated from its 16 BBs. Since H.264/AVC allows tuning the QP at MB level, the more important the MB, the smaller the QP should be applied to its redundant version.

Noticeably, only temporal error propagation is considered in this paper, and the mismatch distortion for intra coding in spatial domain is ignored for simplicity on one hand, and for the less impact of this factor on the total performance on the other hand. In fact, in inter-mode slices, the relative amount of intra predicted MBs, with respect to the other modes, is less than 3% in most of the tested video sequences in this paper.

### B. Proposed Algorithm

As we saw before, different MBs have different importance, this property will be employed while generating the descrip-

tions. First, for each GoP, a single H.264/AVC bitstream is generated, which contains only primary slices encoded with a given  $QP^p$ . This process is represented by block A in Fig. 2. At this coding stage, the information related to the motion compensation and prediction process, such as motion vectors (MVs) and reference frame indexes, is stored. This information will be used to determine the relative importance of each MB in the current GoP, which is accomplished in block B. Moreover, in order to insert a certain amount of redundancy, the rate and distortion data related to each MB should also be stored. It is worth noticing that the primary bitstream can also be a pre-encoded sequence stored in a streaming server. In this case, the previously described information can be collected by partially parsing the primary bitstream. Second, a redundant stream composed of redundant slices is generated (block C) by employing a set of QPs determined in block B.

These values can be evaluated as a function of the error propagation characteristics of the encoded sequence and the network status, which can be seen later. Therefore, it varies from MB to MB, which allows tuning the inserted redundancy at MB level. It is important to notice that the redundant inter-mode MBs are predicted using only the previously encoded primary slices, meaning that redundant slices cannot be used as a reference to predict other redundant slices during the encoding process. Hence, when a redundant slice is decoded and used to replace its primary counterpart, and  $QP^r \neq QP^p$ , a mismatch error will be introduced in the prediction loop of the decoder.

The two unbalanced bitstreams, i.e., the primary and redundant one, can now be reorganized to form two balanced descriptions in block D depicted in Fig. 2. This task is accomplished by interleaving primary and redundant slices so as to create two H.264/AVC bitstreams which contain alternatively the primary and the redundant representation of each slice, as shown in Fig. 3. This polyphase-at-slice-level approach is similar to the one used in [16]. To guarantee that the two bitstreams can be independently decoded, the crucial information contained in the sequence parameter set and picture parameter set is included into both descriptions. Compared to the actual video data, the amount of duplicated information has little impact on the total bitrate. The final output of this encoding process are two balanced descriptions that are compliant with the baseline and extended profile of H.264/AVC, and in general case, they are transmitted across two independent physical or virtual channels, characterized by a certain PLR.

If baseline or extended profile of H.264/AVC is used, a simple pre-processing stage is required at the decoder side. The pre-processing stage merges the two descriptions into one stream by parsing the header of each slice in order to determine the slice position within the merged stream. In this process, slices received out of their play-out deadlines will be neglected. After this stage, the resulted stream is input to the standard H.264/AVC decoder. If the primary description of a given slice is lost or damaged and its redundant version is correctly received, being a compliant H.264/AVC slice, it can be decoded yielding inferior yet acceptable quality. Only in the case that both descriptions are lost, the decoder must invoke a concealment algorithm. It is worth noticing that, as in

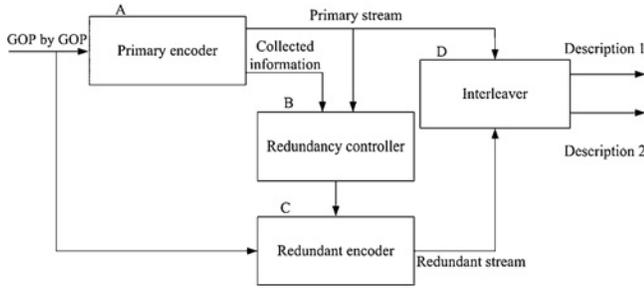
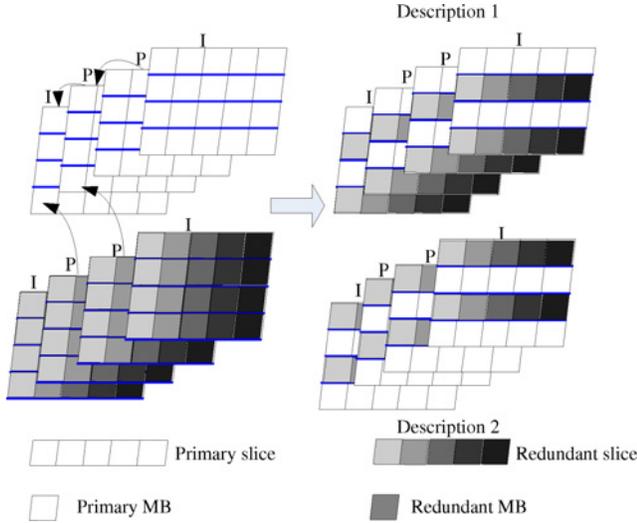


Fig. 2. Block diagram of the proposed encoder.

Fig. 3. Generation of two descriptions compatible with H.264/AVC by interleaving slices of a primary H.264/AVC bitstream (represented by white) and a redundant H.264/AVC bitstream (represented by different levels of gray to indicate different  $QP'$ ).

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 description reception.

If other profiles of H.264/AVC are used, the pre-processing stage needs to merge the two descriptions and output one stream which is compatible with the adopted profile. In detail, the pre-processing stage will do the following operations:

- 1) it passes the primary slice without alteration if it is received without error, and it will drop its redundant version if it is also received;
- 2) if only the redundant slice is received, its header will be transcoded so as to make it appear as a primary slice.

### III. REDUNDANCY TUNING

The efficiency of an MDC scheme mainly attributes to its redundancy tuning tactics. In this section, we put forward an analytical model of the propagated distortion, which allows determining the importance of each MB. Based on the importance of each MB, we propose an algorithm to tune the redundancy effectively. The tuning process is analyzed in the unknown PLR and Bernoulli losses channel, respectively. In addition, an algorithm to insert a given amount of redundancy is presented.

#### A. Analytical Model of the Propagated Distortion

In the following, the smallest block unit of analysis will be based on BB. This is because the discrete cosine transform used in H.264/AVC is basically  $4 \times 4$  pixel, and as we saw before, all the block sizes used for motion compensation are a multiple of  $4 \times 4$ . We will use the notation  $p_{\beta,f}(i, j)$  to denote a pixel in  $BB(\beta, f)$ , with  $i = 1, \dots, 4$  and  $j = 1, \dots, 4$  being the row and column indexes relative to the upper left corner of that BB.

Let us assume that the pixels of  $BB(\beta, f)$  have been encoded into a primary and redundant slice yielding the following distortion<sup>1</sup>  $d_b^p(\beta, f) = \sum_{i=1}^4 \sum_{j=1}^4 \{p_{\beta,f}^p(i, j) - p_{\beta,f}^o(i, j)\}^2$

and  $d_b^r(\beta, f) = \sum_{i=1}^4 \sum_{j=1}^4 \{p_{\beta,f}^r(i, j) - p_{\beta,f}^o(i, j)\}^2$ , respectively.

The superscripts  $o$ ,  $p$ , and  $r$  stand for original, primary, and redundant version of the pixel, respectively; the subscript  $b$  refers to BB. Given the predictive nature of the hybrid video encoding, we can write that a predicted BB in the frame  $f+1$ , say  $BB(\gamma, f+1)$ , as  $p_{\gamma,f+1}^p(i, j) = p_{\gamma,f}^p(i', j') + e_{\gamma,f+1}^p(i, j)$  and  $p_{\gamma,f+1}^r(i, j) = p_{\gamma,f}^r(i'', j'') + e_{\gamma,f+1}^r(i, j)$ , where  $e_{\gamma,f+1}^p$  and  $e_{\gamma,f+1}^r$  represent the prediction residual that are encapsulated into the data field of the primary and redundant slice, respectively. In addition,  $p_{\gamma,f}^p(i', j')$  and  $p_{\gamma,f}^r(i'', j'')$  represent the motion-compensated pixels in the frame  $f$ , these are used to encode the primary and redundant version of corresponding pixels in  $BB(\gamma, f+1)$ , respectively. From now on, we will omit the indexes  $i, j, i', j', i'', j''$  to simplify the notation.

When no losses occur in the previous frames nor in the current frame  $f$ , the total distortion of frame  $f$  can be written as  $d_f = \sum_{\beta \in f} d_b^p(\beta, f)$ . First, assume that only one primary slice is lost in the current GoP. In this case, the redundant version will be used to substitute the missed one, which causes the following mismatch distortion  $d_b^m(\beta, f) = \sum_{i=1}^4 \sum_{j=1}^4 \{p_{\beta,f}^r - p_{\beta,f}^p\}^2$ . Then, the distortion of  $BB(\beta, f)$  becomes as follows:

$$\begin{aligned} \tilde{d}_b(\beta, f) &= \sum_{i=1}^4 \sum_{j=1}^4 \{p_{\beta,f}^r - p_{\beta,f}^p\}^2 \\ &= \sum_{i=1}^4 \sum_{j=1}^4 \{p_{\beta,f}^r - p_{\beta,f}^p + p_{\beta,f}^p - p_{\beta,f}^o\}^2 \\ &= d_b^m(\beta, f) + d_b^p(\beta, f) + 2 \sum_{i=1}^4 \sum_{j=1}^4 \{(p_{\beta,f}^r - p_{\beta,f}^p)(p_{\beta,f}^p - p_{\beta,f}^o)\} \\ &= d_b^m(\beta, f) + d_b^p(\beta, f) + \\ &2 \sum_{i=1}^4 \sum_{j=1}^4 \{(p_{\beta,f}^r(p_{\beta,f}^p - p_{\beta,f}^o)) - p_{\beta,f}^p(p_{\beta,f}^p - p_{\beta,f}^o)\}. \end{aligned} \quad (1)$$

The accent tilde over a variable indicates that the variable has been altered due to the mismatch or propagated error. Assuming that  $\sum_{i=1}^4 \sum_{j=1}^4 \{(p_{\beta,f}^r(p_{\beta,f}^p - p_{\beta,f}^o)) - p_{\beta,f}^p(p_{\beta,f}^p - p_{\beta,f}^o)\} \approx 0$ , we get

$$\tilde{d}_b(\beta, f) = d_b^r(\beta, f) \approx d_b^p(\beta, f) + d_b^m(\beta, f). \quad (2)$$

<sup>1</sup>Unless otherwise noted, we will adopt the squared Euclidean norm to measure the distortion, i.e.,  $d = \|\mathbf{X}, \mathbf{Y}\|^2 = \sum_{i=1}^n (x_i - y_i)^2$  with  $\mathbf{X} = [x_1, x_2, \dots]$  and  $\mathbf{Y} = [y_1, y_2, \dots]$ .

The last assumption is justified by the fact that the quantized error of the primary version of a pixel, is uncorrelated with its quantized version at high coding rate [22] (i.e.,  $\sum_{\forall i, \forall j} p_{\beta, f}^p (p_{\beta, f}^p - p_{\beta, f}^o) = 0$ ). To a higher extent, the quantized redundant version of a pixel will not be correlated with the quantized error of the primary version [i.e.,  $\sum_{\forall i, \forall j} p_{\beta, f}^r (p_{\beta, f}^p - p_{\beta, f}^o) = 0$ ].

It is worth pointing out that the high rate quantization theory is widely used in source coding in order to derive closed-form solutions, and the optimal system designed following this theory still has good performance in low-rate scenarios. In fact, the high rate quantization theory is generally accurate enough at rates down to 2 bit per sample [23].

As we know, the mismatch distortion may propagate to the following frames, so the distortion of, say  $BB(\gamma, f+1)$ , becomes

$$\begin{aligned} & \tilde{d}_b(\gamma, f+1) \\ &= \sum_{i=1}^4 \sum_{j=1}^4 \{\tilde{p}_{*,f} + e_{\gamma, f+1}^p - p_{\gamma, f+1}^o\}^2 \\ &= \sum_{i=1}^4 \sum_{j=1}^4 \{\tilde{p}_{*,f} - p_{*,f}^p + p_{*,f}^p + e_{\gamma, f+1}^p - p_{\gamma, f+1}^o\}^2 \\ &= d_b^p(\gamma, f+1) + \sum_{i=1}^4 \sum_{j=1}^4 \{\tilde{p}_{*,f} - p_{*,f}^p\}^2 \\ &+ 2 \sum_{i=1}^4 \sum_{j=1}^4 \{(\tilde{p}_{*,f} - p_{*,f}^p)(p_{*,f}^p + e_{\gamma, f+1}^p - p_{\gamma, f+1}^o)\} \end{aligned} \quad (3)$$

where  $\tilde{p}_{*,f}$  is the value of the motion-compensated pixel in the frame  $f$ , which is used as a reference to generate the compressed version of  $p_{\gamma, f+1}^o$  in  $BB(\gamma, f+1)$ . This value is  $\tilde{p}_{*,f} = p_{*,f}^p$  if no error has occurred at that pixel position, whereas, it becomes  $\tilde{p}_{*,f} = p_{*,f}^r$  if the redundant version is used instead of its primary one. Consequently, if the BB used as reference for  $BB(\gamma, f+1)$  does not overlap with the missed and recovered BB [i.e.,  $BB(\beta, f)$ ], the second and third term of (3) will become zero. However, if it overlaps  $BB(\beta, f)$  and assuming that the mismatch distortion is uniformly distributed over the pixels of  $BB(\beta, f)$ , we can write that  $\sum_{i=1}^4 \sum_{j=1}^4 \{\tilde{p}_{*,f} - p_{*,f}^p\}^2 = S(\beta, \gamma) d_b^m(\beta, f)$ , with  $S(\beta, \gamma)$  being the percentage of overlap between the reference block for  $BB(\gamma, f+1)$  and the recovered version of  $BB(\beta, f)$ . The assumption that the mismatch distortion is uniformly distributed is justified by the relatively small size of BBs, which makes it reasonable to assume that the majority of BBs have smooth contents, i.e., pixels of one BB are nearly similar, and consequently the error is almost uniform. Under the un-correlation hypothesis, given by the high rate quantization theory, we can write that  $\sum_{i=1}^4 \sum_{j=1}^4 \{(\tilde{p}_{*,f} - p_{*,f}^p)(p_{*,f}^p + e_{\gamma, f+1}^p - p_{\gamma, f+1}^o)\} \approx 0$ . This assumption is justified by the fact that neither the quantized redundant version (i.e.,  $\tilde{p}_{*,f}$ ) of a pixel nor the quantized primary one (i.e.,  $p_{*,f}^p$ ) is correlated with its quantized error. To a higher extent, the difference between two quantized versions of the pixels in the frame  $f$  (i.e.,  $\tilde{p}_{*,f} - p_{*,f}^p$ ) will not be correlated with the quantized error in the next frame  $f+1$  (i.e.,  $p_{*,f}^p +$

$e_{\gamma, f+1}^p - p_{\gamma, f+1}^o$ ). Given the previous assumption, (3) becomes

$$\tilde{d}_b(\gamma, f+1) \approx d_b^p(\gamma, f+1) + S(\beta, \gamma) d_b^m(\beta, f). \quad (4)$$

The term  $S(\beta, \gamma) d_b^m(\beta, f)$  represents the amount by which the distortion of  $BB(\gamma, f+1)$  gets increased, due to the mismatch distortion  $d_b^m(\beta, f)$  at  $BB(\beta, f)$ . In (4), the term  $S(\beta, \gamma)$  acts as a scaling factor to  $d_b^m(\beta, f)$ . If  $S(\beta, \gamma) = 1$ , the mismatch distortion will not be attenuated while propagating to  $BB(\gamma, f+1)$ , which could be the case of skip-mode MB. On the contrary, when  $S(\beta, \gamma) = 0$ , there is no direct prediction path linking  $BB(\beta, f)$  to  $BB(\gamma, f+1)$ , which implies that  $\tilde{d}_b(\gamma, f+1) = d_b^p(\gamma, f+1)$ . A pictorial example on how to evaluate  $S(\beta, \gamma)$  is reported in Fig. 1, where the reference BB for  $BB(6, 2)$  partially overlaps  $BB(5, 1)$ , and the overlap area is  $a$ , in this case  $S(5, 6) = a/16$ . By the same propagation mechanism, the distortion of a BB at frame  $f+2$ , say  $BB(\delta, f+2)$ , may get increased due to the indirect distortion propagation from  $BB(\beta, f)$  to  $BB(\delta, f+2)$ , through  $BB(\gamma, f+1)$ . Thus we can get

$$\tilde{d}_b(\delta, f+2) \approx d_b^p(\delta, f+2) + S(\beta, \gamma) S(\gamma, \delta) d_b^m(\beta, f). \quad (5)$$

At this point, we can evaluate the total amount of distortion that propagates from  $BB(\gamma, f+1)$  to the frame  $f+2$ , due to the mismatch distortion at  $BB(\beta, f)$

$$\sum_{\forall \delta \in f+2} S(\beta, \gamma) S(\gamma, \delta) d_b^m(\beta, f) = d_b^m(\beta, f) S(\beta, \gamma) \sum_{\forall \delta \in f+2} S(\gamma, \delta). \quad (6)$$

If a given  $S(\gamma, \delta)$  is zero, this implies that there is no indirect prediction path that links  $BB(\delta, f+2)$  to  $BB(\beta, f)$ , through  $BB(\gamma, f+1)$ . By generalizing the previous result, we can evaluate the total contribution of the propagated distortion, due to the loss of the primary version of  $BB(\beta, f)$  on the current GoP

$$\begin{aligned} & d_b^m(\beta, f) \left\{ \sum_{\forall \gamma \in f+1} S(\beta, \gamma) \left\{ \sum_{\forall \delta \in f+2} S(\gamma, \delta) \left\{ \dots \left\{ \sum_{\forall \psi \in N} S(\chi, \psi) \right\} \dots \right\} \right\} \right\} \\ &= d_b^m(\beta, f) w_b(\beta, f) \end{aligned} \quad (7)$$

where  $N$  represents the GoP length and  $w_b(\beta, f)$  represents the weight of the propagated distortion in the current GoP, due to the loss of the primary version of  $BB(\beta, f)$ . It is worth noticing that the previous equation does not take into account the distortion at  $BB(\beta, f)$ . At this point we can conclude that, once the redundant version of  $BB(\beta, f)$  is used instead of its counterpart, the distortion associated with  $BB(\beta, f)$ , can be evaluated as the contribution of the distortion at  $BB(\beta, f)$  itself and the propagated distortion, that is

$$\tilde{d}_b(\beta, f) = d_b^r(\beta, f) + d_b^m(\beta, f) w_b(\beta, f). \quad (9)$$

Under the additive model assumption [16], mitigating the loss of a group of BBs that constitute one MB in the frame  $f$ , leads to the distortion as follows:

$$\tilde{d}_g(g, f) = \sum_{\forall \beta \in g} \tilde{d}_b(\beta, f) = \sum_{\forall \beta \in g} d_b^r(\beta, f) + \sum_{\forall \beta \in g} d_b^m(\beta, f) w_b(\beta, f) \quad (10)$$

where the subscript  $g$  indicates MB and the notation  $MB(g, f)$  will be used to denote an MB in the frame  $f$  with index  $g$ .

Since H.264/AVC allows tuning the QP at MB level, defining the distortion at MB level becomes particularly important to tune the redundancy in the proposed approach. For simplicity, assume that the terms  $d_b^m(\beta, f) = d_b^m$  are constant, we have

$$\begin{aligned} \tilde{d}_g(g, f) &= \sum_{\forall \beta \in g} d_b^r(\beta, f) + d_b^m \sum_{\forall \beta \in g} w_b(\beta, f) \\ &= d_g^r(g, f) + d_g^m w_g(g, f) \end{aligned} \quad (11)$$

with  $d_g^r(g, f) = \sum_{\forall \beta \in g} d_b^r(\beta, f)$ ,  $d_g^m = |g|d_b^m$ , and  $w_g(g, f) = \frac{1}{|g|} \sum_{\forall \beta \in g} w_b(\beta, f)$ ;  $|g|$  is the cardinality of the set  $g$ , i.e.,  $|g| = 16$ . It is worth noticing that  $d_g^m \leq d_g^r(g, f)$ .

### B. Evaluation of MB Importance

As we saw before, to evaluate the relative impact of each MB on the total GoP distortion, the prediction paths need to be determined for all the BBs in each MB. All the information required to evaluate the MB importance will be collected on BB basis. For inter MBs, the MVs and the indexes of the reference frame will be stored for all the BBs, during the coding stage that generates the primary version of the sequence. If block size different from  $4 \times 4$  are used for motion compensation, all its BBs will have the same MV and reference frame index; for example, a  $16 \times 16$ -mode MB is composed of 16 BBs having the same MV. The previous process will also be applied for skip-mode MB, whereas, for intra-mode MB, the reference frame index will be set to null. The MVs and the reference frame indexes allow determining the percentage of overlap between the reference block and current block. These values will be used to determine the weight of each MB in the current GoP by using a backward approach that starts from the last frame, and moves frame by frame, toward the first frame. The backward approach is equivalent to evaluating the most inward summation in (8) and then moving outward, which allows tracing direct and indirect prediction paths.

The proposed procedure to evaluate  $w_b(\beta, f)$  is shown using pseudocode in Procedure 1. To simplify its description, we assume that only the first previous frame is used to predict the current one. From a practical standpoint, It is worth indicating that less complex approaches can be adopted to evaluate the importance metric of the MBs.

On the completion of this procedure, it becomes possible to evaluate the weights of the MBs as  $w_g(g, f) = \frac{1}{|g|} \sum_{\forall \beta \in g} w_b(\beta, f)$ .

Take Fig. 1 as an example, where  $N = 3$  and  $BB(10, 3)$  uses  $BB(5, 2)$ ,  $BB(6, 2)$ ,  $BB(9, 2)$ , and  $BB(10, 2)$  as references. The overlap area between  $BB(10, 3)$  and  $BB(6, 2)$  is  $1 \times 1$ , hence the scale factor is  $S(6, 10) = 1/16$ . Consequently, the partial contribution of  $BB(10, 3)$  to  $w_b(6, 2)$  can be evaluated using an accumulation process as  $w_b(6, 2) = w_b(6, 2) + S(6, 10)w_b(10, 3) = w_b(6, 2) + S(6, 10)$ , where  $w_b(10, 3)$  is initialized as 1. Similarly, the weight of other BBs can be obtained.

---

### Procedure 1: Evaluation of $w_b(\beta, f)$

---

**Given**  $N$  the GoP length

**Given**  $B$  the number of BBs in one frame

**Initialize**  $w_b(\alpha, f)$  to 1 for  $\alpha = 1, \dots, B$ , for  $f = 1, \dots, N$

**for**  $f = N$  **downto** 2 **do**

**for**  $\alpha = 1$  **to**  $B$  **do**

**for**  $\beta = 1$  **to**  $B$  **do**

Retrieve  $B_{ref}$ , i.e., the reference BB in  $f - 1$

for  $BB(\beta, f)$ , using the stored MVs.

Evaluate  $S(\alpha, \beta)$  as the percentage of

overlap between  $B_{ref}$  and  $BB(\alpha, f - 1)$  as

$w_b(\alpha, f - 1) = w_b(\alpha, f - 1) + S(\alpha, \beta) \times w_b(\beta, f)$

**end for**

**end for**

**end for.**

---

### C. Allocating the Redundancy Between Redundant MBs

In this section, the encoding parameters of the redundant MBs will be derived based on (11) without taking into account the network status, so the main objective is to determine the optimal rate allocating among the redundant MBs. This is useful for a scenario where the PLR is unknown or highly variable. To simplify (11), let us consider the worst case for the mismatch distortion, in other words  $d_g^m \approx d_g^r(g, f)$ . This approximation is accurate when the primary and redundant version largely differ, i.e., for the low redundancy case [24]. To further verify the correctness of this result, we found that if  $QP^p = 22$  and  $QP^r \geq 45$ , the average of  $d_g^m/d_g^r(g, f)$ , evaluated over all the MBs, is higher than 90%. In this case, we can use the following approximation  $\tilde{d}_g(g, f) \approx d_g^r(g, f)[1 + w_g(g, f)]$ .

Suppose  $d_g^r(g, f)$  is a monotonic decreasing function with respect to the rate, which is a common assumption for the rate distortion (RD) curve. Then an optimal allocation of the redundant rate requires the redundant MBs to have similar slope on the RD curves [25], which means that

$$\frac{\partial \{d_g^r(g_i, f_i)[1 + w_g(g_i, f_i)]\}}{\partial r_g(g_i, f_i)} = \frac{\partial \{d_g^r(g_k, f_k)[1 + w_g(g_k, f_k)]\}}{\partial r_g(g_k, f_k)}; \forall i, k \quad (12)$$

where  $r_g(g_i, f_i)$  is the rate of  $g$  measured in bits. In general, and for practical values of the PLR, the majority of received packets are primary packets. This means that it is possible to assume that the impact of using redundant slices on the prediction paths is minimal, in other words, the prediction paths do not depend on the rate of the redundant slices. Consequently, (12) becomes

$$[1 + w_g(g_i, f_i)] \frac{\partial d_g^r(g_i, f_i)}{\partial r_g(g_i, f_i)} \approx [1 + w_g(g_k, f_k)] \frac{\partial d_g^r(g_k, f_k)}{\partial r_g(g_k, f_k)} \quad \forall i, k. \quad (13)$$

Given the monotonic property of the RD curve, this formula means that the larger the weight of a given MB, the larger the bit rate should be devoted to it. This result matches the intuitive expectation, that MB with higher impact on the propagated distortion should be better protected. By plugging

the following H.264/AVC RD function [26]

$$\frac{\partial D}{\partial R} = -0.85 \times 2^{\left(\frac{QP-12}{3}\right)} \quad (14)$$

into (13), we get the following equation

$$QP^r(g_i, f_i) = 3 \log_2 \left( \frac{1 + w_g(g_k, f_k)}{1 + w_g(g_i, f_i)} \right) + QP^r(g_k, f_k) \quad (15)$$

where  $QP^r(g_i, f_i)$  represents the QP of the redundant MB  $g_i$ . This equation establishes the relationship between any two redundant MBs. If the value of  $QP^r(g_i, f_i)$ , suggested by (15) is out of the QP range defined by H.264/AVC, i.e.,  $[0, 51]$ , the  $QP^r(g_i, f_i)$  should be thresholded to either 0 or 51. It is also worth noticing that it is not convenient to have  $QP^r(g_i, f_i)$  less than its primary counterpart, i.e.,  $QP^r(g_i, f_i) < QP^p(g_i, f_i)$ . In fact, although it may seem beneficial to have a redundant version of an MB better than its primary counterpart, any difference in the encoding quality will introduce mismatch error in the prediction loop.

As for the value of  $QP^p(g_i, f_i)$  that determines the central performance of the system, we will address the case of constant  $QP^p(g_i, f_i)$ , which is usually the most common scenario for video coding. In fact, many experiments have indicated that using constant QP for the entire video sequence typically results in good performance, in terms of both average peak signal-to-noise ratio (PSNR) and achieving small distortion variation across the sequence [27].

In this case, the relationship between the redundant and primary MBs can be determined using the following approach. Let us assume that  $MB(g_l, f_l)$  is the least important MB in the current GoP, in terms of propagated distortion, i.e.,  $w_g(g_l) \leq w_g(g_i) \forall g_i$ , and that we want to guarantee a minimum level of quality (given by  $QP^r(g_l, f_l)$ ) to be delivered by the redundant version of  $MB(g_l, f_l)$ . Hence, (15) can be rewritten as follows:

$$QP^r(g_i, f_i) = 3 \log_2 \left( \frac{1 + w_g(g_l, f_l)}{1 + w_g(g_i, f_i)} \right) + QP^p(g_l, f_l) + \Delta QP \quad (16)$$

where  $\Delta QP = QP^r(g_l, f_l) - QP^p(g_l, f_l) \geq 0$ . Different amount of redundancy can be achieved by tuning  $\Delta QP$ . For example,  $\Delta QP = 0$  means that the two generated descriptions are identical, which is equivalent to saying that the redundancy is 100%.

#### D. Redundancy Tuning with Respect to PLR

In this section, we will address a widely studied scenario, usually for benchmarking comparisons, of transmitting the video data over a memoryless channel with PLR =  $p$ . It is worth mentioning that the following proposed approach requires less strict condition, regarding the stationarity of the network status. In fact, we will assume that  $p$  is constant over the GoP. From a practical standpoint, this assumption is not critical, since the estimating of  $p$  is usually obtained through control packets mechanism, e.g., real time control protocol (RTCP) packets, which are periodically sent to probe the network status. The interval between RTCP packets is required to be larger than 5 s [28], which is usually of the same order of the GoP length for video streaming applications over

unreliable links. In order to achieve optimal insertion of the redundancy at the encoder side, which maximizes the expected quality, we can estimate the expected decoder distortion at encoding time. In [29], the expected distortion is estimated at frame level and it is used in the context of source/channel rate control and adaptive intra mode selection. A similar approach has also been used in [16] and [30]. In this section, the estimation of the expected distortion in presence of redundant slices, follows the similar method employed in [16].

We will assume that each slice fits into one maximum transfer unit (MTU). In this case, we can evaluate the expected distortion of the  $s$ th slice in the frame  $f$  and over the whole GoP as follows:

$$\bar{d}_s = (1 - p)d_s^p(s, f) + p(1 - p)\tilde{d}_s(s, f) + p^2d_s^c(s, f) \quad (17)$$

where  $d_s^p(s, f)$  and  $\tilde{d}_s(s, f)$  are the distortion contribution over the GoP due to the employment of primary and redundant slice, respectively. The term  $d_s^c(s, f)$  is the distortion when both representations of  $s$  are lost and it depends on the adopted concealment strategy, but not on the rate of the redundant version of  $s$ . The latter property means that  $d_s^c(s, f)$  does not directly affect the redundancy optimization problem. Given that it can be neglected for low  $p$ , we will drop this term from the optimization task. Assuming that the errors caused by the loss of a random number of slices, are uncorrelated, we can evaluate the total expected distortion of the GoP by summing all the contribution of all the slices that belong to the current GoP:  $\bar{D} = \sum_{s \in \text{GoP}} \bar{d}_s$ . It is worth noticing that similar additive models have been used in [29] and [30] to estimate the expected distortion over a GoP when mitigating losses by concealment, whereas, in [16], the additive model is used in the redundant slice context. The problem of minimizing  $\bar{D}$  can be formulated as a constrained minimization of

$$\bar{D} = \sum_{f=1}^N \sum_{s \in f} ((1 - p)d_s^p(s, f) + p(1 - p)\tilde{d}_s(s, f)) \quad (18)$$

given the overall rate per GoP as follows:

$$R = \sum_{f=1}^N \sum_{g \in f} (r_g^p(g, f) + r_g^r(g, f)) \quad (19)$$

where  $r_g^p(g, f)$  and  $r_g^r(g, f)$  are the rates, measured in bits, of the primary and the redundant representations of  $MB(g, f)$ . Given the additive distortion model,  $d_s^p(s, f) = \sum_{g \in s} d_g^p(g, f)$  and  $\tilde{d}_s(s, f) = \sum_{g \in s} \tilde{d}_g(g, f)$ , where  $\tilde{d}_g(g, f)$  is given by (11), this problem can be solved by means of the standard Lagrangian approach, leading to the equation as follows:

$$\frac{\partial d_g^p(g, f)}{\partial r_g^p(g, f)} = p(1 + w_g(g, f)) \frac{\partial d_g^r(g, f)}{\partial r_g^r(g, f)}. \quad (20)$$

The QPs for H.264/AVC, working in the open-loop rate control mode, can be worked out by combining the latter equation with (14) as follows:

$$QP^r(g, f) = QP^p(g, f) - 3 \log_2 (p(1 + w_g(g, f))). \quad (21)$$

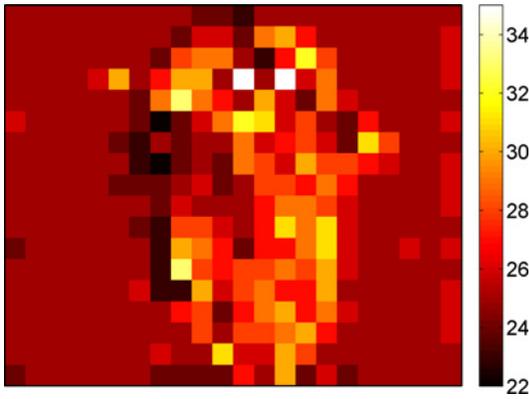


Fig. 4. QP of the redundant MBs for the first frame of CIF *Foreman* sequence.

This equation states that the optimal QP for the redundant version of  $MB(g, f)$ , varies as a function of  $w_g(g, f)$  and the PLR. In particular, the larger the weight of an MB, the smaller the quantization step should be used to guarantee higher protection. On the contrary, smaller  $\rho$  requires larger  $QP^r(g, f)$ , i.e., smaller redundancy.

Fig. 4 shows  $QP^r(g, f)$  for the first frame of *Foreman* sequence, with  $QP^p(g, 1) = 22$  and  $N = 11$ . It can be seen that most of the background areas, being largely used for prediction (i.e., has large value for  $w_g(g, f)$ ), are quantized with small quantization step to ensure low mismatch error in case of an eventual loss of a primary slice. On the contrary, the parts of the frame with non-translation movement are quantized with larger quantization steps, because these areas are less likely to be used as references for the prediction of the successive frames. Allocating more redundancy for the background area seems unnecessary, however, mitigating the contingent loss of a primary MB by replacing it with its redundant version may result in longer propagation paths and larger distortions if the two versions largely differ.

As for the objective of using a closed loop rate-control technique that can optimally partition the rate budget between the primary and redundant version of the MBs, it will be left for future investigation.

#### E. Insertion of Predefined Amount of Redundancy

In this section, we address the problem of inserting a predefined amount of redundancy by tuning  $\Delta QP$  in (16). The simplest and most direct approach to achieve this objective is to generate a number of redundant streams by testing all the possible  $\Delta QP$ , i.e.,  $0 \leq \Delta QP \leq 51 - QP^p$ . We select the redundant stream that best matches the predefined redundancy. Clearly, this exhaustive search approach is time and resource consuming. For this reason, we propose the following algorithm by estimating  $\Delta QP$  that leads to the insertion of the desired amount of redundancy. The main idea behind the proposed mechanism is to estimate the RD curve of each MB by exploiting the collected information during the encoding process of the primary stream. This allows estimating the rate of the redundant MBs for any feasible  $\Delta QP$ ; therefore, it becomes possible to estimate the rate of the redundant stream without encoding the sequence. For the

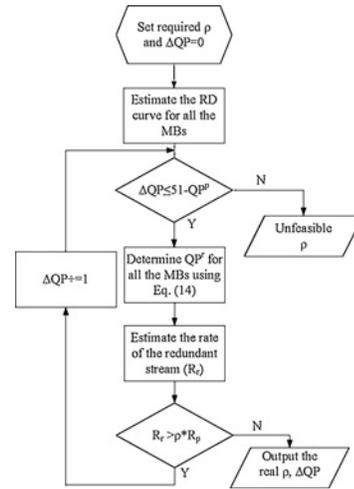


Fig. 5. Insertion mechanism for a given amount of redundancy.

purpose of estimating the RD curve, the approximation [31] can be used as follows:

$$D = \alpha e^{\beta R} \quad (22)$$

where  $\alpha$  and  $\beta$  are unknown parameters that can be determined, for each MB, by using the rate and distortion information of the primary MB. And the slope information can be determined by evaluating (14) at  $QP^p$ .

The flow chart of the proposed approach is shown in Fig. 5, where  $R_p$  and  $R_r$  represent the total rate of the primary and redundant streams, respectively, and  $\rho$  denotes the given redundancy.

#### IV. EXPERIMENTAL RESULTS

The proposed scheme is implemented in the JM9.4 H.264/AVC reference software to evaluate its effectiveness. Different types of video sequences are employed to verify the correctness of the assumptions in the modeling process. In the following results, the PSNR is measured for the luminance component. Unless otherwise noted, the results measured by the expected PSNR are obtained with 200 independent transmission trials, with Bernoulli channel model, i.e., the packets are independently and randomly lost with probability  $p$ . The performance evaluation of the proposed algorithm for channels with memory is left for future investigation.

Since the compressed video is going to be transmitted across a packet network, it becomes important to adopt a data partitioning strategy that yields a certain degree of error resilience. To this end, each frame in the video sequence is partitioned into a certain number of slices to guarantee that each compressed H.264/AVC network access level unit is smaller than the MTU of the network. In fact, the slice concept defined in H.264/AVC allows independent decoding of slices belonging to the same frame by preventing intra prediction between these slices. Consequently, packet loss will appear as partial picture corruption at the decoder side, thus improving the performance of the concealment stage. In the following simulations, the MTU is set to different sizes so as to enable fair comparison with other results under the

same transmission constraints. The size of primary slices is constrained to allow encapsulating each slice into one MTU, whereas, redundant slices are generated following the same frame-to-slice partitioning map of the primary bitstream. In other words, each redundant slice covers the same spatial area covered by its primary counterpart and each redundant slice is encapsulated into one MTU.

The temporal concealment available in JM9.4 for the inter-mode slices is used when both the primary and redundant version of a slice are lost, whereas the lost intra-mode slices are concealed by simply copying the co-located slices in the previous frame. In order to focus on the benefits of using the proposed algorithm, we do not use any other error resilience tools available in H.264/AVC, such as flexible MB ordering, data partition, and others. Nevertheless, the optimization and the joint benefits of other features available in H.264/AVC are beyond the scope of this paper.

As for the video coding, we use the IPPP GoP structure without bi-prediction-mode and the number of reference frames is set to 5.

*A. Results of CIF Sequences*

The first set of simulations is carried out using the following common intermediate format (CIF) 4:2:0 sequences, *Foreman* and *Coastguard* at 30 f/s, and the MTU here is set to 400 bytes.

Fig. 6 reports the expected PSNR versus the total coding rate for the proposed algorithm and RS-MDC. Here, the first 90 frames of *Foreman* are chosen to allow fair comparison with the results reported in [16]. The parameters of  $p$  and  $N$  are  $(p, N) \in \{(0.01, 45), (0.05, 21), (0.10, 11)\}$ . In order to quantify the impairment due to the inserted redundancy, we also report the error-free SDC curve without redundant slices. In this case, the longest GoP is used, i.e.,  $N = 45$ , which represents the performance upper-bound. The redundancy of the proposed algorithm is tuned via  $QP^r$  and as a function of  $p$  in (21). Here,  $QP^p$  for all the primary MBs is selected in the interval [22], [38] so as to get different RD points. The reported results demonstrate that the proposed scheme outperforms RS-MDC at all rates and all the PLRs. This is mainly due to the more effective mechanism for the redundancy tuning, which allows discriminating between MBs according to their importance. Moreover, it is worth noticing that the result of the proposed algorithm with  $p = 0.01$  nearly halves the gap between RS-MDC and the error-free SDC curve, which demonstrates the superior performance of the proposed approach.

Fig. 7 provides the expected PSNR results for *Coastguard* sequence. The first 90 frames are coded using the following parameters  $(p, N) \in \{(0.01, 45), (0.05, 21), (0.10, 11)\}$ . For comparison, the results of SLRD-MDC [19] are included. SLRD-MDC considers the distortions of each slice and the transmission conditions in the redundancy tuning step. However, it does not differentiate the importance of each MB. It can be seen that the proposed scheme outperforms the compared one in almost all the cases.

In order to compare the proposed approach with RS-MDC in terms of their central and side PSNR, Fig. 8 reports the side and central PSNR curves for *Foreman* sequence. In these simulations, the redundancy of the proposed algorithm

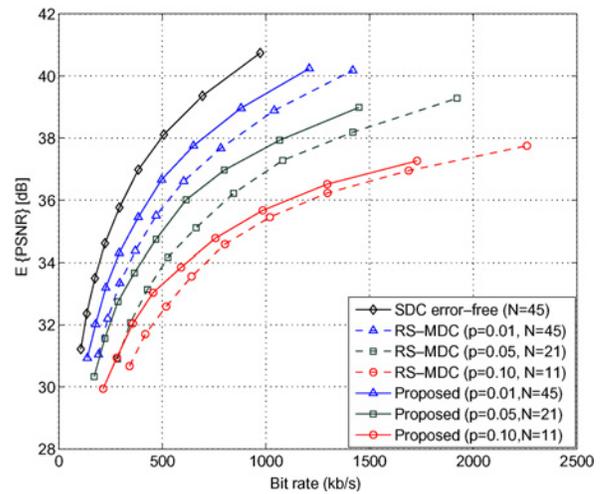


Fig. 6. Average PSNR versus coding rate for CIF *Foreman* sequence at 30 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 45, 21, 11$ .

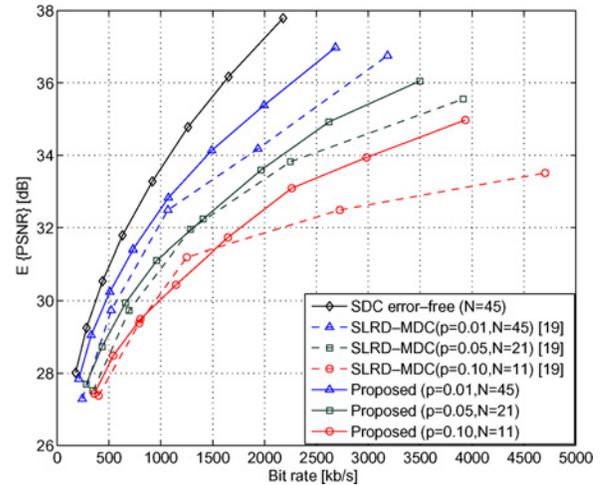


Fig. 7. Average PSNR versus coding rate for CIF *Coastguard* sequence at 30 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 45, 21, 11$ .

is tuned so as to have similar central performance to the results of RS-MDC [16]. In this case, both algorithms have nearly similar redundancy. Although both algorithms work under nearly similar conditions, the proposed approach yields superior performance for all the GoP length and all the level of redundancy, which demonstrates the effectiveness of the proposed redundancy tuning mechanism.

In Fig. 9, the results for *Foreman* sequence of the prediction-based spatial polyphase transform (PSPT) scheme [7], are included for comparison. In PSPT, the input sequence is sampled to generate subframes, and a hybrid interpolation algorithm that exploits adjacent subframes is used. For these simulations, the MTU is fixed to 100 bytes, and  $PLR = 0.10$ . The PSPT results are obtained with bilinear and hybrid prediction approach [7]. The results show that the proposed scheme outperforms the PSPT significantly. In addition, the proposed redundancy tuning strategy is more flexible.

Table I reports the results of the predefined redundancy insertion mechanism (described in Section III-E) for CIF *Foreman* sequence at 30 f/s with  $QP^p = 22$  and 30. The results demonstrate that the obtained redundancy is close to the

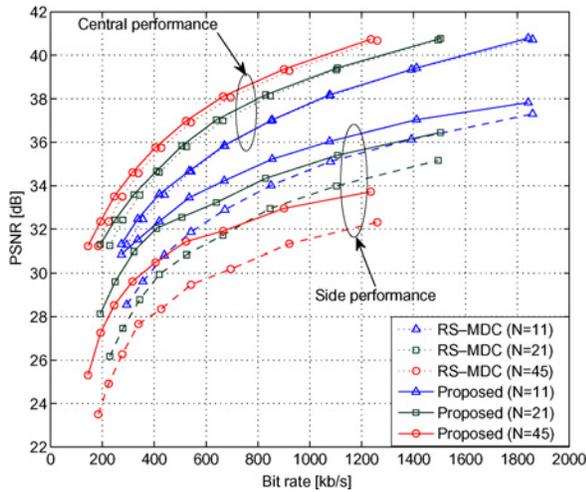


Fig. 8. Central and side performance versus coding rate for CIF *Foreman* sequence at 30 f/s.

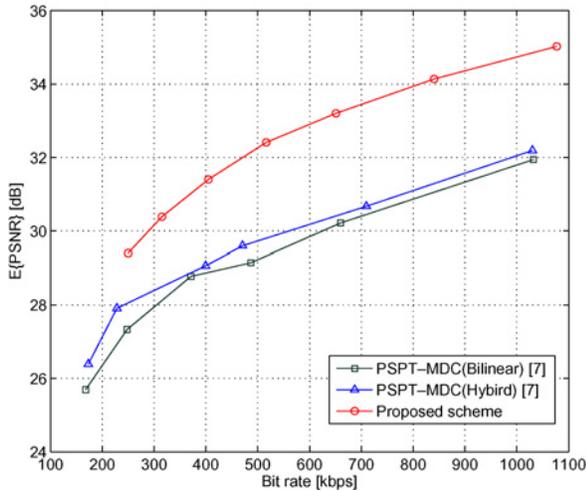


Fig. 9. Average PSNR versus coding rate for CIF *Foreman* sequence at 30 f/s, with  $p = 0.10$  and 100 bytes per packet.

TABLE I  
COMPARISON BETWEEN THE INSERTED AND REQUIRED REDUNDANCY VALUE

Redundancy	$QP=22$				$QP=30$			
	37.9	34.2	31.0	28.1	45.8	41.5	37.6	34.4
Required (%)								
Obtained (%)	40.0	36.7	31.1	24.9	42.5	39.0	33.0	25.9

given value especially at high redundancy. On one hand, this is due to the fact that the proposed approach estimates the RD curves of the MBs using the collected information during the generation of the primary sequence. Consequently, we get a good approximation of the RD curve at high redundancy where the redundant version is close to the primary one. On the other hand, the RD model [31] becomes more accurate at high rate.

### B. Results of QCIF Sequence

The second set of experiments is carried out using quarter common intermediate format (QCIF) sequences in order to assess the algorithm performance with lower spatial and

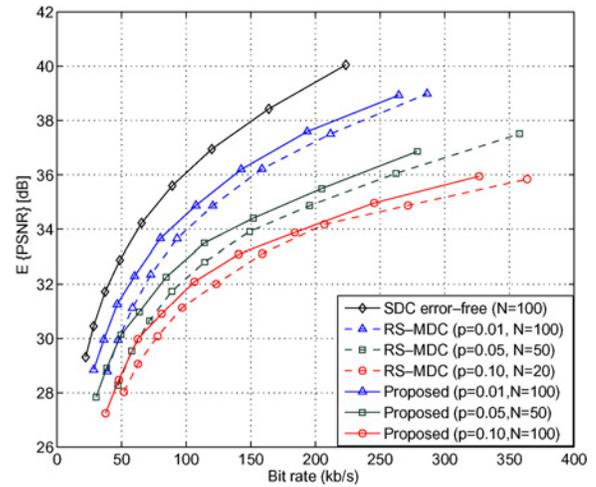


Fig. 10. Average PSNR versus coding rate for QCIF *Foreman* sequence at 15 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 100, 50, 20$ .

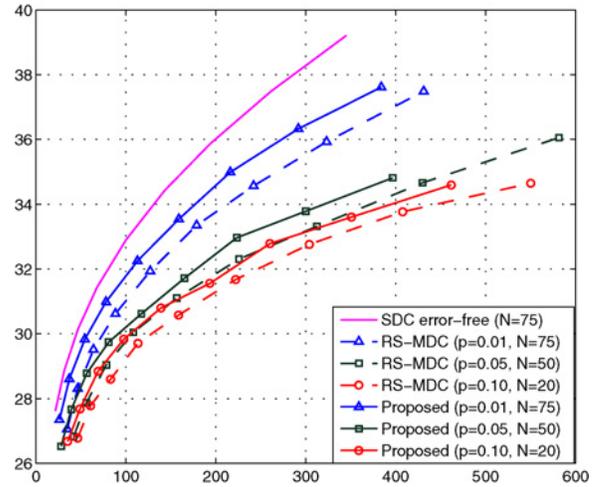


Fig. 11. Average PSNR versus coding rate for QCIF *Coastguard* sequence at 15 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 75, 50, 20$ .

temporal resolution. All of the frames are selected in these experiments.

Figs. 10 and 11 report the results of QCIF *Foreman* and *Coastguard* sequence, respectively. The two results reported in these figures are obtained using the following parameters:  $(p, N) \in \{(0.01, 100), (0.05, 50), (0.10, 20)\}$  and  $(p, N) \in \{(0.01, 75), (0.05, 50), (0.10, 20)\}$ , which are the same conditions used for RS-MDC in [16]. In the same figures, we report the results of the error-free SDC scheme. The reported results demonstrate the effectiveness of the proposed scheme, for small frame size and different sequence content. We can notice that the gain with respect to RS-MDC, is smaller than that obtained for the CIF *Foreman* sequence. This can be explained by the fact that MBs for the CIF format are more smoother and uniform than the corresponding QCIF one, it means that it is possible to have better discrimination and classification of the MBs based on their content in the CIF format.

To compare with the results in [18], another set of experiment is carried out using QCIF *Foreman* sequence at 7.5 f/s, and the results are demonstrated in Fig. 12. In order to have a fair comparison with [18], the MTU is fixed to

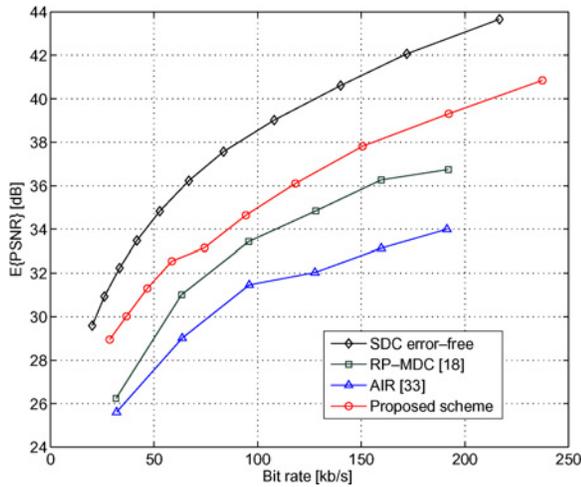


Fig. 12. Average PSNR versus coding rate for QCIF *Foreman* sequence at 7.5 f/s, with  $p = 0.05$  and 1400 bytes per packet.

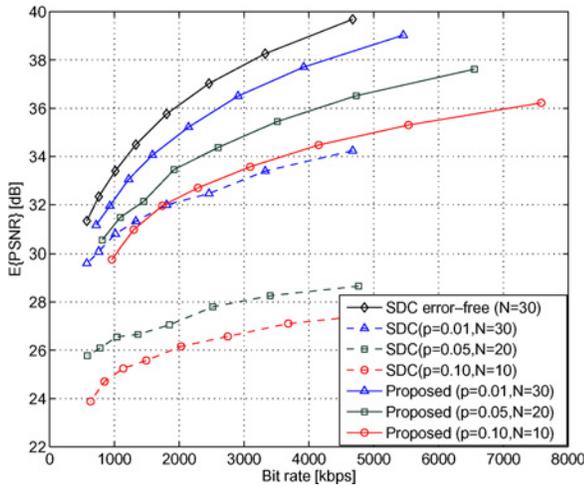


Fig. 13. Average PSNR versus coding rate for 4CIF *Soccer* sequence at 30 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 30, 20, 10$ .

1400 bytes and the 5% packet loss pattern is selected from the error patterns for Internet experiments specified in Q15-I-16r1 [32]. The results are obtained using the entire loss patterns containing 10 000 binary characters. In addition, the results of the error-free SDC and the adaptive intra refresh (AIR) scheme [33], are also included for comparison. In the latter scheme, intra-mode MBs are inserted according to the source distortion and the expected channel distortion. The results indicate that the proposed scheme outperforms RP-MDC and AIR at least 0.8 dB and 3 dB, respectively, which demonstrates the effectiveness of the proposed scheme.

### C. Results of 4CIF Sequence

Finally, Figs. 13 and 14 provide the results of 4CIF ( $704 \times 576$ ) sequences *Soccer* and *Harbor*, respectively. All the frames are adopted and coded at 30 f/s with  $(p, N) \in \{(0.01, 30), (0.05, 20), (0.10, 10)\}$ , and the MTU size is 1000 bytes. For comparison, the results of standard H.264 bitstreams obtained at the same conditions, are also included. The results shows the effectiveness of the proposed scheme on the high resolution video formats.

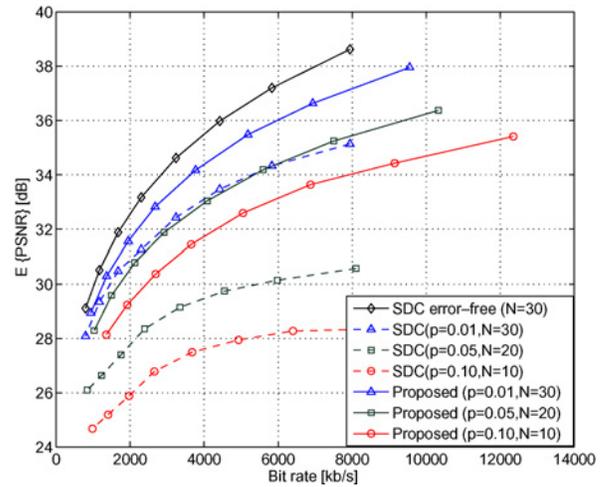


Fig. 14. Average PSNR versus coding rate for 4CIF *Harbor* sequence at 30 f/s, with  $p = 0.01, 0.05, 0.10$ , and  $N = 30, 20, 10$ .

## V. CONCLUSION

This paper presented an MDC scheme that allowed allocating the redundancy at MB level. The allocating is achieved by coding the redundant version of each MB while taking into account its relative importance. A closed-form solution to the redundancy allocation problem is presented, which considers both the transmission conditions and the error propagation paths. The proposed solution is obtained for the high rate case, while the solution for the low rate case is left for future research. The experimental results demonstrated that the algorithm consistently outperforms the state-of-the-art H.264/AVC multiple description approaches. Since H.264/AVC recommendation does not specify a normative approach to tune the QP of redundant MBs, the proposed scheme can be generalized to the context of tuning the quantization parameters of redundant MBs.

## REFERENCES

- [1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.
- [2] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Info. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.
- [3] V. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Oct. 1999, pp. 812–816.
- [4] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 351–367, Mar. 2001.
- [5] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and M. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Oct. 1999, pp. 837–841.
- [6] N. Franchi, M. Fumagalli, and R. Lancini, "Flexible redundancy insertion in a polyphase down sampling multiple description image coding," in *Proc. IEEE ICME*, vol. 2, Aug. 2002, pp. 605–608.
- [7] Z. Wei, C. Cai, and K.-K. Ma, "A novel H.264-based multiple description video coding via polyphase transform and partial prediction," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, vol. 1, Dec. 2006, pp. 151–154.
- [8] E. Akyol, A. M. Tekalp, and M. R. Cinvaral, "A flexible multiple description coding framework for adaptive P2P video streaming," *IEEE J. Selected Topics Signal Process.*, vol. 1, no. 2, pp. 231–245, Aug. 2007.
- [9] C.-W. Hsiao and W.-J. Tsai, "Hybrid multiple description coding based on H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 76–87, Jan. 2010.
- [10] 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 J. Selected Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.

- [11] R. Puri, K. Lee, K. Ramchandran, and V. Bharghavan, "Forward error correction (FEC) codes based multiple description coding for internet video streaming and multicast," *Signal Process.: Image Commun.*, vol. 16, no. 8, pp. 745–762, May 2001.
- [12] D. Comas, R. Singh, A. Ortega, and F. Marques, "Unbalanced multiple-description video coding with rate-distortion optimization," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 81–90, 2003.
- [13] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding using motion vector estimation," in *Proc. IEEE Int. Conf. Image Process.*, vol. 5, Sep. 2004, pp. 3237–3240.
- [14] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding with three motion compensation loops," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2005, pp. 2719–2722.
- [15] C.-C. Su, H. H. Chen, J. Yao, and P. Huang, "H.264/AVC-based multiple description video coding using dynamic slice groups," *Signal Process.: Image Commun.*, vol. 23, no. 9, pp. 677–691, Oct. 2008.
- [16] T. Tillo, M. Grangetto, and M. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 59–70, Jan. 2008.
- [17] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [18] I. Radulovic, P. Frossard, Y.-K. Wang, M. Hannuksela, and A. Hallapuro, "Multiple description video coding with H.264/AVC redundant pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 144–148, Jan. 2010.
- [19] L. Peraldo, E. Baccaglini, E. Magli, G. Olmo, R. Ansari, and Y. Yao, "Slice-level rate-distortion optimized multiple description coding for H.264/AVC," in *Proc. IEEE Int. Conf. Acou. Speech Signal Process.*, Mar. 2010, pp. 2330–2333.
- [20] T. Schierl, T. Wiegand, and M. Kampmann, "3GPP compliant adaptive wireless video streaming using H.264/AVC," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2005, pp. 696–699.
- [21] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.
- [22] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [23] D. L. Neuhoff, "The other asymptotic theory of lossy source coding," in *Proc. Coding Quantiz. DIMACS Series Discrete Math. Theoretic. Comput. Sci.*, vol. 14, Oct. 1993, pp. 55–65.
- [24] T. Tillo, M. Grangetto, and M. Olmo, "On modeling mismatch errors induced by different quantizers," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 785–788, Nov. 2007.
- [25] V. K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 9–21, Sep. 2001.
- [26] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.
- [27] B. Xie and W. Zeng, "A sequence-based rate control framework for consistent quality real-time video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 56–71, Jan. 2006.
- [28] Y. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RTP: A Transport Protocol for Real-Time Applications*, document RFC 1889, Internet Engineering Task Force, Jan. 1996.
- [29] Z. He, H. Cai, and C. W. Chang, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
- [30] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 716–732, Jun. 2006.
- [31] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.
- [32] Y.-K. Wang, S. Wenger, and M. M. Hannuksela, *Common Conditions for SVC Error Resilience Testing*, document P206, Joint Video Team, Aug. 2005.
- [33] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proc. Int. Conf. Image Proc.*, vol. 1, Oct. 2004, pp. 163–166.



**Chunyu Lin** was born in Liaoning, China. He received the B.S. degree from Bohai University, Liaoning, the M.S. degree from Yanshan University, Hebei, China, in 2002 and 2005, respectively, and the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2010.

From 2009 to 2010, he was a Visiting Researcher with the ICT Group, Delft University of Technology, Delft, The Netherlands. His current research interests include areas of image/video compression and robust transmission.



**Tammam Tillo** (S'02–M'06) was born in Damascus, Syria. He received the Engineer Diploma degree in electrical engineering from the University of Damascus, Damascus, in 1994, and the Ph.D. degree in electronics and communication engineering from the Politecnico di Torino, Turin, Italy, in 2005.

From 1999 to 2002, he was with the Souccar for Electronic Industries, Damascus. In 2004, he was a Visiting Researcher with the EPFL, Lausanne, Switzerland. From 2005 to 2008, he was a Post-

Doctoral Researcher with the Image Processing Laboratory, Politecnico di Torino. For few months, he was an Invited Research Professor with the Digital Media Laboratory, Sungkyunkwan University, Suwon, Korea. In August 2008, he joined Xi'an Jiaotong-Liverpool University, Suzhou, China. His current research interests include areas of robust transmission, image and video compression, and hyperspectral image compression.

Dr. Tillo is a member of the technical program committees for several international conferences.



**Yao Zhao** (M'05) received the B.S. degree from Fuzhou University in 1989 and the M.E. degree from Southeast University in 1992, both from the Department of Radio Engineering, and the Ph.D. degree from the Institute of Information Science, Beijing Jiaotong University (BJTU), Beijing, China, in 1996.

He became an Associate Professor with BJTU in 1998 and became a Professor in 2001. From 2001 to 2002, he worked as a Senior Research Fellow with the Information and Communication Theory

Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands. He is currently the Director of the Institute of Information Science, Beijing Jiaotong University. Currently, he is leading several national research projects from 973 Program, 863 Program, and the National Science Foundation of China. His current research interests include image/video coding, digital watermarking, and image/video analysis and understanding.

Dr. Zhao received the National Science Fund of China for Distinguished Young Scholars and the National Outstanding Young Investigator Award of China, both in 2010.



**Byeungwoo Jeon** (S'88–M'92–SM'02) received the B.S. degree (magna cum laude) in 1985 and the M.S. degree in 1987 from the Department of Electronics Engineering, Seoul National University, Seoul, Korea, and the Ph.D. degree from the School of Electrical Engineering, Purdue University, West Lafayette, IN, in 1992.

From 1993 to 1997, he was with the Signal Processing Laboratory, Samsung Electronics, Seoul, where he was responsible for the research and development of video compression algorithms, design

of digital broadcasting satellite receivers, and other MPEG-related research for multimedia applications. Since September 1997, he has been with the Faculty of the School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea, where he is currently a Professor. From 2004 to 2006, he was a Project Manager with Digital TV and Broadcasting, Korean Ministry of Information and Communications, where he supervised all digital TV-related research and development in Korea. He has authored many papers in the areas of video compression, pre/post-processing, and pattern recognition. He also holds more than 50 issued patents (Korean and international) in these areas. His current research interests include multimedia signal processing, video compression, statistical pattern recognition, and remote sensing.

Dr. Jeon is a member of Tau Beta Pi and Eta Kappa Nu. He is a member of SPIE, IEEK, KICS, and KOSBE. He also regularly participates and contributes to international standardization activities, e.g., ITU-T VCEG and ISO/IEC MPEG. He was a recipient of the IEEK Haedong Paper Award in Signal Processing Society, Korea, in 2005.