SCHOOL OF ENGINEERING - STI
ELECTRICAL ENGINEERING INSTITUTE
SIGNAL PROCESSING LABORATORY

Pascal Frossard

EPFL - FSTI - IEL - LTS
Station 11
Switzerland-1015 LAUSANNE

Phone: +4121 6932601
Fax: +4121 6937600
e-mail: pascal.frossard@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

MULTIPLE DESCRIPTION VIDEO CODING WITH H.264/AVC REDUNDANT PICTURES

Ivana Radulovic, Pascal Frossard, Ye-Kui Wang, Miska Hannuksela, Antti Hallapuro

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory

Technical Report LTS-2008-001

February 15, 2008

Part of this work has been submitted to IEEE Transactions on Circuits and Systems for Video Technology
This work has been supported by the Swiss National Science Foundation grant PP-002-68737.

Multiple Description Video Coding with H.264/AVC Redundant Pictures

Ivana Radulovic, Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratories - LTS4
Lausanne, Switzerland
{ivana.radulovic, pascal.frossard}@epfl.ch

Ye-Kui Wang, Miska Hannuksela, Antti Hallapuro
Nokia Research Center
Tampere, Finland
{ye-kui.wang, miska.hannuksela, antti.hallapuro}@nokia.com

Abstract

Joint source and channel coding has proven benefits for efficient transmission of information with delay and complexity constraints. In particular, Multiple Description Coding offers interesting solutions for error resilient multimedia communications as well as for distributed streaming applications. We propose in this paper a scheme based on H.264/AVC for the encoding of image sequences into multiple descriptions. The pictures are split into multiple coding threads, and redundant pictures are inserted periodically to increase the resilience to loss and reduce the error propagation. We propose an end-to-end distortion model that features the influence of the coding parameters for primary and redundant pictures and we derive the optimal coding strategy for given transmission conditions. Extensive experiments demonstrate that the proposed scheme outperforms baseline solutions based on loss- and content-adaptive intra coding. Finally, we show how the decoder can reduce the distortion by efficient combination of primary and redundant pictures, if both are available at the decoder.

I. Introduction

With the tremendous and continuous growth in size and capacity of packet networks such as the Internet on one side, and the excellent performance of image/video compression techniques on the other side, there has been recently a rapid development of multimedia services and applications such as video conferencing or Internet Protocol TV (IPTV). These applications typically require a good media quality, even with varying transmission bandwidth and tight timing constraints that prevent the retransmission of lost or late packets.

Joint source and channel coding (JSCC) as well as streaming strategies that exploit the network diversity have been shown to provide elegant solutions to offer a sustained quality to the users in the absence of guarantee from the transmission channels. On the one hand, JSCC targets the minimization of the end-to-end distortion, which consists of the source and channel distortion. It efficiently splits the streaming rate into source rate that drives the source distortion, and channel rate that controls the resiliency to transmission losses. On the other hand, distributed streaming tries to exploit the availability of multiple sources or peers, or multiple transmission channels for increasing the media quality. In this case, the redundancy between the different streams has to be controlled such that the end-to-end distortion is minimized without wasting bandwidth with packet replicates.

Multiple description coding (MDC) [1] can bring great benefits in such cases, as it permits to generate several representations of the video information and ensures that they are all useful for increasing the quality at the decoder. It leaves an adaptive amount of redundancy between the different descriptions, such that the most important information is duplicated in different transmission packets or sent from different sources. The least important information is then distributed among the different representations, such that they all contribute to decreasing the overall distortion. Multiple description coding has emerged recently as a promising alternative for video streaming, due to its improved performance compared to JSCC schemes based on Forward Error Correction, particularly when the channel conditions are not accurately estimated [2]. In addition, it extends naturally to solving the problem of streaming with network diversity, since the different descriptions can be sent on different channels and improve the end-to-end performance of the streaming system.

In this paper, we propose a standard compatible MDC video scheme. We build on our previous work [3] and use H.264/AVC *redundant pictures* to provide robustness to transmission errors. The video information is split into several encoding threads, and redundant pictures are inserted to reduce the error drift in case of packet loss. We provide a rate-distortion analysis that permits to control the stream redundancy with respect to the expected transmission conditions by adapting the quantization parameters in the redundant pictures. We finally show how the decoding quality can be further improved by a proper handling of the different versions of received pictures available at the decoder. Extensive simulations demonstrate that the proposed scheme outperforms state-of-the-art single- and two-description video coding schemes in terms of average quality, as well as quality variation and resiliency to incorrect estimation of the channel state.

This rest of the paper is organized as follows. We first provide an overview of MDC video coding in Section II. Section III then describes the proposed scheme in details, while Section IV proposes a redundancy and rate-distortion analysis of the MDC video encoder. We compare its performance with state-of-the-art techniques in section V. We finally discuss decoder improvements in section VI.

This work has been partly supported by the Swiss National Science Foundation grant PP-002-68737.

II. Related work

Multiple description coding techniques are generally based on four main families of methods, namely information spitting, quantization, unequal error protection, and redundant expansions. We describe briefly the research work in the different categories, with a special emphasis on video coding schemes.

One of the most popular algorithms based on information splitting is the so-called multiple state video coding (MSVC) [4], which is similar to video redundancy coding proposed in [5]. The input video sequence is split into subsequences of odd and even frames and each subsequence is independently coded, with its own prediction process and state. With this solution, even if one description is completely lost, another one can be independently decoded and reconstructed at half of the frame rate. Moreover, lost frames in one description can be reconstructed by interpolation from the neighboring frames in another description. Numerous other multiple description video coding schemes are based on information splitting in the temporal ([6]–[11]), spatial ([12], [13]) and the frequency domain ([14]–[16]). As the performance of a splitting technique is highly dependent on the sequence content and network characteristics, several works have proposed to combine the different splitting techniques for improved performance ([17], [18]). Finally, a combination of MDC with multiresolution representations and layered coding has been addressed in [19]–[23].

Successful solutions have also been implemented by unequal error protection and channel coding. The most important parts of the information streams are therefore replicated in several descriptions, while the least important information is distributed in different packets, without strong channel protection [24], [25]. The works of Chou et al. ([26], [27]) and Taal et al. ([28], [29]) further consider the combination of scalable video coding in combination with unequal error protection. Unequal error protection usually permits to generate a fairly large number of descriptions, possibly at the expense of low granularity or decoding delay.

Furthermore, MDC principles can be efficiently combined with redundant signal expansions. In [30], [31], Zakhor et al. use the matching pursuit algorithm to generate two descriptions of video sequences, where the redundancy between the descriptions is controlled with a number of atoms repeated in both descriptions. The same principle, combined with Multiple Description Scalar Quantization, can be found in [32] and [33]. In [34], [35] a scheme for multiple description scalable video coding based on a motion-compensated redundant analysis has been proposed.

A few other techniques have also been proposed for Multiple Description video coding. Multiple description scalar quantization has been applied in [36], [37] to jointly quantize the DCT coefficients or coefficients in a motion-compensated temporal filtering based encoder. Correlating transforms have been used in [19], [38]. Finally, coset codes can be used for the generation of multiple descriptions when the prediction in multiple description video coding is considered as a variant of the Wyner-Ziv coding problem [39].

Overall, information splitting methods are probably a good choice for applications that favor compliance with video coding standard for easier deployment. In particular, extensions of the MSVC scheme [4] may provide standard bitstreams for each description. In this paper, we propose to use redundant pictures in H.264/AVC since they typically offer an interesting solution to extend and improve the MSVC scheme with increased resiliency to error propagation. Parallel work of Tillo et al. [10] also proposes an MDC video coding scheme based on redundant pictures, where descriptions are however not completely independent. Our work proposes an extended rate-distortion analysis for independent descriptions, as well as an improved decoding solution that tries to exploit efficiently all the information available at the decoder.

III. MDC with redundant pictures

A. Redundant pictures in H.264/AVC

Redundant pictures (RP) are one error resilient tool included in H.264/AVC. According to the standard, each picture may be associated with one or more (up to 127) RPs, which a decoder can reconstruct in case a primary picture (or parts thereof) is missing. If a primary picture is correctly received, it is suggested by the standard that RPs are discarded by the decoder.

The standard is very flexible regarding the design of RPs. For example, it permits a redundant picture to be of the same size as the corresponding primary picture, but also does not exclude the possibility to encode only its most important parts (such as region of interest). Moreover, H.264/AVC does not define *how* to generate redundant pictures, as long as a decoded redundant picture is visually similar to the corresponding decoded primary picture. One prominent scenario is to use the same coding mode decisions for producing redundant pictures compared to the primary ones, except that the redundant picture is more coarsely quantized. Quantization is driven by the quantization scale factors Q_p and Q_r for the primary and redundant pictures, respectively. These factors have to be smaller than 51 and the quantization step-size is typically doubled when the factors increase by 6. Usually, if a primary picture is encoded with the parameter Q_p, the quantizer scale factor for the redundant picture can take any value between Q_p and 51. This is a widely accepted approach in literature [40], [41] and we will also consider it in our work.

Naturally, the quality of redundant pictures should be chosen by taking the network loss rate into account. If the loss rate is very low, the probability that a primary picture is lost and has to be replaced by the corresponding RP is also low, and therefore it does not make much sense to waste a lot of bits for encoding the RPs. This is why RPs should be quantized coarsely for low loss rates. On the other hand, as the loss rate increases, better quality of RPs becomes more advisable. Clearly, at the

fixed total rate, this comes at a price of reducing the quality of primary pictures. However, in the average sense, having better quality RPs is more beneficial, since now the probability that a primary picture is lost and a RP is used is higher. This will be discussed more in detail later, after we establish the rate-distortion model.

B. Proposed scheme: MSVC-RP

We extend the MSVC scheme [4] and increase the resiliency to temporal propagation of errors by the addition of redundant pictures. The proposed coding scheme (MSVC-RP) is illustrated in Figure 1. The input video sequence is split into sequences of odd and even source pictures. When encoding, each primary picture in the even/odd description is predicted only from other pictures of the same description, typically the previous picture. In addition, redundant pictures are included in the bitstream of each description. These RPs carry the information from the alternate description. In the time domain, they are placed such that they can replace a lost primary picture. Unlike the primary pictures, which use the previous primary frames from the *same thread* as a reference, redundant pictures are predicted from the previous frame in the input sequence. Redundant pictures are coded as P pictures and each primary frame has its redundant version. Redundant pictures are not used as a reference for any subsequent picture. The descriptions are sent possibly over two different lossy links, or over one link in an interleaved fashion. We can see that the resulting streams are independent, and therefore a reconstruction at a full frame rate is possible even with one of them only.

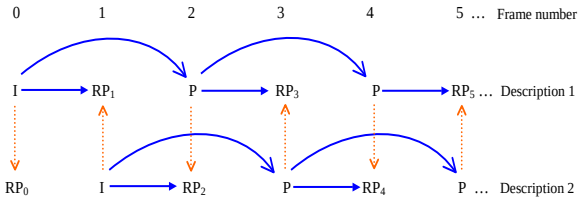

Fig. 1. The proposed scheme for MDC video.

At the receiver, if primary pictures are received error-free, the standard suggests that the RPs should be discarded. In our work, we will first follow this approach and we will treat the RPs as pure redundancy in case a primary picture is received. In this way we will also keep the decoding process as simple as possible, which can be of great importance for delay-sensitive applications. In Section VI, we will eventually improve the decoding process with efficient handling of all the received pictures.

If a primary picture (or parts thereof) has been lost, the corresponding redundant picture is reconstructed and used to replace its missing parts. Typically, replacing the lost parts of a primary picture with the same content, but more coarsely quantized, creates much smaller artifacts than if the missing parts are concealed with the information from the neighboring macroblocks from the same and/or subsequent frames. Finally, if both primary and redundant parts of a picture are lost, the missing information is reconstructed using an error concealment algorithm, e.g. by coping the closest available previous frame from either description. After the necessary discarding/replacement/concealment, the two descriptions are subsequently interleaved to produce the final reconstruction.

Obviously, the quantization parameters of primary and redundant pictures have to be properly selected in the MSVC-RP encoder, so that the reconstructed video at the receiver has the best possible quality. The optimal ratio between Q_p and Q_r depends on the expected network loss rate. We analyze the rate-distortion performance of MSVC-RP in the next section, and we formulate an end-to-end distortion minimization problem whose solution leads to the optimal choice of Q_p and Q_r under given bitrate constraints.

IV. Rate-distortion analysis

In this section, we study the redundancy that is introduced by MSVC-RP, and we present a model for the end-to-end distortion that will permit to select the coding parameters as a function of the channel conditions. We extract the parameters for our model from pre-encoded sequences. We further assume that the average distortion D_{av} in the presence of losses can be written as the sum of the source rate distortion D_S (the distortion due to quantization), and the average distortion due to losses, Δ, i.e., $D_{av} = D_S + \Delta$ [42]. We present below a source rate-distortion analysis of the MSVC-RP and we eventually compute the distortion due to missing information parts.

A. Redundancy in MSVC-RP

MSVC-RP obviously introduces redundancy for increased robustness compared to single description coding scheme. We can distinguish between the two sources of redundancy. First, there is the redundancy due to the fact that the frames in each

description are now temporarily further apart, which clearly requires more bits for encoding the same content with the same quality. The same kind of redundancy is introduced in MSVC scheme [4]. In addition, encoding of redundant pictures requires additional bitrate, which clearly depends on how fine (or coarse) they are quantized. The redundancy can therefore be expressed as:

$$\rho = \frac{R}{R_{SD}} - 1 = \frac{R_{MD} + R_r}{R_{SD}} - 1 \tag{1}$$

where R_{MD} denotes the necessary rate for encoding even and odd threads of frames, R_r represents the redundant picture rate, while the R_{SD} denotes the single description coding rate (prediction based on the previous frame and no redundant picture included).

Figure 2 shows the introduced redundancy for the Foreman QCIF sequence, compared to both single description coding and MSVC. In this case, we fix $Q_p = 20$ (which corresponds to the total single description rate of 222.6 kbits/s), and we vary Q_r from Q_p to 50. We can see that, compared to MSVC, the amount of redundancy varies from 2.85%, when $Q_r = 50$, up to 80.34%, when $Q_p = Q_r = 20$. Compared to the single description case, the redundancy varies from 37.46% to 140.5%. As we will see later, this redundancy however serves as an efficient error resilience tool on lossy channels.

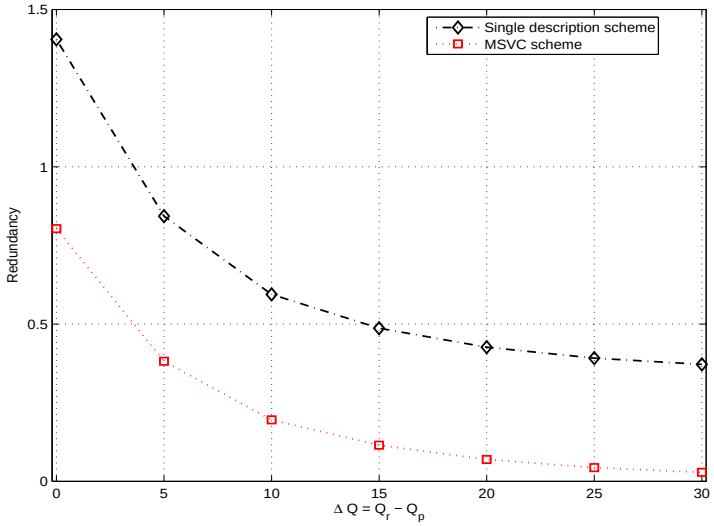

Fig. 2. Relative rate redundancy introduced in our scheme, compared to MSVC [4] and the single description case. Sequence: Foreman QCIF, 7.5 fps.

B. Source rate-distortion model

We derive now a model for the source rate-distortion characteristics. Suppose that we are given the total rate of R bits. A part of this rate, R_p, is used to encode primary pictures, while the remaining $R_r = R - R_p$ is used for encoding redundant pictures. To express the source distortions as functions of the bit rate, we use the model given in [42], according to which we can write the following expression for the distortion corresponding to primary pictures only:

$$D_p = \chi_p \; R_p^{\xi_p}, \tag{2}$$

In an analog way, we can write the distortion for redundant pictures as $D_r = \chi_r \; R_r^{\xi_r}$. D_p and D_r respectively denote the source distortions averaged over entire sequences, while χ_p, χ_r, ξ_p and ξ_r are directly related to the encoding scheme and the video sequence content, and are extracted from the corresponding bitstreams. Figure 3 illustrates this model for the *Foreman* and *News* QCIF sequences, where we use the mean square error (MSE) as a distortion measure.

Next, we model the peak signal-to-noise ratio ($PSNR_p$) for the primary pictures as a function of the quantization parameter Q_p. We again use a widely accepted linear model [43]:

$$PSNR_p = \alpha_p \; Q_p + \beta_p \tag{3}$$

and its validation for the *Foreman* QCIF and *News* QCIF sequences is given on Figure 4. As the PSNR is given by $10 \; log_{10} \frac{255^2}{D_p}$, we obtain from Eq. (2):

$$Q_p = \frac{10 \; log_{10} \frac{255^2}{\chi_p} + 10 \; \xi_p \; log_{10} R_p - \beta_p}{\alpha_p}, \tag{4}$$

which indicates that the quantizer scale factor can be approximated as a logarithmic function of the source rate. Finally, a similar relation holds for the redundant pictures.

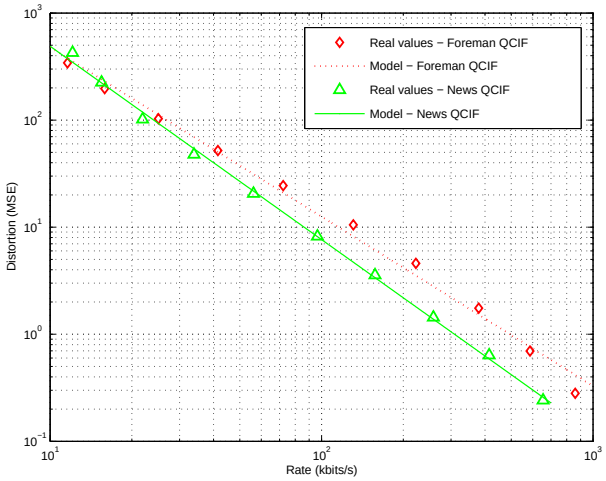

Fig. 3. Verification of the model given by Eq. (2). Sequences: Foreman QCIF and News QCIF.

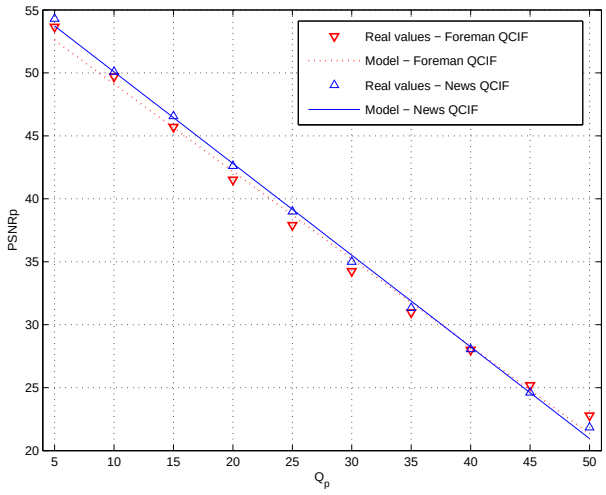

Fig. 4. Verification of the model given by Eq. (3). Sequences: Foreman QCIF and News QCIF.

C. Channel distortion

The degradation due to missing packets is driven by the loss probability p, the size of the area damaged by one loss (it corresponds to one slice in our scenario), and the temporal loss propagation factor η, which depends on the content of the video sequence and quantization parameters.

To establish our channel distortion model, we distinguish between three cases: 1) primary picture (or a slice, if a picture contains more slices) is received, 2) primary picture is lost but the corresponding redundant picture is received and 3) both primary and redundant pictures are lost. We assume that the average distortion in the presence of losses can be written as a weighted sum of the three distortions corresponding to the cases explained above:

$$\begin{aligned} D_{av} &= (1-p)\ D_p + p(1-p)\ \eta_1 D_r + p^2\eta_2 D_0 \\ &= (1-p)\chi_p R_p^{\xi_p} + p(1-p)\eta_1\chi_r R_r^{\xi_r} + p^2\eta_2 D_0 \end{aligned} \tag{5}$$

Here D_0 represents the average distortion when a complete frame is missing and is replaced by the previous decoded frame, while η_1 and η_2 are the temporal error propagation factors that reflect the increase in distortion if a lost primary picture is replaced by its redundant version or copied from the previous frame respectively. Consequently, our model assumes that a correctly received primary picture, although with possible erroneous references, does not induce any error propagation for later frames. Typically, η_1 depends on the ratio between Q_p and Q_r and the sequence activity, while η_2 can be approximated as a function of sequence activity and Q_p. Modeling η_1 as a polynomial of second degree of $\frac{Q_p}{Q_r}$ can provide a reasonably good approximation, as can be seen on Figure 5, which shows the results for the *Foreman* and *News* QCIF sequences. Furthermore, we model η_2 as a quadratic function of Q_p, as illustrated on Figure 6.

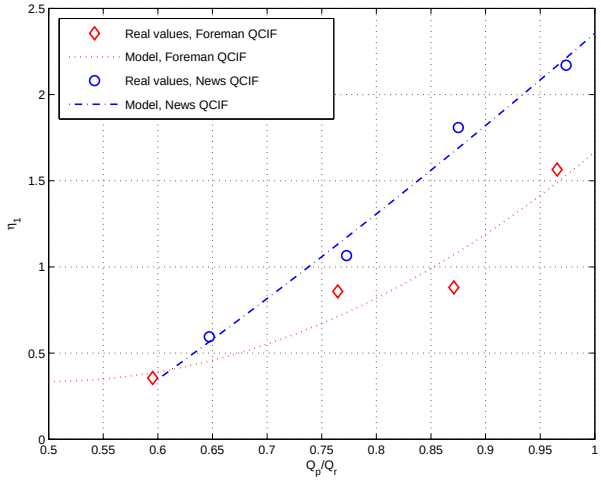

Fig. 5. Approximation of the parameter η_1 as a second degree polynomial of $\frac{Q_p}{Q_r}$. Sequences: Foreman QCIF and News QCIF.

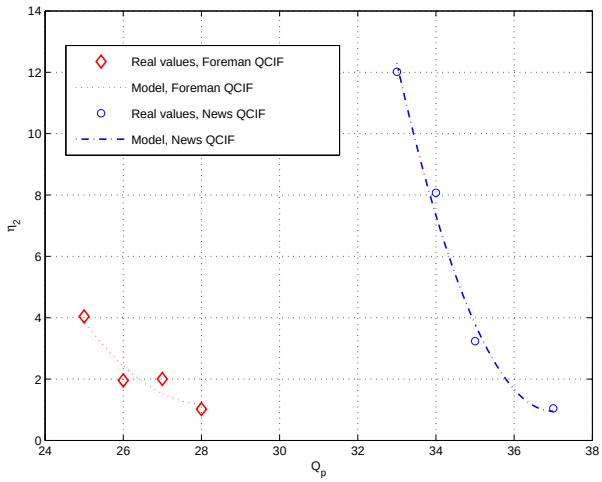

Fig. 6. Approximation of the parameter η_2 as a quadratic function of Q_p. Sequences: Foreman QCIF and News QCIF.

Even if the model is relatively simple, Eq. (5) provides a very good approximation of the end-to-end distortion, which is helpful for selecting the encoding parameters of primary and redundant pictures. We validate the model first with the *Foreman* QCIF sequence encoded at 7.5 fps and 144 kbits/s. For this sequence, the following combinations of encoding parameters give the desired bitrate: $\{(Q_p = 25, Q_r = 42), (Qp = 26, Q_r = 34), (Q_p = 27, Q_r = 31), (Q_p = 28, Q_r = 29), (Q_p = 29, Q_r = 29)\}$. All the bitstreams are affected by losses taken from loss patterns given in [44] that correspond to loss rates of 3%, 5%, 10% and 20% respectively. Average distortions for multiple combinations and the corresponding models are depicted on Figure 7. We can see that the proposed model fits very well the actual distortion values.

We further test the model on the News QCIF sequence encoded at 10 fps and 48 kbits/s. The following combinations of encoding parameters lead to the desired bitrate: $\{(Q_p = 33, Q_r = 51), (Qp = 34, Q_r = 44), (Q_p = 35, Q_r = 40), (Q_p = 36, Q_r = 38), (Q_p = 37, Q_r = 38)\}$. Again, we measure the average distortion when all the bitstreams are subject to losses taken from loss patterns corresponding to loss rates of 3%, 5%, 10% and 20% respectively. Average distortions of several combinations as well as the corresponding models are depicted on Figure 8 and confirm the validity of the end-to-end distortion model.

D. Rate allocation problem

The above rate-distortion model estimates the end-to-end distortion, as a function of the encoding parameters Q_p and Q_r, and the expected loss probability. We can therefore formulate the following distortion minimization problem. Given the total available rate R and the packet loss ratio on the network, p, find the coding rates of primary and redundant pictures R_p^* and R_r^*, respectively, such that $R_p^* + R_r^* \leq R$ and D_{av} given by Eq. (5) is minimized. We can write

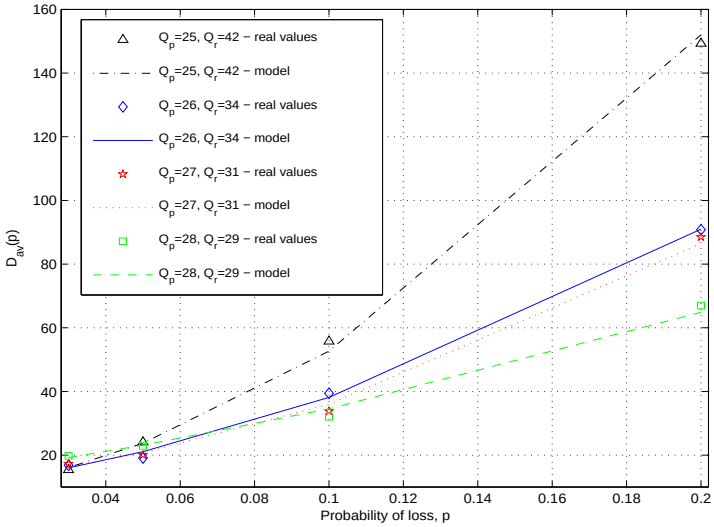

Fig. 7. Verification of the model given by Eq. (5). Sequence: Foreman QCIF at 7.5 fps and 144 kbits/s.)

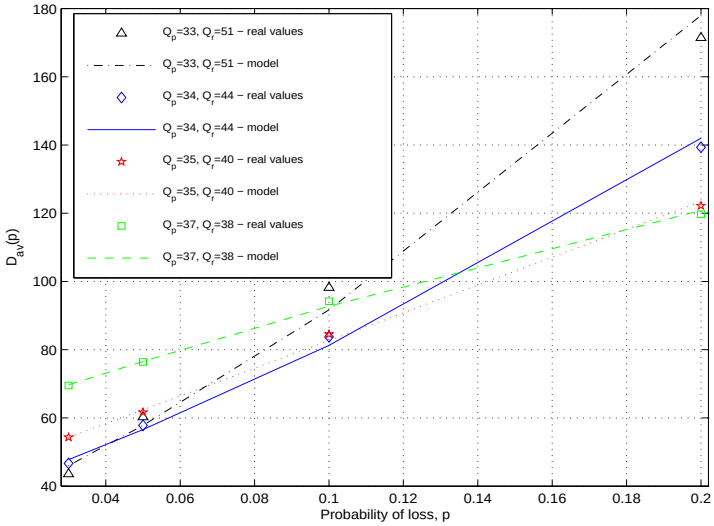

Fig. 8. Verification of the model given by Eq. (5). (Sequence: News QCIF at 10 fps and 48 kbits/s.)

$$(R_p^*, R_r^*) = \underset{R_p \geq 0, R_r \geq 0, R_p + R_r \leq R}{\arg\min} \{(1-p)\ \chi_p\ R_p^{\xi_p} + p(1-p)\ \eta_1\ \chi_r\ R_r^{\xi_r} + p^2 \eta_2\ D_0\}. \quad (6)$$

Note that the choice of optimal source rates is equivalent to the selection of the optimal quantizer scale factors Q_p and Q_r due to the relation of Eq. (4). Due to the nature of Eq. (6) that is actually the sum of two concave functions, the optimal choice of the encoding parameters based on the end-to-end distortion model is straightforward. It can even be reduced to solving the optimization problem for one single variable since the optimal solution is to fully use the available rate R. For example, one can solve the above problem for the variable R_p by setting $R_r = R - R_p$ and by differentiating Eq. (6) and taking its first derivative as an optimal solution for R_p. We compare this solution to the optimal solution obtained by full search over the space of (Q_p, Q_r) in the next section.

V. Performance evaluation

In this section, we compare our scheme with three solutions proposed in the literature. We start by describing the testing conditions. Then, we compare the average performance as a function of packet loss ratio on the network. Further on, we investigate how the PSNRs evolve on a frame-by-frame basis, by applying the same random sequence of packet losses on the different error resilient solutions. Finally, we examine the robustness of the schemes to erroneous estimation of the channel loss probability.

p	$Q_p^{opt} - model$	$Q_r^{opt} - model$	$Q_p^{opt} - real$	$Q_r^{opt} - real$	$D_{av}^{real} - D_{av}^{model}$
3%	25	35	25	42	0.72
5%	26	34	26	34	-1.98
10%	27	32	28	29	-2.47
20%	28	30	28	29	2.13

TABLE I
OPTIMAL QUANTIZATION PARAMETERS THAT MINIMIZE THE AVERAGE DISTORTION, AS A FUNCTION OF p. SEQUENCE: FOREMAN QCIF AT 7.5 FPS AND 144 KBITS/S.

A. Testbed

Our testbed corresponds to the common error resilience testing conditions specified in JVT-P206 [45], which specifies the required testing sequences, together with the corresponding bitrates and frame rates, as well as the bitstream packetization. The NAL unit size is limited to 1400 bytes, and the maximal size of each slice is chosen such that it can be fit in one NAL unit. Therefore, depending on the bitrate and the sequence format, there may be more slices per frame. Finally, an overhead of 40 bytes for the RTP/UDP/IPv4 headers is also taken into account when calculating the total bitrates.

We compare our MSVC-RP with three state-of-the-art schemes:

- MSVC scheme: the author in [4] consider several error concealment strategies when an entire frame is lost. In our work we only consider the simple scheme, where a lost frame is replaced with the closest possible received frame from either description, similarly to [17].
- *Adaptive intra refresh* (AIR) scheme [6], which takes into account both the source distortion and the expected channel distortion (due to losses) and chooses an optimal mode for each macroblock based on Lagrange optimization. Therefore, it is likely to place intra macroblocks in more "active" areas.
- *Random intra refresh* (RIR) scheme, which increases the robustness to losses by randomly inserting macroblocks whose number is proportional to a packet loss rate [46].

To have a fair comparison, we fix the total bit rates for all the schemes to be equal. In case of loss, parts or entirely lost pictures are replaced with their redundant versions taken from the alternate description in our MSVC-RP implementation. If both primary and redundant picture are lost, we copy the temporally closest decoded picture from either description. For the other schemes, in case of partial frame losses, the missing pieces are copied from the corresponding places in the previous pictures. If an entire picture is lost, we copy the entire previous picture, as it is implemented in the MSVC-RP scheme. In addition, only the first frames in all the video sequences are encoded as I pictures.

To simulate losses on the network, we used four lossy patterns, included in the ITU-T VCEG Q15-I-16 [44], that correspond to average packet losses of 3%, 5%, 10% and 20%. The pattern files were obtained from the real-world experiments on the Internet backbone between one sender and three reflector sites. We have tested all the sequences specified in JVT-P206 and at all the loss rates. To obtain statistically meaningful results, all the bitstreams are concatenated and tested with the entire loss patterns containing 10000 characters, for all the packet loss rates. We show here only the results for the three sequences: News QCIF, Foreman QCIF and Stefan CIF. Similar results have been obtained for other sequences, and can be found in [47].

B. Optimal solution vs model-based

We first evaluate the performance of the model-based optimization strategy, relying on the Eq. (6), and we compare it to the one given by an optimal selection of the coding parameters. The optimal values are obtained by full search over all combinations of quantization parameters that satisfy the total bitrate constraint.

Table I shows the optimal quantization parameters Q_p and Q_r, obtained by the model-based optimization or by full search and real measurements, for the *Foreman* QCIF sequence encoded at 7.5 fps and 144 kbits/s. In addition, the last column in this Table shows the difference between the actual distortion obtained by measurements and the one obtained with the distortion model. Similarly, Table I shows the same results for the *News* QCIF sequence at 10 fps and 48 kbits/s. We can see that the optimal quantization parameters obtained from the model generally match the real optimal parameters very well. In addition, we can see that the difference between the real and modeled values for distortions is very small, which confirms that an effective choice of the coding parameters can be found by the distortion model.

We can observe that the values of the quantizer scale factor of the redundant pictures, Q_r, decreases when the loss rate increases, as expected. When the losses are very high (20%), the primary and redundant pictures are even coded with very similar quantization parameters. The increase in quality of redundant pictures comes clearly at the expense of decreasing the quality of primary pictures when the overall bit rate is constrained. This however improves the average distortion, since the probability of using the redundant pictures becomes significant. On the other hand, when the loss rate is low, the optimal allocation tends to give as much rate as possible to primary pictures, while the redundant pictures are made very coarse. In this case, the system avoids wasting bits on the redundant pictures that are unlikely to be used in the decoding process.

p	$Q_p^{opt} - model$	$Q_r^{opt} - model$	$Q_p^{opt} - real$	$Q_r^{opt} - real$	$D_{av}^{real} - D_{av}^{model}$
3%	33	51	34	43	-2.45
5%	34	44	35	41	1.23
10%	35	40	36	39	2.49
20%	36	38	37	38	2.7

TABLE II
OPTIMAL QUANTIZATION PARAMETERS THAT MINIMIZE THE AVERAGE DISTORTION, AS A FUNCTION OF p. SEQUENCE: NEWS QCIF AT 10 FPS AND 48 KBITS/S

C. End-to-end distortion analysis

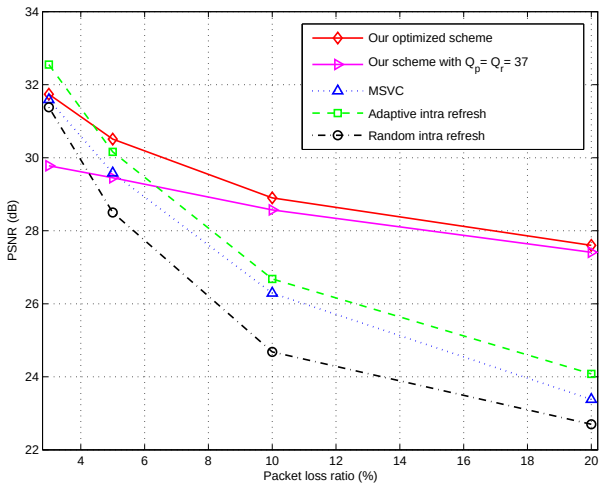

Fig. 9. Average PSNR, for four schemes and four loss patterns. Sequence: News QCIF, 10 fps, 48 kbits/s.

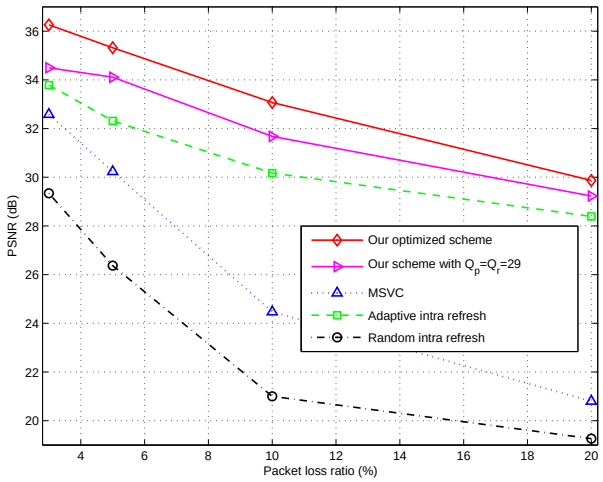

Fig. 10. Average PSNR, for four schemes and four loss patterns. Sequence: Foreman QCIF, 7.5 fps, 144 kbits/s.

We analyze here the performance of the different error resilient coding solutions in terms of average distortion, for different loss ratios. The average PSNR is illustrated in Figures 9, 10 and 11 for different test sequences (i.e., *News* QCIF at 10 fps and 48 kbits/s, Foreman QCIF at 7.5 fps and 144 kbits/s, and Stefan CIF at 512 kbits/s respectively). We compare our optimal MSVC-RP solution, with the MSVC, RIR and AIR schemes, as well as with MSVC-RP with maximal redundancy (i.e., $Q_p = Q_r$). It can be seen that the MSVC-RP scheme performs generally the best at all packet loss rates, and that the AIR scheme also provides an efficient solution at either low or high packet loss rate, depending on the activity in the vide sequence. At 10% loss probability, the MSVC-RP scheme outperforms the AIR and MSVC schemes by approximately 2.2 and 2.7 dB for the News sequence, and the gain reaches 3 dB and 8 dB respectively for the Foreman sequence. The quality gain due to MSVC-RP generally increases with the loss rate, since the redundancy offered by the design of two descriptions is really

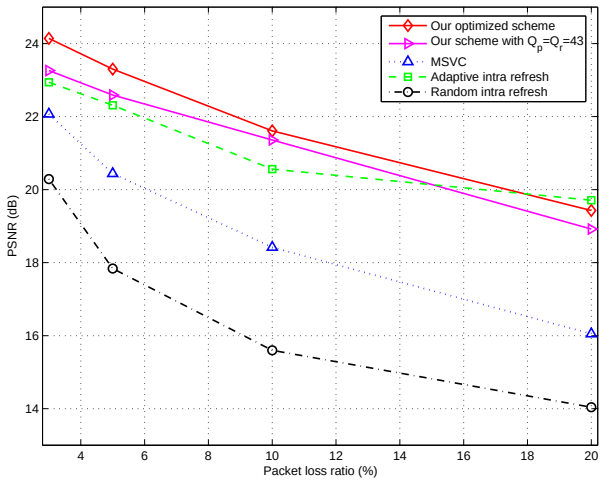

Fig. 11. Average PSNR, for four schemes and four loss patterns. Sequence: Stefan CIF, 30 fps, 512 kbits/s.

beneficial in this case, compared to joint coding with only one coding thread. For the complex sequences like Stefan encoded at medium bitrate, we can see that the performance of the MSVC-RP stays close to the AIR scheme, due to the limitations of the simple error concealment method that is unable to provide a sustainable quality when the loss of one description becomes frequent. On the contrary, the coding of Intra blocks in areas of high activity helps to improve the quality for the AIR scheme at high loss rate.

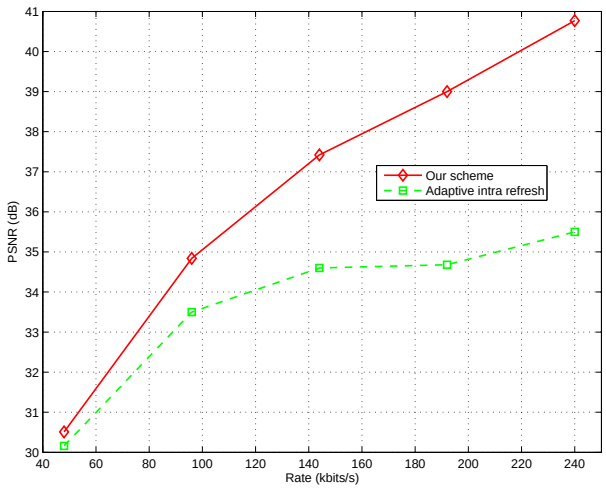

Fig. 12. Average PSNR, as a function of encoding rate, when PLR = 5%. Sequence: *News* QCIF, 10 fps.

We analyze the performance of the proposed scheme on a wider range of rate constraints, and we compare it to the AIR scheme. Figures 12 and 13 shows the average PSNR as a function of the rate constraint R, for the *News* and *Foreman* sequences respectively, when the loss rate is equal to $p = 5\%$. We can see that our approach gives the best performance in the whole range of bitrates, from 0.4 dB at 48 kbits/s to 5.3 dB at 240 kbits/s for the *News* sequence, and from 0.6 dB at 32 kbit/s up to 2.7 dB at 192 kbits/s for the *Foreman* sequence. Moreover, the gain increases as the bitrate increases. We have obtained very similar results for all the other loss rates.

Finally, Figure 14 presents the temporal evolution of the PSNR for the different encoding schemes, for the same loss trace. The error pattern is taken from a random entry in the error pattern file. The MSVC-RP scheme generally gives the best decoding quality. We can also notice that the AIR succeeds to catch up with our scheme, but with big variations in quality and with the performance similar to MSVC-RP in short intervals. The MSVC scheme performs very bad before the scene change in the last frames. Then it recovers, thanks to inserted intra macroblocks after the scene change, but the frame-by-frame quality varies in significant amounts, up to 12 dB between two consecutive frames. Overall, it can be observed that the variations of quality for the MSVC-RP scheme are much smaller than for the other schemes. This illustrates the benefits of the design of two descriptions that can decoded independently. Similar results have been observed for other loss rates, and other sequences [47].

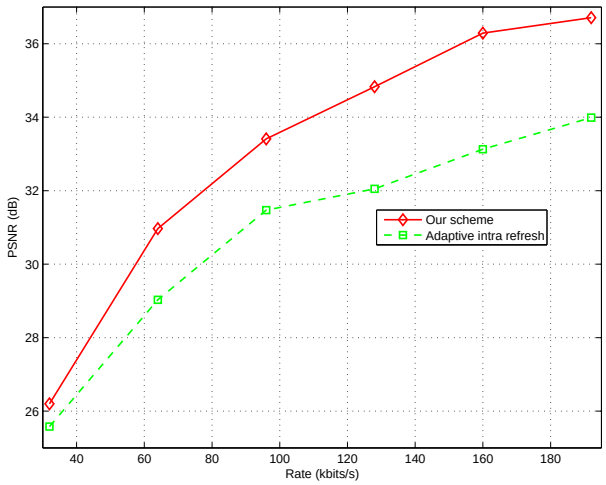

Fig. 13. Average PSNR, as a function of encoding rate, when PLR = 5%. Sequence: *Foreman* QCIF, 7.5 fps.

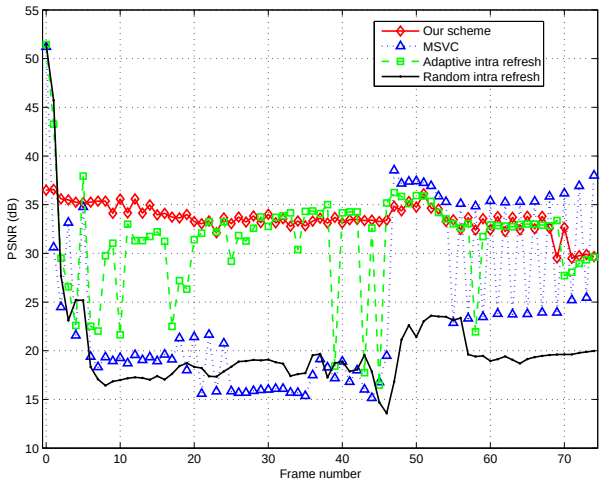

Fig. 14. Reconstructed video quality, on a frame basis, when PLR = 10%. Sequence: Foreman QCIF, 7.5 fps, 144 kbits/s.

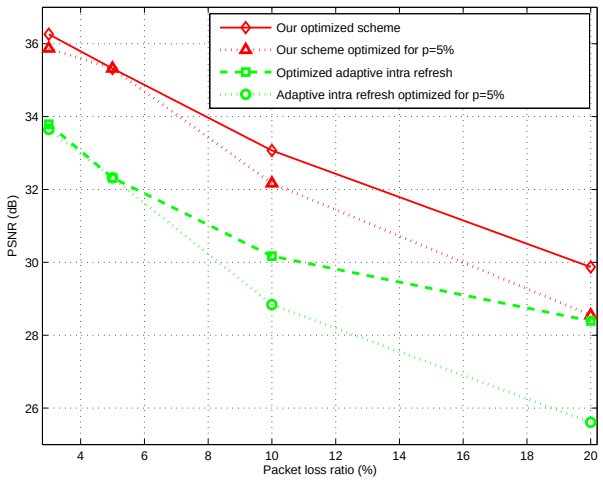

Fig. 15. Actual and minimal distortion vs. the actual PLR, when all the schemes are optimized for PLR = 5% (Foreman QCIF, 7.5 fps, 144 kbits/s).

D. Incorrect loss rate estimation

We analyze here the robustness of the encoding schemes to incorrect loss rate estimation. We compare the MSVC-RP and the AIR approaches that are optimized for a given loss ratio p, but where the actual loss rate is different from the expected one.

This is actually a common situation in practical scenarios. Figure 15 presents the end-to-end quality for the *Foreman* sequence at 7.5 fps and 144 kbits/s, when all the schemes are optimized for $p = 5\%$, but when the actual loss ratio varies from 3% to 20%. For the sake of completeness we also plot the best performance of MSVC-RP and AIR at each loss ratio. The difference between the optimized and actual performance for both schemes are 0.39 dB and 0.14 dB respectively, when $p = 3\%$. Not surprisingly, the gap between the optimized and actual performance increases as the actual loss ratio moves away from 5%. At $p = 10\%$ these gaps for both schemes are 0.9 dB and 1.33 dB respectively, while at $p = 20\%$ the corresponding gaps are 1.32 and 2.78 dB. Therefore, we can conclude that MSVC-RP is more robust to unknown network conditions. This can be a very desirable property, especially if the sender cannot change the encoding parameters as fast as the network conditions change. A similar behavior is observed for the other sequences, with even a larger gain for the MSVC-RP scheme when the sequence has low complexity, like the *News* sequence.

VI. Improved frame reconstruction

A. Combination of pictures

In the first part of this paper, we used to discard redundant pictures if the corresponding primary pictures are available at the decoder. This solution has an advantage of simplicity, but it is clearly suboptimal. Although a primary picture is correctly received, it may happen that its reference frames are affected by losses, which causes error propagation that also affects the primary picture. At the same time, it may happen that the thread from which a redundant picture is decoded is error-free or less affected by transmission errors. In these scenarios, choosing a redundant instead of a primary picture may be beneficial. This especially makes sense if the quantization parameters for primary and redundant pictures of the same original picture are very similar, which further induces similar visual qualities for both frames. Since primary and redundant pictures are decoded from different threads, the transmission error propagates only in one thread or description. We can therefore choose to use either frame in case of loss in the reference frames, depending on the which coding thread is less affected.

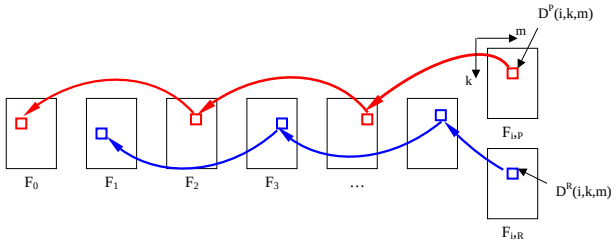

Fig. 16. Macroblock rate-distortion modeling: primary and redundant pictures from frame i are decoded from different threads. We can model the average distortion on each macroblock by tracking the losses that occurred in the corresponding threads.

We propose in this section to improve the decoding quality by combining the primary and redundant pictures at the decoder. We propose a rate-distortion model, which estimates the distortion of each macroblock, by tracking the state of reference areas in previous frames. In more details, let us consider a time instant i and suppose that both primary and redundant pictures are received, as depicted in Figure 16. We can determine the state of the macroblocks in primary and redundant pictures with coordinates (k, l) (where k is the height and l is the width of a macroblock), by tracking the motion from the reference frame along the corresponding coding threads. That means that we can track which reference frames are affected by losses and at which time instance that happened. The decoding state of a spatial area in a reference frame can be either (i) correctly received, (ii) taken from a redundant picture, or (iii) copied from the previous frame. In the last two cases, we can model the resulting error propagation, and decide for each macroblock of a frame if we should choose a primary or a redundant picture for the final reconstruction, in the case where both descriptions are received. The performance of the MSVC-RP scheme can be improved by more than 1 dB at high loss rates by adaptive decoding.

B. Macroblock distortion model

We now propose simple rate-distortion models for capturing the impact of loss propagation. We can distinguish two types of distortions due to error propagation, ΔD_r and ΔD_c, which correspond to using a redundant picture and temporal concealment, respectively, for decoding reference frames. Generally, both distortions are decreasing along time, when the decoder moves further away from the affected frame. We illustrate the behavior of the distortion due to temporal error propagation in Figure 17 that shows the impact of replacing the primary frame 15 with a redundant frame, on the next frames of the encoding thread. The results correspond to the Foreman QCIF sequence encoded with $Q_p = 25$ and $Q_r = 42$. We can see that the distortion due to error propagation indeed attenuates with time, and that the decay is close to exponential. This rather intuitive behavior corresponds to the observations reported in other works [48], [49]. We therefore assume that the distortions present an

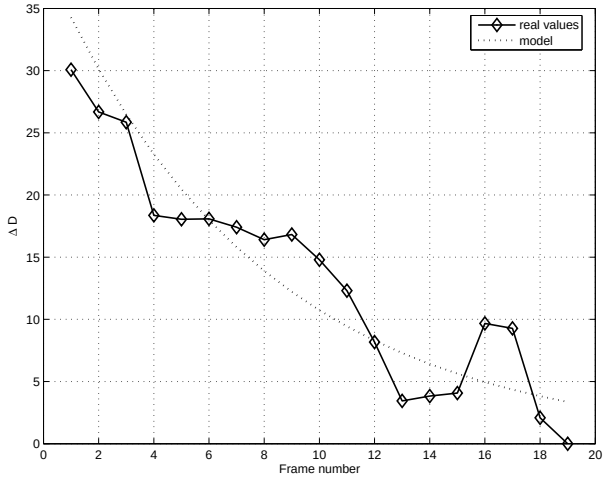

Fig. 17. Propagation of error $\Delta D_r(j)$ in the same encoding thread when the 15^{th} primary frame is replaced by a redundant pictures in the Foreman QCIF sequence encoded with $Q_p = 25$ and $Q_r = 42$ ($j = \frac{i-15}{2}$).

exponential decay with time. We denote by $\Delta D_r(i-m)$ and $\Delta D_c(i-n)$ the distortion observed in frame i due to a loss in frame n, which has been replaced by the corresponding redundant picture or respectively concealed. We can write

$$\Delta D_r(i-m) \approx e^{a_r \cdot (i-m)+b_r}, i \geq m \tag{7}$$

$$\Delta D_c(i-n) \approx e^{a_c \cdot (i-n)+b_c}, i \geq n, \tag{8}$$

where a_r, a_c, b_r, b_c are parameters that depend on the sequence content and encoding rates. These parameters can be estimated at the encoder, by simulating the corresponding losses of primary and/or redundant slices, and can be communicated to the decoder. We want now to approximate the average distortion in each macroblock that represents the basic unit for motion estimation. We assign the following indicator values to each decoded pixel:

- "C", if both primary and redundant slices carrying information about this pixel are lost
- "R", if a primary slice is lost but the corresponding redundant slice is received
- "G", otherwise.

These indicators permit to track errors over time, and the resulting distortion can be estimated based on the history of the macroblock references. We propose the following distortion model for computing the average end-to-end distortion of a macroblock with coordinates (k,l) in the primary picture i:

$$D^p(i,k,l) = \chi_p R_p^{\xi_p} + \sum_{j=1}^{\lfloor \frac{i}{2} \rfloor} \{[I(i-2j,k_j^p,l_j^p) == R] \cdot \Delta D_r(i-2j) + [I(i-2j,k_j^p,l_j^p) == C] \cdot \Delta D_c(i-2j)\} \tag{9}$$

where

$$k_j^p = k + \sum_{t=1}^{j} mv_y(i-2(t-1),k_{t-1}^p,l_{t-1}^p) \tag{10}$$

$$l_j^p = l + \sum_{t=1}^{j} mv_x(i-2(t-1),k_{t-1}^p,l_{t-1}^p) \tag{11}$$

and

$$k_0^p = k \tag{12}$$

$$l_0^p = l, \tag{13}$$

where " == " is the equality operator. The first part of the distortion is due to source distortion, while the second part represents the additional distortion due to error propagation. The indicators report the possible errors in reference frames from the same coding thread. Note that Intra macroblocks reset the temporal error propagation, in which case the indicator functions are reset to 'G' for the corresponding pixels. The parameters $mv_x(i,k,l)$ and $mv_y(i,k,l)$ denote the horizontal and vertical components of a motion vector for the macroblock with coordinates (k,l) in the frame i. Finally, we can write the expected distortion for macroblocks in redundant frames in a very similar way.

C. Performance analysis

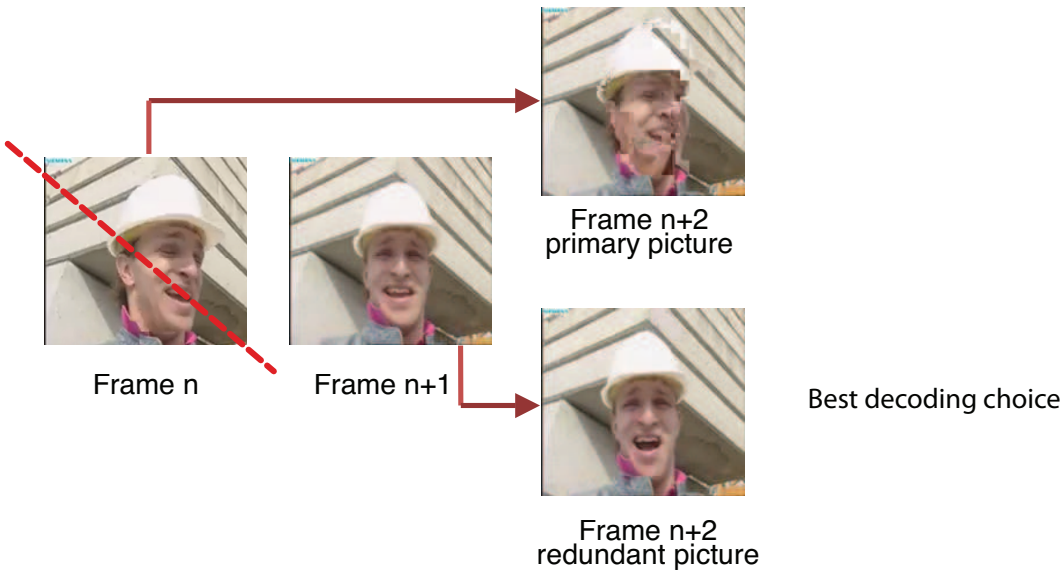

Fig. 18. Reconstruction of the $n+2^{th}$ frame from the Foreman QCIF sequence ($Q_p = 28, Q_r = 29$) when its both primary and redundant picture are received. Here the n^{th} frame, used as a reference for the $n+2^{th}$ primary frame, is entirely lost, while the $n+1^{th}$ frame, used as a reference for the $n+2^{th}$ redundant frame is correctly received. As can be seen, choosing the redundant picture can greatly improve the quality and reduce the error propagation.

We analyze here the performance of the improved decoding based on combination of redundant and primary pictures, with the simple rate-distortion model proposed above. We first show in Figure 18 that the choice of a redundant picture can improve the quality of the decoded sequence. In the case where both primary and redundant pictures are available, it is not always effective to discard the redundant picture, and to use only the primary picture. According to our distortion model, the primary picture, although it has been correctly received, has a much higher average distortion due to errors in previous frames of the same coding thread. Therefore, the redundant picture is chosen for decoding, which also improves the decoding quality of the subsequent frames in the same thread.

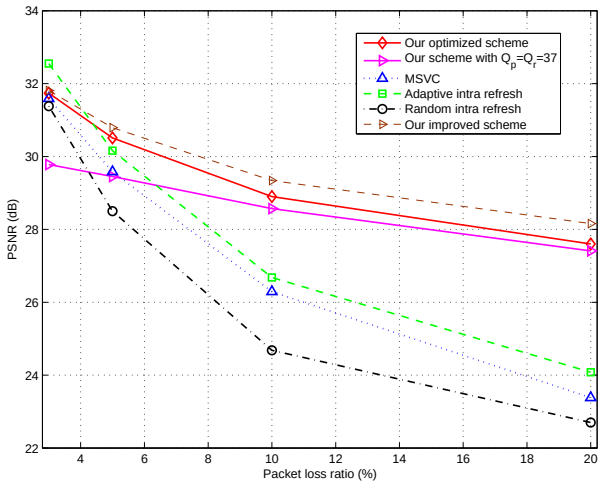

Fig. 19. Minimal achievable average distortion, as a function of a probability of loss, p. Sequence: News QCIF, 10 fps, 48 kbits/s.

We report on Figures 19 and 20 the benefits of the improved decoding process in terms of average distortion for the *News* QCIF sequence encoded at 48 kbits/s and 10 fps and *Foreman* QCIF sequence encoded at 144 kbits/s and 7.5 fps, respectively. We can see that the PSNR quality improvement ranges from 0.07 dB when $p = 3\%$ to 0.56 dB at 20% loss rate for the simple *News* sequence. In the *Foreman* sequence that contains more temporal activity, we can see that the gain ranges from 0.07 dB when $p = 3\%$ to 1.14 dB when $p = 20\%$. In general, the improvement at low loss rates is rather small, and gets more important at high loss rates. As the loss rate gets higher, it becomes very likely that an entire frame can be lost, in which

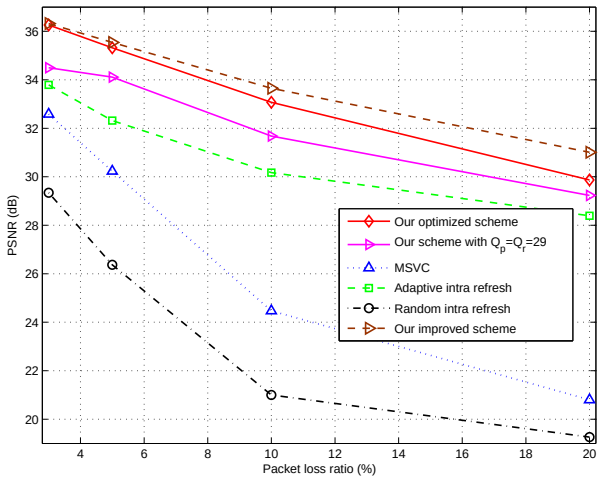

Fig. 20. Minimal achievable average distortion, as a function of a probability of loss, p. (Sequence: Foreman QCIF, 7.5 fps, 144 kbits/s).

case a serious quality degradation can be seen in subsequent frames. At the same time, the probabilities that both threads are seriously affected stays small, so that the possibility of choosing the frame to decode becomes beneficial.

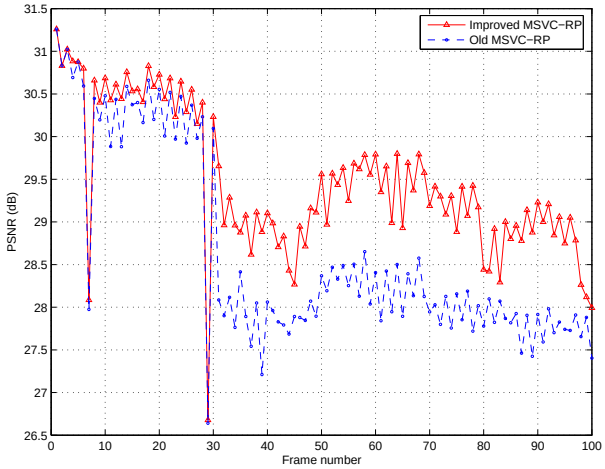

Fig. 21. Frame-by-frame behavior of the improved and old MSVC-RP. Sequence: News QCIF, 10 fps, 48 kbits/s.

We finally represent in Figure 21 the temporal evolution of the decoding quality in one specific realization of the loss process when $p = 20\%$. It shows that the choice of a redundant picture that has a correct reference is a much better option than choosing a primary picture whose reference has been lost. In this specific scenario, the improved MSVC-RP scheme provides an average gain of 0.83 dB over the old scheme. Discarding the redundant pictures by default is therefore not optimal, as the additional information provided by these pictures can be very helpful against temporal error propagation.

VII. Conclusions

In this paper, we have proposed a simple and standard compatible Multiple Description Video Coding scheme based on redundant pictures. We analyzed the rate-distortion behavior of the proposed solution in lossy scenarios, which permits to derive a strategy for the choice of the encoding parameters that control the redundancy and minimize the end-to-end distortion. Compared to state-of-the-art error resilient coding solutions, the proposed scheme offers significant gains in terms of average PSNR quality, fewer temporal fluctuations in the picture quality, and improved robustness to bad estimation of the loss probability in the network. We finally propose to improve the decoding process by choosing, on the macroblock level, between the primary and redundant macroblocks, based on their expected distortions, which proves to be beneficial at high loss rates.

References

[1] V.K. Goyal, "Multiple Description Coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep. 2001.
[2] Y. Wang, A.R. Reibman and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[3] I. Radulovic, Y.-K. Wang, S. Wenger, A. Hallapuro, M.M. Hannuksela and P. Frossard, "Multiple description H.264 video coding with redundant pictures," in *Proceedings of the ACM MV 2007*, Sept. 2007, pp. 37–42.

[4] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Proceedings of Visual Communications: Image Processing*, pp. 392–409, Jan. 2001.

[5] S. Wenger, "Video Redundancy Coding in H.263+," in *Proceedings of Workshop on Audio-Visual Services for packet networks*, 1997.

[6] Y. Zhang, W. Gao, H. Sun, Q. Huang and Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proceedings s of International Conference on Image Processing*, vol. 1, Oct. 2004, pp. 163–166.

[7] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 438–452, June 2002.

[8] J. G. Apostolopulous and S. J. Wee, "Unbalanced Multiple Description Video Communication Using Path Diversity," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, Oct. 2001, pp. 966 – 969.

[9] C.-S. Kim and S.-U. Lee,, "Multiple description motion coding algorithm for robust video transmission," in *Proceedings of IEEE Int. Symp. on Circuits and Systems*, Mar. 2000, pp. 717–720.

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

[11] D. Comas and R. Singh and A. Ortega and F. Marques, "Unbalanced multiple-description video coding with rate-distortion optimization," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 81–90, 2003.

[12] N. Franchi, M. Fumagalli, R. Lancini and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 321–334, Mar. 2005.

[13] D. Wang, N. Canagarajah and D. Bull, "Slice group based multiple description video coding using motion vector estimation," in *Proceedings of IEEE International Conference on Image Processing*, vol. 5, Oct. 2004, pp. 3237 – 3240.

[14] A.R. Reibman, H. Jafarkhani, M.T. Orchard and Y. Wang, "Multiple description video using rate-distortion splitting," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2001, pp. 978–981.

[15] D. Comas, R. Singh, A. Ortega, "Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, 2001, pp. 581–586.

[16] K.R. Matty and L.P. Kondi, "Balanced multiple description video coding using optimal partitioning of the DCT coefficients," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 928–934, July 2005.

[17] B. Heng, J. Apostolopoulos, J.S. Lim, "End-to-end rate-distortion optimized MD mode selection for multiple description video coding," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. Article ID 32592, 2006.

[18] S. Somasundaram and K.P. Subbalakshmi, "3-D multiple description video coding for packet switched networks," in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 1, July 2003, pp. 589–592.

[19] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard and R. Puri, "Multiple-description video coding using motion-compensated temporal prediction," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 193–203, Mar. 2002.

[20] N.V. Boulgouris, E. Zachariadis, A.N. Leontaris, M.G. Strintzis, "Drift-free multiple description coding of video," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, 2001, pp. 105–110.

[21] N.V. Boulgouris, K.E. Zachariadis, A. Kanlis and M.G. Strintzis, "Multiple description wavelet coding of layered video," in *Proceeding of International Conference on Image Processing*, vol. 4, Oct. 2004, pp. 2263–2266.

[22] H. Wang and A. Ortega, "Robust video communication by combining scalability and multiple description coding techniques," in *Proceedings of SPIE*, vol. 5022, May 2003, pp. 111–124.

[23] Y. Su, T. Tao, J. Lu and J. Wang, "Channel-optimized video transmission over WCDMA system," in *Proceedings of Vehicular Technology Conference (VTC)*, 2002, pp. 265– 269.

[24] R. Puri, K. Ramchandran, K.W. Lee, and V. Bharghavan, "Application of FEC based multiple description coding to Internet video streaming and multicast," in *Proceedings of IEEE Packet Video Workshop*, May 2000.

[25] J. Kim, R.M. Mersereau and Y. Altunbasak, "Distributed video streaming using multiple description coding and unequal error protection," *IEEE Trans. on Image Processing*, vol. 14, no. 7, pp. 849–861, July 2005.

[26] P. Chou, H. Wang and V. Padmanabhan, "Layered multiple description coding," in *Packet Video Workshop*, Apr. 2003.

[27] P.A. Chou and K. Ramchandran, "Clustering source/channel rate allocations for receiver-driven multicast under a limited number of streams," in *Proceedings of IEEE International Conference on Multimedia and Expo*, July 2000, pp. 1221–1224.

[28] J.R. Taal, J.A. Pouwelse and R.L. Lagendijk, "Scalable multiple description coding for video distribution in P2P networks," in *Proceedings of Picture Coding Symposium*, Apr. 2006.

[29] J.R. Taal, R.L. Lagendijk and S. Li, "Fair rate allocation of scalable multiple description video for many clients," in *Proceedings of SPIE Visual Commun. Image Processing Conference*, vol. 5960, July 2005, pp. 2172–2183.

[30] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 566–575, June 2002.

[31] T. Nguyen and A. Zakhor, "Matching pursuits based multiple description video coding for lossy environments," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, Sep. 2003, pp. 57–60.

[32] H.-T. Chan, C.-M. Fu and C.-L. Huang, "A new error resilient video coding using matching pursuit and multiple description coding ," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 1047 – 1052, Avg. 2005.

[33] G. Karabulut and A. Yongacoglu, "Multiple Description Coding Using Orthogonal Matching Pursuit," in *Proceedings of Med-Hoc-Net*, June 2004, pp. 529–534.

[34] T. Petrisor, C. Tillier, B. Pesquet-Popescu and J.-C. Pesquet , "Redundant multiresolution analysis for multiple description video coding," in *Proceedings of MMSP*, Oct. 2004, pp. 95–98.

[35] C. Tillier, T. Petrisor and B. Pesquet-Popescu, "A motion-compensated overcomplete temporal decomposition for multiple description scalable video coding," *EURASIP Journal on Image and Video Processing*, vol. 2007, no. Article ID 31319.

[36] V.A. Vaishampayan, S. John,, "Interframe balanced multiple description video compression," in *Proceedings of Packet Video 99*.

[37] F. Verdicchio, A. Munteanu, A. I. Gavrilescu, J. Cornelis and P. Schelkens, "Embedded multiple description coding of video," *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 3114–3130, Oct. 2006.

[38] Y.-C. Lee, Y. Altunbasak and R.M. Mersereau, "A two-stage multiple description video coder with drift-preventing motion compensated prediction," in *Proceedings of International Conference on Image Processing*, vol. 3, June 2002, pp. 557–560.

[39] A. Jagmohan, A. Sehgal, N. Ahuja, "Two-channel predictive multiple description coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, Sep. 2005, pp. 670–673.

[40] C. Zhu, Y.-K. Wang, M.M Hannuksela and H. Li, "Error resilient video coding using redundant pictures," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 2006, pp. 801–804.

[41] S. Rane, A. Aaron and B. Girod, "Systematic lossy forward error protection for error-resilient digital video broadcasting - a Wyner-Ziv coding approach," in *Proceedings of the International Conference on Image Processing*, vol. 5, Oct. 2004, pp. 3101–3104.

[42] P. Frossard and O. Verscheure, "Joint source/fec rate selection for quality-optimal mpeg-2 video delivery," *IEEE Trans. on Image Processing*, vol. 10, no. 12, pp. 1815–1825, Dec. 2001.

[43] M. Hamdi, J.W. Roberts and P. Rolin, “Rate Control for VBR Video Coders in Broad-Band Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1040–1051, Aug. 1997.
[44] S. Wenger, “Proposed error patterns for Internet Experiments,” *ITU-T VCEG document Q15-I-16*, Oct. 1999.
[45] Y.-K. Wang, S. Wenger and M.M. Hannuksela, “Common conditions for SVC error resilience testing,” *JVT document P206*, Aug. 2005.
[46] P. Haskell, D. Messerschmitt, “Resynchronization of motion compensated video affected by ATM cell loss,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Mar. 1992, pp. 545–548.
[47] I. Radulovic, “Balanced multiple description coding in image communications,” *PhD thesis, EPFL*, Dec. 2007.
[48] J. Chakareski, P.A. Chou and B. Aazhang, “Computing rate-distortion optimized policies for streaming media to wireless clients,” in *Proceedings of the IEEE Data Compression Conference*, Mar. 2002, pp. 53–62.
[49] S. Kanumuri, P.C. Cosman, A.R. Reibman and V.A. Vaishampayan, “ Modeling packet-loss visibility in MPEG-2 video,” *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341– 355, Apr. 2006.