Error Resilient Video Coding Using Redundant Pictures

Chunbo Zhu, Ye-Kui Wang, *Member, IEEE*, Miska M. Hannuksela, *Member, IEEE*, and Houqiang Li

***Abstract*—This paper presents several error resilient video coding methods based on redundant pictures. We combine redundant picture coding with reference picture selection and reference picture list reordering to prevent error propagation in motion compensated video coding. A hierarchical redundant picture allocation method is employed to make a tradeoff between error resilience and coding efficiency. For improved end-to-end rate-distortion performance in packet loss environment, three adaptive redundant picture allocation methods are further developed, utilizing characteristics of the input video content. Simulation results show that the adaptive redundant picture coding methods can achieve average luma peak-signal-to-noise improvements up to 3.5 dB compared to the loss-aware rate distortion optimized intra macroblock refresh algorithm implemented in the H.264/AVC Joint Model (JM). The proposed redundant picture coding methods are standard-compliant and do not introduce any additional end-to-end delay, therefore suit for low-delay applications such as video telephony and video conferencing. Due to the good error resilience performance, some of the proposed redundant picture coding methods have been adopted and integrated into the JM.**

***Index Terms*—Error control coding, error resilience, redundant picture, video coding.**

I. Introduction

IT IS WELL KNOWN that motion-compensated temporal prediction is applied in most video coding schemes. A significant advantage of predictive coding in video coding is that very high compression efficiency can be achieved. However, many video communication systems undergo transmission errors. Because of predictive coding, transmission errors will not only affect the decoding quality of the current picture but also be propagated to the following predictively coded pictures directly or indirectly referring to the picture with transmission errors. Without control of temporal error propagation, image quality may become seriously degraded or completely corrupted.

Manuscript received May 31, 2007; revised January 19, 2008. First published September 23, 2008; current version published January 21, 2009. The work of C. Zhu and H. Li was supported by National Science Foundation of China (NSFC) General Program under Contract 60572067, in part by the NSFC Key Program under Contract 60736043, and in part by the 863 Program under Contract 2006AA01Z317. This paper was recommended by Associate Editor H. Sun.

C. Zhu and H. Li are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, 230027 Hefei, China (e-mail: cbzhu@mail.ustc.edu.cn; lihq@ustc.edu.cn).

Y.-K. Wang and M. M. Hannuksela are with Nokia Research Center, Tampere FIN-33721, Finland (e-mail: ye-kui.wang@nokia.com; miska.hannuksela@nokia.com).

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.2008.2005802

Techniques for preventing temporal error propagation include interactive methods and noninteractive methods. Interactive methods refer to techniques where the decoder generates feedback messages about corrupted decoded areas and the encoder adjusts encoding according to the feedback message, e.g., feedback-based reference picture selection [1]. Noninteractive methods do not involve interaction between the encoder and the decoder. For systems where feedback information cannot be used in a large scale, e.g., broadcasting or multicasting with a huge number of receivers, noninteractive methods should be employed to prevent temporal error propagation. It should be noted that in these systems, transmission error rate estimation by the sender is still possible, and many noninteractive methods rely on this estimation to perform error resilient coding in an adaptive manner. Noninteractive methods include forward error correction (FEC) [2], which belongs to channel coding, and intra refresh (in the granularity of either macroblock or picture) [3], which is a source coding method. A general review of source coding error resilience methods can be found in [4].

An inherent characteristic of FEC is its increased end-to-end delay, which is unfriendly to low-delay applications, particularly conversational applications, such as video telephony and video conferencing. Intra-refresh techniques do not increase end-to-end delay, but the intra picture itself is more sensitive to errors due to the large amount of coded bits when compared to typical inter-coded pictures. This paper targets at conversational applications wherein low end-to-end delay is a must and smooth bit rate is desired, and focuses on source-coding error resilience methods.

The recent video coding standard H.264/AVC [5] supports many error resilient coding mechanisms, including reference picture selection (RPS) [1] and redundant pictures. In this paper, we first propose a method that combines the two techniques. Then we propose a hierarchical allocation of redundant pictures [6]. This redundant picture coding method has been adopted by the Joint Video Team (JVT) into the H.264/AVC Joint Model (JM). In this method, the allocation of redundant pictures is fixed as long as the group-of-pictures (GOP) size and the sub-GOP hierarchy are fixed. After that, we present three algorithms to perform allocation of redundant pictures in an adaptive fashion, according to the contents of the coded pictures, so to increase the reproduced video quality under error-prone conditions. For adaptive allocation of redundant pictures, we first propose a method based on the mean absolute motion vector value of each coded picture, and then a method according to a potential error propagation distortion of each coded picture, and at last a rate-distortion-optimized method, wherein adaptive coding of redundant pictures is treated as an optimization problem for coding mode selection.

The remainder of this paper is organized as follows. In Section II, related work on error resilient source coding techniques is reviewed. The proposed redundant picture coding methods, including a hierarchical and three adaptive redundant picture allocation methods are presented in Section III. Section IV provides the experimental results. Finally Section V concludes the paper.

II. Related Work

When talking about video error resilience or error propagation prevention, we are concerned with techniques that limit the scope of the damage to the reconstructed video quality resulted from transmission errors. This may be done by either channel coding or source coding. This work focuses on error resilient source coding techniques. While there are a lot of source coding layer error resilient video coding technologies [4], [7], [8], in this section we discuss two families of them, intra refresh and reference picture selection (RPS). As far as we know, intra refresh is among the most important error resilient coding methods in practical applications and has been extensively researched. Therefore, we use intra refresh as an anchor to demonstrate the performance of our proposed redundant picture coding methods. Reference picture selection is one of the bases for our proposed redundant picture coding methods.

A. Intra-Refresh

To prevent or minimize the effect of error propagation, insertion of intra pictures [3] can be employed to weaken the inter-picture prediction. However, it is costly to code an entire picture by intra-coding, because the size of intra-coded pictures is always much larger than that of inter-coded pictures. The balance between coding efficiency and error resilience needs to be compromised. Though adjusting the period of intra coded pictures is possible, more research efforts have been dedicated to intra macroblock refresh, where the problem to solve is to decide how many and which macroblocks to be intra coded. Zhang *et al.* first treated this problem as optimal coding mode selection of macroblocks [9]. They presented an algorithm to calculate the overall distortion introduced by quantization, error propagation and error concealment for optimal switching between inter-coding and intra-coding. Later, Cote *et al.* proposed in [10] an algorithm for a rate-distortion optimized mode selection and synchronization marker insertion scheme by taking into account both the channel conditions and the error concealment method used by the decoder. In [11], a flexible intra macroblock update algorithm was investigated to optimize the expected rate-distortion performance. It can choose optimal coding modes and reference pictures of macroblocks with a new method on choosing Lagrange multiplier in packet loss environments. The works in [9]–[11] are so-called loss-aware rate-distortion optimized intra refresh algorithms, which are currently the best known way for determining both the correct number and placement of intra-macroblocks for error resilience. Compared to intra picture refresh, intra macroblock refresh can avoid coding of large-sized intra pictures. However, because of the remaining inter-coded macroblocks for each picture, temporal error propagation cannot be completely prevented.

B. Reference Picture Selection (RPS)

An efficient way to prevent temporal error propagation without insertion of intra pictures/macroblocks is feedback-based reference picture selection (RPS) [1], [4]. In earlier RPS methods such as the one in [12], the basic assumption is that there is a feedback channel between the encoder and the decoder. If the encoder knows the position of erroneously decoded pictures or parts of pictures according to feedback information, the encoder can choose properly decoded pictures as references for latter pictures. In such a way, error propagation can be entirely stopped after a delay of several pictures since transmission of feedback messages requires time. Comparing to intra picture or macroblock refresh, the coding efficiency penalty for using inter-coding with RPS is much lower if reference pictures are not far away. The results in [13] show that the RPS algorithms are better than conventional intra update algorithms in terms of error resilience. However, RPS needs interaction between encoder and decoder and depends on the round-trip time of network. It is also shown in [13] that the performance of RPS suffers when the network round-trip time increases.

III. Coding of Redundant Pictures

H.264/AVC includes a feature called redundant pictures. A redundant picture is a coded representation of a primary picture or a part of a primary picture. The decoder should not decode redundant pictures when the corresponding primary picture is correctly received and can be correctly decoded. However, when the primary picture is lost or cannot be correctly decoded, a redundant picture can be utilized to improve the decoded video quality, if the redundant picture or part of it can be correctly decoded. A redundant picture can be coded as an exact copy of the primary picture, or with different coding parameters. Redundant pictures even do not have to cover the entire region represented by the primary pictures.

In [14], exact-copy redundant pictures were encoded for unequal protection of key pictures, which use relatively long-term reference pictures for inter prediction, such that the average temporal prediction chain is shortened for reduced temporal error propagation. Redundant pictures can also be encoded with some quality degradation by using larger quantization parameters than primary pictures, such that fewer bits will be used to represent redundant pictures. The method called systematic lossy error protection (SLEP), reported in [15]–[19], belongs to this category. In SLEP, FEC is also applied, which increases end-to-end delay. In [20], Multiple Description Coding (MDC) was realized using redundant pictures in an H.264/AVC compatible manner.

In this section, we introduce some new redundant picture coding algorithms, which combine redundant picture with RPS, reference picture list ordering (RPLR) and several redundant picture allocation methods. We first investigate a method that combines redundant picture coding with RPS and RPLR in a standard-compliant way. Then, a hierarchical allocation method is introduced to unequally protect primary pictures and make a trade-off between error resilience and coding efficiency. After that, we present three adaptive redundant picture allocation

methods, taking advantages of the characteristics of the video content.

A. RPS-Based Redundant Picture Coding

Feedback-based RPS is an efficient way to prevent error propagation in low-delay video transmission. However, when there is no interaction between the encoder and the decoder, feedback-based RPS cannot be directly used. In this paper, we propose a solution which combines RPS with redundant picture coding without relying on encoder-decoder interaction.

Our idea is that the encoder encodes the primary picture and redundant pictures with different selected reference pictures, and the decoder selects to decode the primary picture or a redundant picture for which the used reference pictures are correct in the decoder side. For example, if the latest reference picture in decoding order is lost, typically the primary picture cannot be correctly reconstructed. In this case, a redundant picture not referring to the latest reference picture can be used for constructing the current picture; hence temporal error propagation from the latest reference picture to the current picture and the following picture can be stopped.

More specifically, in the encoder side, for a primary picture to be protected, a number of redundant pictures are coded. The number of redundant pictures for the primary picture can be decided according to the estimated or known error rate; the higher the error rate, the larger should be the number. Reference pictures for redundant coded pictures are selected as following. Assume that the list of primary and redundant pictures are P, $R_{d0}, R_{d1}, \ldots, R_{dn}$, where P is the primary picture and R_{di} is its ith redundant picture. R_{d0} must use only a subset of reference pictures that may be used by P, and any $R_{di} (i \geq 1)$ must use only a subset of reference pictures that may be used by R_{di-1}, by excluding a few latest reference pictures in decoding order. For example, if the list of available reference pictures is $R_0, R_1, \ldots, R_m$, which are in reverse bitstream or decoding order, the primary picture may use all the available reference pictures, the first redundant picture must not use the first $n_1 (n_1 > 0)$ reference pictures in the list, the second redundant picture must not use the first $n_2 (n_2 > n_1)$ reference pictures, and so on.

In the decoder side, when decoding a coded picture, if the primary picture can be correctly reconstructed (i.e., all data pertaining to the primary picture is correctly received and all the referenced samples used for inter prediction is correctly reconstructed), the primary picture is decoded; otherwise erroneous image regions in the primary picture will be replaced with the corresponding regions in the first correctly decoded redundant picture when possible. Note that, a picture is correct only when the entire picture is received by the decoder and its reference pictures are correct. With the above decoding process, the selection of a redundant picture to be decoded can be done without parsing/decoding further beyond the slice headers.

Fig. 1 illustrates an example where 2 redundant pictures using different reference pictures are coded for a primary picture. The arrows indicate the prediction reference relationship; the pointed-to object is used by the pointed-from object for prediction reference. If the redundant picture R_{d0} can be correctly decoded, temporal error propagation due to any data loss in reference picture R_0 can be stopped. If the redundant picture R_{d1} can be correctly decoded, temporal error propagation due to any loss in reference pictures R_0 and R_1 can be stopped.

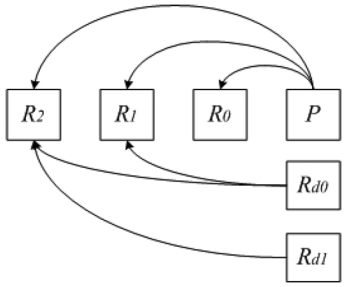

Fig. 1. Redundant pictures use selected reference picture(s) of the primary picture.

To enable the decoder to easily derive which reference pictures have been referred to, the encoder uses RPLR commands to put the to-be-referred reference pictures in the beginning of the reference picture list if one or more than one latest reference pictures are not to be used. In addition, the encoder sets the maximum reference index equal to the number of to-be-referred reference pictures. This way, the decoder can derive which reference pictures are used without syntax change to the H.264/AVC standard. Therefore, the redundant picture coding method is standard-compliant such that the encoded bitstream can be decoded by any H.264/AVC decoder of the Baseline or Extended profile, which supports redundant picture coding.

This RPS-based redundant picture coding method is the basis of all the following redundant picture allocation methods. It can prevent errors propagated from incorrect reference pictures. The proposed method impose no restriction to macroblock mode selection for redundant picture coding, thus the mode of an MB may differ between a primary coded picture and its redundant coded picture. We must point out that since reference pictures of a primary picture and its redundant picture(s) are different in our method, the error-free reconstruction of a redundant picture is not identical to that of the primary picture. Thus, error drifting exists if the primary picture is replaced with a redundant picture as reference picture for subsequent pictures.

B. Hierarchical Redundant Picture Allocation (HRP)

It is obvious that the strength of error robustness is proportional to the number of coded redundant pictures. However, the more redundant pictures are coded, the lower coding efficiency will be. Therefore, we have developed a hierarchical redundant picture allocation method to solve this traditional compromise problem—better coding efficiency or better error resiliency.

The hierarchical allocation of redundant picture is as follows. An input video sequence is first divided into GOPs. The first picture of each GOP, either intra-coded or inter-coded, is referred to as a key picture and has one redundant picture referencing the key picture of the previous GOP. This picture is thus better protected against errors similarly as in [14], compared to those pictures for which no redundant pictures are coded. If no further division of the GOP is carried out and no more redundant pictures are coded for the GOP, the hierarchy depth is defined as equal to 0. The redundant picture of the first picture in the

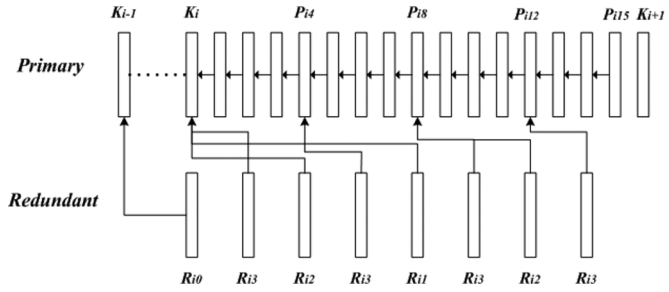

Fig. 2. Hierarchical allocation of redundant pictures.

GOP is associated with hierarchy level 0. If a pre-defined hierarchy depth is 1, then each GOP is further divided into two sub-GOPs (these are referred to as level-1 sub-GOPs). The first picture in the first level-1 sub-GOP is also the first picture of the GOP, hence it has already been coded with a redundant picture. For the second level-1 sub-GOP, the first picture is coded with a redundant picture (with hierarchy level 1), referencing the first picture in the first level-1 sub-GOP. If the pre-defined hierarchy depth is 2, then each level-1 sub-GOP is further divided into two level-2 sub-GOPs. The first picture in the first level-2 sub-GOP is also the first picture of the level-1 sub-GOP, hence has already been coded with a redundant picture. For the second level 2 sub-GOP, the first picture is coded with a redundant picture (with hierarchy level 2), referencing the first picture in the previous level-2 sub-GOP. And so on so force. Assuming that the GOP length is L, then the maximum depth of hierarchy n satisfies $0 \leq \mathrm{n} \leq \lceil \log_2 L \rceil$. If $n = \lceil \log_2 L \rceil$, then every primary picture in the GOP is coded with a redundant picture and the minimum sub-GOP length is 1.

Fig. 2 depicts an example where the GOP size is 16 pictures and half of the pictures are coded with redundant pictures according to the above allocation method. The maximum depth of hierarchy is 3. Arrows indicate the prediction reference relationship as in Fig. 1. K_i denotes the key picture in GOP i, P_{im} denotes the mth primary picture in GOP i, and R_{il} denotes a redundant picture with hierarchy level l in GOP i. The redundant picture R_{i0} of key picture K_i is coded using K_{i-1} of GOP $i-1$ as reference picture. Note that key pictures do not have to be intra-coded.

C. Adaptive Redundant Picture Allocation

The redundant picture coding methods above combine redundant picture coding with RPS, reference picture list reordering (RPLR) and a hierarchical allocation method. The main purpose of using RPS-based redundant picture is to protect a series of pictures from error propagation which may be caused by an earlier erroneously decoded picture. However, using of fixed hierarchical allocation may not reach optimal error resiliency performance, because the content of video pictures is not considered. For any coded video sequence, some coded pictures may be more sensitive to transmission errors than other pictures, and those coded pictures should be protected better than others.

In this sub-section, we introduce three adaptive redundant picture allocation methods taking into account the video content. The first two methods are based on two thresholds—average motion vector value and potential error propagation distortion [21]. One redundant picture can be allocated to a primary picture if a pre-defined threshold is exceeded. For improved rate-distortion performance, we have developed the third method, which is a rate-distortion-optimized method to adaptively allocate redundant pictures. We model the estimation of overall rate-distortion and propose an algorithm to calculate expected end-to-end distortion of GOPs including primary pictures and redundant pictures. Then, redundant pictures are adaptively coded for optimal R-D performance under an estimated target packet loss rate.

1) Thresholding by Mean Absolute Motion Vector Value (MAMV-ARP): Errors propagate temporally due to the recursive structure of the decoder. This temporal error propagation is typical for hybrid video coding that relies on motion compensated prediction in the inter-coding mode. It is very important to consider the impact of motion for the design of an error resilient video encoder since motion has a significant influence on the sensitivity of the decoder to errors. For this simple approach, we consider that the mean absolute motion vector value (MAMV) reflects the impact of motion to decoded picture quality, if few blocks are intra-coded in inter-frames. The MAMV of each primary picture, for which a redundant picture is decided to be coded, is calculated and compared to a threshold T. If the MAMV is larger than T, a redundant picture is coded for the corresponding primary picture; otherwise, no redundant picture is coded. The coded redundant picture uses the previous key picture or the previous primary picture which is coded with a redundant picture for inter prediction reference.

The MAMV is calculated by averaging the absolute motion vector value for all the 4×4 blocks in the coded picture as follows:

$$\overline{\mathrm{MV}} = \sum_{i=0}^{N-1} \left(|\mathrm{MV}X_i| + |\mathrm{MV}Y_i| \right) \tag{1}$$

where $\overline{\mathrm{MV}}$ denotes the MAMV, N is the total number of 4×4 blocks of the coded primary picture, and $\mathrm{MV}X_i$ and $\mathrm{MV}Y_i$, respectively, are the horizontal and vertical components of the motion vector of the ith block.

2) Thresholding by Potential Error Propagated Distortion (PEPD-ARP): To optimize the overall performance of a video transmission system, it is important to take into account the effect of error propagation. There are several approaches in the literature to investigate macroblock-level end-to-end rate-distortion optimization [11], [21], [22]. The basic assumption of these works is that the overall distortion is caused by two different types of errors, quantization and transmission errors. More specifically, let D_d be the overall end-to-end distortion. If the channel is error-free, D_d will be equal to the source distortion D_s that is introduced by quantization in the encoder. If the channel is error-prone, the overall distortion will be larger than the source distortion due to error propagation from reference frames. In the latter case, the overall distortion of correctly received inter-coded macroblock contains source distortion and potential error-propagated distortion of its reference blocks. If the current macroblock is lost, the overall distortion will be the error-concealed distortion which contains potential error-propagated distortion of the concealed macroblock and

the distortion between concealed macroblock and reconstructed macroblock.

Thus, the overall distortion of one macroblock can be denoted as a weighted sum of three types of distortion:

$$D_d = (1-p)(D_s + D_r) + pD_c \tag{2}$$

where p denotes an estimated packet loss rate, D_s denotes quantization distortion, D_r denotes the potential error-propagated distortion introduced by reference blocks, and $D_c = D_{c\_rec} + D'_p$ denotes error-concealed distortion which consists of potential error-propagated distortion of the concealed macroblock D'_p and the distortion between concealed macroblock and reconstructed macroblock D_{c_rec}.

D_r indicates the potential error-propagated distortion of the blocks from which the current macroblock is predicted. D_r is defined as

$$D_r = \sum_{m=1}^{4} w^m D_p^m \tag{3}$$

where w^m denotes a weighting factor applied to each reference block according to the overlapped area pointed to by the motion vector of one block, and D_p^m is potential error-propagated distortion of the mth overlapped reference block.

Based on above definitions of different types of distortion, the potential error-propagated distortion is defined as

$$D_p = (1-p)D_r + p\left(D_{c\_rec} + D'_p\right) \tag{4}$$

More details about the definition of potential error-propagated distortion can be found in [21]. In our approach, D_p is calculated in 4×4-block level. When D_p of all blocks in a picture have been obtained, an average D_p is then calculated.

The definition of D_p reflects error propagation sensitivity of a block. Pictures with larger average D_p are often more sensitive to transmission errors than other pictures. These pictures should be better protected against errors. Thus, for our PEPD-ARP method, we use average D_p as a measurement to allocate redundant pictures.

To calculate D_p and to allocate redundant pictures for this approach, the encoder must follow the following steps.

1) For the first frame of a sequence, set $D_p = 0$;
2) Calculate the D_p of each block according to (4). Note that, if the primary picture is a key picture, and if it is intra coded, let $D_r = 0$. So we get

$$D_p = p\left(D_{c\_rec} + D'_p\right) \tag{5}$$

   Moreover, D_r of intra-coded blocks in any frame must be 0.
3) After encoding a whole picture, average the D_p of all blocks and compare the average D_p with a pre-defined threshold D_T. If $D_p > D_T$, a redundant picture is coded for the primary picture; otherwise, no redundant pictures is coded.

Notice that, when a primary picture is coded with a redundant picture, the primary picture and the redundant picture usually have different D_p values due to their different used reference pictures. Assuming the primary picture is correctly received at the decoder with a probability of $(1-p)$, the error-propagated distortion D_p of the entire frame is identical to that of the primary picture. Otherwise, if the primary picture is lost and its redundant picture is correctly decoded with a probability of $p(1-p)$, D_p of the entire frame is identical to that of the redundant picture. If the redundant picture is also lost with a probability of p^2, D_p of the entire frame will be identical to error-concealed distortion. Thus, when a redundant picture is encoded, D_p of the entire frame must be calculated as a weighted sum of the above three distortions

$$D_p = (1-p)D_{p\_p} + p(1-p)D_{p\_r} + p^2\left(D_{c\_rec} + D'_p\right) \tag{6}$$

where D_{p_p} is the D_p of the primary picture and D_{p_r} is the D_p of the redundant picture which are calculated by (4) independently. The coded redundant picture uses the previous key picture or the previous primary picture which has been coded with a redundant picture for inter prediction reference.

The threshold D_T for this approach depends on the channel error rate and the characteristics of the video content. We have tested different D_T under different packet loss rates and found that if given a proper estimation of the packet loss rate, the best D_T which is set for the minimum overall end-to-end distortion is approximately linearly related to the packet loss rate p

$$D_T = pD_{p0} \tag{7}$$

where D_{p0} is a constant value that only depends on the characteristics of the input video sequence and the error concealment method used in the decoder. If advanced error concealment methods are used in the decoder, the value of D_{p0} can be decreased.

3) Rate-Distortion Optimized Adaptive Redundant Picture Allocation (RDO-ARP): In this part, a rate-distortion optimized adaptive redundant picture allocation method, denoted as RDO-ARP, is investigated to optimize the end-to-end rate-distortion performance with an estimated target packet loss rate. In this approach, the allocation problem is treated as a sub-GOP-level coding mode selection problem, for which a rate-distortion model is formulized. Note that, the definition of sub-GOP for RDO-ARP is different from that for HRP. In this approach, a current sub-GOP consists of pictures from the current picture to the last picture of the current GOP in coding order. When selecting the best coding mode of the current picture, the estimated rate-distortion performance of the current sub-GOP is optimized as a whole.

a) R-D model: When a primary picture is encoded, we have to decide whether it needs to be protected against transmission errors since it will influence the R-D performance of subsequent pictures for an error-prone transmission. With our redundant picture coding scheme, this problem is equivalent to a sub-GOP-level mode selection problem. Assume that there are two coding mode candidates when a primary picture is being coded: $mode1$ represents "not to code a redundant picture," and $mode2$ represents "to code a redundant picture." Then the total R-D cost of all the subsequent pictures (including the current

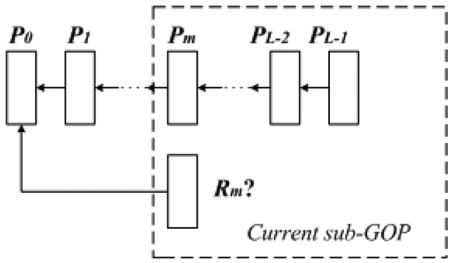

Fig. 3. Example of GOP and sub-GOP structure for RDO-ARP. The current sub-GOP contains the current picture P_m and all subsequent pictures within the GOP.

picture) within the current sub-GOP for the two modes can be denoted as

$$\begin{aligned}\text{RD\_cost(mode1)} &= D_1 + \lambda R_1 \\ \text{RD\_cost(mode2)} &= D_2 + \lambda R_2\end{aligned} \tag{8}$$

where D_1 and D_2 denote the total end-to-end distortions of all the pictures within the current sub-GOP for mode 1 and mode 2, respectively. R_1 and R_2 are the total bits of the pictures for mode 1 and mode 2, respectively. $\lambda(\lambda > 0)$ is a Lagrange multiplier. Therefore, the mode decision can be denoted as

$$\text{Best\_mode} = \arg\min\left[\text{RD\_cost(mode1)}, \text{RD\_cost(mode2)}\right]. \tag{9}$$

If $\text{RD\_cost(mode1)} > \text{RD\_cost(mode2)}$, a redundant picture is coded for the primary picture; otherwise, no redundant picture is coded. From (8) and (9), mode 2 will be selected only if (10) holds

$$D_1 - D_2 > \lambda(R_2 - R_1). \tag{10}$$

Assuming the mode decision here for the current picture does not affect possible coding of redundant pictures for other pictures in current sub-GOP, we have $R_2 - R_1 \approx R_{\text{rp}}$, where R_{rp} denotes the rate of a corresponding redundant picture. The mode decision criterion is derived as

$$D_1 - D_2 > \lambda R_{\text{rp}}. \tag{11}$$

Therefore, a redundant picture will be coded only if (11) holds.

b) R-D optimized allocation: Fig. 3 illustrates a GOP and sub-GOP structure. Assume that we are currently coding a primary picture P_m in the GOP. The picture number of the current primary picture is m; the GOP length is L; the picture number of each inter-coded picture in the GOP is in a range from 1 to $L-1$. The mode decision problem is: whether to code a redundant picture R_m for P_m or not. For our RDO-ARP method, the R-D performance of the current sub-GOP that consists of P_m and all subsequent pictures within the GOP is optimized as a whole.

With the proposed R-D mode selection model, we need to compute the overall end-to-end distortion of all the pictures within the current sub-GOP. There remains one problem that, due to the coding order, the distortion of latter pictures cannot be directly calculated when current picture is being encoded. Rather, the total distortion has to be estimated for both mode 1 and mode 2.

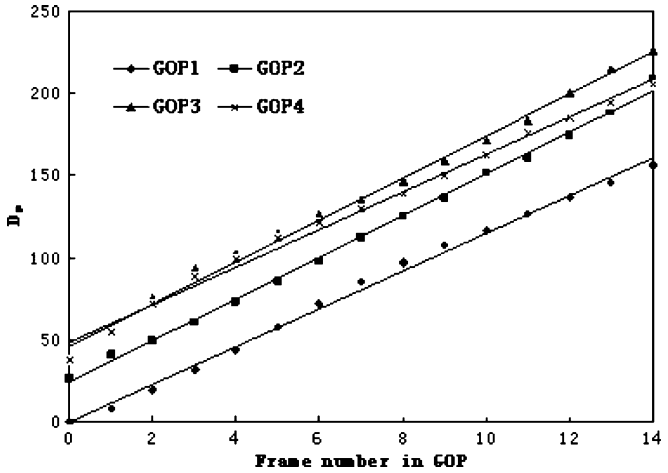

Fig. 4. D_p (in mean square error) versus frame number for the first 4 GOPs of the test sequence *Foreman* with GOP length 15 and estimated packet loss rate 10%.

Using the definition of the overall distortion in (2), we define the total end-to-end distortion D_{total} from P_m to P_{L-1} within the current sub-GOP as

$$\begin{aligned}D_{\text{total}} &= \sum_{i=m}^{L-1} D_d(i) \\ &= \sum_{i=m}^{L-1} \left\{(1-p)\left[D_s(i)+D_r(i)\right] + p\left[D_{c\_rec}(i)+D'_p(i)\right]\right\}\end{aligned} \tag{12}$$

where $D_d(i)$ denotes the end-to-end distortion of picture P_i, $D_s(i)$ denotes the source distortion, $D_r(i)$ and $[D_{c\_rec}(i) + D'_p(i)]$ denote the referenced error-propagated distortion and error-concealed distortion of picture P_i, respectively. From the definition of D_p in (4) and assuming that error concealment operates by copying the reconstructed pixels at the co-located positions in the previous picture, we have

$$D_p(i) = (1-p)D_r(i) + p\left[D_{c\_rec}(i) + D_p(i-1)\right] \tag{13}$$

where $D_p(i)$ is the error-propagated distortion of P_i. Then applying (13) into (12), we have

$$D_{\text{total}} = \sum_{i=m}^{L-1} \left\{(1-p)D_s(i) + D_p(i) - pD_p(i-1)\right\}. \tag{14}$$

For any $i > m$, $D_p(i)$ cannot be directly calculated when the current picture P_m is being encoded. Fortunately, it can be estimated, since it approximately linearly increases as the prediction chain grows if the GOP length is moderate and only few blocks are intra-coded. This linear relation can be seen from Figs. 4 and 5.

Therefore, for a limited-sized GOP with few intra-coded blocks, we have

$$D_p(i+1) = D_p(i) + D_{\text{delta}}, \quad 0 \le i < L-1 \tag{15}$$

where D_{delta} is constant within one GOP and only depends on the estimated packet loss rate and the characteristics of the input

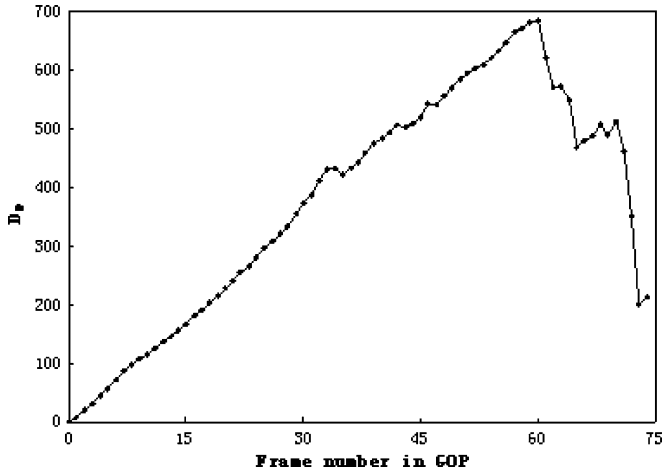

Fig. 5. D_p (in mean square error) versus frame number for the first GOP of the test sequence *Foreman* with GOP length 75 and estimated packet loss rate 10%. Comparing with Fig. 4, it can be inferred that the smaller is the GOP length, the better is the linear fit of D_p within a GOP.

video sequence. Then (16) and (17) can be obtained by applying (15) into (14) and replacing D_{total} with D_1 and D_2, respectively

$$\begin{aligned} D_1 = &\sum_{i=m}^{L-1}\{(1-p)D_s(i)\} + [(L-m)-p(L-m-1)] \\ &\times D_{p1}(m) - pD_{p1}(m-1) \\ &+ \frac{1}{2}\left[(L-m)(L-m-1)\right. \\ &\qquad \left.-p(L-m-1)(L-m-2)\right] D_{\text{delta}} \end{aligned} \tag{16}$$

$$\begin{aligned} D_2 = &\sum_{i=m}^{L-1}\{(1-p)D_s(i)\} + [(L-m)-p(L-m-1)] \\ &\times D_{p2}(m) - pD_{p2}(m-1) \\ &+ \frac{1}{2}\left[(L-m)(L-m-1)\right. \\ &\qquad \left.-p(L-m-1)(L-m-2)\right] D_{\text{delta}} \end{aligned} \tag{17}$$

where $D_{p1}(m)$ and $D_{p2}(m)$ denote the error-propagated distortion of picture P_m for mode 1 and mode 2, respectively. Therefore, we have

$$\begin{aligned} D_{p1}(m) &= D_{p\_p}(m) \\ D_{p2}(m) &= (1-p)D_{p\_p}(m) + p(1-p)D_{p\_r}(m) \\ &\quad + p^2\left(D_{c\_rec} + D'_p\right) \end{aligned} \tag{18}$$

According to (6) and knowing that for the previous picture P_{m-1}, the coding mode of P_m does not affect its distortion, we have

$$D_{p2}(m-1) = D_{p1}(m-1). \tag{19}$$

Then from (16)–(19) and neglecting the p^2-terms (the value of p is typically small), we have

$$D_1 - D_2 \approx [(L-m)-p(L-m-1)]\,[pD_{p\_p}(m) - pD_{p\_r}(m)]. \tag{20}$$

Finally, the mode decision criterion is

$$p\,[(L-m)-p(L-m-1)]\,[D_{p\_p}(m) - D_{p\_r}(m)] > \lambda R_{\text{rp}}. \tag{21}$$

If (21) holds, a redundant picture is coded for the primary picture; otherwise, no redundant picture is coded.

c) Lagrange multiplier selection: According to (8), the cost of each mode can be calculated as

$$\text{Cost} = D + \lambda R \tag{22}$$

where Cost denotes the R-D cost, D denotes the estimated total distortion, R denotes the coded bit-rate and $\lambda(\lambda > 0)$ is a Lagrange multiplier.

Further combining (14) and (22), and let the derivative of Cost to R be zero, we have

$$\lambda = -\frac{dD}{dR} = -\sum_{i=m}^{L-1}\left\{(1-p)\frac{dD_s}{dR}\right\} = (1-p)(L-m)\lambda_0 \tag{23}$$

where λ_0 is the error-free Lagrange multiplier. The error-free Lagrange multiplier used in H.264/AVC reference software is defined as

$$\lambda_0 = 0.85 \times 2^{\frac{q}{3}} \tag{24}$$

where q denotes the quantization parameter.

IV. Experimental Results

The proposed redundant picture coding algorithms have been integrated into the Joint Model (JM) version 10.2 [23] of H.264/AVC. The loss-aware rate-distortion-optimized intra refresh (LA-RDO) algorithm [11] implemented in the JM was used as the anchor. For our proposed methods, up to one redundant picture was coded for each primary picture. For less overhead, the QP of redundant pictures was set greater than QP of primary pictures by 6.

A. Simulation Environment

In our simulations, all pictures of a sequence were encoded once, and the resulting bitstreams were concatenate in order to generate a bitstream containing 4000 coded pictures. The overhead of 40 bytes IP/UDP/RTP headers per packet was taken into account in calculating of the bit-rate (the target transport protocols are IP/UDP/RTP). Slices were encoded with sizes up to 1400 bytes, and each slice was assumed contained in one packet. After a bitstream was generated, the loss simulator in [24] was used to simulate packet losses during transmission. For all the tested video sequences, 3%, 5%, 10%, and 20% average packet loss rates were tested. Then the lossy bitstream was decoded.

For better error robustness, constrained intra prediction was turned on for all testing cases including the anchor. Each inter-coded picture used 1 reference picture. RD optimization was turned on and no intra-macroblock update method was used for redundant picture coding methods. The error concealment method applied was frame copy for all testing cases, i.e., the corrupted image is replaced by corresponding pixels from the previous frame.

B. Objective Decoded Quality

Our simulations were targeted at low-delay application without encoder-decoder interaction and we verified the proposed algorithms based on two scenarios: 1) Coding with periodically intra-coded key pictures and 2) Coding with first intra-coded picture and the remainder inter-coded.

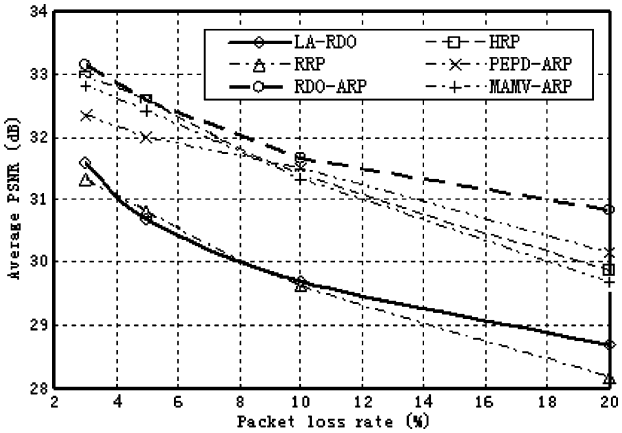

Fig. 6. Average PSNR (dB) versus packet loss rate (%) for testing sequence *News* QCIF at 10 fps and 48 kbps for simulation case 1). The mean absolute value of motion vectors threshold for MAMV-ARP is 2.0. The D_{p0} for PEPD-ARP is 1.4.

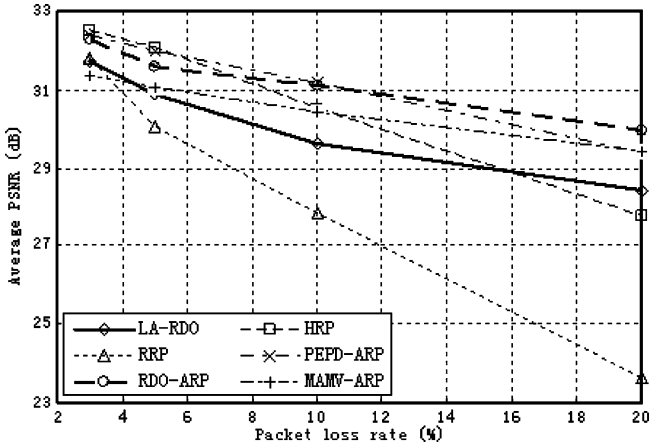

Fig. 7. Average PSNR (dB) versus packet loss rate (%) for testing sequence *Foreman* QCIF at 7.5 fps and 64 kbps for simulation case 1). The mean absolute value of motion vectors threshold for MAMV-ARP is 5.0. The D_{p0} for PEPD-ARP is 3.2.

1) Coding With Periodically Intra-Coded Key Pictures: For some scenarios of applications such as broadcast and video conference, frequent random access points must be provided for newcomers to join in. Therefore, we periodically inserted intra-coded pictures as random access points for this case. We have tested sequences *News* QCIF 10 fps at 48 kbps, *Foreman* QCIF 7.5f ps at 64 kbps and 144 kbps, and *Paris* CIF 15 fps at 384 kbps. The period of intra-pictures was set to 1 second for *News* and *Paris* and 2 seconds for *Foreman*. The JVT common test condition for error resilience tests in [25] was applied. For this case, the numerically lowest constant quantizer was used for a whole sequence that stays within the bit rate constraints. Moreover, for better coding efficiency, the QP of redundant pictures for all redundant picture coding methods was set larger than the QP of primary pictures by 6. Other experimental settings were as follows.

- *LA-RDO*: LA-RDO was simulated as the anchor. For optimal error resilience of LA-RDO, encoded bitstreams were optimized for an estimated target packet loss rate same as that used in the loss simulator, and the number of decoders was 30.
- *HRP*: HRP stands for the hierarchical redundant picture allocation method described in Section III-B. The GOP length was set equal to the period of intra pictures and the first intra picture (key picture) of each GOP was coded with a redundant picture referencing the first intra-picture of the previous GOP. The maximum depth of hierarchy to allocate redundant pictures was set to 1 for *News*, 2 for *Foreman* and *Paris*. In other words, 2 redundant pictures were coded every 10 primary pictures for *News* and 4 redundant pictures were coded every 15 primary pictures for *Foreman* and *Paris*.
- *RRP*: RRP stands for repeated redundant picture. For RRP, redundant pictures were periodically allocated to primary pictures. The period of redundant pictures was set to 5 for *News*, 4 for *Foreman* and *Paris*. In other words, 1 redundant picture was coded every 5 primary pictures for *News* and every 4 primary pictures for *Foreman* and *Paris*. Each redundant picture was a copy of the corresponding primary picture. This method is often used for coding redundant pictures and therefore, it was compared with our proposed redundant picture coding methods.

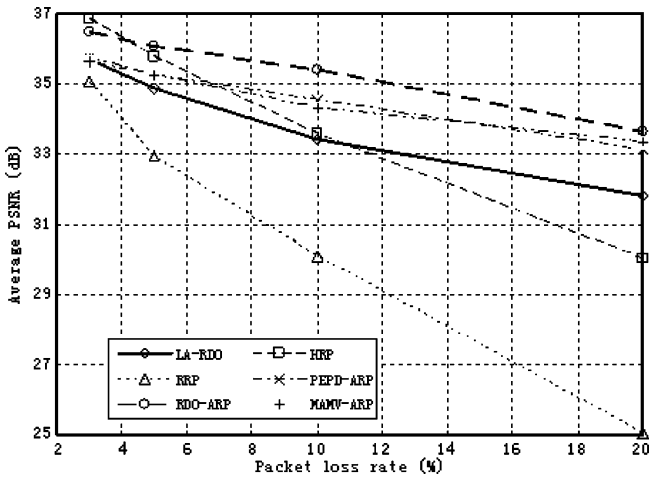

Fig. 8. Average PSNR (dB) versus packet loss rate (%) for testing sequence *Foreman* QCIF at 7.5 fps and 144 kbps for simulation case 1). The mean absolute value of motion vectors threshold for MAMV-ARP is 5.0. The D_{p0} for PEPD-ARP is 2.2.

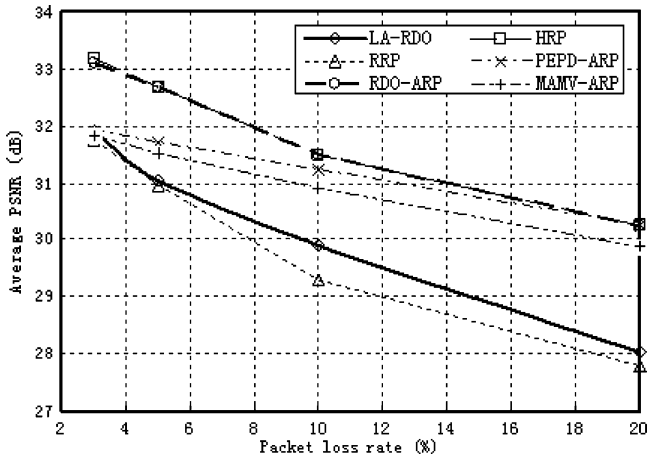

Fig. 9. Average PSNR (dB) versus packet loss rate (%) for testing sequence *Paris* CIF at 15 fps and 384 kbps for simulation case 1). The mean absolute value of motion vectors threshold for MAMV-ARP is 1.2. The D_{p0} for PEPD-ARP is 0.9.

TABLE I
ENCODED AND DECODED RESULTS OF LA-RDO AND RDO-ARP FOR CASE 1) WITH 10% PACKET LOSS RATE BOTH IN THE ENCODER AND LOSS SIMULATOR

Seq.	Target Bitrate (kbps)	Method	QP	Coded Bitrate (kbps)	Redundancy (%)	Enc. time (s)	Dec. PSNR (dB)
News	48	LA-RDO	35	48	35.09	172.3	29.7
		RDO-ARP	33	45.69	**13.52**	**120.3**	**31.67**
Foreman	64	LA-RDO	34	63.94	45.68	232.8	29.6
		RDO-ARP	32	63.18	**31.92**	**146.3**	**31.11**
Foreman	144	LA-RDO	28	139.6	48.74	263.4	33.41
		RDO-ARP	26	141.9	**35.31**	**158.7**	**35.38**
Paris	384	LA-RDO	33	371.2	38.97	2437	29.91
		RDO-ARP	32	367.4	**30.35**	**1033**	**31.5**

- *MAMV-ARP*, *PEPD-ARP* and *RDO-ARP*: Proposed adaptive redundant picture allocation methods which are described in Section III-C. For these methods, the estimated target packet loss rate used in the encoder was the same as that used in the loss simulator. The GOP length and the period of intra-pictures were the same as that for HRP. The first intra picture (also key picture) of each GOP was coded with a redundant picture referencing the first intra-picture of the previous GOP. For any other redundant picture coded for the current GOP, the key picture of the current GOP was used as reference.

Figs. 6–9 show the average luma PSNR values of case 1). We used constant QP for all testing sequences in the encoder so that the coded bit-rate was nearest to but would not exceed the target bit-rate. The average luma PSNR of 4000 decoded pictures were measured. For *News* 48 kbps, *Foreman* 64 kbps and 144 kbps, RDO-ARP performs the best among the anchor and the other redundant picture coding methods. For *News* 48 kbps, RDO-ARP outperforms the anchor by 1.6 $\sim$ 2.2 dB at 3% $\sim$ 20% packet loss rates. For *Foreman* 64k bps and 144 kbps, RDO-ARP outperforms the anchor by 0.6 $\sim$ 2.0 dB at 3% $\sim$ 20% packet loss rates. It can also be seen that the higher packet loss rate, the better error robustness RDO-ARP shows. For *Paris* 384 kbps, HRP outperforms the anchor by 1.3 $\sim$ 2.2 dB at 3% $\sim$ 20% packet loss rates. RDO-ARP has similar performance to HRP but show about 0.08 dB drop at 3% packet loss rate. For all testing sequences, RRP, for which a redundant picture is an exact copy of a corresponding primary picture, shows the worst error resiliency performance. The reason is that although a lost primary picture may be recovered by a repeated redundant picture, the redundant picture cannot prevent or partially mitigate propagated errors from earlier decoded pictures.

In Table I we compare the encoded and decoded results of RDO-ARP with those of LA-RDO at 10% estimated packet loss rate (used in encoding) and 10% simulated packet loss rate (used in packet loss simulation). As can be seen from the table, the proposed RDO-ARP introduces less redundancy into the bit-streams but improves the error robustness. Note that, the redundancy bit-rate is given by subtracting the normal bit-rate from the total bit-rate. The normal bit-rate was obtained by coding each sequence with no error resilient techniques with the same QP. As our method introduces less redundancy compared to the anchor method (LA-RDO), higher quantization accuracy (i.e., smaller QP) can be used at the same target bit-rate. We also compared the computational complexities for encoding an entire sequence between RDO-ARP and LA-RDO. The testing platform was Microsoft Windows XP 2.0 with a CPU of Pentium IV 3.0 GHz and 512M RAM. As can be seen from the table, RDO-ARP shows lower computing complexity than the anchor method.

2) Coding With First Intra-Coded Picture and the Remainder Inter-Coded: For some applications, intra-coded pictures with large size of bits should not be inserted frequently, to allow for smooth bit rate that affects required buffer size and end-to-end delay. In this set of simulations, all pictures except for the very first picture were inter-coded. Most of the simulation conditions were the same as in 1), except for the following.

1) Only the first picture of each sequence was coded as an intra-picture.
2) The key picture (also the first picture) of each GOP was inter-coded with one redundant picture referencing the key picture of the previous GOP. If any other redundant picture was coded for the GOP according to a proposed algorithm, the redundant picture would use the key picture of the current GOP as reference.
3) Each sequence was coded with different QP values and the average luma PSNR was measured under different bit-rates.
4) As the inefficiency of RRP has been exemplified, we did not simulate RRP for case 2).

Figs. 10–13 show the R-D curves of the testing sequence *News* under 3%, 5%, 10% and 20% packet loss rates. As shown, RDO-ARP still performs the best among the anchor and other redundant picture coding methods. Up to 3.5 dB gain can be obtained with RDO-ARP comparing to the anchor at 20% packet loss rate.

We can also observe from Figs. 6–13 that the simple redundant picture coding method HRP performs closely to RDO-ARP for low packet loss rate, e.g., 3%. This indicates that for systems where high packet loss rates will not occur or do not need to be handled by source coding, implementation of the simple HRP method is preferable than RDO-ARP.

C. Subjective Decoded Quality

Fig. 14 shows an example of the subjective decoded quality for case 1). 10% packet loss rate was simulated for the sequence *Foreman* at 144 kbps. Four frames in each reconstructed sequence were picked up. The 5th frame was error-free and the

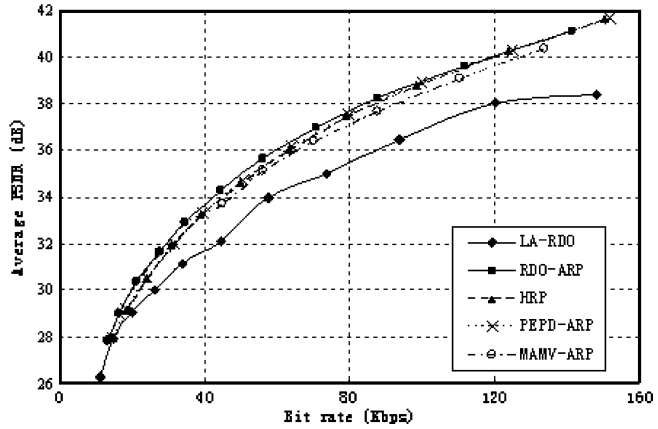

Fig. 10. Average PSNR (dB) versus bit rate (Kbps) for testing sequence *News* QCIF at packet loss rate 3% for simulation case 2). The D_{p0} for PEPD-ARP is 2.4 with estimated packet loss rate $p = 3\%$. The mean absolute value of motion vectors threshold for MAMV-ARP is 3.2.

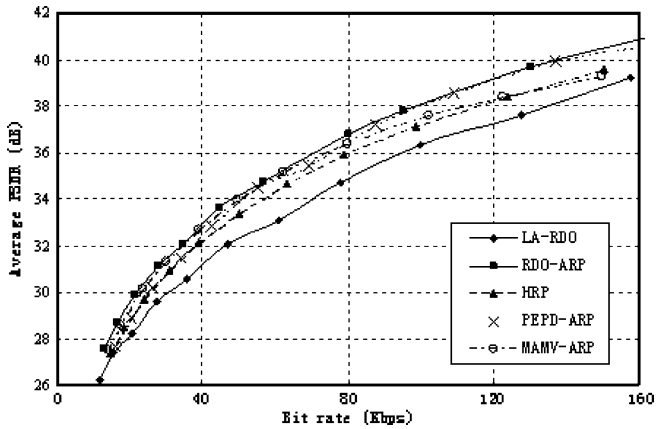

Fig. 11. Average PSNR (dB) versus bit rate (Kbps) for testing sequence *News* QCIF at packet loss rate 5% for simulation case 2). The D_{p0} for PEPD-ARP is 1.6 with estimated packet loss rate $p = 5\%$. The mean absolute value of motion vectors threshold for MAMV-ARP is 2.0.

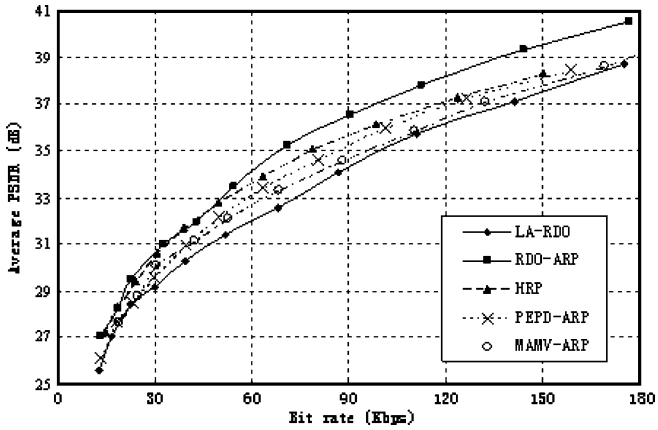

Fig. 12. Average PSNR (dB) versus bit rate (Kbps) for testing sequence *News* QCIF at packet loss rate 10% for simulation case 2). The D_{p0} for PEPD-ARP is 1.2 with estimated packet loss rate $p = 10\%$. The mean absolute value of motion vectors threshold for MAMV-ARP is 1.2.

6th frame was hit by a transmission error. After error concealment in the decoder, the visual quality for RDO-ARP was much

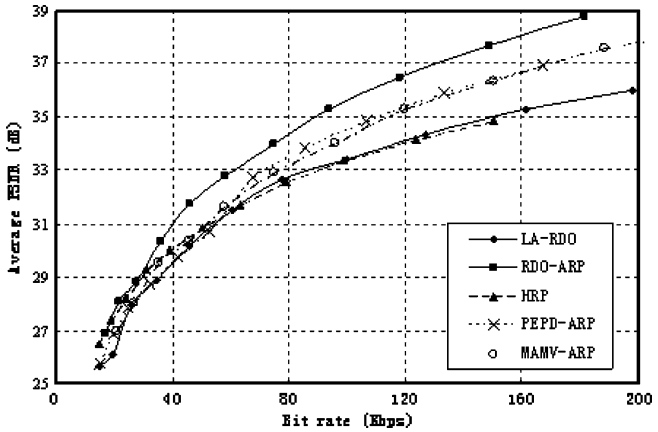

Fig. 13. Average PSNR (dB) versus bit rate (Kbps) for testing sequence *News* QCIF at packet loss rate 20% for simulation case 2). The D_{p0} for PEPD-ARP is 0.8 with estimated packet loss rate $p = 20\%$. The mean absolute value of motion vectors threshold for MAMV-ARP is 0.6.

Fig. 14. Subjective results of *Foreman* 144 kbps, QCIF resolution, for case 1). 10% packet loss rate was simulated. From left to right: frame No.5 ∼ No.8. Top row: LA-RDO. Bottom row: RDO-ARP.

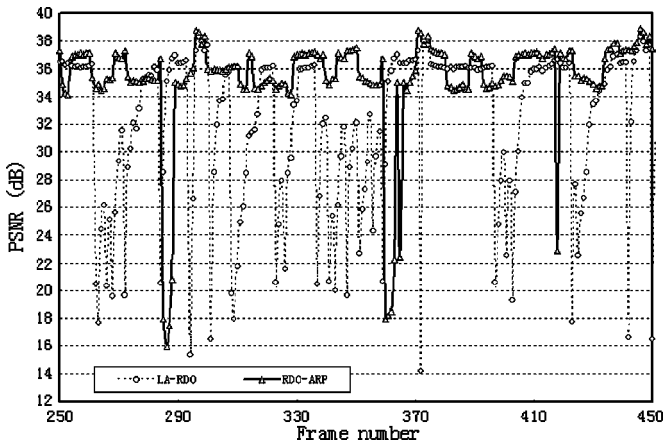

Fig. 15. PSNR of 200 decoded frames of sequence *Foreman* QCIF at 144 kbps and 7.5 fps with 10% packets loss for case 1).

better than that for the anchor. It can also be seen that the anchor could not prevent the error in the 6th frame propagating to the 7th and 8th frames. With RDO-ARP, the decoder could replace the erroneous primary frame with a properly decoded redundant frame, so error propagation was mitigated. In Fig. 15 we gave the comparison of the PSNR of 200 decoded frames under the same testing conditions as Fig. 14. As can be seen, RDO-ARP shows fewer sudden drops of PSNR than LA-RDO. Therefore, smoother visual quality can be achieved for users.

V. Conclusion and Future Work

Several methods for encoding, decoding and allocation of redundant pictures for improved error resilience were presented in this paper. As noninteractive error resilience tools, the methods can be applied in application scenarios wherein feedback cannot be used in a large scale. Furthermore, the methods do not introduce additional end-to-end delay hence are applicable to applications that require low end-to-end delay. Simulation results demonstrated that the methods significantly outperform the rate-distortion optimized loss-aware intra refresh algorithm implemented in the H.264/AVC reference software.

It should be noted that error drifting may exist when a redundant picture is used for inter prediction reference, as the decoding reconstruction of a redundant picture is typically different from that of the corresponding primary picture. A possible future work is to apply the H.264/AVC SP/SI coding technique to make the redundant picture and the corresponding primary picture have identical decoding construction, such that the error drifting is prevented.

References

[1] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.

[2] P. Frossard, "FEC performance in multimedia streaming," *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 122–124, Mar. 2001.

[3] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proc. ICASSP 92*, San Francisco, CA, vol. 3, pp. 545–548.

[4] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 4, pp. 61–82, Jul. 2000.

[5] *Advanced Video Coding for Generic Audiovisual Services* ITU-T Rec. H.264, May 2003.

[6] C. Zhu, H. Li, Y.-K. Wang, and M. M. Hannuksela, *Coding of Redundant Pictures for Improved Error Resilience,* JVT-R058, Jan. 2006.

[7] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 7, pp. 867–877, Nov. 1998.

[8] K.-P. Lim, G. Sullivan, and T. Wiegand, Text *Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods,* JVT-R095, Jan. 2006.

[9] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Selected Areas in Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[10] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the Inernet," *Signal Process.: Image Commun.*, vol. 15, no. 1, pp. 25–34, Sep. 1999.

[11] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," presented at the Packet Video Workshop 2002, Pittsburgh, PA, Apr. 2002.

[12] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. Global Telecommun. Conf.*, Nov. 1996, vol. 3, pp. 1503–1508.

[13] Y.-K. Wang, C. Zhu, and H. Li, "Error resilient video coding using flexible reference frames," in *Proc. SPIE VCIP*, Jul. 2005, vol. 5960, pp. 691–702.

[14] Y.-K. Wang, M. M. Hannuksela, and M. Gabbouj, "Error resilient video coding using unequally protected key pictures," in *Proc. Int. Workshop VLBV03*, Madrid, Spain, Sep. 2003, pp. 290–297.

[15] S. Rane, P. Baccichet, and B. Girod, Systematic *Lossy Error Protection Based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering*, JVT-S025, Apr. 2006.

[16] P. Baccichet, S. Rane, and B. Girod, "Systematic lossy error protection based on H.264/AVC redundant slices and flexible macroblock ordering," presented at the Packet Video Workshop (PV 2006), Hangzhou, China.

[17] S. Rane, P. Baccichet, and B. Girod, *Progress Report on CE6: Systematic Lossy Error Protection Based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering*, JVT-T093, Jul. 2006.

[18] S. Rane, P. Baccichet, and B. Girod, "Modeling and optimization of a systematic lossy error protection system based on H.264/AVC redundant slices," presented at the Picture Coding Symposium (PCS 2006), Beijing, China.

[19] S. Rane, P. Baccichet, and B. Girod, *Progress Report on CE9: Systematic Lossy Error Protection Using H.264/AVC Redundant Slices,* JVT-U057, Oct. 2006.

[20] I. Radulovic, Y.-K. Wang, S. Wenger, A. Hallapuro, and M. M. Hannuksela, *Multiple Description Coding Using AVC Redundant Pictures*, JVT-W054, Apr. 2007.

[21] 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 Process.*, Oct. 24–27, 2004, vol. 1, pp. 163–166.

[22] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, Jun. 2000.

[23] *The H.264/AVC Joint Model*, ver. 10.2 [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/

[24] Y. Guo, H. Li, and Y.-K. Wang, *SVC/AVC Loss Simulator Donation,* JVT-Q069, Oct. 2005.

[25] S. Wenger, *Common Conditions for Wire-Line, Low Delay IP/UDP/RTP Packet Loss Resilient Testing,* ITU-T VCEGN79r1, Sep. 2001.

Chunbo Zhu received the B.S. degree in electronic engineering and information science from University of Science and Technology of China, Hefei, China, in 2003, where he is currently working toward the Ph.D. degree in signal and information processing.

He was working as an Intern for Microsoft Research Asia, Beijing, China, during September 2006–February 2007. His research interests include image/video processing, image/video coding, error control, video adaptation, and computer vision.

Ye-Kui Wang (M'02) received the BS degree in industrial automation from Beijing Institute of Technology, Beijing, China, in 1995 and the Ph.D. degree in electrical engineering from the Graduate School at Beijing, University of Science and Technology of China, Beijing, China, in 2001.

He is currently a Principal Member of Research Staff in Nokia Research Center, Tampere, Finland. From February 2003 to April 2004, he was a Senior Design Engineer in Nokia Mobile Phone. Before joining Nokia, he worked as a Senior Researcher from June 2001 to January 2003 in Tampere International Center for Signal Processing, Tampere University of Technology, Finland. His research interests include video coding and transport, particularly doing those in an error resilient and scalable manner. He has co-authored over 200 technical standardization contributions and about 40 academic papers as well as co-invented over 60 issued and pending patents in the fields of multimedia coding, transport and application systems.

Dr. Wang has been an active contributor to different standardization organizations, including ITU-T VCEG, ISO/IEC MPEG, JVT, 3GPP SA4, IETF and AVS. He has acted as an Editor for several (draft) standard specifications, including ITU-T Rec. H.271, and the MPEG file format and the IETF RTP payload format for the scalable video coding (SVC) standard. He has also been a chair of the special session of Scalable Video Transport in the 15th International Packet Video Workshop (2006).

Miska M. Hannuksela (M'02) is a Research Leader and the head of the Media Systems and Transport Team in Nokia Research Center, Tampere, Finland.

He has more than ten years of experience in video compression and multimedia communication systems. He has been an active delegate in international standardization organizations, such as the Joint Video Team, the Digital Video Broadcasting Project, and the 3rd Generation Partnership Project. His research interests include scalable and error-resilient video coding, real-time multimedia broadcast systems, and human perception of audiovisual quality. He has authored more than 20 international patents and several tens of academic papers.

Houqiang Li received the B.S., M.S. and Ph.D degrees from the Department of Electronic Engineering and Information Science (EEIS), University of Science and Technology of China (USTC), Hefei, China, in 1992, 1997, and 2000, respectively.

From November 2000 to November 2002, he did postdoctoral research in the Signal Detection Laboratory, USTC. Since December 2002, he has been on the faculty of the Department of EEIS, USTC, where he is currently an Associate Professor. His current research interests include image and video coding, image processing, and computer vision.