ADAPTIVE REDUNDANT PICTURE FOR ERROR RESILIENT VIDEO CODING

Chunbo Zhu[1], *Ye-Kui Wang*[2], *and Houqiang Li*[1]

[1]University of Science and Technology of China, Hefei, China
[2]Multimedia Technologies Laboratory, Nokia Research Center, Tampere, Finland

ABSTRACT

We present several efficient adaptive redundant picture coding methods for error resilient video coding. In our previous work, redundant picture coding in combining with reference picture selection, reference picture list reordering and hierarchical redundant picture allocation was proposed. This paper investigates how to allocate redundant pictures more efficiently according to the content characteristics of the primary pictures. Simulation results show that the adaptive redundant picture coding methods can achieve average PSNR improvements around 2 to 4 dB compared to the loss-aware rate distortion optimized (LA-RDO) intra macroblock refresh implemented in H.264/AVC Joint Model (JM). This paper also asserts that the methods do not introduce any additional end-to-end delay, therefore suit for low-delay applications such as video telephony and video conferencing, which demand better error resilience than other applications.

Index Terms—video coding, error resilience, redundant picture, rate distortion optimization

1. INTRODUCTION

It is well know that motion-compensated temporal prediction is applied in most video coding methods. 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 following predictively coded pictures. Without control of temporal error propagation, image quality may become seriously degraded or completely corrupted.

Techniques for preventing temporal error propagation include interactive methods and non-interactive methods. Interactive methods refer to techniques where the recipient transmits information about corrupted decoded areas and/or transport packets to the transmitter. Such methods include feedback based intra refresh and reference picture selection. Non-interactive methods do not involve interaction between the transmitter and the receiver. For systems where feedback information cannot be used, non-interactive methods have to be employed to prevent temporal error propagation. Non-interactive methods include forward error correction (FEC) [1], which is done in transport coding layer, and intra refresh (in the granularity of either macroblock (MB) or picture), which is a source coding layer method. For example, Stockhammer et al. provided a loss-aware rate-distortion optimized intra refresh solution [2]. A general review of source coding error resilience methods can be found in [3].

An inherent characteristic of FEC is its long encoding and decoding delay, which makes it impractical for low-delay applications. Intra-refresh techniques can avoid such long delay, but they are not always error-robust because the intra data itself is more sensitive to errors due to its large size of coded bits, which also need large buffer size. This paper is targeted at the non-interactive applications which need low end-to-end delay and smooth bit rate that affects required buffer size, such as video conferencing. Therefore, FEC and intra-refresh techniques cannot be used when we limit our discussion to the low-delay source coding error-resilient techniques.

The newest video coding standard H.264/AVC is capable of utilizing two methods called reference picture selection (RPS) and redundant picture. In our previous work [4], we proposed a method combining the two techniques to code redundant pictures, which was proved to be able to efficiently prevent error propagation. We also proposed a hierarchical allocation of redundant pictures. The redundant picture coding method in [4] was then adopted by JVT into the H.264/AVC Joint Model. However, the allocation of redundant pictures is fixed as long as the group-of-picture (GOP) size and the sub-GOP hierarchy are fixed. In conclusion, the prior art methods do not consider the contents of video pictures for allocation of redundant pictures. 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. When redundant pictures are applied, more redundant pictures should be allocated to those pictures that are more sensitive to transmission errors. However, there lacks such a way to allocate redundant pictures adaptively.

This paper discloses an algorithm 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. We first proposes an adaptive redundant picture allocation method according to a mean absolute motion vector value of each coded picture (MAMV-ARP), or according to a potential error propagation distortion of each coded picture (PEPD-ARP). Then we propose an adaptive redundant picture allocation method (RDO-ARP) that optimizes the estimated rate-distortion performance of a group-of-pictures (GOP), wherein a Lagrange multiplier selector is used.

2. ALGORITHM DESCRIPTION

The redundant picture coding method in [4] combines redundant picture coding with RPS, reference picture list reordering (RPLR) and a hierarchical allocation of redundant pictures. The allocation method is fixed without considering the content of the input video. We propose two adaptive redundant picture allocation methods in this paper, adaptive allocation (ARP) and rate-distortion optimized allocation (RDO-ARP), utilizing the characteristics of the input video. More detailed information can be found in [5]. In the following, we assume that pictures are allocated into groups of pictures (GOPs). The first picture of each GOP is referred to as a key picture. However, it should be mentioned that the methods are also applicable when no GOP allocation is performed.

2.1. Adaptive redundant picture allocation (ARP)

A straight forward way of utilizing the video content to allocate redundant pictures is to set thresholds for some video features which reflect the sensitivity of a video sequence to errors. A redundant picture can be allocated to a primary picture when the chosen threshold is exceeded. We introduce two approaches using this threshold criterion. The first is based on the mean absolute motion vector value, so-called MAMV-ARP. The second is based on potential error propagation distortion [6], so-called PEPD-ARP.

2.1.1. ARP according to mean absolute motion vector value (MAMV-ARP)

For this simple approach, the mean absolute motion vector value of each primary picture is calculated and compared to a threshold T. If the value is larger than T, a redundant picture is coded for the corresponding primary picture; otherwise, no redundant picture is coded. Preferably, the coded redundant picture uses the previous key picture, in decoding order, for inter prediction reference. Alternatively, the coded redundant picture uses the previous primary picture with a redundant picture for prediction reference.

The value of mean absolute motion vector is calculated by averaging the absolute motion vector value for all the 4x4 blocks in the coded picture.

2.1.2. ARP according to potential error propagation distortion (PEPD-ARP)

For this approach, we measure the potential error propagation distortion of a block defined in [6]:

$$D_p = (1-p)D_r + p(D_c + D_p^{'}) \tag{1}$$

Where p denotes an estimated packet loss rate, D_r denotes the distortion introduced by an erroneous reference picture, D_c denotes the concealment distortion and D_p' denotes the D_p of the concealed block. D_r is defined in [6] as:

$$D_r = \sum_{m=1}^{4} w^m D_p^m \tag{2}$$

Where D_p^m is error propagated distortion of m-th reference block of current block. w^m denotes a weighting factor applied to each reference block according to the overlapped area pointed to by the motion vector of current block.

The definition of D_p can be considered as a potential error propagated distortion. It indicates the error propagation feature of a block. Frames with larger average D_p are often more sensitive to transmission errors. These frames should be specially protected against errors. Thus, we use D_p as a measurement to allocate redundant pictures.

To calculate D_p and to allocate redundant pictures in our approach, the encoder must follow the steps:

1) For the first frame of a sequence, set D_p=0;

2) Calculate the D_p of each block. Note that, if the primary picture is a key picture, and if it is intra coded, let D_r=0. So we get:

$$D_p = p(D_c + D_p^{'}) \tag{3}$$

Moreover, D_r of intra blocks in any frame must be zero.

3) After encoding a whole frame, average the D_p of all blocks and compare the average D_p with a threshold D_T. If $D_p > D_T$, a redundant picture will be coded for the primary picture; otherwise, no redundant pictures will be coded;

Notice that, when a redundant picture is encoded, D_p for the consequent frame must be calculated by:

$$D_p = (1-p)D_{p\_primary} + p(1-p)D_{p\_redundant} + p^2(D_c + D_p') \tag{4}$$

Where $D_{p\_primary}$ is the D_p of the primary picture and $D_{p\_redundant}$ is the D_p of the redundant picture.

Preferably, the coded redundant picture uses the previous key picture or, alternatively, the previous primary picture which has a redundant picture for inter prediction reference.

2.2. Rate-distortion optimized adaptive redundant picture allocation (RDO-ARP)

In this sub-section, a new adaptive redundant picture allocation is introduced to optimize the end-to-end distortion. It is called rate-distortion optimized adaptive redundant picture allocation (RDO-ARP.) We treat the redundant picture allocation to be a coding mode selection problem and formulize the end-to-end distortion of different coding modes. Rate-distortion optimization is performed for mode decision and a Lagrange multiplier is derived.

2.2.1. R-D model for mode decision

Assuming that there are two coding mode candidates when a primary picture is currently coded: mode 1 represents 'not to code a redundant picture', and mode 2 represents 'to code a redundant picture'. The total R-D cost of all the latter frames within one GOP for the two modes can be represented by:

$$\begin{aligned} RD\text{cost}(\text{mode1}) &= D_1 + \lambda R_1 \\ RD\text{cost}(\text{mode2}) &= D_2 + \lambda R_2 \end{aligned} \tag{5}$$

Where D_1 denotes the total end-to-end distortion of all the frames following current primary picture within same GOP for mode 1, and D_2 for mode 2. R_1 and R_2 are the total bits of latter frames for mode 1 and mode 2 respectively. λ is a Lagrange multiplier. Therefore, the mode decision problem can be represented by:

$$\text{Best\_mode} = \arg\min[RD\cos t(\text{mode1}),\ RD\cos t(\text{mode2})] \tag{6}$$

This means if the following holds,

$$RD\cos t(\text{mode1}) > RD\cos t(\text{mode2}) \tag{7}$$

a redundant picture will be coded for the primary picture; otherwise, no redundant picture will be coded.

From (5), (6) and (7), mode 2 will be selected only if

$$D_1 - D_2 > \lambda(R_2 - R_1) \tag{8}$$

In other words, a redundant picture will be coded if (8) holds. Note that,

$$R_2 - R_1 = R_{rp} \tag{9}$$

Where R_{rp} denotes coded bits of a corresponding redundant picture. Finally, from (8) and (9), we have:

$$D_1 - D_2 > \lambda R_{rp} \tag{10}$$

This means a redundant picture will be coded only if (10) holds.

2.2.2. Distortion computation and RDO-ARP allocation

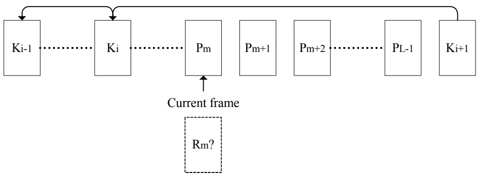

Fig. 1. An example of GOP structure.

Fig.1 shows a GOP structure. In this figure, the key picture K_i of current GOP i uses the key picture K_{i-1} of previous GOP as reference. Assume that we are now coding a primary picture P_m in i-th GOP. The picture number of current primary picture is m; the GOP length is L; the picture number of each inter-picture in the GOP is in a range from 1 to $L-1$. The mode decision problem is: whether to code a redundant picture for P_m or not.

We define the total end-to-end distortion D_{total} from primary frame P_m to P_{L-1} within one GOP as:

$$D_{total} = \sum_{i=m}^{L-1} D_d(i) = \sum_{i=m}^{L-1} \{(1-p)[D_s(i) + D_r(i)] + pD_c(i)\} \tag{11}$$

Where $D_d(i)$ denotes the end-to-end distortion of i-th frame in the GOP. $D_s(i)$ denotes the source distortion introduced by quantization of i-th frame. $D_r(i)$ and $D_c(i)$ denote the reference distortion and concealed distortion of i-th frame which are defined as in (1).

Assuming the concealment method is frame copy, we have

$$D_p(i) = (1-p)D_r(i) + p[D_c(i) + D_p(i-1)] \tag{12}$$

Where $D_p(i)$ is the distortion due to transmission errors of the i-th frame, which is also referred to as the error propagated distortion. Then we have

$$D_{total} = \sum_{i=m}^{L-1} [(1-p)D_s(i) + D_p(i) - pD_p(i-1)] \tag{13}$$

For any $i>m$, $D_p(i)$ cannot be directly obtained when the current picture P_m is encoded. However, we can estimate it since it approximately linearly increases as the prediction chain grows within a GOP. This is supported by our experiments in [5].

Therefore, when we are handling a limited-sized GOP with few intra coded blocks, we have

$$D_p(i+1) = D_p(i) + D_{delta}, \qquad 0 \le i < L-1 \tag{14}$$

Where D_{delta} is a constant that only depends on the estimated packet loss rate and the characteristics of the input video sequence. Then we obtain (15) and (16) by applying (14) into (13) and replacing D_{total} with D_1 and D_2:

$$\begin{aligned} D_1 = &\sum_{i=m}^{L-1}(1-p)D_s(i) \\ &+[(L-m)-p(L-m-1)]D_{p1}(m) - pD_{p1}(m-1) \\ &+\frac{1}{2}[(L-m)(L-m-1)-p(L-m-1)(L-m-2)]D_{delta} \end{aligned} \tag{15}$$

$$\begin{aligned} D_2 = &\sum_{i=m}^{L-1}(1-p)D_s(i) \\ &+[(L-m)-p(L-m-1)]D_{p2}(m) - pD_{p2}(m-1) \\ &+\frac{1}{2}[(L-m)(L-m-1)-p(L-m-1)(L-m-2)]D_{delta} \end{aligned} \tag{16}$$

where $D_{p1}(m)$ and $D_{p2}(m)$ denotes the error propagated distortion of m-th frame for mode 1 and 2 respectively.

According to (4) and knowing that for frame P_{m-1}, the coding mode of P_m does not affect its distortion, we have

$$\begin{aligned} &D_1 - D_2 = \\ &[(L-m)-p(L-m-1)][pD_{p\_primary} - p(1-p)D_{p\_redundant} - p^2(D_c + D_p')] \end{aligned} \tag{17}$$

Finally, the mode decision is determined by:

$$p[(1-p)(L-m)+p][D_{p\_primary}(m) - (1-p)D_{p\_redundant}(m) - p(D_c + D_p')] > \lambda R_{rp} \tag{18}$$

This means only if (18) holds, a redundant picture will be coded for the corresponding primary picture.

2.2.2. Lagrange multiplier selection

According to (5), the cost of each mode can be calculated as

$$Cost = D + \lambda R \tag{19}$$

Where *Cost* denotes the cost; D denotes the estimated distortion; R denotes the coding bit rate and λ is the Lagrange multiplier.

Further combining (13) and (19), and let the derivative of *Cost* to R be zero, we have

$$\lambda = -\frac{dD}{dR} = -\sum_{i=m}^{L-1}(1-p)\frac{dD_s}{dR} = (1-p)(L-m)\lambda_0 \qquad (20)$$

Where λ_0 is the error-free Lagrange multiplier.

3. SIMULATION RESULTS

In our simulations, we have integrated our approach into the Joint Model JM10.2. We coded all pictures of a sequence once, and concatenated the resulting packet stream in order to form a 4000-pictures sequence. We also used the numerically lowest constant quantizer for the whole sequence that stays within the bit rate constraints. The overhead of 40 bytes IP/UDP/RTP headers per packet was taken into account. After a bitstream was generated, a loss simulator [7] was used to simulate packet loss during transmission. For all sequences, 3%, 5%, 10%, and 20% average packet loss rate were tested. The loss-aware rate-distortion-optimized intra refresh (LA-RDO) algorithm [3] in the JM was used as the anchor, for which the bitstreams were optimized for a target packet loss rate of 5%.

We did simulations targeting at low-delay applications without feedback, and based on two scenarios: A) Coding with periodical intra-coded key pictures and B) Coding with first intra-coded and the rest inter-coded pictures. Fig. 2 and Fig. 3 show the simulation results for the two scenarios respectively. The settings for the simulations were as follows: HRP which was proposed in [4] was tested and compared to the proposed methods in this paper. For HRP and ARP, the key picture (also the start picture) of each GOP was intra-coded or inter-coded referencing the key picture of last GOP; the period of key pictures was set equal to the frame rate; the redundant picture of the key picture always used the key picture of last GOP as reference. For PEPD-ARP and RDO-ARP, all redundant pictures except the redundant picture of the key picture, used the key picture within the same GOP as reference. Other primary pictures used the latest coded primary picture as reference.

4. CONCLUDING REMARKS

Several adaptive redundant picture coding methods were presented for improved error resilience. As non-interactive error resilience tools, the methods can be applied in application scenarios wherein feedback is not used and low end-to-end delay is required. Simulation results demonstrated that the rate-distortion optimized adaptive redundant picture allocation method (RDO-ARP) outperforms the optimal loss-aware intra refresh algorithm and other redundant picture allocation methods.

ACKNOWLEDGEMENT

The work of Chunbo Zhu and Houqiang Li are partially supported by NSFC General Program under contract No. 60572067 and 863 Program under contract No. 2006AA01Z317.

REFERENCES

[1] P. Frossard, “FEC Performance in multimedia streaming,” *IEEE Communications Letters*, vol. 5, no. 3, pp. 122-124, Mar. 2001.

[2] T. Stockhammer, D. Kontopodis, and T. Wiegand, “Rate-distortion optimization for JVT/H.26L coding in packet loss environment”, *Packet Video Workshop 2002*, Pittsburgh, PY, USA, Apr. 2002.

[3] Y. Wang, S. Wenger, J. Wen, and A.K. Katsaggelos, “Error resilient video coding techniques,” *IEEE Signal Processing Magazine*, vol. 17, issue 4, pp. 61 - 82, Jul. 2000.

[4] C. Zhu, Y.-K. Wang, M.M. Hannuksela, and H. Li, “Error Resilient Video Coding Using Redundant Pictures,” presented on *International Conference on Image Processing* (ICIP), Atlanta, USA, Oct. 2006.

[5] C. Zhu, Y.-K. Wang, and H. Li, “Adaptive redundant picture coding,” JVT-U114, Oct. 2006.

[6] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, “Error resilient video coding in H.264 encoder with potential distortion tracking,” *International Conference on Image Processing* (ICIP), vol. 1, pp.163 – 166, Oct. 2004.

[7] Y. Guo, H. Li and Y.-K. Wang, “SVC/AVC loss simulator donation”, JVT-Q069, Oct. 2005.

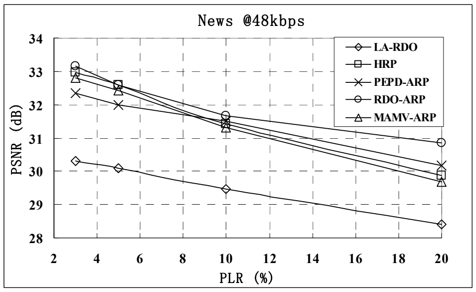

Fig. 2. PSNR vs. packet loss rate (PLR) curves of “*News*” at 48kbps, 10fps and QCIF resolution for scenario A.

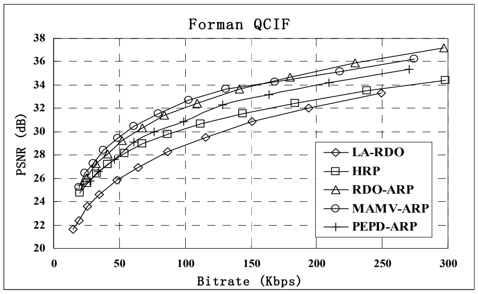

Fig. 3. PSNR vs. bitrate curves of “*Foreman*” at 20% packet loss rate, 7.5 fps and QCIF resolution for scenario B.