Joint Data Partition and Rate-Distortion Optimized Mode Selection for H.264 Error-Resilient Coding

Yuan Zhang
Department of TV Engineering
Communication University of China
yzhang@cuc.edu.cn

Wen Gao
Graduate School
Chinese Academy of Sciences
wgao@jdl.ac.cn

Debin Zhao
School of Computer Science, Harbin
Institiue of Technology, Harbin, China
dbzhao@jdl.ac.cn

Abstract—**Data partitioning (DP) is an efficient error-resilient video coding tool. Its contribution to performance improvement in the error-prone environment arises from the superior error concealment mechanisms that are available with the help of protected data partitions. Since error-concealment in terms of DP is closely related to coding mode, it is desirable to have an optimized coding mode selection scheme. However, the existing coding mode selection techniques usually assume that the same error-concealment mechanism is used for a block when it is lost, and the associated distortion also remains the same. Obviously, this assumption is not true when DP involves. In this paper, a generalized end-to-end distortion model is proposed for the rate-distortion optimized coding mode selection, which fully utilizes the superior error-concealment mechanism in terms of DP. The proposed distortion model is also advantageous in the suppression of approximation errors caused by pixel average operations such as sub-pixel interpolation and deblocking filter. Therefore, it can lead to a low-complexity solution for real-time applications such as live streaming.**

Keywords—Error resilience; rate distortion optimization; data partitioning; H.264/MPEG-4 AVC

Topic area—Multimedia Processing

I. INTRODUCTION

With the rapid development of network technology, the bandwidth is no longer the bottleneck of real-time video transmission. However, without the guarantee of QoS, the packet loss is still inevitable, which may result in severe quality degradation due to the error propagation along the hybrid-coded video [1]. Error-resilient video coding tools are usually used to remove or reduce the error propagation [2]. Typically, error-resilient coding tools are closely related to the video codec. H.264 is the up-to-date video coding standard, which provides not only the open structure for high coding efficiency, but also a number of error resilient features such as data partitioning (DP).

DP provides the ability to separate more important and less important syntax elements into different packets of data, and enables the application of unequal error protection (UEP) and other types of improvement of error/loss robustness [3]. It is clear that the performance improvement arises from the superior error concealment mechanisms that are available with the help of protected data partitions. In fact, the error propagation will not be significantly reduced without intra refreshment. Therefore words, it is more desirable to jointly utilize DP with intra refreshment in error-resilient video coding. Then, a further problem arises here. How to select the optimum coding mode when considering the DP features?

In the past years, many rate-distortion (R-D) optimized coding mode selection algorithms have been proposed [4]-[8]. ROPE is the most recognized one, which estimates the expected sample distortion by keeping track of the first- and second- order moments of the reconstructed pixel value at decoder side. However, since ROPE is very sensitive to the approximation errors, it has to perform the intensive computing to guarantee the accuracy when pixel-averaging operations (e.g. sub-pixel motion-compensated prediction) involved [8]. Actually, an error-robust rate-distortion optimization (ER-RDO) method has been adopted in the H.264 reference software [9], in which the distortion of an MB is computed as the average over the K distortions by decoding this MB K times based on the erroneous reference frames. It is no doubt that ER-RDO also leads to very high complexity when simulating multiple decoders in encoder.

Nevertheless, DP is seldom considered in the existing R-D optimized mode selection schemes. These RDO techniques usually assume that the same error-concealment scheme is used for an MB when it is lost, and the associated distortion estimated at the encoder also remains the same regardless of the coding mode. However, this assumption is not true when DP involves, and accordingly, the estimated end-to-end distortion is not accurate enough. Previously, we have proposed an end-to-end distortion model by taking the overall distortion as the sum of source, error-propagated and error-concealment distortion items, which can suppress the approximation errors when the pixel average operations involve [10]. In this paper, we further extend it to be a generalized end-to-end distortion model including the consideration of DP.

The rest of this paper is organized as follows. In Section 2, we describe the proposed end-to-end distortion model in detail. In Section III, we present the proposed R-D optimized mode selection algorithm. The experimental results are presented in Section IV. Finally, Section V concludes this paper.

II. End-to-End Distortion with Data Partition

In the proposed model, we assume that three data partitions *A*, *B* and *C* are used. *A* contains the header information such as MB types, quantization parameters, and motion vectors, which are more important than the remaining slice data. *B* contains transform coefficients of the intra-coded blocks, which can stop the further error propagation. *C* contains coefficients of the inter-coded blocks. Compared to *A* and *B*, *C* is less important. Nevertheless, both *B* and *C* depend on *A*. When *A* is lost, *B* and *C* become useless. On the other hand, when *B* and *C* are lost, the available header information (e.g., MB types and motion vectors from DP *A*) can still be used to in error concealment.

Now, we define some notations used in the derivation of the proposed end-to-end distortion model. Let f_n^i be the original value of pixel *i* in frame *n*, and let $\hat{f}_n^i$ and $\tilde{f}_n^i$ be the reconstructed values in the encoder and decoder, respectively. Let $\hat{r}_n^i$ be the reconstructed residue in the encoder, i.e. $\hat{f}_n^i = \hat{f}_{ref}^j + \hat{r}_n^i$, when it references pixel *j* in frame *ref*. Suppose the transmission error rates of partitions *A*, *B* and *C* are known as p_A, p_B and p_C, respectively.

For the *i*th pixel in an intra block, when *A* or *B* is lost, the decoder copies from the pixel in the previous frame in frame *n*-1 at the same spatial position. Then, we can represent $\tilde{f}_n^i$ as:

$$\tilde{f}_n^i = \begin{cases} \hat{f}_n^i & w.p. \quad (1-p_A)(1-p_B) \\ \tilde{f}_{n-1}^i & w.p. \quad (1-p_A)p_B \\ \tilde{f}_{n-1}^i & w.p. \quad p_A \end{cases}, \qquad (1)$$

Then, we can derive the expectation of end-to-end distortion in the decoder to be:

$$\begin{aligned} d(n,i) &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (1-p_A)(1-p_B)E\{(f_n^i - \hat{f}_n^i)^2\} + (1-p_A)p_B E\{(f_n^i - \tilde{f}_{n-1}^i)^2\} + p_A E\{(f_n^i - \tilde{f}_{n-1}^i)^2\} \\ &= (1-p_A)(1-p_B)E\{(f_n^i - \hat{f}_n^i)^2\} \\ &\quad + (1-p_A)p_B(E\{(f_n^i - \hat{f}_{n-1}^i)^2\} + E\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\}) \\ &\quad + p_A(E\{(f_n^i - \hat{f}_{n-1}^i)^2\} + E\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\}) \\ &= (1-p_A)(1-p_B)d_s + ((1-p_A)p_B + p_A)(d_{ec\_prev\_o} + d_{ep\_prev}) \end{aligned} \qquad (2)$$

where d_s denotes the source distortion, $d_{ec\_prev\_o}$ indicates the mean square error (MSE) between the original and error-concealment values in the encoder, representing the original previous error-concealment distortion. d_{ep_prev} denotes the error-propagated distortion from previous frame. The third equality in (2) bases on the assumption that effects of source distortion in the encoder and error-propagated distortion in the decoder are additive.

For the *i*th pixel in inter mode, we assume that in the case of *A* is lost, the decoder copies from pixel *i* in frame *n*-1. In the case of *A* is received but *C* is lost, the decoder copies from pixel *k* in the reference frame using the correct motion vector from *A*. Then, we can represent $\tilde{f}_n^i$ as:

$$\tilde{f}_n^i = \begin{cases} \hat{r}_n^i + \tilde{f}_{ref}^k & w.p. \quad (1-p_A)(1-p_C) \\ \tilde{f}_{ref}^k & w.p. \quad (1-p_A)p_C \\ \tilde{f}_{n-1}^i & w.p. \quad p_A \end{cases}, \qquad (3)$$

Then, the expectation of end-to-end distortion in the decoder is:

$$\begin{aligned} d(n,i) &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (1-p_A)(1-p_C)E\{(f_n^i - (\hat{r}_n^i + \tilde{f}_{ref}^k))^2\} + (1-p_A)p_C E\{(f_n^i - \tilde{f}_{ref}^k)^2\} \\ &\quad + p_A E\{(f_n^i - \tilde{f}_{n-1}^i)^2\} \\ &= (1-p_A)(1-p_C)(E\{(f_n^i - \hat{f}_n^i)^2\} + E\{(\hat{f}_{ref}^k - \tilde{f}_{ref}^k)^2\}) \\ &\quad + (1-p_A)p_C(E\{(f_n^i - \hat{f}_{ref}^k)^2\} + E\{(\hat{f}_{ref}^k - \tilde{f}_{ref}^k)^2\}) \\ &\quad + p_A(E\{(f_n^i - \hat{f}_{n-1}^i)^2\} + E\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\}) \\ &= (1-p_A)(1-p_C)(d_s + d_{ep\_ref}) + (1-p_A)p_C(d_{ec\_ref\_o} + d_{ep\_ref}) \\ &\quad + p_A(d_{ec\_prev\_o} + d_{ep\_prev}) \end{aligned} \qquad (4)$$

where d_{ep_ref} denote the error-propagated distortion from the reference frame. $d_{ec\_ref\_o}$ indicates the original referenced error-concealment distortion for the inter mode. The third equality in (4) also bases on the assumption that the effects of original error-concealment distortion in the encoder and error-propagated distortion in the decoder are additive.

It can be seen that the end-to-end distortion is always the sum of three distortion items, including the distortion induced when the information related to the selected mode are received correctly, the distortion induced when only A received, and the distortion induced when A is lost. Note that the source distortion and error-concealment distortions are readily calculated. The remained problem is how to calculate the error-propagated distortion. Without losing the generality, we derive the formula to calculate d_{ep} as follows. For the *i*th pixel in terms of an intra mode, d_{ep} is:

$$\begin{aligned} d_{ep} &= E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\} \\ &= (1-p_A)p_B E\{(\hat{f}_n^i - \tilde{f}_{n-1}^i)^2\} + p_A E\{(\hat{f}_n^i - \tilde{f}_{n-1}^i)^2\} \\ &= ((1-p_A)p_B + p_A)(E\{(\hat{f}_n^i - \hat{f}_{n-1}^i)^2\} + E\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\}) \\ &= ((1-p_A)p_B + p_A)(d_{ec\_prev\_r} + d_{ep\_prev}) \end{aligned}, \qquad (5)$$

where $d_{e\_prev\_r}$ indicates the MSE between the reconstructed and error-concealment values in the encoder, representing the reconstructed previous error-concealment distortion.

For the *i*th pixel in inter mode, the d_{ep} is:

$$\begin{aligned} d_{ep} &= E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\} \\ &= (1-p_A)(1-p_C)E\{(\hat{f}_{ref}^k - \tilde{f}_{ref}^k)^2\} + (1-p_A)p_C E\{(\hat{f}_n^i - \tilde{f}_{ref}^k)^2\} \\ &\quad + p_A E\{(\hat{f}_n^i - \tilde{f}_{n-1}^i)^2\} \\ &= (1-p_A)(1-p_C)E\{(\hat{f}_{ref}^k - \tilde{f}_{ref}^k)^2\} \\ &\quad + (1-p_A)p_C(E\{(\hat{f}_n^i - \hat{f}_{ref}^k)^2 + E\{(\hat{f}_{ref}^k - \tilde{f}_{ref}^k)^2\}) \\ &\quad + p_A(E\{(\hat{f}_n^i - \hat{f}_{n-1}^i)^2\} + E\{(\hat{f}_{n-1}^i - \tilde{f}_{n-1}^i)^2\}) \\ &= (1-p_A)(1-p_C)d_{ep\_ref} + (1-p_A)p_C(d_{ec\_ref\_r} + d_{ep\_ref}) \\ &\quad + p_A(d_{ec\_prev\_r} + d_{ep\_prev}) \end{aligned} , \qquad (6)$$

where $d_{ec\_ref\_r}(n, i)$ indicates the reconstructed referenced error-concealment distortion. Since the reconstructed error-concealment distortion can also be readily calculated, the calculation of d_{ep} only depends on the availability of the error-propagated distortions from its previous frames. Note that d_{ep} of the first frame that is typically coded as an intra frame can be directly derived without considering the error propagation. Hence d_{ep} of the following frames can be recursively calculated frame by frame. In particular, the end-to-end distortion of the current frame is first calculated by referencing the error-propagated distortions of the previous frames. And then, the error-propagated distortion of the current frame is updated as well and stored as a distortion map. The impact of future error propagation is considered through the distortion map directly.

III. R-D Optimized Coding Mode Selection

The hybrid video coding usually contains a number of coding modes in the MB coding. The coding mode in H.264 can vary from the block partition of 4x4 to the whole block of 16x16 with respect to the different prediction types. Besides, the multiple references structure in H.264 also increases the coding options in the macroblock coding. Assume o denotes a candidate coding option that is the combination of coding mode and reference frame. The best coding option of macroblock m in frame n can be selected as the one having the minimum coding cost $J(n, m, o)$ throughout the candidate coding options, where

$$J(n,m,o) = D(n,m,o) + \lambda R(n,m,o) \ . \qquad (7)$$

Note that data partition is always used with unequal error protection (UEP). Therefore, the rate should involve the redundant bits used by error protection. Suppose the channel error rate p_A, p_B and p_C are known as a priori in the encoder. According to the (2) and (3), the overall end-to-end distortion of a macroblock can be defined as the sum of distortions of all contained pixels. In the proposed distortion model, the recursive calculation of the end-to-end distortion requires the definition of distortion map for the storage of error-propagated distortions at each frame. Besides the number of distortion maps that depends on the number of reference frames, the resolution of the distortion map is also a practical problem, which can be at either pixel level or block level. In this paper, we prefer the block-level solution due to the following reasons. On the one hand, the block-level implementation can reduce the computing complexity and memory cost. On the other hand, it can also increase the robustness against the effects of sub-pixel motion-compensated prediction (MCP).

In particular, we propose that the element in the distortion map corresponds to the minimum block size in MCP (i.e. 4x4 block in H.264). Suppose block m in frame n references block m_j in frame ref. Since m_j may not always correspond to a single element in the distortion map due to sub-pixel MCP, we derive D_{ep_ref} by weight-averaging the error-propagated distortions of the overlapped blocks. Specifically, we have:

$$D_{ep\_ref}(n,m) = \sum_{l=1}^{4} w_l D_{ep}(ref, m_l) \ , \qquad (8)$$

where w_l indicates the ratio that the overlapped region between m_l and m_j.

Since the error-concealed distortion introduced when DP A of the current MB is lost is the same for intra and inter mode, it is unnecessary to be calculated in mode selection. In the case of intra mode, only the original previous error-concealed distortion should be calculated, while the error-propagated distortion from previous frame can be derived from the stored distortion map directly. In the case of inter mode, the additive compute complexity come from the calculation of d_{ep_ref} and the original referenced error-propagated distortion. After the current frame is encoded, the corresponding distortion map is derived according to (5) and (6) for the coding of future frames. Since the calculation is based on the block level, the proposed algorithm only introduces very little extra computing complexity.

IV. Experimental Resultes

Extensive experiments have been carried out to verify the performance of the proposed algorithm. The testing platform is the H.264 reference software JM9.5. We compare the proposed method (Exp1) with three other coding schemes, including ER-RDO without data partition (Exp2), our previous model [10] but without data partition (Exp3), and data partition without intra refreshment (Exp4). In particular, for ER-RDO without data partition, it is good enough to operate about K =30 decoders in the encoder [9].

The testing results of two sequences, Foreman (144kbps, 7.5fps, QCIF) and Coastguard (384kbps, 15fps, CIF) are reported in this paper. Only the first frame is encoded as an I frame, and the left frames are encoded as P frames. There are 4 packets per frame for QCIF and 9 packets for CIF. It is well known that DP should be jointly used with unequal error protection. In our test, it is achieved by sending the

RTP packet with the header partition twice (for the 3% loss rate cases) or three times (for the 5%, 10%, and 20% loss rate cased). Of course, the rate control should be modified so that the complete packet stream, including the multiple header partitions, fits into the bit rate budget. The 40 bytes of IP/UDP/RTP headers per packet have been taken into account. These bitstreams are decoded after simulating the packet loss under the loss rates 3%, 5%, 10% and 20%. The packet loss situation is simulated according to the error resilience testing conditions specified in [11]. The bitstreams are decoded multiple times. The number of decoding runs is selected to have totally at least 8000 packets for each sequence in our test.

Fig. 1 and Fig. 2 show the PSNR results for the two testing sequences under the different packet loss rates. It can be seen that the proposed algorithm outperforms the other three algorithms almost in all cases. However, for Coastguard, it is a little bit worse than our previous algorithm without data partitioning at packet loss rate of 5%. As we know, in the case of packet loss rate of 5%, A is transmitted three times in our test. The high redundant rate due to the UEP of A will more or less scarify the coding efficiency. Furthermore, the proposed algorithm only adds less than 5% extra complexity to the original encoder.

V. Conclusion and Future Work

In this paper, we have presented a generalized end-to-end distortion model for rate-distortion optimized coding mode selection when data partitioning involves. The proposed distortion model takes the overall distortion as the sum of source, error-propagated and error-concealment distortion items. Each distortion item is derived in terms of the superior error concealment mechanisms that are available with the help of protected data partitions. Moreover, the proposed end-to-end distortion model can suppress the approximation errors caused by pixel average operations, which leads to the low-complexity solution for real-time applications. Nevertheless, there still remains some future work, e.g. the joint source-channel rate allocation for the optimized overall transmission efficiency.

Acknowledgement

This work is supported by opening project of Beijing Key Laboratory of Multimedia and Intelligent Software in Beijing University of Technology.

References

[1] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1012-1032, June 2000.

[2] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, pp. 974-997, May 1998.

[3] S.Wenger, "H.264/AVC over IP", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 645-656, July 2003.

[4] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 966-976, June 2000.

[5] G. Cote, S. Shirani, F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952-965, June 2000.

[6] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y.-Q. Zhang and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: Theory and application," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 977-995, June 2000.

[7]. Z. H. He, J. F. Cai, C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 511-523, June 2002.

[8] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation", in *Proc. ICIP*, Barcelona, Spain, Sep. 2003, pp. 469-472

[9] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," in *Proc. PVW*, Pittburgh, PY, Apr.2002

[10] Y. Zhang, W. Gao, H.F. Sun, Q.M. Huang, Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proc. ICIP*, Singapore, Oct. 2004, pp. 163-166

[11] S. Wenger, "Common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing," ITU-T VCEG document VCEG-N79r1, Sep. 2001.

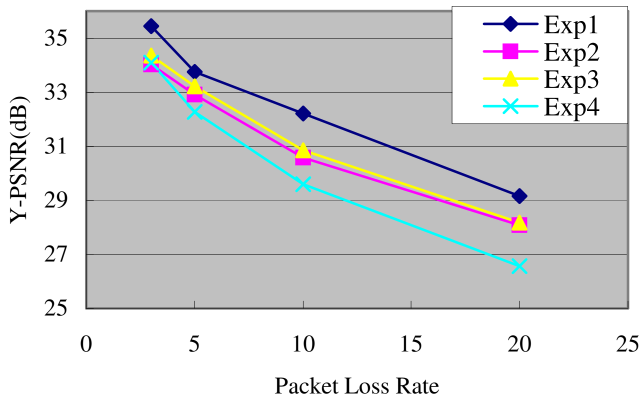

Fig. 1 Simulation results for the Foreman sequence.

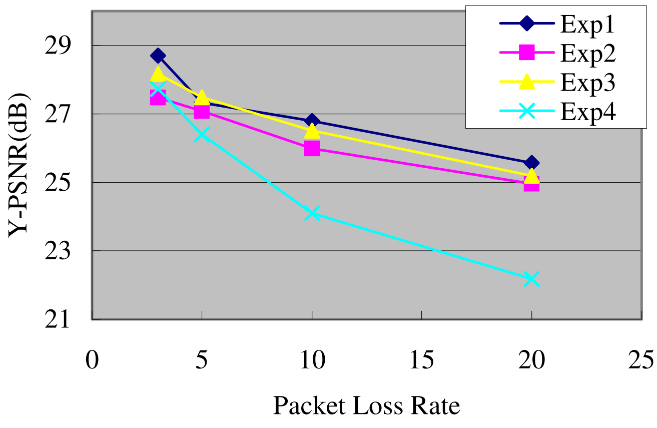

Fig. 2 Simulation results for the Coastguard sequence.