ERROR RESILIENCE VIDEO CODING IN H.264 ENCODER WITH POTENTIAL DISTORTION TRACKING

Yuan Zhang[1, 2], *Wen Gao*[1], *Huifang Sun*[3], *Qingming Huang*[1], *Yan Lu*[4]

[1] Graduate School, Chinese Academy of Sciences, Beijing 100080, China
[2] Department of TV Engineering, Beijing Broadcasting Institute, Beijing 100024, China
[3] MERL - Mitsubishi Electric Research Laboratories, Murray Hill, NJ
[4] Microsoft Research Asia, Beijing 100080, China

ABSTRACT

In this paper, an efficient rate-distortion (RD) model for H.264 video encoder in packet loss environment is presented. The encoder keeps tracking the potential error propagation on a block basis by taking into account the source characteristics, network conditions as well as the error concealment method. The end-to-end distortion invoked in this RD model is estimated according to the potential error-propagated distortion stored in a distortion map. The distortion map in terms of each frame is derived after the frame is encoded, which can be used for the RD-based encoding of the subsequent frames. Since the channel distortion has been considered in the proposed RD model, the new Lagarangian parameter is derived accordingly. The proposed method outperforms the error robust rate-distortion optimization method in H.264 test model better in terms of both transmission efficiency and computational complexity.

1. INTRODUCTION

The transmission of compressed video over the existing packet-switched networks presents many new challenges, e.g. how to tackle the problem caused by packet loss. In the packet-loss environment, transmitting the hybrid-coded video may suffer from the error propagation and lead to the well-known drifting phenomenon. To tackle this problem, error resilience video coding tools and error concealment algorithms have been devised at the encoder and decoder, respectively. These two types of methods are usually jointly used in the transmission of the hybrid-coded video over packet-loss networks.

In standard-compliant techniques, intra coding can suppress the error propagation at the cost of reducing the coding efficiency. The main problem of inserting more intra-coded macroblocks (MB) in the encoder to make the bitstream more resilient to potential errors is on how to optimize the trade-off between the coding efficiency and the suppression of potential errors. Towards this goal, many researches have been done on the topic of adaptively inserting intra blocks in a coded frame/sequence.

The earliest algorithm was developed to randomly place intra MBs [1]. Then, the content-adaptive coding mode selection scheme was proposed to intra-code the MBs at regions with high activity [2]. To further improve the performance, rate-distortion optimization (RDO) algorithms have been proposed, wherein the main problem arises from the end-to-end distortion estimation. In [3], a simple block-based distortion estimation model has been proposed, which is not accurate enough because it does not consider the potential errors from all previous frames. In [4] and [5], the more accurate methods have been proposed to estimate the end-to-end distortion at pixel level precision. However, both methods have high computational complexity and implementation cost.

H.264 is the most up-to-date video coding standard. One of its main applications targets at the error prone network video streaming. An error robust rate distortion optimization (ER-RDO) method, has been developed in the H.264 test model for video coding in packet loss environment [6][7]. The decoded MB distortion is computed as the average over the K distortions by decoding this MB K times based on the erroneous reference frames. The expected decoder distortion can be estimated accurately in the encoder if K is chosen large enough. However, the high computational complexity and make it impractical when K is increased very large.

To tackle these problems, in this paper, we propose a novel RD model for H.264 video encoding in packet loss environment. The major contribution is that the end-to-end distortion is estimated in a more reasonable way. Briefly, a block-based distortion map is defined to store the potential errors of the current frame that may propagate to the future frames. In other words, when coding the current frame, the potential channel distortions of its reference frames have been known as a priori. The potential distortion of a block is defined by jointly considering the potential distortion of its reference frame (that is zero for intra block) and the distortion of the current error-concealed block. In addition, how to select

0-7803-8554-3/04/$20.00 ©2004 IEEE.

the appropriate Lagrange parameter of the RD model is also discussed.

The rest of this paper is organized as follows. In section 2, we describe the proposed rate-distortion model in detail, including the end-to-end distortion estimation, the derivation of the Lagrange parameter and the implementation of the proposed RD model in H.264 encoder. In section 3, the simulated results are presented. Finally, section 4 concludes this paper.

2. PROPOSED RATE-DISTORTION MODEL

In the hybrid video coding systems, the coding process is usually performed on the MB basis. For coding a MB, the key point is to select a proper coding option composed of coding mode and reference frame (if multiple reference frames are used). Suppose *O* denotes the set of all selectable coding options for an MB. In the case of error-free environment, for coding MB *m* in frame *n*, we choose the coding option o^* that minimizes the cost given by:

$$J(n,m,o) = D(n,m,o) + \lambda R(n,m,o)\,, \tag{1}$$

where $D(n,m,o)$ and $R(n,m,o)$ denote the source distortion and the rate in terms of the coding option *o*, respectively, and λ denotes the Lagrange parameter.

Concretely, after computing the distortion and the rate with coding option *o* in terms of Lagrange parameter λ, we choose the coding option o^* that yields the best rate-distortion tradeoff for that MB. Parameter λ reveals the relation between the rate and distortion. Therefore, the choice of the Lagrange parameter becomes a key point in RD-based video encoding. In the error-free environment, it is closely tied to the quantization parameter Q via the formula [8]:

$$\lambda = \begin{cases} 0.85 \cdot Q^2, & for\ H.263 \\ 0.85 \cdot 2^{Q/3}, & for\ H.264 \end{cases}. \tag{2}$$

However, in the error-prone environment, both the potential channel distortion and the source distortion should be taken into account to achieve the best transmission efficiency. Therefore, how to estimate the end-to-end distortion also remains a major issue in the RD-based video encoding for the packet-loss networks. Since the potential channel distortion is also considered, the relation between the rate and the overall distortion changes as well. Accordingly, the Lagrange parameter λ should be properly selected. The following subsections will first describe the proposed RD model with end-to-end distortion estimation, and then derive the new Lagrange parameter for the proposed RD model.

2.1. End-to-end distortion estimation

The distortion of the current MB under consideration consists of two parts. Assuming that the MB is correctly received, for the intra-coded MB, only the source distortion due to the quantization errors contributes to the overall distortion; for the inter-coded MB, besides the source distortion, the channel distortion due to the error propagation from the previous frames may also contribute to the overall distortion. In the latter case, the overall distortion of the correctly received inter-coded MB is the quantization distortion plus the potential error-propagated distortion of the blocks from which it is predicted. The latter one is referred to as the referenced error-propagated distortion hereafter. Assuming that the current MB is lost, the distortion is the error-concealed distortion that includes the potential error-propagated distortion of the blocks from which the current MB is concealed.

Thus, the overall distortion of the *m*th MB in the *n*th frame with the candidate coding option *o* is represented by:

$$\begin{aligned} D(n,m,o) = {} & (1-p)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) \\ & + pD_{ec}(n,m) \end{aligned}\,, \tag{3}$$

where *p* denotes the packet loss rate; $D_s(n,m,o)$ and $D_{ep\_ref}(n,m,o)$ denote the source and the referenced potential error-propagated distortions, respectively; and $D_{ec}(n,m)$ denote the error-concealed distortion in case that the current MB is lost. Since $D_{ec}(n,m)$ is independent of the coding option, it is unnecessary to be calculated in the coding mode selection. The calculation of $D_{ep\_ref}(n,m,o)$ is not straightforward. The key point is about how to estimate the potential error-propagated distortion.

To tackle the above problem, we define a distortion map D_{ep} for each frame on a block basis (e.g. 4x4). The first frame is intra-coded without considering the error propagation. Therefore, the distortion map of the first frame can be derived directly. For coding the following frames, the distortion maps of the previous frames indicating the possible influence of the propagated errors are referenced to select the proper coding mode of each MB. After coding one frame, the distortion map is derived as well. The following presents how to derive $D_{ep\_ref}(n,m,o)$ according to the distortion map.

The referenced error-propagated distortion of the current MB is computed by weighted averaging the potential error-propagated distortion of the surrounding blocks that overlap with the motion-compensated blocks in the reference frame(s), as illustrated in Fig. 1. In detail, the referenced error-propagated distortion $D_{ep\_ref}(n,m,o)$ is calculated by:

$$\begin{aligned} D_{ep\_ref}(n,m,o) &= \sum_{k=1}^{K} D_{ep\_ref}(n,m,k,o) \\ &= \sum_{k=1}^{K} \sum_{l=1}^{4} w_l D_{ep}(n_l,m_l,k_l,o) \end{aligned}\,, \tag{4}$$

where $D_{ep\_ref}(n,m,k,o)$ denotes the referenced potential error-propagated distortion of the *k*th block in the current MB. As shown in the (4), $D_{ep\_ref}(n,m,k,o)$ is calculated by weighed averaging the potential error-propagated distortion $D_{ep}(n_l,m_l,k_l,o_l)$ of the block k_l in the MB m_l of the reference frame n_l. The weight w_l is the proportion ratio of the overlap area. Since there is no reference frame for intra-coded MB, the propagated errors from previous frames can be suppressed, and therefore $D_{ep\_ref}(n,m,o)$ in terms of the intra mode equals to zero.

After the current frame is coded, the associated distortion map D_{ep} is then derived. For each block in the current frame, $D_{ep}(n,m,k)$ with the selected optimal coding mode $o*$ is computed by:

$$\begin{aligned} D_{ep}(n,m,k) = (1-p)D_{ep\_ref}(n,m,k,o*) + \\ p(D_{ec\_rec}(n,m,k,o*) + D_{ec\_ep}(n,m,k)) \end{aligned} . \quad (5)$$

The latter term composed of D_{ec_rec} and D_{ec_ep} denotes the error-concealed distortion that would propagate to the following frames. In detail, $D_{ec\_rec}(n,m,k,o*)$ is the distortion between the error-concealed block and the reconstructed block in terms of coding option $o*$. The reconstructed block instead of the original one is used because only the distortion relative to the reconstructed block will propagate to the subsequent frames.

The calculation of $D_{ec\_ep}(n,m)$ depends on the employed error concealment method at the decoder side. In the H.264 non-normative decoder, the lost MB is reconstructed by copying some blocks from the previous frames. In this case, $D_{ec\_ep}(n,m)$ can be derived from the potential error-propagated distortions of these blocks. Notice that the distortion map of the first frame is only related to D_{ec_rec}. After the distortion map is derived, it is stored for the encoding of the following frames.

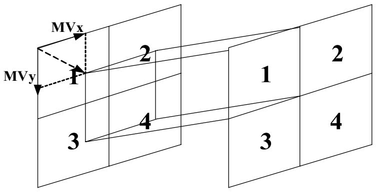

Fig.1. Overlapped blocks in terms of inter prediction.

2.2. Lagrange parameter decision

In this subsection, we present how to select a proper Lagrange parameter for the proposed RD model in packet loss environment. Based on the selection of the Lagrange parameter in the error-free environment in [8], we derive the new Lagrange parameter as follows.

If assuming high-resolution quantization, it is well known that the source distortion $D_s(R)$ conform to:

$$D_S(R) = \beta \bullet 2^{-\alpha R}, \quad (6)$$

where β is a constant depending on the variance of the source. Further assuming that the distortion-to-quantizer relation is at sufficiently high rates, the source probability distribution can be approximated as uniform within each quantization interval:

$$D_s(\Delta) = \frac{\Delta^2}{12}. \quad (7)$$

Combining (6) and (7), we obtain:

$$R(\Delta) = \frac{1}{\alpha}\log_2\left(\frac{\beta}{D_s(\Delta)}\right) = \frac{1}{\alpha}\log_2\left(\frac{\beta}{\Delta^2/12}\right). \quad (8)$$

According to the (3) and (7), we obtain:

$$D(\Delta) = (1-p)(\frac{\Delta^2}{12} + D_{ep\_ref}) + pD_{ec}. \quad (9)$$

Further combining the derivatives for Δ in the (8) and (9), we obtain the new Lagrange parameter

$$\lambda' = -\frac{dD(R)}{dR} = -\frac{dD}{d\Delta}\frac{d\Delta}{dR} = (1-p)\frac{\alpha \ln 2}{12}\Delta^2 = (1-p)\lambda, \quad (10)$$

where λ is the Lagrange parameter in the error-free environment defined in the (2).

2.3. Implementation in the H.264 encoder

In this subsection, we present the proposed RD-based coding for H.264 encoder in the packet-loss environment. Since the channel distortion of the B frame would not propagate to the following P frames, it is unnecessary to store the potential error-propagated distortion for the B frame to. The coding mode selection of B frames can be the same as that in the P frames. For simplicity, we assume that B frames are not used. In H.264, the coding mode of P frames is selected to be one of the 2 intra modes and 8 inter modes. In terms of inter mode, the reference can be from one of the several previous frames. We assume a simple error concealment scheme at the decoder side is used. If a MB is lost, the decoder simply copies the co-located MB in the previous decoded frame.

Notice that the distortions introduced when the current MB is lost are independent of the coding option. Supposing the packet loss rate p is known at the encoder side, according to the (1), (3) and (10), the coding option can be selected with:

$$\begin{aligned} o*(n,m) &= \arg\min_{o\in O}((1-p)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) + \\ &\quad pD_{ec}(n,m) + (1-p)\lambda R) \\ &= \arg\min_{o\in O}(D_s(n,m,o) + D_{ep\_ref}(n,m,o) + \lambda R) \end{aligned} . \quad (11)$$

After the current frame is encoded, the distortion map is derived according to the (5) for instructing the encoding of the future frames.

3. SIMULATED RESULTS

Some experiments have been carried out to verify the performance of the proposed algorithm. The testing platform is the H.264 reference software JM7.3. The default ER-RDO algorithm is taken as a reference in the comparison. According to [6], it is good enough to operate about K=30 decoders in the encoder for ER-RDO. Two sequences, i.e. Foreman in QCIF format at 15 frames per second (fps) and Paris in the CIF format at 30fps, are used. Only the first frame is encoded as I frame, and the left frames are encoded as P frames. Each packet contains one row of MBs in transmission, e.g. 11 MBs in QCIF format. The packet loss situation is simulated according to the error resilience testing conditions specified in [9].

Fig. 2 shows the YPSNR-Bitrate curves under the packet loss rate 10% and 20%, respectively. The results show that the proposed algorithm outperforms ER-RDO in terms of transmission efficiency in all cases. In other words, the proposed RD model works more stable rather than the statistics from 30 times decoding. Furthermore, compared to the ER-RDO algorithm that adds 30 times decoding to the encoder, the proposed algorithm introduces much less extra computational complexity, which makes it more practical in real applications.

4. CONCLUSION

In this paper, we have presented an efficient rate-distortion model for H.264 video encoder in the packet loss environment. The encoder keeps tracking the distortion on a block basis while taking into account the source characteristics, network conditions as well as the error concealment method. The proposed model reveals the inherent relationship between the potential error-propagated distortion and the characteristics of the input source video data. Compared to the error robust rate-distortion optimization method in H.264 test model, the proposed model performs better in terms of both transmission efficiency and computational complexity. Furthermore, although this algorithm is proposed for H.264 encoder, it is also feasible to be used in other standard-compliant video encoders.

5. ACKNOWLEDGEMENT

This research has been supported by NSFC (60333020), 973 Program of China (2001CCA03300), 863 Program of China (2002AA119010) and Bairen Project of CAS.

REFERENCES

[1] G. Cote and F. Kossentini, “Optimal intra coding of blocks for robust video communication over the Internet,” in Image Commun. Sept. 1999, pp.25-34

[2] P. Haskell and D.Messerschmitt, “Resynchronization of Motion-Compensated Video Affected by ATM Cell Loss,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, volume 3, 1992, pp.545-548

[3] G. Cote, S. Shirani, F. Kossentini, “Optimal mode selection and synchronization for robust video communications over error-prone networks,” in IEEE Journal on Selected Areas in Communications, Vol. 18, No. 6, June 2000, pp. 952-965

[4] D. Wu, et al, “An end-to-end approach for optimal mode selection in Internet video communication: Theory and application,” in IEEE Journal on Selected Areas in Communications, Vol. 18, No. 6, June 2000, pp.977-995

[5] R. Zhang, S. L. Regunathan, and K. Rose, “Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience,” in IEEE Journal on Selected Areas in Communications, Vol. 18, No. 6, June 2000, pp. 966-976

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

[7] ITU-R Rec. H.264 | ISO/IEC 14496-10 AVC Draft Text, Joint Video Team document JVT-E146D37, Oct. 2002

[8] T. Wiegand and B. Girod, “Lagrangian multiplier selection in hybrid video coder control,” in Proc. ICIP2001, Thessaloniki, Greece, October 2001.

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

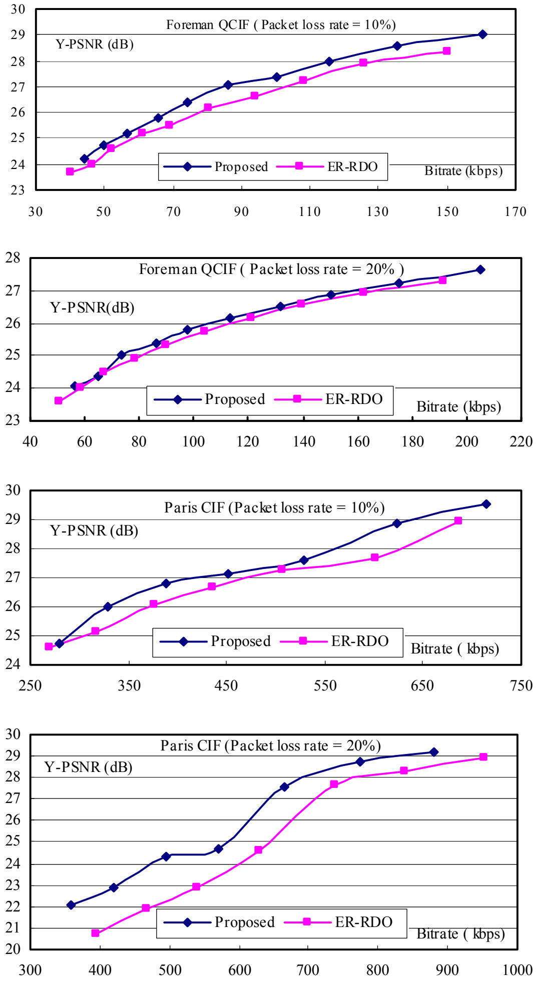

Fig. 2. Experimental results on Foreman and Paris sequences.