Optimum End-to-End Distortion Estimation for Error Resilient Video Coding

Yuan Zhang[1,2], Qingming Huang[1], Yan Lu[3], and Wen Gao[1]

[1] Graduate School, Chinese Academy of Sciences, 100080 Beijing, China
{yzhang, qmhuang, wgao}@jdl.ac.cn
[2] Beijing Broadcasting Institute, 100024 Beijing, China
[3] Microsoft Research Asia, 100080 Beijing, China
t-yanlu@microsoft.com

Abstract. End-to-end distortion estimation plays an important role in error-resilient video coding. The intuitive method is to simulate the decoding process many times at the encoder, as used in the H.264 test model. However, the computing complexity is very high. In this paper, an optimum end-to-end distortion estimation scheme is proposed. Concretely, the correlations of the potentially propagated errors of the different frames are modeled based on the theoretical analysis, and then a block-based potential distortion tracking scheme with very low computing complexity is proposed. Statistics show that the gaps of the estimated distortions between the proposed algorithm and the H.264 test model become smaller and smaller when the times of simulated decoding in H.264 test model increases. In other words, the proposed algorithm is more accurate than H.264 test model. Moreover, an improved rate-distortion optimization algorithm based on the optimum end-to-end distortion estimation is proposed, wherein the rate control is also jointly utilized.

1 Introduction

Transmitting the hybrid-coded video over the packet-switched networks with packet losses often suffers from the error propagation and leads to the well-known drifting phenomenon [1]. To tackle this problem, error-resilient video coding and error concealment algorithms have been devised at the encoder and the decoder, respectively [2]. Intra block refreshment is one of the most popular error-resilient coding schemes. In standard-compliant techniques, intra coding can suppress the error propagation at the cost of reduced coding efficiency. The main problem is about how to achieve the optimum transmission efficiency while considering both the absolute coding efficiency and the suppression of potential errors. Towards this goal, many researchers have been focused on the methods of adaptively inserting intra blocks into the coded frame/sequence.

The early algorithms were developed to randomly place intra MBs [3], or periodic intra-code contiguous blocks in a frame [4]. The intra refresh frequency was determined in a heuristic way and the intra-coding was applied uniformly to the whole frame. Then the content-adaptive coding mode selection scheme

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 513–520, 2004.

was proposed to intra-code the MBs at regions with high activity [5]. To further improve the performance by jointly considering the network condition and the error concealment, rate-distortion (RD) optimized algorithms have been proposed with the theme of achieving an optimum trade-off between the distortion and the bit rate. Rate distortion optimization (RDO) scheme is well known in source coding, in which the distortion only refers to the quantization errors. However, in the error-prone environment, transmission errors inevitably increase the actual distortion of the reconstructed frame. Therefore, the channel distortion should also be considered in the RDO-based video coding.

The recent work proposed in [6] developed a statistical model to estimate the channel distortion and decide the intra refresh rate before coding each frame. However, the sequence/frame-level adaptive intra-refreshing algorithm considers only the overall R-D behavior of the whole video/frame, which lacks the accuracy in the local end-to-end distortion estimation. In [7], a recursive optimal per-pixel estimate (ROPE) algorithm is proposed to estimate the end-to-end distortion at pixel level. Although it works well for the H.263 codec, its extension to the up-to-date H.264 recommendation is not straightforward due to the high complex prediction schemes employed in H.264, such as intra-prediction, in-loop filter and sub-pixel motion compensation. Consequently, 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 [8][9]. 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 implementation cost make it impractical when K is increased.

In this paper, we propose an optimum end-to-end distortion estimation scheme for error-resilient video coding. The basic end-to-end distortion model is derived according to the theoretical analysis at the pixel level. The correlations of the potential distortions among different frames are derived. Then, the block-level implementation based on the theoretical model is proposed to tackle the influence of sub-pixel motion compensation. In the proposed scheme, a distortion map is defined to store the potential errors of each frame that may propagate to its future frames. In other words, when coding the current frame, the potential channel distortions of its reference frames have been known as a priori. Based on the proposed end-to-end distortion estimation scheme, an improved RDO-based intra/inter mode selection algorithm is developed, in which a new Lagrange parameter in RDO coding is employed. Since the video codec is usually associated with some rate control scheme in application, the rate control technique in H.264 test model is jointly implemented with the proposed error resilient video coding scheme.

The rest of this paper is organized as follows. The end-to-end distortion model is described in Section 2. Simulation results as well as the discussion are also presented. In Section 3, a joint rate-distortion optimization method for H.264

video encoding in packet loss environment is described. Afterwards, experimental results are presented. Section 4 concludes this paper.

2 End-to-End Distortion Estimation

2.1 Theoretical Model

Let f_n^i be the original value of pixel i in the nth video frame, and $\widehat{f}_n^i$ be the corresponding reconstruction value at the encoder side. Suppose $\widehat{f}_{ref}^j$ and $\widehat{r}_n^i$ denote the predicted value (i.e. pixel j in reference frame ref) and the quantized residual, respectively. Then $\widehat{f}_n^i$ is given by $\widehat{f}_n^i = \widehat{f}_{ref}^j + \widehat{r}_n^i$. Let $\widetilde{f}_n^i$ be the reconstructed value at the decoder, which is a random variable for the encoder. Assume that the temporal error concealment is used to reconstruct this pixel in case of packet loss. Let this replacement be the pixel k in the previous frame $n-1$, denoted as $\widetilde{f}_{n-1}^k$. If packet loss rate is p, then we have:

$$\widetilde{f}_n^i = (1-p)(\widetilde{f}_{ref}^j + \widehat{r}_n^i) + p\widetilde{f}_{n-1}^k. \tag{1}$$

Therefore, the expected overall distortion of pixel i in frame n at the decoder is:

$$\begin{aligned} d_n^i &= E\left\{(f_n^i - \widetilde{f}_n^i)^2\right\} = (1-p)E\left\{(f_n^i - \widehat{r}_n^i - \widetilde{f}_{ref}^j)^2\right\} + pE\left\{(f_n^i - \widetilde{f}_{n-1}^k)^2\right\} \\ &\approx (1-p)E\left\{(f_n^i - \widehat{f}_n^i)^2\right\} + (1-p)E\left\{(\widehat{f}_{ref}^j - \widetilde{f}_{ref}^j)^2\right\} + pE\left\{(f_n^i - \widetilde{f}_{n-1}^k)^2\right\} \\ &= (1-p)d_s + (1-p)d_{ep\_ref} + pd_{ec}, \end{aligned} \tag{2}$$

where d_s denotes the quantization distortion, d_{ep_ref} denotes the error propagated distortion from the predicted pixel in the reference frame, and d_{ec} denotes the error concealment distortion for pixel i in frame n. For a pixel in the intra-coded MB, its predict range is restricted in the current slice. If the MB is not lost, its predictor is not lost too. Therefore, the reconstructed pixel value at the encoder and the decoder are the same and therefore d_{ep_ref} is zero. For a pixel in the inter-coded MB, even in case that there are not any channel errors, it still suffers from the propagated distortion through the motion compensation path. Now the key point is about how to obtain the error-propagated distortion. Obviously, after the current frame has been encoded, the error-propagated distortion d_{ep} can be achieved by:

$$\begin{aligned} d_{ep} &= E\left\{(\widehat{f}_n^i - \widetilde{f}_n^i)^2\right\} = E\left\{[(\widehat{f}_n^i - (1-p)(\widetilde{f}_{ref}^j + \widehat{r}_n^i) - p\widetilde{f}_{n-1}^k]^2\right\} \\ &= (1-p)E\left\{(\widehat{f}_n^i - \widehat{r}_n^i - \widetilde{f}_{ref}^j)^2\right\} + pE\left\{(\widehat{f}_n^i - \widetilde{f}_{n-1}^k)^2\right\} \\ &\approx (1-p)E\left\{(\widehat{f}_{ref}^j - \widetilde{f}_{ref}^j)^2\right\} + pE\left\{(\widehat{f}_n^i - \widehat{f}_{n-1}^k)^2\right\} + pE\left\{(\widehat{f}_{n-1}^k - \widetilde{f}_{n-1}^k)^2\right\} \\ &= (1-p)d_{ep\_ref} + p(d_{ec\_rec} + d_{ec\_ep}), \end{aligned} \tag{3}$$

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} is the newly

incurred distortion between the error concealed pixel and the reconstructed pixel at the encoder. And if the pixel is error concealed, then the error-propagated distortion of the error concealed pixel d_{ec_ep} is inherited. In order to obtain the end-to-end distortion, the error concealment distortion for pixel i in frame n is computed by:

$$\begin{aligned} d_{ec} &= E\Big\{(f_n^i - \widetilde{f}_{n-1}^k)^2\Big\} \\ &\approx E\Big\{(f_n^i - \widehat{f}_n^i)^2\Big\} + E\Big\{(\widehat{f}_n^i - \widetilde{f}_{n-1}^k)^2\Big\} \\ &= d_s + d_{ec\_rec} + d_{ec\_ep}. \end{aligned} \tag{4}$$

Combining (2), (3) and (4), the end-to-end distortion can be computed by the linear combination of the source distortion d_s and the error-propagated distortion d_{ep}. Note that if sub-pel motion compensation is used, the reference sample could point to a sub-pel position. Further considering the use of in-loop filter, the computation of D_{ep_ref} becomes more complex. To tackle these problems, we propose a block-based algorithm to track the potential distortion.

2.2 Block-Based Potential Distortion Tracking

Assume a distortion map D_{ep} is defined for each frame on a block basis (e.g. 4x4). The first frame in a sequence is coded with intra mode without considering the error propagation. Then the distortion map of the first frame is derived for coding the subsequent frames. For coding the other frames, the distortion maps of their previous frames indicating the possible influence of propagated errors are referenced to select the proper coding modes. After coding each frame, the associated distortion map is derived accordingly. The referenced error-propagation distortion of the kth block in the mth MB of the nth frame, denoted as $D_{ep\_ref}(n, m, k)$, is computed by weighting the potential error-propagated distortion of the surrounding blocks in the reference frames that overlap with the motion-compensated blocks. Since there is not reference frame for the intra-coded MB, the propagated errors from the previous frames are suppressed and therefore $D_{ep\_ref}(n, m, k)$ in terms of the intra mode equals to zero.

After the current frame is coded, the distortion map D_{ep} in terms of the current frame is then derived according to (3). The calculation of $D_{ec\_ep}(n, m, k)$ and $D_{ec\_rec}(n, m, k)$ 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 in the previous frames. In this case, $D_{ec\_ep}(n, m, k)$ can be derived from the potential error-propagated distortions of these blocks.The distortion map is stored for the following frame encoding. Notice that for coding the first frame, the referenced distortion maps are null, which indicates that the first frame does not suffer from the propagated errors. Moreover, the distortion map of the first frame is only associated with error-concealed distortion D_{ec_rec}. The end-to-end distortion of the kth block in the mth MB of the nth frame is computed as the sum of source distortion and the error-propagated distortion.

2.3 Simulations and Discussion

To test the performance of the proposed distortion estimation scheme, we simulate packet loss in H.264 video coding and use this scheme to estimate the end-to-end distortion. Each row of macroblocks is transmitted in a separate packet. If a MB is lost, the decoder copies the co-located MB in the previous frame. This simple concealment method is used in both encoder and decoder. The mode selection is performed according to the ER-RDO algorithm. Only the first frame is encoded as I frame, the left frames are encoded as P frame. The number of reference frames is one. The Foreman QCIF video sequence is simulated 30,100,500 times and the average frame-level decoder distortion is computed. The estimation result is shown in Fig. 1. The accuracy of experimental results suggests that the proposed distortion model is consistent to the statistical K-time decoding at the encoder side, however, the added complexity is much lower than the latter.

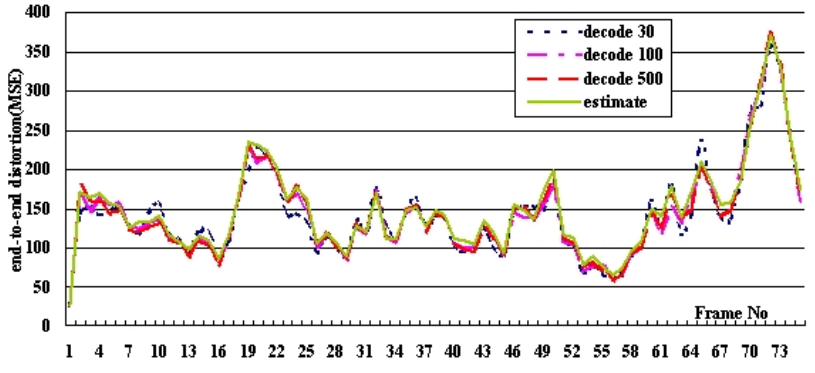

Fig. 1. end-to-end distortion estimation results for Foreman sequence, r=64kbps, packet loss rate=10%

3 RDO-Based Error Resilient Coding

3.1 RDO-Based Error Control

Suppose O denotes the set of all selectable coding options in terms of an MB. The coding option o^* of the mth MB in the nth frame is selected to be the one that minimizes the cost given by:

$$J(n, m, o) = D(n, m, o) + \lambda R(n, m, o) \tag{5}$$

According to the previous section, since the potential channel distortion is also considered in the expected overall distortion, the relation between the rate and the overall distortion changes as well. Accordingly, the Lagrange parameter λ

should be properly selected. Similar to the Lagrange parameter selection scheme for the error-free environment in [10], we derive the new Lagrange parameter as $(1-p)\lambda$, where λ is the Lagrange parameter in the error-free environment. Then, the proposed RDO-based error-resilient coding for H.264 encoder in the packet-loss environment is performed as follows. Since the channel distortion of the B frame would not propagate to the following P frames, it is unnecessary to define the distortion map for the B frame to store the potential error-propagated distortion. The coding mode selection of B frames can be the same as that in the P frame coding. 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 while 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 (2) and (5), the coding option can be selected with

$$o^*(n,m) = \underset{o\in O}{argmin}((1-p)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) + pD_{ec}(n,m) + (1-p)\lambda R) = \underset{o\in O}{argmin}(D_s(n,m,o) + D_{ep\_ref}(n,m,o) + \lambda R). \quad (6)$$

After the current frame is encoded, the distortion map is derived according to the (3) for the encoding of the future frames.

3.2 Discussion of Rate Control

While we investigate the error resilience feature of the proposed R-D based model in H.264/AVC, we also analyze the effect of the rate control scheme adopted in the H.264/AVC test model [11]. Basically, rate control resolves two main problems, i.e. bit allocation and quantization parameter adjustment. In H.264 test model, it consists of three tightly consecutive components: GOP level rate control, picture rate control and the optional basic unit level rate control. The basic unit is defined as a group of successive MBs in the same frame. For detailed algorithm of the basic unit level rate control, we refer to [11].

Supposing a slice/packet is taken as a basic unit, rate control can be easily jointly implemented with the coding mode selection scheme described in the previous subsection. The remained problem is whether or not the accuracy and coding efficiency of rate control can satisfy the requirements. Obviously, the rate allocation at GOP level does not have effect. The R-D model in picture/basic unit layer is:

$$R = C_1 \times \frac{MAD}{Q_{step}} + C_2 \times \frac{MAD}{Q_{step}^2} - M, \quad (7)$$

where M is the total number of header bits and motion vector bits, and C_1 and C_2 are two adaptively adjusted coefficients. The key point is about how

to estimate the MAD prior to doing the current basic unit rate control. In the traditional RDO-based framework, it is proven that the current MAD can be estimated with the linear regression model using the actual MAD of the previous picture or the co-located basic units in previous picture. In the proposed coding mode selection scheme, this MAD-based linear regression method is still applicable because the same mode selection scheme is used for each picture and the statistics of MAD remains very similar. Simulations have shown that the basic unit level rate control (i.e. taking the slice as the basic unit) has the very similar coding efficiency to the fixed quantization parameter coding.

3.3 Experimental Results

Some experiments have been carried out to verify the performance of the proposed algorithm. Two algorithms are compared: the proposed coding mode decision algorithm and ER-RDO. The testing platform is the H.264 reference software JM7.5c [11]. In the default ER-RDO algorithm, K=500 decoders are simulated in the encoder. Only the first frame is encoded as I frame, and the following frames are encoded as P frames. Five coded sequences are generated for each algorithm: Foreman@64kbps (QCIF, 7.5fps, and denoted as Fore_64k), Foreman@144kbps (QCIF, 7.5fps, and denoted as Fore_144k), Hall Monitor@32kbps (QCIF, 10fps, and denoted as Hall_32k), Paris@144kbps (CIF, 15fps, and denoted as Paris_144k) and Paris@384kbps (CIF, 15fps, and denoted as Paris_384).

The packet loss situation is simulated according to the error resilience testing conditions specified in [12]. The coded sequences were decoded after packet loss simulation under packet loss rates 3, 5, 10 and 20%. Note that there are 4 packets per frame for QCIF and 9 packets for CIF. The 40 bytes of IP/UDP/RTP headers per packet have been taken into account. The simple previous frame copy error concealment method is used in all simulations.The average YPSNR values of the coded sequences except the first frame under different packet loss rate are shown in Table 1. The results show that the proposed algorithm outperforms ER-RDO in terms of transmission efficiency in all cases. Moreover, the most significant difference lies in the computing complexity. In these experiments, the running time of ER-RDO is about as 25 times long as the original algorithm without error control, whereas the running time of the proposed algorithm is very similar to the original algorithm.

Table 1. Comparison results of average PSNR (in dB) at different packet loss rates

Sequences	LossRate: 3%		LossRate: 5%		LossRate: 10%		LossRate: 20%	
	Proposed	ER-RDO	Proposed	ER-RDO	Proposed	ER-RDO	Proposed	ER-RDO
Fore_64k	30.31	30.21	29.48	29.42	27.60	27.46	25.58	25.50
Fore_144k	34.36	34.23	33.22	33.15	30.85	30.60	28.17	28.11
Hall_32k	33.58	33.25	33.44	33.09	32.29	31.87	31.19	30.93
Paris_144k	27.51	26.71	27.01	26.15	25.93	25.59	24.84	24.54
Paris_384k	33.08	32.68	32.29	31.72	33.74	33.34	29.33	28.98

4 Conclusion

An optimized rate-distortion model for H.264 video encoder in the packet loss environment has been presented in this paper. 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.

References

1. Stuhlmuller, M., Farber, N., Link, N., Girod, B.: Analysis of Video Transmission over Lossy Channels. IEEE J. Selected Areas in Communications, Vol. 18. 6 (2000) 1012-1032
2. Wang, Y., Zhu, Q. F.: Error Control and Concealment for Video Communication: A Review. Proc. IEEE, Vol. 86. 5 (1998) 974-997
3. Cote, G., Kossentini, F.: Optimal Intra Coding of Blocks for Robust Video Communication over the Internet. Image Commun.. 9 (1999) 25-34
4. Zhu, Q. F., Kerofsky, L.: Joint Source Coding, Transport Processing and Error Concealment for H.323-based Packet Video. Proc. SPIE, VCIP'99, Vol. 3653. San Jose, CA (1999) 52-62
5. Haskell, P., Messerschmitt, D.: Resynchronization of Motion-Compensated Video Affected by ATM Cell Loss. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 3. (1992) 545-548
6. He, Z. H., Cai, J. F., Chen, C. W.: Joint Source Channel Rate-Distortion Analysis for Adaptive Mode Selection and Rate Control in Wireless Video Coding. IEEE Trans. Circuits and Systems for Video Technology, Vol. 12. 6 (2002) 511-523
7. Zhang, R., Regunathan, S. L., Rose, K.: Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience. IEEE J. Selected Areas in Communications, Vol. 18. 6 (2000) 966-976
8. Stockhammer, T., Kontopodis, D., Wiegand, T.: Rate-Distortion Optimization for JVT/H.26L Coding in Packet Loss Environment. Proc. PVW. Pittburgh, PY (2002)
9. ITU-R Rec. H.264 — ISO/IEC 14496-10 AVC: Draft Text. Joint Video Team document JVT-E146D37 (2002)
10. Wiegand, T., Girod, B.:Lagrangian Multiplier Selection in Hybrid Video Coder Control. Proc. ICIP2001. Thessaloniki, Greece (2001)
11. http://bs.hhi.de/ suehring/tml/download/jm75c.zip
12. Wenger, S.: Common Conditions for Wire-line, Low Delay IP/UDP/RTP Packet Loss Resilient Testing. ITU-T VCEG document VCEG-N79r1 (2001)