Rate-Distortion Optimized Bit Allocation for Error-Resilient Transcoding

Minghui Xia Anthony Vetro Bede Liu
Huifang Sun

TR-2004-047 May 2004

Abstract

A key issue in video transcoding for transmission in an error-prone environment is to balance the rate used for the video source with bits allocated for error resilience such that the end-to-end distortion is minimized under the given rate constraint and channel condition. The approach presented in this paper differs from previous works by accounting for the inter-frame dependence in both video source requantization and error propagation of motion compensated video. Based on the rate-distortion models developed in this paper, an optimal Group-of-Picture based bit allocation scheme is proposed. We also propose a sub-optimal scheme that is suitable for a real-time implementation. Both the optimal and the suboptimal scheme achieve better PSNR performance than the fixed heuristic bit allocation scheme.

201 Broadway, Cambridge, Massachusetts 02139

Proc. IEEE Int’l Symposium on Circuits and Systems, May 2004

RATE-DISTORTION OPTIMIZED BIT ALLOCATION FOR ERROR RESILIENT VIDEO TRANSCODING

Minghui Xia, Anthony Vetro, and Bede Liu

ABSTRACT

A key issue in video transcoding for transmission in an error-prone environment is to balance the rate used for the video source with bits allocated for error resilience such that the end-to-end distortion is minimized under the given rate constraint and channel condition. The approach presented in this paper differs from previous works by accounting for the inter-frame dependence in both video source requantization and error propagation of motion compensated video. Based on the rate-distortion models developed in this paper, an optimal Group-of-Picture based bit allocation scheme is proposed. We also propose a sub-optimal scheme that is suitable for a real-time implementation. Both the optimal and the sub-optimal scheme achieve better PSNR performance than the fixed heuristic bit allocation scheme.

1. INTRODUCTION

Video communication through wireless channels is still a challenging problem mainly due to the limitations in bandwidth and the presence of channel errors. The two primary tools used for error-resilience source coding are resynchronization marker insertion and intra-block insertion. Resynchronization is achieved by inserting markers periodically so that when an error occurs, decoding can be restarted at the point where the synchronization marker has been placed. The insertion of intra blocks is used to provide a temporal localization of errors by decreasing the temporal dependency in the coded video sequence.

Error resilience video transcoding has been addressed by several researchers [1, 2, 3, 4] in which the resilience methods described above have been considered. One important aspect that is neglected by these approaches is the issue of inter-frame dependency. Bit allocation or coding mode selection is often optimized only for the current macroblock (MB) or the current frame. Another problem that has not been sufficiently addressed is a jointly optimal solution for video bit-rate reduction and error resilience insertion. In this paper, we address the optimal bit allocation between video source rate and the error resilience insertion by establishing rate-distortion (R-D) models to characterize each transcoding component: video source requantization, resynchronization marker insertion, and intra refresh. The proposed models are novel in that inter-frame dependency is included in both video source model and error resilience model, and also the error resilience model takes into account the process of error concealment. Another contribution of this paper is a proposed sub-optimal bit allocation scheme that achieves near-optimal performance and has been implemented to enable real-time video transcoding. We adopt a drift-free cascaded video transcoding architecture to validate the effectiveness of the proposed models and demonstrate the performance of the proposed real-time optimization technique.

2. PROBLEM STATEMENT

The problem addressed in this paper is to minimize the end-to-end distortion of a coded video bitstream subject to rate constraints, where the overall rate budget is allocated among different components that contribute to the rate. Let K denote the number of components. In this paper, three distinct components are considered: video source, intra refresh in inter-coded frames and resynchronization marker insertion. The whole problem can then be solved as a constrained minimization problem, for which a Lagrangian optimization approach [5] is taken to minimize:

$$\sum_{k=1}^{K} d_k(\omega_{\mathbf{k}}) + \lambda \sum_{k=1}^{K} r_k(\omega_{\mathbf{k}}) \tag{1}$$

where d_k and r_k are the distortion and rate of each component, λ is the Lagrangian multiplier, and $\omega_{\mathbf{k}}$ are the specific parameters used in the allocation, i.e., quantization parameters, intra refresh rate and frequency of synchronization marker insertion. A bisection algorithm can be used to obtain the optimal λ used to solve this problem [5], but this is computationally expensive. Also, obtaining accurate R-D sample points is still an issue. In the following sections, we will establish R-D models for each component so that we do not have to rely on obtaining actual R-D sample points from simulation. We then propose a low-complexity bit allocation scheme that utilizes these models.

3. VIDEO SOURCE R-D MODEL

This section develops an R-D model for a coded video source that accounts for inter-frame dependency.

3.1. R-D Model for Intra-Coded (I) Frame

A R-D model for video quantization has been reported in [6], based on decomposing an I frame into independent identically distributed (i.i.d.) Gaussian sources. However, it has been shown in [7] that a decomposition based on generalized Gaussian model is more accurate. Furthermore, coarse quantization will often violates the model assumption as in [6]. We have discovered from experiment that by introducing two additional parameters, β and γ, a more accurate R-D model can be established to overcome the two issues mentioned above. For an I frame, we have:

$$D_0 = \sigma_0^2 e^{-\beta_0 R_0^{\gamma_0}} \tag{2}$$

M. Xia and B. Liu are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA. ({mxia,liu}@ee.princeton.edu) A. Vetro is with Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139 USA

where R_0 and D_0 are the rate and distortion of the frame as a result of requantization, σ_0^2 is the total variance of the signal, β_0 and γ_0 are model parameters that need to be estimated from two sample points.

3.2. R-D Model for Inter-Coded (P) Frame

We model the inter-frame dependency by changing the variance of the kth frame σ_k^2 to $\sigma_k^{\star 2}$:

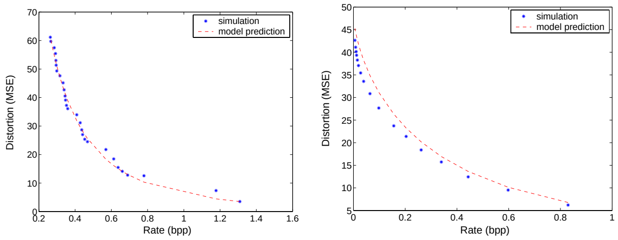
$$D_k = \sigma_k^{\star 2} e^{-\beta_k R_k^{\gamma}} = (\sigma_k^2 + \alpha_k D_{k-1}) e^{-\beta_k R_k^{\gamma}}, \qquad (3)$$

where $\sigma_k^{\star 2} = \sigma_k^2 + \alpha_k D_{k-1}$. The term $\alpha_k D_{k-1}$ models the dependence between the current and the previous frame. It captures the quantization error propagation effect caused by motion compensation. D_{k-1} denotes the extra quantization residue error produced when the previous frame is requantized with a larger quantization scale, and α_k denotes the propagation ratio, which is determined by the amount of motion compensation.

3.3. Model Accuracy

The model parameters can be accurately estimated through a two-pass recoding process in which we generate three R-D samples for each frame. We have observed that γ_k and α_k do not change much within a given sequence. Therefore it is sufficient to estimate them once at the start of a sequence. For parameters that do change with time, i.e., σ_k and β_k, their values are updated at each frame.

The model has been tested extensively on various sequences. However, due to limitations in space, we can only show the result for one sequence. The results shown here are consistent with other sequences that have been tested. We encode the *Foreman* sequence with a GOP size of 12 and with quantization scale $Q_I = Q_P = 3$. After estimating the model parameters, we requantize the sequence using an arbitrarily selected quantization scales: $Q_I = 8, Q_P(frames\,\#1 - \#10) = 16, Q_P(frame\,\#11) \in [4, 31]$. The R-D plots of the requantized I-frame and the last P-frame in the GOP are compared with the model estimates in Fig.1. Note that since we account for inter-frame dependency in our model, the model estimates for the last P-frame rely on accurate estimates for all preceding frames. The plots show that the model is able to accurately estimate the actual R-D values.



(a) I frame R-D (Foreman frame #0)

(b) P frame distortion (Foreman frame #11) test I: $Q_I = 8, Q_P(\#1 - \#10) = 16, Q_P(\#11)$=4-31

Fig. 1. Requantization RD model test on Foreman sequence

4. ERROR-RESILIENCE R-D MODELS

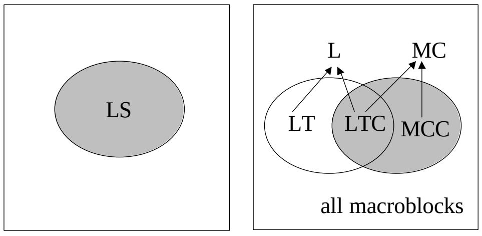
The rate consumed by resynchronization marker can be simply calculated from the number of bits in the resynchronization header and the resychronization marker spacing. The rate consumed by intra-fresh can be calculated from the intra-refresh rate and the average rate increase by replacing an inter-coded MB with an intra-coded MB. In the following, we will first present the test environment for the error resilience coding. The distortion models for resynchronization and intra-refresh will then be established and verified through experiment.

4.1. Test Environment

We adopt a Binary Symmetric Channel model, which assumes independent bit error P_e in a bitstream. Error is recovered by resynchronizing to the frame headers or to the added slice resynchronization markers. For an intra-coded MB, spatial concealment is employed by copying the MB from its immediate above neighbor. For an inter-coded MB, temporal concealment is employed where the motion vector of the lost MB is set to be the median of the motion vectors selected from its immediate left, upper and upper-right neighbors. The MB in the previous frame that this motion vector is referencing to is then copied to the current location.

4.2. Overall Distortion Due to Channel Error

The overall distortion for an I or P frame caused by channel errors can be decomposed as illustrated in Fig.2. The rectangle denotes the set of all MBs in a frame. The distortion mainly comes from two parts: distortion from lost MBs, which are referred to as L MBs, and distortion propagated from previous corrupted MBs through motion compensation, which are referred to as MC MBs. The lost MBs can be further decomposed into three categories: intra-coded MBs lost and concealed with spatial concealment (LS MBs), inter-coded MBs lost and concealed with temporal concealment (LT MBs), and inter-coded MBs lost and concealed with temporal concealment but the replacement MB itself was corrupted (LTC MBs). Note that LTC MBs define the intersection of L MBs and MC MBs. The MCC MBs refer to MBs that are received correctly, but reference the previous corrupted MBs through motion compensation.



(a) I frame

(b) P frame

Fig. 2. Distortion compositions from channel error

Let Y_l denote the number of MBs lost in a frame, Y_{mc} the number of MBs corrupted through motion compensation, and the total number of MBs in a frame by M. Then, the average number of corrupted MBs in a frame, $E[Y]$, can be expressed as:

$$E[Y] = E[Y_l] + E[Y_{mc}] - E[Y_{ltc}], \tag{4}$$

where $Y_{ltc} = Y_l \bigcap Y_{mc}$. We assume this intersection is proportional to the number of lost MBs and the number of inter-coded MBs corrupted through motion compensation. Subsequently,

$$E[Y_{ltc}] = E[Y_l \bigcap Y_{mc}] \approx \frac{E[Y_l] \cdot E[Y_{mc}]}{M}, \tag{5}$$

The total average distortion (measured in MSE) can therefore be calculated by:

$$D = \begin{cases} \frac{1}{M}\{E[Y_l] \cdot D_s\}, & I \\ \frac{1}{M}\{E[Y_{lt}] \cdot D_t + E[Y_{ltc}] \cdot D_{tc} + E[Y_{mcc}] \cdot D_{mc}\}, & P \end{cases}$$

where Y_{mcc} is the number of MCC MBs as depicted in Fig.2, and D_s, D_t, D_{tc}, D_{mc} are the average spatial concealment distortion related to $Y_l, Y_{lt}, Y_{ltc}, Y_{mcc}$ respectively.

4.3. Error Propagation Distortion

The probability that a single MB is corrupted through motion compensation can be computed by:

$$p_{mc} = \rho\theta_1 + [1 - (1-\rho)^2]\theta_2 + [1 - (1-\rho)^4]\theta_3, \tag{6}$$

where ρ is the probability of one MB corrupted in the previous frame, and $\theta_1, \theta_2, \theta_3$ denote the proportion of MBs that reference one MB, two MBs, and four MBs in the previous frame respectively. If we denote the proportion of intra-coded MBs by η, then we have $\theta_1 + \theta_2 + \theta_3 + \eta = 1$. From this relation, it is clear that a higher value of η will yield a lower value of p_{mc}, which is expected.

As in [1], we may adopt a n-step Markov model to estimate error propagation through motion compensation. However, the model is too computationally intensive to be implemented in real time. Therefore, we consider replacing the n-step Markov model with a 1-step Markov model, and Y_{mc} can be obtained by:

$$E\{Y_{mc}\} = M \cdot p_{mc}. \tag{7}$$

It follows that we can express the average distortion due to motion compensation at frame n by

$$D_{mc}(n) = \rho \cdot (1 - \eta) \cdot D(n-1) \tag{8}$$

where $D(n-1)$ is the average total distortion in the previous frame $n-1$.

4.4. Error Concealment Distortion

The probability that one MB is lost in a video frame p_l can be modeled by p_{sl}, the probability that a video packet (or slice) is lost. Let the channel bit error rate (BER) be denoted by P_e and the average slice length in bits by L_s. Then,

$$p_l = p_{sl} = 1 - (1 - P_e)^{L_s} \tag{9}$$

It then follows that the average number of lost MBs in frame n, $E[Y_l(n)]$, is simply $p_l \cdot M$.

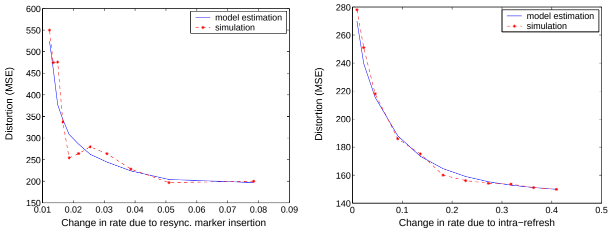
D_s represents the difference between the lost MB and its above neighboring MB and can be estimated by calculating pixel differences among correctly reconstructed MBs in the same frame. D_t is the difference between the lost MB and the MB copied from the previous frame and can similarly be estimated by calculating pixel differences between correctly reconstructed MBs in adjacent frames. D_{tc} can be approximated by an addition of motion compensation corruption to D_t:

$$D_{tc} = D_t + D_{mc}, \tag{10}$$

where D_{mc} is given in Eq.(8) and $E[Y_{ltc}]$ is given in Eq.(5).

4.5. Model Accuracy

Fig. 3(a) shows a test of the R-D model for resynchronization marker insertion as a function of marker spacing (or video packet length). For the simulation, we used 100 frames of QCIF format *Foreman* sequence, coded at 10 frames/second. The marker spacing varies from 130 bits to 1300 bits. The simulation is performed with a $BER = 10^{-4}$. Fig. 3(b) shows a test of the R-D model as a function of changes in the intra-refresh rate. It can be seen that the proposed models accurately predicts the actual distortion.



(a) Resync. marker insertion R-D model

(b) Intra-refresh R-D model

Fig. 3. Test of error resilience R-D model accuracy

5. SUB-OPTIMAL BIT ALLOCATION THROUGH R-D DERIVATIVE EQUALIZATION

To enable real-time implementation for bit allocation, a technique to determine a sub-optimal operating point is proposed. We refer to this technique as an R-D derivative equalization scheme. This scheme is based on the fact that optimal bit allocation is achieved at the point where the slopes of the R-D function for each component[1] are equal [8].

If we start from a operation point close to an optimal point, our objective is to continually adjust the the operating point in the direction of the optimal point. To achieve this, we need to start from an close-to the optimal point, and find a strategy to move towards an optimal point. The first step is easy since we can carry out the optimal scheme for the first GOP and assume that few-second initial delay is acceptable for a transcoder. For the second step, we compare the local derivatives of each R-D curve and adjust the bits allocated to each component accordingly. Given that the rate budget has not changed, we have deduced that reallocating a change in rate, ΔR, from the component with the smallest absolute derivative value to the component with the largest absolute

---

[1] In this paper, component refers to the three transcoding components: requantization, synchronization marker insertion and intra-refresh.

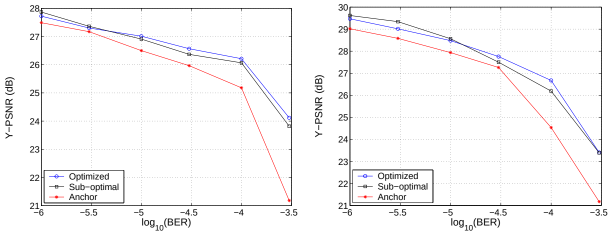
derivative value is a close approximation to the optimal solution. The proof to this approximation, as well as a means to determine an optimal value of ΔR, could not be included here due to space limitations.

6. EXPERIMENT RESULT

Our experiment is conducted using 300 frames of the *Foreman* and *Coastguard* sequences in QCIF format. The original coding is performed with an MPEG-2 encoder provided by MPEG Software Simulation Group [9]. The original bit rate is set at 384 kb/s and the original frame rate is set at 10 frames/second[2]. A GOP size of 10 frames is used; B-frames are not included. We then recode the video sequence to an MPEG-4 format with lower bit-rate, while keeping the output frame rate the same. Bit-rate reduction and error resilience insertion is achieved by requantization, resynchronization marker insertion and intra-refresh.

With the above simulation conditions, optimal bit allocation is achieved by applying the Lagrangian optimization algorithm described in Section 2 using the proposed R-D models in Section 3 and 4. The optimal scheme models the entire R-D curve and performs a complete estimation of the model parameter. On the other hand, the sub-optimal scheme only models the R-D curves locally to obtain local derivatives, and performs a simplified model parameter estimation as described in Section 3.3. We also simulate an anchor scheme that recodes the sequence with a fixed, but reasonably good set of error resilience parameters. In this scheme, we fix the resynchronization marker insertion to be every 11 MBs and assign an intra-refresh rate of 20%.

Fig.4 shows the comparison of our sub-optimal bit allocation scheme with the optimal scheme and the anchor using *Coastguard* and *Foreman* sequences recoded at 64kbps. It can be observed that the sub-optimal scheme only performs slightly worse than the optimal scheme and both schemes outperform the anchor. This test confirms that the proposed sub-optimal scheme provides performance similar to the optimal scheme, but with much lower complexity.



(a) *Coastguard* sequence

(b) *Foreman* sequence

Fig. 4. Video PSNR at 64 kb/s in error prone transmission channel: Anchor vs. Optimized vs. Sub-optimal Bit Allocation

7. CONCLUSION

We have proposed new R-D models that consider inter-frame dependancy for optimal bit allocation in error resilient video transcoding. Sub-optimal scheme has also been proposed for realtime implementations. The proposed algorithms demonstrated superior performance over fixed error-resilient transcoding schemes.

8. REFERENCES

[1] G. Reyes, A. R. Reibman, S. F. Chang, and J. C.-I. Chuang, “Error-resilient transcoding for video over wireless channels,” *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1063–1074, 2000.

[2] G. Cote, S. Shirani, and 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, 2000.

[3] 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, no. 6, pp. 966–976, 2000.

[4] S. Dogan, A. Cellatoglu, M. Uyguroglu, A. H. Sadka, and A. M. Kondoz, “Error-resilient video transcoding for robust internetwork communications using gprs,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 453–464, 2002.

[5] H. Everett, “Generalized lagrange multiplier method for solving problems of optimum allocation of resources,” *Operations Research*, vol. 11, pp. 399–417, 1963.

[6] H. M. Hang and J. J. Chen, “Source model for transform video coder and its application — part i: Fundamental theory,” *IEEE Transactions on Circuits, System and Video Technology*, vol. 7, no. 2, pp. 287–298, 1997.

[7] J. Minguillon and J. Pujol, “Uniform quantization error for laplacian sources with applications to jpeg standard,” in *Mathematics of Data/Image Coding, Compression, and Encryption*. Proceedings of the SPIE, 1998, vol. 3456, pp. 77–88.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.

[9] MPEG Software Simulation Group, “Mpeg-2 codec, version 1.2,” in *ftp://ftp.mpeg.org/pub/mpeg/mssg/*, Jul. 1996.

---

[2] Note that MPEG-2 does not code video at 10 frames/second. We achieve this rate by dropping 2 out of 3 frames before encoding.