Optimal Intra Coding of Macroblocks for Robust H.263 Video Communication over the Internet

Guy Côté and Faouzi Kossentini
Department of Electrical and Computer Engineering,
University of British Columbia,
Vancouver BC V6T 1Z4, Canada
E-mail: guyc@ece.ubc.ca

September 17, 1998

Abstract

Reliable transmission of compressed video in a packet lossy environment cannot be achieved without error recovery mechanisms. We describe an effective method for increasing error resilience of video transmission over packet lossy networks such as the Internet. Intra coding (without reference to a previous picture) is a well known technique to eliminate temporal error propagation in a predictive video coding system. For example, randomly intra coding of the macroblocks within an H.263 video coder increases error resilience to packet loss, as shown in this paper. However, when the error concealment used by the decoder is known, intra encoding following a method that optimizes the tradeoffs between compression efficiency and error resilience is a better alternative. In this paper, we present a rate-distortion optimized mode selection method for packet lossy environments that takes into account the error concealment method used at the decoder. We present results for different packet loss rates and typical packet sizes of the Internet that illustrate the advantages of the proposed method.

1 Introduction

Transmission of compressed video over noisy channels presents many challenges. Video compression algorithms remove unnecessary redundancy, however, some level of redundancy is necessary to achieve error robustness. Thus, it is important to carefully add redundancy to the coded video data. Shannon's separation theorem has been applied in this field of research, where video compression and transmission optimizations have been performed independently. This is optimal only if infinite coding delays are assumed. However, most video communication applications require relatively high reliability and small delays in video transmission. Thus, joint source-channel coding of video signals is necessary for achieving high error resilience and maintaining low delay communication, simultaneously.

Emerging video coding standards now include provisions for error resilience, especially MPEG-4 [1] and H.263+ [2]. When using a feedback channel, very effective error recovery mechanisms can be implemented [3]. In this paper, we propose a method that does not require a feedback channel. Feedback channels introduce additional delay and complexity, and they are usually also affected by channel errors. Moreover, they may not be available in many scenarios such as video broadcasting or multi-point communication, common in Internet applications.

Standard-based video coding algorithms employ inter-picture prediction to reduce temporal dependencies. However, an error in the compressed video bit stream will propagate unless picture information is coded without reference to a previous picture. We call these two coding modes *inter* and *intra*, respectively. Coding a complete video frame in the intra mode is an efficient method to stop error propagation due to temporal prediction. Unfortunately, intra coded frames require many more bits than inter coded frames, yielding undesirable long delays when transmitting video over a fixed bit rate channel. Therefore, higher performance levels are expected by intra coding only some of the macroblocks within a frame. While intra coding of macroblocks is still very expensive in terms of bits, we will see that it is effective in terms of error recovery. In fact, very good tradeoffs between compression efficiency and error resilience can be achieved.

Randomly choosing to code macroblocks in all frames in the intra mode at a certain frequency may introduce unnecessary redundancy. In order to only intra code macroblocks that cannot be appropriately concealed at the decoder, we propose a rate-distortion (RD) optimized mode selection method that employs both the error concealment distortion and the quantization distortion in the minimization criterion. When a video encoder is aware of the concealment technique used by the decoder, it can perform the same error concealment at the encoder and choose the coding mode (inter/intra) that yields the best RD tradeoffs. The performance of this method is directly affected by the effectiveness of the error concealment method used, but the described techniques are applicable to any error concealment method. In [4], Lee proposes a multiple description technique to protect macroblocks that are badly concealed. The performance of the error concealment is measured in terms of incurred distortion, without taking into account the incurred rate for the added redundancy of the second description.

The rest of the paper is divided as follows. First, we present the proposed framework and error recovery mechanisms used for comparison purposes. We follow with a description of the proposed intra updating method using RD optimized mode selection. Finally, experimental results are presented using typical packet sizes of the Internet.

2 Proposed Video Coding Framework

Many different methods have been proposed to increase error resilience of coded video in packet lossy networks. In this section, we present a framework where the various error resilience methods can be compared. We first describe the packetization specific to the Internet and show that a Group of Blocks[1] (GOB) packetization is reasonable for such a network. We then discuss the error concealment method and the random intra updating method used in this work.

2.1 Packetization of H.263 Bit Streams for the Internet

Video communication over the Internet usually employs the Real-time Transport Protocol (RTP) [6]. The payload for H.263 bit streams is defined in [13] and the profile for audiovisual conferencing is defined in [14]. Together, these documents form a packetization structure defining an RTP packet, containing an RTP header, H.263 RTP payload header and H.263 RTP payload (compressed data stream). The RTP Control Protocol (RTPC) is employed in particular to monitor the quality of service. The RTP does not provide guaranteed QoS and usually operates on top of the UDP and IP. High packet loss rates can be expected if real-time communication is to be maintained. With RTP, it is possible to divide the video packets into GOBs. The H.263 payload header does not require additional bits, since the 2-byte prefix of the H.263 synchronization marker is removed and replaced with the H.263 RTP payload header, which is exactly 2 bytes in the basic operation mode used here. The additional overhead required is simply the

[1] A GOB represents a complete line of macroblocks

RTP header, which can be as low as 96 bits in size.

If every GOB is packetized in an RTP packet, the overhead required by RTP packetization is at least 8.6 kbps at QCIF resolution and 10 fps. This may seem excessive in terms of additional bits, but error resilience is greatly improved by using such a packetization scheme. Packetizing a complete video frame in one RTP packet reduces the overhead, but if packet loss occurs synchronization between the encoder and decoder in lost until the next received intra frame. Moreover, error concealment for missing GOBs is much easier and more accurate than error concealment for missing frames.

2.2 Error Concealment for Missing GOBs

The error concealment method used in this work is based on the TCON model described in H.263 Test Model TMN-10 [5]. We next summarize this method.

Before errors can be concealed at the decoder, they must be detected. The decoder can receive information from the multiplexing layer regarding the status of a packet or it can detect that a packet is missing from the packet numbering (e. g.the sequence number in RTP [6]). The decoder can choose to drop the complete packet if errors are indicated or detected. If a packet received with errors is passed to the video decoder, errors can be detected using syntactic or semantic violations of the video bit stream [1]. These include:

- motion vector outside of allowable range,
- invalid VLC table entry,
- DCT coefficient out of range, and
- number of DCT coefficients in a block exceeding 64.

Once an error is detected, the decoder searches for the next synchronization point. Error recovery and concealment is then performed. We assume that a synchronization point is present at the beginning of every GOB. No error recovery is here attempted, i. e., if an error is detected in a GOB, the complete GOB is discarded and error concealment is performed for the missing GOB.

Error concealment is performed as follows. Motion vectors of the missing macroblocks are copied from the macroblock above when available, otherwise set to *zero*. Then the macroblock from the previous frame at the same spatial location is motion compensated with this motion vector and copied to the current location in the current frame.

Many other error concealment techniques have been proposed in the literature, and an excellent review is available in [7]. However, many of these techniques require substantial additional complexity that can be tolerated in still image decoding but not in real-time video decoding. The method used in this work provides efficient error concealment and requires very little additional computational complexity. Only the error detection and motion compensation operations are necessary. Therefore, real-time video decoding can still be easily maintained.

2.3 Random Intra Coding of Macroblocks

In a typical low bit rate video sequence, errors in many macroblocks can efficiently be concealed. Therefore, randomly intra coding all macroblocks in a frame would not be efficient. We propose an intra coding pattern where only macroblocks that contain texture information (i. e. excluding macroblocks that are skipped or only motion compensated) are intra coded. Therefore, most of the active regions and/or regions of interest will then be intra updated.

Randomly updating macroblocks using the intra coding mode has been suggested in [8, 9]. In [8], Haskell proposes that the portion of macroblocks to be intra updated in a coded frame is chosen based on the life expectancy of the errors. He states that this life expectancy depends on the intra macroblock refresh rate but is fairly independent of the probability p that a macroblock is in error. Also, a method is proposed to only update blocks with high activity (variance). In [9], Naka proposes that intra updating each macroblock every 5 frames is sufficient for a BER of 0.01%.

In this work, we develop a relationship between the probability that a macroblock is corrupted by errors, p, and the intra updating frequency, I_{freq}, based on our experimental results. To obtain this relationship, we encode different video sequences with a fixed bit rate and vary the intra refresh rate over four different macroblock loss rates $p = 0\%$, 5%, 10% and 20%. Decoder Y-PSNR results are presented for the sequence PARIS encoded at 64 kbps in Figure 1. Other sequences give very similar results. It can be observed that the frequency of the intra updating can be approximated by

$$I_{freq} = \frac{1}{p}. \tag{1}$$

For example, for a probability of macroblock loss p of 20%, each candidate coded macroblock should be updated once in every 5 coded frames.

3 Proposed Intra Updating Method

RD optimized video coding provides an efficient means for coding mode selection. A summary of research work on RD optimized mode selection for an error free environment can be found in [10]. Here, three coding modes are considered: *skip*, *inter*, and *intra*. The skip mode is a special case of inter mode where no information is transmitted, and the macroblock is simply repeated from the spatially corresponding macroblock in the previous frame. Independently for every macroblock, we choose the mode that minimizes the Lagrangian given by

$$J = D + \lambda R, \tag{2}$$

that is, the coding mode that yields the best RD tradeoffs for the macroblock. Using

$$\lambda = 0.85 \times \left(\frac{Q}{2}\right)^2 \tag{3}$$

has been shown to provide good RD tradeoffs [10], where Q is the quantization step size of the macroblock.

Using the above method, the coding mode selection is only optimal if the video bit stream is received without errors at the decoder. When errors are present, temporal prediction will allow errors to propagate if the inter mode is chosen. Using the intra coding mode will stop error propagation, but at a higher coding rate cost.

If we know the error concealment method employed by the decoder and error rates of the network, we can achieve better tradeoffs between compression efficiency and error resilience. First, we can attribute the distortion to two sources: distortion D_q caused by quantization error, and distortion D_c remaining after error concealment. Assuming a macroblock error rate of p, we minimize the Lagrangian

$$J = (1-p)D_q + pD_c + \lambda R. \tag{4}$$

Here, the rate R is the rate at which the coded sequence is transmitted, and is the same as in Equation (2).

For a given macroblock, two distortions are computed for all three coding modes considered: the coding distortion D_q and the concealment distortion D_c. Then D_q is weighted by the probability $(1-p)$ that this macroblock is received without error, and D_c is weighted by the probability p that the same macroblock is lost and concealed. Using this above minimization, good RD tradeoffs can be achieved subject to the probability of error rate and concealment constraints. The error concealment method will directly affect the mode decision. A better error concealment method than the one employed here will give better RD performance given the same probability of error rate.

Note that by minimizing Equation 4, regions that are usually well concealed will most probably not be coded in the intra mode. If a given macroblock is perfectly concealed, then $D_c = D_q$ and Equations (2) and (4) are therefore equivalent.

4 Experimental Results

An H.263 video coder [11] implementing the Test Model TMN-10 [5] specifications with RD optimizations [12] is used in all simulations. A packet of video consists of a complete GOB, with a synchronization marker at the beginning of every GOB (packet). For each Packet Loss Rate (PLR), 25 simulations are performed and the average luminance PSNR (Y-PSNR) is computed. Video source material is coded at QCIF spatial resolution (176×144 pixels) and temporal resolution of 10 frames per seconds. All video sequences are coded at 64 kbps using the TMN-10 rate control method.

4.1 Random Intra refresh with Error Concealment

In this section, we present results for the proposed random intra refresh pattern and frequency for a given PLR with the error concealment method described in Section 2.2 at the decoder. Results are presented for the video sequence **Foreman** in Figure 2 where *con* means concealment is used at the decoder and *I-MB* means random intra updating is used at the encoder. It can be observed that intra coding alone gives better results than error concealment alone. Combining random intra coding and error concealment at the decoder greatly improves video quality, by as much as 7 dB at a PLR of 20% for this video sequence.

4.2 RD Optimized Mode Selection with Error Concealment

In this section, we present results for our proposed method described in Section 3, where the error concealment is also applied at the encoder and mode decision is based upon Equation (4). If RTP is used, the value p can be obtained from the receiver report mechanism of RTPC [6]. The same error concealment technique is always employed at the decoder. We compare the proposed method with random intra coding at the encoder and present results for the video sequences **Foreman** in Figure 3 and **Coastguard** in Figure 4. As much as 2 dB can be gained using our new method. Figure 5 shows the Y-PSNR of the first 50 decoded frames of the sequence **Foreman** at a PLR of 20% for random intra updating and the RD method. The same error pattern was used for both cases. Our method maintains a higher reproduction quality throughout. The subjective quality is also significantly improved, as seen in Figure 6, where the decoded frame number 40 of the sequence **Foreman** is shown. Again, the same error pattern was used for the random and RD optimized methods.

In the random updating case, every coded macroblock is updated once every five frames. Using our method, only macroblocks significantly affected by packet losses and/or where error concealment fails will be more often intra coded. These regions are usually regions of interest in a typical videophone or video conferencing application. Low activity regions and static backgrounds are usually well concealed, and need not be coded accurately. This can be observed in Figure 6 where the background in the random update case is of better quality than in the RD case, but the facial expression is much better reproduced in the RD case. At 20% packet loss rates, our method still provides very useful video, whereas the usability in the other reproduced video is questionable.

5 Conclusions

We have shown that the use of intra coding for video transmission in a packet lossy network such as the Internet can significantly improve video reproduction quality. Random intra coding of the macroblocks with simple error concealment was shown to offer satisfactory results. However, when error concealment is considered with an RD optimized framework, more intelligent coding mode selection is possible, and such can significantly improve error resilience to packet losses. We proposed a method that yields very good RD tradeoffs given the packet loss rate and error concealment constraints. The proposed method does not require the use of a feedback channel. Information about the packet loss rate can be obtained from the network or a priori assumed. The negotiation between encoder and decoder on the error concealment method can be achieved by external means.

References

[1] R. Talluri, "Error resilient video coding in the MPEG-4 standard," *IEEE Comm. Magazine*, vol. 26, no. 6, pp. 112–119, June 1998.

[2] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. on Circuits and Syst. Video Technol.*, vol. 8, no. 6, Nov. 1998.

[3] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. IEEE, Special Issue on Video for Mobile Multimedia*, 1999, to appear.

[4] M.R. Pickering W.S. Lee, M.R. Frater and J.F. Arnold, "A diversity-based scheme for reducing error propagation in video," in *International Conference on Image Processing*, Santa Barbara, CA, Oct. 1997.

[5] ITU Telecom. Standardization Sector of ITU, "Video Codec Test model near-term, number 10 (TMN10)," *H.263 Ad Hoc Group*, Apr. 1998, Available from ftp://standards.pictel.com.

[6] R. Frederick H. Schulzrinne, S. Casner and V. Jacobson, "RTP: A transport protocol for real-time applications," *RFC 1889*, Jan. 1996.

[7] Y. Wang and Q.F. Zhu, "Error control and concealment for video communication: A review," *IEEE Communication Magazine*, vol. 86, no. 5, pp. 974–997, May 1998.

[8] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *ICASSP-92*, San Francisco, CA, USA, Mar. 1992, vol. 3, pp. 545–548.

[9] N. Naka, S. Adachi, M. Saigusa, and T. Ohya, "Improved error resilience in mobile audio-visual communications," in *IEEE International Conference on Universal Personal Communications*, Tokyo, JAPAN, Nov. 1995, vol. 1, pp. 702–706.

[10] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Proc. Magazine*, Nov. 1998.

[11] G. Côté, B. Erol, M. Gallant and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Trans. on Circuits and Syst. Video Technol.*, vol. 8, no. 6, Nov. 1998.

[12] G. Côté M. Gallant and F. Kossentini, "Description of and results for rate-distortion based coder," in *ITU-T Study Group 16, Video Experts Group, Q15-D-49*, Tampere, Finland, Apr. 1998.

[13] C. Bormann, L. Cline, G. Deisher, T. Gardos, C. Maciocco. D. Newell, J. Ott, G. Sullivan, S. Wenger and C. Zhu, "RTP payload format for the 1998 version of ITU-T Rec. H.263 video (H.263+)," *Internet-Draft*, May 1998.

[14] H. Schulzrinne, "RTP profile for audio and video conference with minimal control," *RFC 1890*, Jan. 1996.

[15] S. Wenger, "Video redundancy coding in H.263+," in *Audio-Visual Services over Packet Networks*, Scotland, UK, Sept. 1997.

List of Figures

1 Performance of intra macroblock refresh with picture loss rate for **Paris**. 15

2 Performance of random intra macroblock refresh with error concealment for **Foreman**. 15

3 Performance of random vs RD optimized intra macroblock refresh with error concealment for the video sequence **Foreman**. 16

4 Performance of random vs RD optimized intra macroblock refresh with error concealment for the video sequence **Coastguard**. 16

5 Performance of random vs RD optimized intra macroblock refresh with error concealment at a PLR of 20% for the first 50 frames of **Foreman**. 17

6 Decoded frame no 40 of the sequence **Foreman**: (a) error-free, (b) random intra updating with PLR of 20%, and (b) RD optimized intra updating with PLR of 20%. 18

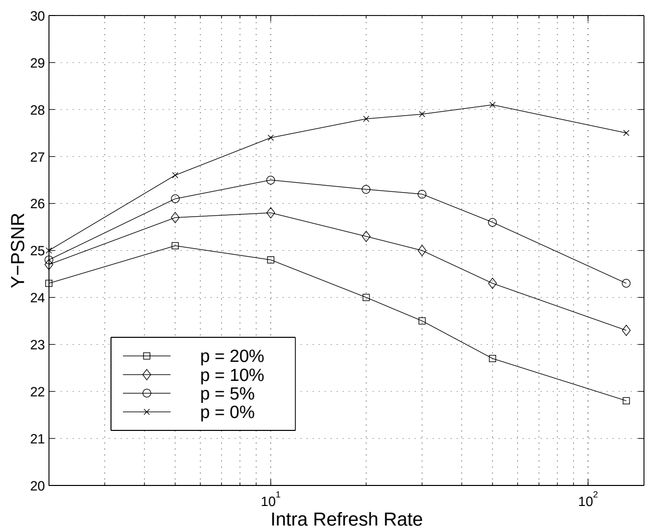

Figure 1: Performance of intra macroblock refresh with picture loss rate for PARIS.

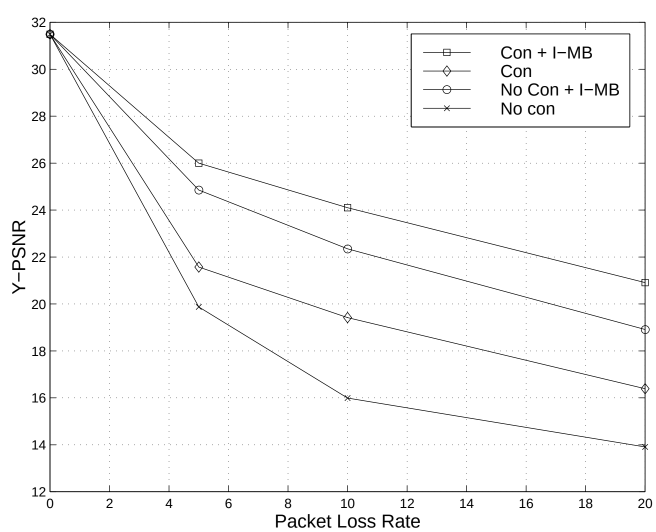

Figure 2: Performance of random intra macroblock refresh with error concealment for FOREMAN.

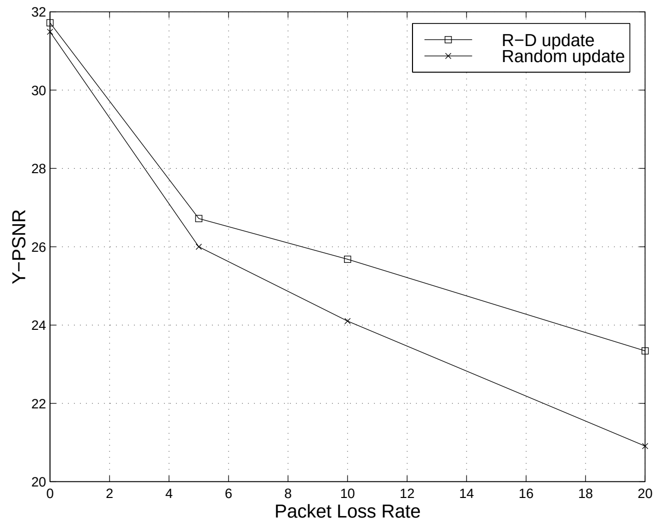

Figure 3: Performance of random vs RD optimized intra macroblock refresh with error concealment for the video sequence **Foreman**.

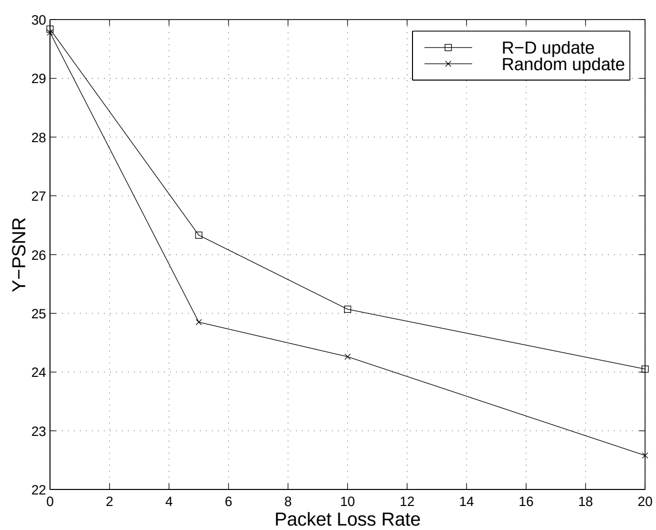

Figure 4: Performance of random vs RD optimized intra macroblock refresh with error concealment for the video sequence **Coastguard**.

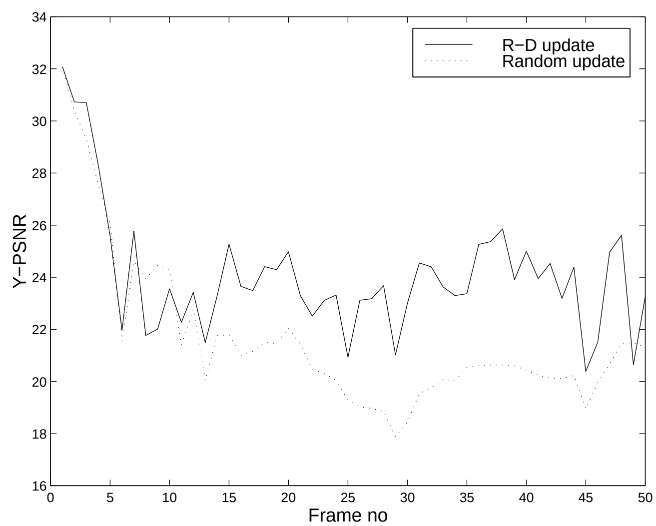

Figure 5: Performance of random vs RD optimized intra macroblock refresh with error concealment at a PLR of 20% for the first 50 frames of FOREMAN.

(a)

(b)

(c)

Figure 6: Decoded frame no 40 of the sequence **Foreman**: (a) error-free, (b) random intra updating with PLR of 20%, and (b) RD optimized intra updating with PLR of 20%.