ERROR RESILIENCE OF VIDEO TRANSMISSION BY RATE-DISTORTION OPTIMIZATION AND ADAPTIVE PACKETIZATION

Yuxin Liu[*], *Paul Salama*[**†], *Edward Delp*[*]

[*]Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN, 47907, U.S.A.

[**]Department of Electrical and Computer Engineering
Indiana University – Purdue University Indianapolis
Indianapolis, IN, 46202 U.S.A.

ABSTRACT

We propose new schemes to introduce error resilience into the compressed video bitstreams for transmission over packet networks. First, we develop an adaptive packetization scheme that prohibits any dependency across packets, for error resilience purposes, while exploiting the dependency within each packet to improve the source coding performance. Secondly, we address a two-layer rate-distortion optimization scheme to serve our packetization method. We also use Forward Error Correction (FEC) coding across packets to provide further error protection. Finally, we present a simplified version of our schemes to make it fully compliant with the current ITU video coding standard – H.263+.

1. INTRODUCTION

In this paper we address the problem of video transmission over packet networks. In particular, our schemes are designed to cope with packet loss during transmission across packet networks. Packet loss can result in quality degradation of the transmitted compressed video stream. This is due to the fact that current video coding standards such as the ITU standard H.263 version 2 (H.263+) adopt motion estimation and differential coding schemes [2]. Such coding schemes introduce dependency between different parts of the bitstream, which results in error propagation when packet losses occur. This problem has attracted great attention recently due to the rapidly growing demand for Internet video streaming services [1,3,4].

2. BACKGROUND

2.1 Error resilience in H.263+

The H.263+ standard specifies sixteen negotiable coding options that are denoted as annexes that further improve coding efficiency and support additional capabilities. Among them Annex K, the Slice Structure mode, Annex N, the Reference Picture Selection mode, and Annex R, the Independent Segment Decoding (ISD) mode, are designed to introduce error resilience elements into the bitstream.

Annex K introduces the slice structure that replaces the original GOB layer in each frame. Every macroblock in a frame is assigned to one and only one slice. A slice can be a rectangular area, in units of macroblocks, or it may contain a sequence of macroblocks in lexicographic order. All the slices in one frame can be encoded in lexicographic order, or in any arbitrary order. In order to do so, Annex K prohibits motion vector prediction, overlapped block motion compensation (OBMC), and the Advanced INTRA coding mode (Annex I) from being implemented across slice boundaries. Annex K provides a more flexible structure, compared to GOBs, so that frames can be segmented into slices as needed. Moreover, the headers of slices can be used as resynchronization points in the bitstream. Notice that Annex K does not prevent dependency across slice boundaries in the reference picture for motion prediction purposes.

By turning on Annex R, dependencies across different segments in one picture can be further prevented. Annex R regards a single GOB, a number of consecutive GOBs, or one slice as one segment. Furthermore, Annex R demands that the segmentation of a frame that adopt motion estimation be the same as that of its reference picture. However, although Annex R allows each segment to be decoded independently at the receiver, the encoding process of the segment is not completely independent as long as motion estimation is adopted. Therefore, Annex R effectively prevents spatial error propagation, but cannot prevent temporal error propagation if a segment contains INTER-mode macroblocks.

2.2 Packetization and Reed-Solomon Coding

[†] This research was supported by a grant from the Indiana 21st Century Research and Technology Fund. Address all correspondence to P. Salama at psalama@iupui.edu.

A memo proposed by the Internet Society has specified an RTP (Real Time Protocol) payload header format for the H.263+ bitstream [7]. RTP packets can be transmitted out of order since the packet header contains its own time stamp, which realizes video transmission in a more flexible manner. It supports packet fragmentation at the GOB or slice boundaries, or at the macroblock boundaries. The packet size is flexible, and usually the maximum is around 1500 bytes. There is a trade-off between error resilience of the packetized bitstream and the packetization overhead. The smaller the packet, the less the information lost as a result of packet loss, but the more overhead introduced.

To further improve the reliability of the packetized bitstream, Forward Error Correction (FEC) by Reed-Solomon coding is often implemented in the application layer for error protection [1,4]. An *(N,k)* Reed-Solomon code can correct up to $(N-k)/2$ symbol errors, and *(N-k)* erasures. Reed-Solomon coding can be applied across packets. By analyzing the packet header, the receiver can exactly locate the lost packets. Therefore, a lost packet is considered as an erasure error by the Reed-Solomon decoder. However, given a total bitrate for the overall video transmission system, using FEC will result in a reduction of the source coding rate.

2.3 Rate-distortion optimization

Since Shannon's theoretical rate-distortion bound might not be practically achieved, an operational rate-distortion optimization method, which uses Lagrange multipliers to find the optimal operating point, is used:

$$J_{opt} = \min_{i \in I_{mode}} (D_i + \lambda R_i), \quad \text{subject to } R \le R_{budget} \qquad (1)$$

where R_{budget} denotes the given overall data rate. A review of rate-distortion optimization techniques for video transmission in both error-free and error-prone environment is given in [5].

A key problem for rate-distortion optimized video coding is how to measure and predict the distortion. There are two kinds of distortion associated with a decoded video signal after transmission over lossy networks: a deterministic distortion caused by quantization on the source coding stage, and a distortion caused by the packet loss. Error resilience elements, FEC coding, together with appropriate packetization scheme all aim at reducing the distortion caused by the lossy channels. The introduction of above schemes will inevitably demand a larger portion of the total data rate, which results in less data rate assigned to pure source coding and hence larger quantization distortion incurred. It is critical to create an appropriate quality metric that creates a compromise between the above two kinds of distortion [1].

3. PROPOSED SCHEMES

3.1 Adaptive packetization

Techniques that address robust video transmission over packet networks need to simultaneously optimize three bit-allocation problems: the bit allocation between source coding and channel coding, the bit allocation that introduces appropriate error resilience into the bitstream, and the bit allocation between the coded bitstream and packetization overhead.

Adopting the INTRA mode prevents error propagation and achieves resynchronization of the bitstream, since an INTRA coded frame is independent of all the other portions of the bitstream from both the encoding and decoding points of view. Nevertheless, INTRA coding is also the most bit-consuming scheme since it does not fully exploit the redundancy within the video signals.

As an alternative, we can exploit ISD (Annex R) in conjunction with the slice structure (Annex K) for the sake of error resilience. With packet fragmentation occurring at the segment boundaries determined by ISD, we can guarantee that each packet can be independently decoded. Nevertheless, the "independency" of ISD is evaluated only from the decoder's perspective, not the encoder, since the dependency still exists across packets by the adoption of motion prediction. If one packet is lost, all the information it carries will not be available, and thus all the packets whose motion information is based on that packet will be seriously affected.

In fact, for video transmission over packet networks, data loss always occurs in units of packets. Therefore, we only need to introduce error resilience to prevent dependency across packets, and we can take advantage of any dependency within each packet to improve the coding efficiency. Therefore, we propose a new packetization scheme, which prohibits any kind of dependency across the boundaries of packet while trying to take full advantage of the dependency within each packet.

The idea is as follows: In the source coding stage, we divide each frame into several segments, as is done in Annex R. We turn on any optional coding mode to exploit the dependency within each segment while treating the segment boundaries in a same way as picture boundaries. In the packetization stage, we place the segment into the same packet as its reference segment (if any). If it cannot be fit into that packet, it will be intra coded instead and a new packet started. For example, if one GOB is taken as one segment, then the encoding and packetization processes obey the following principles:

1. *No dependency across the GOBs in one picture, which prohibits motion prediction, OBMC, and advanced INTRA block prediction across GOB boundaries;*
2. *If there is at least one macroblock in a GOB that is encoded using the INTER mode, it is placed in the same packet as its reference GOB;*
3. *If a GOB cannot fit into the packet containing its reference, all of its macroblocks are encoded using the INTRA mode, and a new packet started;*

4. *The number of GOBs in one packet is constrained by the predefined maximum packet size, and packet fragmentation is always implemented at the GOB boundaries;*
5. *Each GOB can be referenced at most once for motion estimation.*

For an H.263+ encoder with Annex R turned on, our packetization scheme is implemented to packetize a series of consecutive segments having the same position into the same packet. A packet always starts with an INTRA coded segment, or contains an INTRA segment while the remaining segments' motion vectors are obtained from that INTRA segment if backward motion estimation is employed.

3.2 Two-layer rate-distortion optimization

Similar to Annex R, we can take one or several GOBs or any rectangular slice as a segment in our scheme. For simplicity, in this paper we take one GOB as one segment. We propose a two-layer rate-distortion optimized coding scheme to serve as our packetization method. As in [1], we design four modes for each macroblock: INTRA, INTER, INTER4V, and SKIP. In Annex R, the motion vectors are only allowed to refer to the same area as the current segment in the reference picture. We loosen this constraint first to allow the current GOB to refer to any GOB at any position in the reference picture. This is done to improve source coding performance.

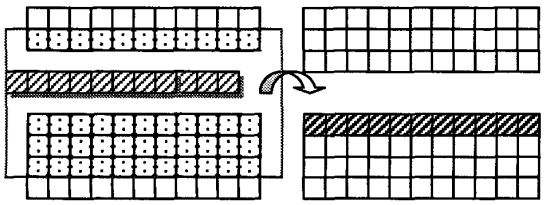

Figure 1: Reference GOB selection by rate-distortion optimization

As shown in Figure 1, for the current GOB (dark shaded blocks in the current picture) to be processed, we decide on the candidate GOBs for motion estimation based on the range of motion vectors. In H.263+, if Annex D (Unrestricted Motion Vector mode) is turned on, the search range can be as large as $[-32,31.5]$ for QCIF pictures. Therefore, we select five GOBs – the GOB located in the same position as the current one and two above and two underneath as the candidate reference GOBs. These are the shaded regions in the reference frame enclosed within the bounding rectangle in Figure 1. For each reference GOB, we implement a first layer rate-distortion optimization scheme that determines the optimal coding mode, whether INTRA, INTER, INTER4V, or SKIP, for each macroblock in the current GOB, given the chosen reference GOB. Note that the search window is limited within the reference GOB area. A second layer rate-distortion optimization is then used to select the final reference GOB out of all the possible GOBs by choosing the one that minimizes the following sum:

$$k_{opt} = \arg\min_{k \in I} \sum_{i \in Curr\_GOB} J_i^{(k)} \qquad (2)$$

where $J_i^{(k)}$ is the optimal rate-distortion Lagrangian obtained in the first layer optimization by Eq. (1), which is obtained based on the kth reference GOB for the ith macroblock in the current GOB. Finally, we encode each macroblock with the optimal mode obtained when the optimal GOB is referenced. We notice that the central area in each picture usually contains more significant information than the rest. Therefore, we start with the central GOBs and proceed to the top and the bottom. For the nine GOBs in a QCIF picture ordered 1 through 9 from the top to the bottom, for example, we process the GOBs in the following order:

$$\{5, 4, 6, 3, 7, 2, 8, 1, 9\}$$

To further improve coding performance, we adopt Annex D and Annex F (Advanced Prediction mode). We extrapolate the edge area of the GOB, interpolate it to generate the half-pixel values, and employ the OBMC scheme. Since the unrestricted motion estimation mode is adopted, we have to signal the information regarding which GOB is selected to be the reference for the current GOB, which makes the proposed scheme not fully compliant with the H.263+ standard.

Considering that the above scheme is not fully compliant with H.263+, we simplify our scheme where the reference GOB is always the one in the same position as the current GOB, which is consistent with Annex R of the standard. From the experimental results present in the next subsection, we will see that this scheme is applicable since the encoder always tends to select the one in the same position except in the case complex motion or when scene changes occur.

4. EXPERIMENTAL RESULTS AND CONCLUSION

In our experiments we chose the *Foreman* sequence due to its complex motion, zoom-in, zoom-out and scene changes. All frames are 4:2:0 YUV QCIF sized frames. The sequence is 400 frames in length. For simplicity, we use PSNR as the distortion metric for each decoded frame.

First we encoded *Foreman* at a date rate of 56 kbps and frame rate of 10 fps with our two-layer rate-distortion encoding method. Error resilience is introduced to the bitstream by our encoder in which a GOB only depends on at most one GOB in the reference picture, which results in 1.5dB loss in PSNR (see Figure 2). We observed that the encoder is much likely to choose the same segment in the reference picture based on the two-layer rate-distortion optimization. Only 11 frames out of the total 130 encoded P-frames included segments referring to an area other than the same segments in the reference picture. Those frames

are around the 80th encoded P-frame for *Foreman* where scene changes occur. Therefore, we can take advantage of Annex R in H.263+, which demands the same segmentation between two I-frames to replace the second layer rate-distortion optimization in our scheme. Notice that each reference GOB was treated in the same way as a reference picture. This included extrapolating the edge area and realizing OBMC. The decoded video quality in PSNR of the simplified scheme is 30.12dB in PSNR with 0.05dB loss compared to the two-layer scheme.

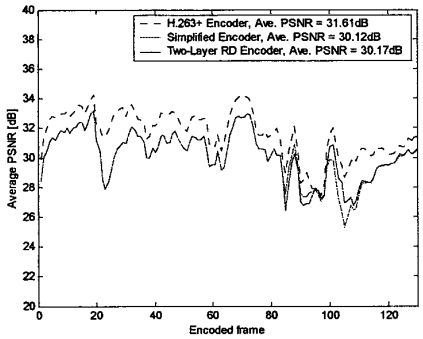

Figure 2: Source encoding for Foreman

Next we use our schemes with error-prone packet networks. It has been shown in [6] that INTRA refresh rate is a critical parameter for error resilience. We map this parameter to the number of GOBs contained in each packet in our simplified scheme. In the packetization stage, we place three GOBs in one packet, which always starts with an INTRA GOB. Therefore, every three frames are fragmented into nine packets, altogether containing nine INTRA GOBs, which is equivalent to setting the INTRA refresh rate to be 1/3 frame, i.e., on average 1/3 of the macroblocks in each frame are forced to be INTRA. Without adopting Reed-Solomon coding, the received video quality is shown in Figure 3, after de-packetization and decoding. The red curve denotes the distortion when no packets are lost, while the blue one denotes the distortion with packet loss at rate 5%. By using Reed-Solomon coding, we can keep the packetized bitstream almost intact at lower packet loss rate while achieving similar performance at higher packet loss rate as in the lower rate case without employing error correction coding.

Notice that for some frames the PSNR drops more than 5dB each, as for the 15th to 21st encoded P frames shown in Figure 3. This is because one packet loss means three consecutive frames suffer in the same location, and our current error concealment method simply copies the same located macroblock in the previous frame if the current macroblock data are not available. We can adopt better error concealment method for future work to improve the performance. For example, we can choose the macroblock in the previous picture with the most matched neighboring area as the neighbors of the current lost macroblock [8]. Alternatively, we might introduce additional redundancy to facilitate error concealment by associating a neighboring segment with the current segment's motion information.

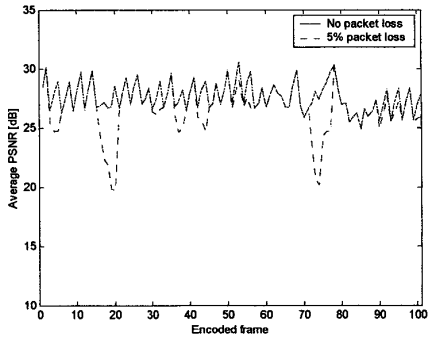

Figure 3: Foreman transmitted over 5% packet loss network

For time sensitive applications such as video conferencing, we have to take into account the delay constraint. Since a packet is composed of segments across consecutive pictures in our method, the larger the packet is, the more delay it will cause. For future work, we will consider the trade-off between delay and error resilience performance. We notice that if FEC is adopted for error protection, our method does not cause additional delay compared to common packetization schemes.

5. REFERENCES

[1] Michael Gallant and Faouzi Kossentini, "Rate-Distortion Optimized Layered Coding with Unequal Error Protection for Robust Internet Video," IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 357-372, Mar. 2001.

[2] ITU-T Recommendation H.263, "Video Coding for Low Bit Rate Communication," Feb. 1998.

[3] Philip Chou and Zhourong Miao, "Rate-Distortion Optimized Streaming of Packetized Media," submitted to *IEEE Trans. Multimedia*.

[4] K. Stuhlmüller, M. Link, and B. Girod, "Robust Internet Video Transmission Based on Scalable Coding and Unequal Error Protection," *Image Communication*, Special Issue on Real-time Video over the Internet, pp. 77-94, 15(1-2).

[5] Antonio Ortega and Kannan Ramchandran, "Rate-Distortion Methods for Image and Video Compression," IEEE Signal Processing Mag., pp. 23-50, Nov. 1998.

[6] Yuxin Liu, Christine Podilchuk, and Edward Delp, "Evaluation of Joint Source and Channel Coding over Wireless Networks," to be submitted.

[7] C. Bormann, et. al., "RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+),"RFC 2429, Oct. 1998.

[8] P. Salama, N. B. Shroff, and E. J. Delp, " Error Concealment in MPEG Video over ATM Networks," *IEEE Journal on Selected Areas in Communication,* vol. 18, no. 6, pp. 1129-1144, June 2000.