A ZERO ERROR PROPAGATION EXTENSION TO H264 FOR LOW DELAY VIDEO COMMUNICATIONS OVER LOSSY CHANNELS

Oztan Harmanci and A. Murat Tekalp*

(harmanci, tekalp)@ece.rochester.edu

Electrical and Computer Engineering Dept., University of Rochester, Rochester, NY 14627 *also with College of Engineering, Koc University, Istanbul, Turkey

ABSTRACT

In this paper, we introduce a method for video transmission over lossy channels. The proposed method is similar to NEWPRED[1] and uses feedback information to stop error propagation. Main operational difference is replication of packet loss and error concealment process at the encoder and using the possibly error concealed frames as reference frames in the motion prediction loop. We have studied the optimization of proposed method under moderate error rates where a both ACK/NACK mode NEWPRED does not efficiently solve the problem. The proposed solution is an adaptive method that optimizes the operation based on the expected error propagation. We present comparison results with NEWPRED under various channel conditions and video sources.

1. INTRODUCTION

Error propagation is one of the important factors that affect the video transmission performance in a lossy environment. A number of methods have been proposed to reduce the impact of error propagation (see [2] for a detailed survey). These methods include (but are not limited to) using error tracking with intra updates to recover the affected areas, confining the error propagation in a limited area by constraining the motion search area and using resynchronization markers to recover from errors.

One of such successfull methods is NEWPRED[1]. In NEWPRED, two types of operational modes exist: ACK and NACK. In the ACK mode, decoder sends an ACK for every successfull packet reception. In this mode, encoder uses only the successfully received areas as reference. In the NACK mode, decoder sends a NACK only when a packet is lost. In this mode, encoder can use any frame as reference. When the encoder receives a NACK, it should switch to ACK mode temporarily until the error propagation is stopped. The mode switching process allows a certain degree of adaptation.

NEWPRED is most efficient in low delay communications. If delay increases, both encoder and decoder need bigger frame buffers. Also, the coding efficiency decreases because encoder has to use frames further back in time, where temporal correlation is lower.

Being developed as a standard compliant method, NEW-PRED has certain limitations. For example, SKIP mode macroblock coding will hardly be used because SKIP mode refers to the immediately previous frame. This requires that feedback delay should be as low as one frame time which is not realistic.

Another limitation is that commonly the search range is limited. If the search area on the reference frame is lost, encoder has to step one frame further back using reference frames even farther back in time. This causes longer motion vectors and bigger frame indices, thus decreases encoder performance.

Finally, NEWPRED cannot make use of the observation that certain losses can be concealed efficiently if data partitioning (DP) is used with appropriate unequal error protection (UEP)[3]. This is because when DP is used, if decoder receives only prediction part, it can still make use of it and most of the time concealment performance is high even though residual data is lost. NEWPRED treats this as complete loss and the successfully concealed region is removed from the search range.

In the next Section, we present the proposed system that overcomes the shortcomings of NEWPRED discussed above. In Section 3, we extend the system and propose a novel method to make it adaptive to channel conditions. Section 4 contains the preliminary experimental results and conclusions are given in Section 5.

2. PROPOSED SYSTEM

In this section, we present our system that is designed to overcome the weaknesses of NEWPRED. We base our system on two assumptions: (i) There is a reliable feedback channel from decoder to encoder. This is not a big assumption because the feedback channel requires very low bitrate. This can be easily combined with channel coding or integrated into Quality of Service. (ii) Encoder perfectly knows



Fig. 1. Block diagram of the proposed system



Fig. 2. Temporal operation of the proposed system

the decoder concealment method. This is a non-normative aspect of video coding and it is not defined by the standards. However, we have seen that in the literature([2]) on lossy video coding, it is common to assume that the encoder knows the decoder's concealment method. We have observed that any method that models the loss error assumes known decoder concealment method. We have not yet found a study that compares what would happen to their system if the assumed decoder model at the encoder side does not match the actual decoder model.

Figure 1 shows the block diagram of the proposed system. At the encoder side, there is an additional decoder that is out of the common video coding loop: it is coupled to the channel and receives feedback from actual receiver. This decoder uses channel feedback to learn the packet losses experienced at the decoder side. Using the packet loss information, it can simulate the concealment process of the actual decoder. Since the encoder side decoder needs to have the feedback information to perform concealment simulation, it runs later than the encoder. This delay is equal to the feedback delay imposed by the network and is called round trip time (RTT). RTT is the time for a packet sent from encoder to arrive at the decoder and come back to encoder again. Feedback information at the decoder is generated at the display time.

The encoder uses modified frames in the frame buffer as reference. When the encoder side decoder receives feedback about a frame, it replicates the errors and concealment process. The generated error-concealed frame overrides the corresponding frame in the frame buffer. Any frame that has not been updated by feedback is not used as a reference frame.

Figure 2 shows the temporal operation of the system. Assume that current time is T_e and current frame to encode is $F_K + N_D$. The feedback message that belongs to frame F_K , which was encoded at $T_e - RTT$, is received at the encoder after a time RTT. Assuming that the RTT jitter is limited, N_D is defined as the maximum RTT in terms of frame time. In such a case, feedback information for all of the frames before N_D th frame will be received and encoder can use any frame before N_D th as reference. Note that N_D is indexed relative to current frame.

Although the proposed system is similar to ACK mode NEWPRED, main differences are as follows:

- The proposed system fixes the reference frames to the N_D th and previous frames. In NEWPRED, it is not possible to tell beforehand which frame will be used because even though RTT is bounded, the motion search area in the reference frame could be lost.
- Another difference is that, in the proposed system the lost and concealed regions can be used for reference, which is an advantage over NEWPRED. This allows the proposed system to efficiently utilize the DP which is important for UEP and successful concealment.
- Finally, our method provides the use of SKIP mode macroblocks by making a SKIP mode macroblock refer to the N_Dth frame instead of the frame that is immediately before the current frame.

The disadvantage of the proposed system is that it is not standard compliant. We change the syntax of SKIP mode and assume that the decoder error concealment method is known at the encoder. Also, to be able to use SKIP mode we extend the syntax and add the value of N_D to the slice header.

3. ADAPTIVITY TO CHANNEL CONDITIONS

Until now, we discussed the system when the channel error is high and fixed the beginning for reference frame search to N_D th. Similarly we discussed NEWPRED in ACK mode.



Fig. 3. Trade-off in choosing N_A when $N_D = 4$

However, it is obvious that when there are no errors in the channel, we will experience a source coding loss as shown in figure3. In the context of NEWPRED, this shortcoming is remedied as follows: Sender runs in NACK mode until an error is signalled from reciever in the form of a NACK message. When the sender receives this NACK, it switches to ACK mode until the error propagation is stopped and then goes back to NACK mode again. Although this provides a certain degree of adaptivity, the process is not that trivial. Consider the case when error rate is moderate. There will be many ACK-NACK mode switches, and during each NACK-to-ACK switch, there will be error propagation. When there are too many mode switches, severe error propagation will be experienced.

A similar problem is investigated in [4] in detail and promising results are presented. Their solution works at macroblock level to model the error propagation. The drawbacks of their method is (i) Optimizing the lossy mode decision and reference picture selection process for each macroblock is computationally expensive. (ii) The assumptions, which their error modelling is based on, is not suitable for reallife. For example, the error propagation model presented uses a simple *copy from previous frame* based error concealment. Better error concealment methods would further increase the complexity of the system. Furthermore their solution assumes that multiple packet losses following each other is highly unlikely. On the contrary, in packet networks, packet losses generally occur in bursts.

The adaptivity algorithm we propose for our system is a simplified form of [4]'s algorithm adapted for our system. We use a frame level error propagation model. Consider the following observations:

• When the channel error rate is high, sender should use N_Dth frame (and previous ones) as the reference

frames thus stop and prevent errors from propagating.

- When the channel error is low, sender should use the immediately available frame (and previous ones) as the reference frames. When a loss is detected, sender should temporarily switch to high error case until the error propagation is stopped and then switch back to low error case.
- For the intermediate error rates, sender should use frame N_A (and previous ones) (where $N_A > 0$ and $N_A < N_D$) as reference frames. When a loss is detected sender should temporarily switch to high error case until the error propagation is stopped and then switch back to low error case.

The third item in the above paragraph is the key observation. It stems from the fact that, for example if we are using N_A 'th frame as the beginning of the reference frames, when an error occurs that error will propagate for only $N_D - N_A$ frames. When N_A is close to N_D , the effect of error propagation will be less because error propagation will be recovered quickly. However, when N_A is close to 0, the loss in encoding efficiency will be low.

This trade-off is presented in figure 3. Loss in encoding efficiency(LIEE) is the decrease in sequence pSNR compared to using 0th frame as beginning of reference frames in a lossless channel, therefore it is zero when N_A is equal to 0. Notice how distortion due to error propagation(DDEP) becomes zero when N_A is equal to N_D . DDEP depends on channel loss rate and the video content whereas LIEE depends only on video content.

Optimal N_A is the point where LIEE plus DDEP is minimized. It is straightforward to calculate the LIEE by simply re-encoding the frame for each N_A . Accurate calculation of DDEP is not possible, because in a videoconferencing situation we do not have access to the future frames. However, we can exploit the temporal correlation in the sequence. If we assume that the video behaviour in the last N_D frames is similar to the upcoming N_D we can use the information from previous behaviour and predict the DDEP.

4. PRELIMINARY EXPERIMENT RESULTS

We have used H264[5] for the experiments. Non-normative aspects such as optimal encoding and decoder concealment process are performed according to [6]. We are using high complexity mode decision and motion search.

The system aims to work in a low delay environment. Therefore, it tries to match the target rate for each frame. To achieve this, a frame is re-encoded with different quantization levels until the frame bitrate matches the target rate. The Lagrange optimization used in macroblock mode decision and motion vector selection inherently uses the quantization level to adjust the bitrate.



Fig. 4. RD curves for Foreman Sequence

We performed the experiments in high loss mode. For the proposed encoder, the reference frames always began from N_D th frame and the NEWPRED simulations are done in full ACK mode. Channel adaptivity comparisons will be included in the final version of the paper.

We used 10Hz QCIF foreman and carphone sequences. We put 3 slices per frame making 33 macroblocks per slice. It is assumed that network RTT and jitter is bounded and the bound is equal to 300milliseconds. This makes $N_D=2$ using a 0 based indexing.

The packet channel loss simulation is performed based on a 2 state Markov model. Packet loss rates that are discussed here are the average loss rates. Channel bitrate is 96Kbps and there is no channel coding involved.

A total of 100 frames are encoded in each simulation. Each simulation is performed 20 times and result is averaged.

Figures 4 and 5 compare the rate distortion(RD) performances of the proposed method and NEWPRED for Foreman and Carphone sequences, respectively. The average loss rate is equal to 0.1. On the average, the proposed system results in about 0.5 dB higher SNR.

We are currently working on adaptivity performance results. They will be presented in the final version of the paper.

5. CONCLUSIONS

In this paper, we presented an algorithm that allows zero error propagation for video communications over lossy channels. We demonstrated experimental results comparing the



Fig. 5. RD curves for Carphone Sequence

proposed algorithm with NEWPRED. We have observed that at high error rates, proposed algorithm performs about 0.5dB better in sequence pSNR.

We are currently working on the channel adaptivity of the proposed system. The final system will adaptively choose the optimal N_A based on input video source and channel conditions.

6. REFERENCES

- LBC-95-033 Telenor RD ITU-T, SG15/WP15/1, "An error resilience method based on backchannel signalling and FEC," 1996.
- [2] W. Wang, S. Wenger, J. Wen, and K.Katsaggelos, "Review of error resilient coding techniques for real-time video communications," *Signal Processing Magazine*, July 2000.
- [3] M. Budagavi, W. Rabiner Heinzelman, J. Webb, and R. Talluri, "Wireless mpeg-4 video communication on dsp chips," *IEEE Signal Processing Magazine*, January 2000.
- [4] Thomas Wiegand, Niko Farber, Klaus Stuhlmuller, and Bernd Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE JSAC*, 2000.
- [5] ISO/IEC JTC1 JVT-I050, "Advanced video coding," .
- [6] ISO/IEC JTC1 JVT-I049, "Joint model reference encoding methods and decoding concealment methods,"