OPTIMIZATION OF 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 study the data partitioning (DP) and its optimization for H264 video coding standard. H264 does not include DP in baseline profile, which is the most suitable profile for low delay, low complexity, and loss prone environments. To analyze the optimization of DP, we first introduce the concept of subchannels to abstract the physical layer. This allows us to move channel coding from application layer to physical layer. Then, we build the video encoder system around NEWPRED[1] so that error propagation and its analysis is eliminated. Finally, we provide macroblock and slice level optimizations that result in optimal mode decisions and Unequal Error Protection (UEP) rates for data partitions. Experimental results show about 0.5dB performance increase compared to no data partitioning.

1. INTRODUCTION

The motivation behind this work is to explore data partitioning in H264 video coding standard. Although in various researches it is established that a UEP scheme combined with DP [2] results in a significant increase in performance compared to no DP case, DP is not included in H264's baseline profile [3]. Baseline profile is the most suitable profile for low delay and low complexity applications. Such applications include video-telephony where the channel may be wireless and therefore the transmission is lossy. For these applications, good error concealment plays an important role in achieving high performance.

Data partitioning allows separation of residual coding data and prediction data[4]. Encoder can enforce better loss protection for the prediction partition. If the decoder fails to receive the residual coding data, it can use the prediction information to perform satisfactory loss concealment.

In literature, UEP is generally used with scalable video coding such as [5]. However, it is observed that data partitioning in non-scalable coding can benefit from UEP too. [2] does an experimental analysis on how to find the optimal UEP rate for data partitioned MPEG-4 video. However [2]'s work is not adaptive to different channel and source video conditions.

Furthermore, for reasons which we will discuss in the upcoming sections, we would like to remove channel coding from the application layer and let the physical layer handle it. Since physical layer is not aware of which packets are more important, it treats all packets equally and uses same error protection for all. To avoid this and fully utilize UEP, we define an interface between application and physical layer that allows application layer to select a packet's channel coding parameters. To achieve this, we will present an abstraction scheme to expose certain functionalities of the physical layer to the application layer.

We approach the problem in a videoconferencing framework. As such, we realize that the problem requires relatively low delay communications and stable short term output bitrate. We choose to use ACK mode NEWPRED to remove the error propagation and help estimate the loss distortion more accurately.

In the next section, we provide a channel abstraction method that enables the application layer to use channel coding in the physical layer. In Section 3, we present the proposed joint source channel coding framework for macroblock and slice level optimizations. In Section 4, we present the experiment results. Conclusions are presented in Section 5.

2. ABSTRACTING THE PHYSICAL LAYER

In various researches ([6], [5], [7]), channel coding is done at the application layer. Although this allows for more flexibility, it has certain limitations. First of all, independent of the application layer, some form of channel coding is already employed in the physical layer([8]). Secondly, since bit errors complicate the routing and signaling process, the physical layer drops a packet if it fails to pass the error detection stage. Therefore, the application layer may not receive packets with bit errors in it. Finally, physical layer is aware of the link state and it is where the necessary channel coding should be done.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Keeping these points in mind, we target a 3GPP-based system and expose required functionalities to the application layer to achieve better cross layer optimization. We follow the physical layer operation as suggested in [9].

We assume that the physical layer has N subchannels (SC_i, $i = 1...N$). Application layer can query the physical layer for available subchannels and subchannel parameters at any time. Following items summarize the functionalities exposed by the physical layer.

1- Each subchannel has a code rate (CR_i) associated with it.

2- Application layer can query the physical layer to learn the packet loss probability ($PLP_i(packetsize, t)$) through SC_i. PLP_i is a function of time and packet size.

3- Application layer can query the physical layer to learn the energy required to transmit a packet ($ERT_i(packetsize, t)$) through SC_i. ERT_i is also a function of time and packet size.

For now, we will not focus on power optimization and therefore only the first two items are of use to us.

By following this abstraction, the application layer can make most of the physical layer functionality without knowing its inner workings. This is possible because the subchannel definition process is completely left to the physical layer. Application layer does not need to know what kind of channel coding is used in physical layer or what parameters are used for a specific subchannel.

3. OPTIMIZING THE JOINT SOURCE CHANNEL CODING

A number of studies ([7] for a detailed survey) exist on estimation and modeling of the error propagation in lossy video communications. In this paper, we would like to perform accurate mode analysis and decision-making. Therefore, we choose ACK mode NEWPRED as the mode of operation and remove the error propagation.

3.1. Optimizing the Macroblock Mode Selection

In [9], loss aware macroblock mode selection is optimized by using N decoders at the encoder side. Then, the average coding and error concealment result is taken as the distortion introduced by the channel. This distortion is used in the Lagrangian optimization process as described in [[10], [11]]. Based on the fact that the N-decoder system must know the channel loss model, we simplify the mode decision process into a probabilistic framework. This allows for significantly reduced complexity.

The proposed mode selection process replaces the N-decoder system as follows. We calculate the expected distortion according to the loss probabilities of the partitions that the macroblock belongs to. Let's say that at the time of encoding macroblock MB_i, the probability that prediction data partition will be lost is P_1 and the probability that residual data partition will be lost is P_2. For mode selection, lossless Lagrange optimization chooses the mode that minimizes the function $L = D_{mb} + \lambda_{mb} R_{mb}$, where D_{mb} is source coding distortion and R_{mb} is the number of bits of the macroblock for a given mode. In the proposed system, we replace D_{mb} with its expected value $E[D_{mb}]$. Since we are using ACK mode NEWPRED there is no error propagation and $E[D_{mb}]$ depends on (i) each partition's loss probability, (ii) decoder loss concealment method and (iii) source coding distortion. There are three cases that should be considered in the computation of $E[D_{mb}]$:

1- None of the partitions are lost. This happens with probability $P_o = (1-P_1)(1-P_2)$. We denote the resulting distortion with D_o (only source coding distortion).

2- Residual data partition is lost and prediction data partition is received. This happens with probability $Pp = (1-P_1)P_2$. We denote the resulting distortion with D_p (concealment using prediction information)

3- Prediction data partition is lost. This happens with probability $P_c = P_1$. It does not matter whether residual data partition is received or lost since residual information is useless without the prediction information. We denote the resulting distortion with D_c. (concealment without any information at all)

$$E[D_{mb}] = P_o D_o + P_p D_p + P_c D_c \quad (1)$$

We pose the new Lagrange optimization problem as

$$E[L_{mb}] = E[D_{mb}] + \lambda_{mb} R_{mb} \quad (2)$$

The macroblock mode that minimizes $E[L_{mb}]$ is chosen as the optimal mode. There are a couple points to note here: (i) The distortions D_o, D_p and D_c are macroblock level distortions. (ii) The optimal mode computed here is the optimal mode for the current pass. However, due to rate control, multiple passes may be needed and thus the optimal macroblock mode may change until the final pass.

3.2. Optimizing the Subchannel Selection: UEP optimization of Data Partitions

3.2.1. Data partitioning in H264

There are three types of partitions in H264: (i) prediction data, (ii) intra macroblocks' residual coefficient data, (iii) inter macroblocks' residual coefficient data. In our experiments, we observed that sometimes the number of intra macroblocks in a slice is too small. In such a case, the size of the 2nd partition is also small. This results in inefficient usage of bandwidth because of the TCP/IP/RTP headers. To avoid such a situation, we combined 2nd and 3rd partitions in one packet. H264 network adaptation layer (NAL) format is still preserved in each partition. Therefore, we treat

the system as if there are only 2 partitions, partition 1 being the prediction information and partition 2 being the residual information.

3.2.2. Optimization of UEP

We assume that each slice is partitioned into two parts: 1st part contains prediction data and 2nd part contains residual data. For simplicity of analysis, equal number of macroblocks are packed into all slices. Therefore, there is a fixed number of slices in each frame. Each slice is partitioned into transmission units(TU) and transmitted through the network independently. Let TU_{fji} denote i'th partition of j'th slice of f'th frame, where $i = 1, 2$. Similar to the source coding, we use a Lagrange optimization scheme to determine the optimal subchannel (SC_{k^*}) for a given TU. The problem can be defined as follows.

Given a set of subchannels SC_k, $k = 1...N$, determine k^* that will minimize the expected distortion $E[D_{TU}]$ for TU_{fji}.

Since we are using NEWPRED, there is no error propagation and $E[D_{TU}]$ depends on; (i) the loss probabilities of $TU_{fn(1,2)}$, where $n < j$, (ii) the loss concealment algorithm and, (iii) the source coding distortion. Similar to the calculation of $E[D_{mb}]$ given in Eqn. 1,

$$E[D_{TU}] = P_o D_{TU_o} + P_p D_{TU_p} + P_c D_{TU_c} \qquad (3)$$

where distortions D_{TU_o}, D_{TU_p} and D_{TU_c} refer to the distortion from all of the macroblocks in the TU.

We assume that the slices are independent from each other as we did in macroblock mode decision process. After encoding one slice, all possible subchannel modes are tested for all its TUs. When there are 2 partitions, optimal subchannel set will be (SC_{k1^*}, SC_{k2^*}) for (TU_{fj1}, TU_{fj2}). The set that minimizes the Lagrange function,

$$E[L_{TU}] = E[D_{TU}] + \lambda_{TU} R_{TU} \qquad (4)$$

is chosen for each TU. If there are N subchannels and 2 partitions, the number of possible subchannel sets is N^2 for one slice (all combinations of 1st and 2nd partitions of each TU through each subchannel).

When there is no data partitioning, a slice becomes only one TU but it still goes under the Lagrange optimization process. If there are N-subchannels, there are only N cases to test for a slice. This is the case of equal error protection (EEP).

As in the macroblock mode selection process, due to rate control, there may be multiple encoding passes and the final optimal subchannel assignments may change.

In our experiments, we the distortion metric used is the sum of squares distortion(SSD) and we used $\lambda_{TU} = \lambda_{mb}$ as suggested in [12].

4. RESULTS

The H264 video codec used in the experiments was written by us. Non-normative aspects such as optimal encoding and decoder concealment process were performed according to [12]. We slightly changed the error concealment algorithm so that concealment of a slice does not depend on the slices that come after it. Rather, it is performed in the order of encoding.This allows us to estimate the loss distortion without encoding future slices. We used high complexity mode decision and motion search as described in [12].

Rate control was done for a low delay environment where stable output rate in the short term is needed. Therefore, we performed per frame rate control. This stable output rate is the combination of both channel and source coding. Figure 1 presents a typical channel and source bitrate allocation during the coding of 100 frames.

Channel was simulated based on [10]. We improved the software suggested in [10] to enable channel coding at different rates. Loss model is a uniform bit error model. Channel coding is done with Reed-Solomon codes to allow for a closed form expression of residual error function. In our experiments, we used Reed-Solomon codes with 6 bit symbols. The block size is about 48 bytes, which corresponds to the size of a physical data unit in [9]. We used 6 subchannels with code rates approximately at: 12/12, 11/12, 10/12, 9/12, 8/12 and 7/12. Total bitrate of channel is 96kbps. Video input is QCIF at 10Hz.

A total of 100 frames were encoded in each simulation and each simulation was performed 20 times and the results were averaged.

Figure 2 shows the comparison between DP and no DP case for carphone and foreman sequences. At low error rates, there is not too much difference. However, as we increase the channel error rate, the channel coding starts consuming enough amount of bandwidth to start causing a visible difference. At the highest bit error rate we tested, which is 0.01, we observe that there is more than 0.5dB increase for carphone and about 0.4dB increase for foreman in sequence peak Signal-to-Noise Ratio(pSNR). This behaviour is due to the following factors:

When DP is not used, the correct protection rate selection for each partition can not be done efficiently. Forcing each partition to have the same protection (although the protection rate is optimized) may cause unnecessary protection of residual coding information, thus waste the bandwidth. At low error rates, required channel coding is also low. Therefore, the unnecessary channel coding of residual data does not waste too many bits. This explains the increasing difference pattern in the figures. Also, error concealment performance is higher with DP compared to no DP.

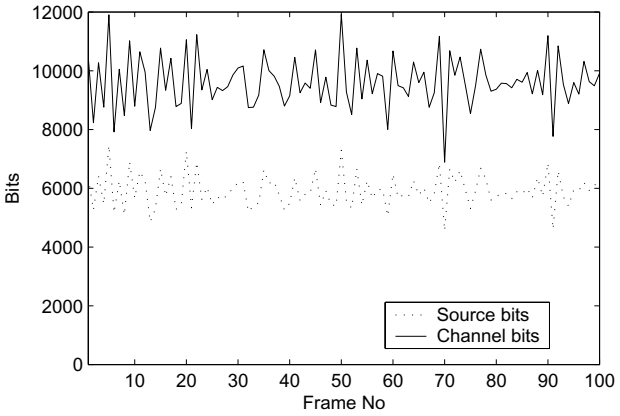

Fig. 1. Short Term Rate Control

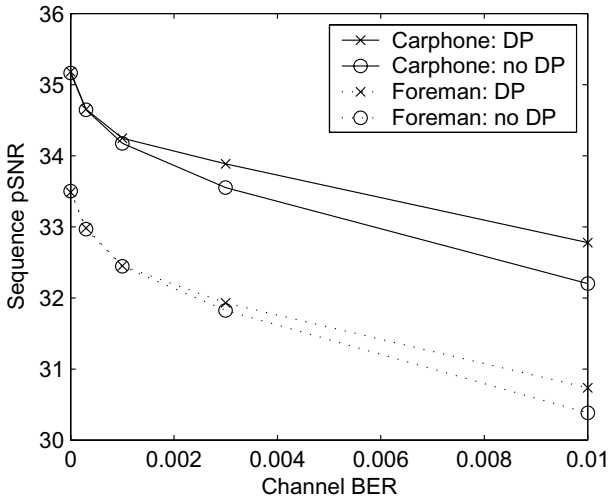

Fig. 2. Results of UEP vs. EEP

5. CONCLUSIONS

We have presented a method to implement and optimize UEP for DP and EEP for no DP using H264 video. We have compared the results and observed that even though there is no error propagation, at high bit error rates we achieved about 0.5dB gain by simply using UEP DP. Although DP is not included in the baseline H264 profile, we have shown that there are significant gains possible from a UEP optimized DP.

6. REFERENCES

[1] LBC-95-033 Telenor RD ITU-T, SG15/WP15/1, "An error resilience method based on backchannel signalling and fec," 1996.

[2] M. Budagavi, W. Rabiner Heinzelman, J. Webb, and R. Talluri, "Wireless mpeg-4 video communication on dsp chips," *IEEE Signal Processing Magazine*, January 2000.

[3] ISO/IEC JTC1 JVT-I050, "Advanced video coding," .

[4] Raj Talluri, "Error resilient video coding in the ISO MPEG-4 standard," *IEEE Communications Magazine*, June 1998.

[5] U. Horn, K. Stuhlmuller, 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*, September 1999.

[6] Thomas Wiegand, Niko Farber, Klaus Stuhlmuller, and Bernd Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE JSAC*, 2000.

[7] W. Wang, S. Wenger, J. Wen, and K.Katsaggelos, "Review of error resilient coding techniques for real-time video communications," *IEEE Signal Processing Magazine*, July 2000.

[8] TIA/EIA/IS-2000.2-C, "Physical layer standard for cdma2000 spread spectrum systems," .

[9] Thomas Stockhammer, Miska M. Hannuksela, and Thomas Wiegand, "H264/AVC in wireless environments," *IEEE Transactions on CSVT*, July 2003.

[10] Thomas Wiegand, Heiko Schwarz, Anthony Joch, Faouzi Kossentini, and Gary J. Sullivan, "Rate-contrained coder control and comparison of video coding standards," *IEEE Transactions on CSVT*, July 2003.

[11] G.J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Communications Magazine*, November 1998.

[12] ISO/IEC JTC1 JVT-I049, "Joint model reference encoding methods and decoding concealment methods," .