Video Transcoding for Resilience in Wireless Channels

Gustavo de los Reyes
AT&T Labs
gdelosreyes@att.com

Amy R. Reibman
AT&T Labs-Research
amy@research.att.com

Justin C. Chuang
*AT&T Labs-Research*
justin@research.att.com

Shih-Fu Chang
Columbia University
sfchang@ctr.columbia.edu

Abstract

We describe a method to maintain an acceptable quality for video transported over wireless networks under time-varying conditions. We use a transcoder to modify the resilience of the encoded bitstream by using source coding techniques to provide the appropriate level of resilience for the prevailing channel conditions. We develop a statistical model for image loss versus resilience and wireless conditions. Simulation results indicate that using a transcoder to adjust the resilience can improve video quality when errors occur without significantly sacrificing quality when there are no errors. Also, simulation results compare favorably to the analytical model.

1. Introduction

Designing video coding systems for wireless channels requires consideration of the harsh conditions to which the video stream is subjected. A typical architecture for video transmission over a mixed wired and wireless channel is shown in Figure 1.
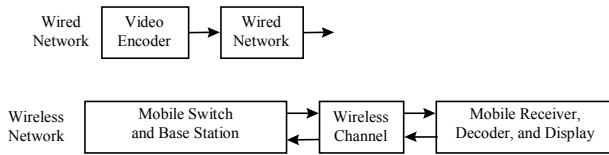
Figure 1. Typical architecture for video transmission over a combined wired and wireless channel.

Wired channels typically have relatively high bandwidth and consistently low bit error rates. This is in contrast to a typical wireless channel in which the bandwidth is generally lower and the error rate much greater and time varying, both within a given call and from call to call.

Ideally, the video system might use feedback from the decoder to encoder [7,9] to adapt its coding parameters (bit-rate or otherwise) to the time-varying conditions. However, this may not be possible for a variety of reasons. First, in multicast, an encoder can only produce a single bitstream, while each receiver may have different capabilities due to the characteristics of both the end terminal and the channel. Second, in applications in which a compressed video bitstream is stored for future delivery, the encoder cannot directly change the resilience of the bitstream. Third, an encoder designed for use on a wired channel is usually not "wireless aware," in that it operates without knowledge of any wireless hops between the encoder and the decoder. There is, and will continue to be, an embedded base of encoders designed to generate bitstreams optimized for reasons other than resilience in wireless channels. Therefore, in this paper, we examine the use of a transcoder to modify an already encoded bitstream for the purpose of improving its resilience prior to transmission over a noisy wireless channel.

2. Resilience

A variety of methods have been proposed to improve the resilience of video transported over wireless channels [5,7,9]. These techniques can be categorized as error detection, spatial and temporal localization, or error concealment. Error detection, including the many channel coding schemes, attempts to detect, and perhaps correct, errors in transmission before they cause problems in the displayed video. Spatial and temporal localization techniques limit the propagation of errors within a frame or to subsequent frames. We do not consider concealment in this paper, although we do assume the decoder uses some form of error concealment.

In this paper, we focus on techniques for spatial and temporal localization. These methods prevent signal errors caused by a bit error or packet loss from propagating within a video frame or to subsequent frames. Spatial error propagation can occur when the decoder loses synchronization while decoding the variable length codes. One approach to combat this effect is to shorten the length of a slice by inserting additional synchronization headers. A shorter slice allows the decoder to re-synchronize more quickly, resulting in less lost data. Another method of spatial localization is to limit spatial error propagation by limiting the reliance on

motion compensation. If predictions are only made within the spatial extent of the current slice as in the Independent Segment Decoding Mode of [3], errors will not propagate outside that slice.

For temporal localization, a primary resilience technique is to transmit more intra-frames (I-frames) or I-blocks. These frames or blocks are coded to be temporally independent of previous frames or blocks and are used as references for subsequent temporal prediction. More frequent I-blocks reduce the duration of error propagation. An alternate technique is to form temporal predictions using a reference frame that is known to be correct [9]. This can limit error propagation without incurring the severe bit-rate overhead of a new I-frame. This technique typically requires additional storage at both encoder and decoder; hence it would not be implemented by a typical video encoder designed for a wired channel. The technique can, however, be accomplished with a transcoder.

To date, these resilience techniques have been studied for use in encoders, but have not been examined in the context of transcoding. Previous transcoders have been designed only to reduce the bit-rate of the video, operating under the implied assumption that resilience can only be increased via channel coding such as FEC and/or ARQ. However, using only channel coding limits the type of resilience that can used and does not take advantage of source coding methods to improve the resilience.

In this paper, we examine a transcoder that modifies the amount of spatial and temporal localization in the video bitstream. This allows the bitstream to be tailored to suit the prevailing characteristics of the wireless channel. The flexibility of using either source coding or channel coding to tailor the amount of resilience to the degree of impairments in the wireless channel will produce the best overall video quality.

3. Transcoder

The transcoder operation is shown in Figure 2. The transcoder decodes the incoming bitstream to the degree required to add resilience. Resilience is added by the transcoder, and, if necessary, the bit-rate is reduced. Then the resilient bitstream is re-quantized (for temporal resilience only) and variable length re-encoded.
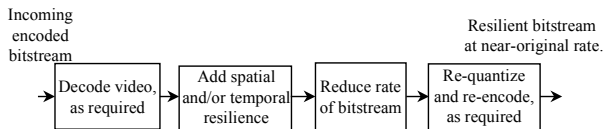

Figure 2. High-level transcoder operation.

We consider here two of the four basic techniques described above: spatial localization using more frequent slices and temporal localization using more frequent I-blocks. The spatial localization technique of adding more slice headers requires only simple bitstream parsing and arithmetic operations. A variable-length decoder parses the bitstream and inserts additional slice start codes where necessary. In addition, motion vectors (MVs) are decoded, a new differential MV is calculated from reference MVs only within the current slice, and the result is variable length encoded.

The transcoder requires more complicated processing to increase the temporal resilience by using more frequent I-blocks. The incoming bitstream is fully decoded, including motion compensation, to create the new I-blocks. The coefficients of those I-blocks to be transmitted are then DCT-coded, re-quantized, scanned, and variable-length encoded.

Temporal localization also can be done in the compressed domain using algorithms from Chang [1]. Compressed domain processing may or may not result in fewer computations than performing an inverse and a forward DCT, depending on the percentage of non-zero motion vectors and non-zero DCT coefficients in the video. For this initial implementation of the transcoder, we consider only the spatial domain approach because computational efficiency is of secondary importance. Future implementations of the transcoder will require more analysis to determine a preferred method.

The temporal and spatial localization techniques improve resilience at the cost of an increase in the overall bit-rate. Therefore, we also use rate-reduction techniques to recover some bit-rate to accommodate that necessary for improved resilience. These techniques fall into two general classes: requantizing coefficients [4,6,8,11] and discarding coefficients [2]. In this paper, we take the simple approach of discarding coefficients to reduce the rate in a given frame by using a zonal mask for the low-frequency coefficients. The size of the mask is selected to provide the required rate reduction.

4. Spatial Resilience Model

In addition to implementing the transcoder in software, we derive in this section a statistical model of the video quality as a function of the spatial resilience (measured by slice length) and the wireless channel conditions. This model can be used in conjunction with simulation results to understand the trade-offs among the various parameters and to select the operating point for the transcoder for good performance in the prevailing channel conditions. In particular, an analytical model reduces reliance on simulation results based on a narrow set of sequences.

The slice length is measured in macroblocks (MBs) as described in [3]. For this analysis, we only use bit error rate (BER) as the indicator of wireless channel conditions. We assume that once the decoder detects an error, it discards the MB containing the error and all subsequent MBs until the next slice header. We further assume that every bit error results in an error that is detected by the decoder. This overestimates the number of lost MBs but does not characterize the distortion caused by errors that cannot be detected from video syntax.

For this spatial resilience model, we extend the work in Liao [5] that determined the probability of block losses as a function of their distance from a synchronization header. From [5], b_s and b_e are the starting and ending bits, respectively, in a MB from the *start* of the header, and P_e is the probability of a bit error (BER). Thus, P(MB n through the end of the slice are lost) = P(no bit errors in MBs 1 to (n-1)) $\mathbf{x}$ P(at least one bit error in MB n), due to the assumed independence of bit errors. Therefore, the expected number of MBs lost per slice is

$$E\ (MB_{loss}/slice)$$

$$= {}_{i=1}\Sigma^{length}\ [(length - i)[(1 - P_e)^{bs} - (1 - P_e)^{be}]\ |\ b_s, b_e]$$

$$= \Sigma_{bs,\ be}\ \{_{i=1}\Sigma^{length}\ [(length - i)$$
$$[(1 - P_e)^{bsi} - (1 - P_e)^{bei}]\}\ P(b_s = a_1, b_e = a_2)$$

where b_{si} and b_{ei} are the starting and ending bit positions, respectively, for each MB i in a slice having a number of MBs equal to length. For initial calculations, we assume the same mean length in bits for each MB and a constant slice header of 40 bits.

The result of the spatial resilience model is shown in Figure 3. This figure shows strong non-linearities across both slice length and BER. One goal of the research is to take advantage of these non-linearities to select an operating point for the transcoder that produces higher video quality while minimizing the increase in bit-rate due to increased resilience.
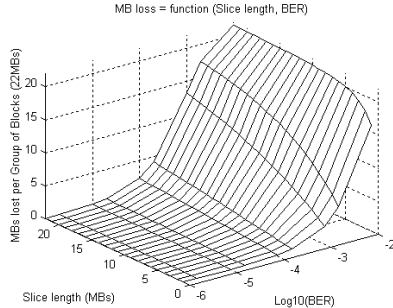
Figure 3. Effect of spatial resilience and wireless channel conditions on macroblock (MB) loss.

5. Simulation Results

In this section we use simulations to show how video quality varies as a function of spatial and temporal resilience and wireless channel conditions. For this paper, we consider the effects of spatial and temporal resilience separately (except as noted in Section 5.3, below). As before, we only use BER as the indicator of wireless channel conditions. The simulation setup is shown in Figure 4. The Draft H.263 Recommendation [3] is used as the basis for this research. A standard H.263 encoder [10] produces an encoded video bitstream. The transcoder adds resilience as described above to produce a transcoded bitstream at near the original bit-rate. Random bit errors are introduced at this point to simulate the wireless channel. A standard H.263 decoder [10] produces video output.
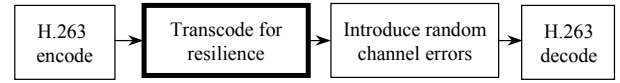

Figure 4. Setup for resilience simulation experiments.

The sequence Stefan coded at 384 kbps and 10 frames/sec is used for all simulations. Twenty simulations using different random number seeds are run at each condition. Peak Signal to Noise Ratio (PSNR) averaged over the twenty runs is used to measure the video quality.
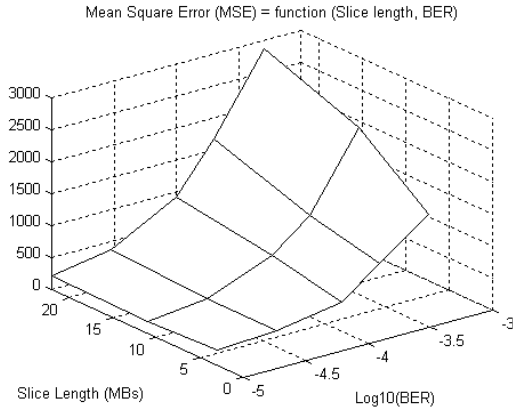
5.1. Spatial Resilience

For spatial resilience, slices of length 3, 11, and 22 MBs are added to the incoming bitstream. (A Group of Blocks (GOB) contains 22 MBs [3].) Channel conditions are simulated for BERs of 10^{-5}, 3×10^{-5}, 10^{-4}, 2×10^{-4}, and 5×10^{-4}. The results are shown in Figure 5, where BER=10^{-6} is used to denote the case with no errors. A BER of 10^{-3} is too severe for the decoder as implemented. At this BER, the decoder is not able to decode the entire video bitstream. As noted below, increased temporal resilience can accommodate higher BERs.
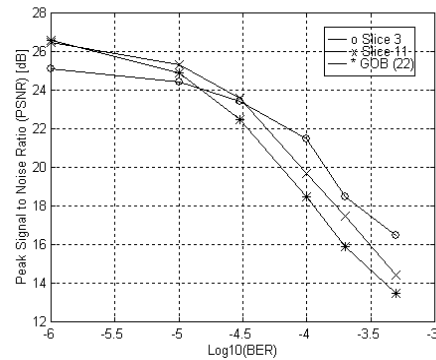
Figure 5. PSNR vs. slice length and BER.

The results show that video quality as measured by PSNR is improved by introducing spatial resilience, even after the necessary rate reduction for a slice length of 3 and 11 MBs. In addition, during error-free transmission, video quality is only minimally reduced by the increased resilience as shown. This reduction in quality is significantly less visible than the degradation caused by errors.

These results, re-plotted in Figure 6 using Mean Square Error (MSE), can be compared to the analytical results shown in Figure 3. The simulation results confirm the predicted non-linearities.


Figure 6. Effect of spatial resilience and wireless channel conditions on video quality (from simulations).

5.2. Temporal Resilience

For temporal resilience, we vary the I-block rate to change the resilience. We characterize the I-block rate using an equivalent I-frame rate by coding new I-blocks such that each macroblock is updated at the desired I-frame rate. For example, if the desired I-frame rate is every 5 frames, then 20% of the macroblocks in each frame are transcoded to be I-blocks. The actual I-block rate can be higher depending on the original encoding of the blocks for each frame. Results in Figure 7 show I-frame rates of 3, 5, 10, 20, and 80 frames. The rate of 80 frames corresponds to the original video sequence. Channel conditions are simulated for BERs of 10^{-5}, 10^{-4}, 5×10^{-4}, and 10^{-3}. Again, BER=10^{-6} corresponds to the case with no errors. With this degree of added temporal resilience, the decoder can tolerate the higher BER and decode the entire video sequence. All simulations are done with the slice length equal to 22 MBs.

These results also confirm the ability of the transcoder to improve the video quality by increasing temporal resilience. The video quality degradation without errors is more severe than for spatial resilience, although a judicious choice of the correct operating point can minimize this gradual degradation.
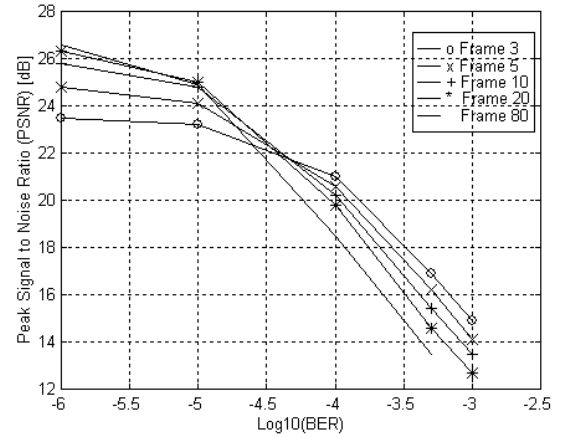
Figure 7. PSNR vs. I-frame rate and BER.

5.3. Combined Spatial-temporal Resilience

In a preliminary study on the impact of combining spatial and temporal resilience, we simulated a single condition: a slice length of 5, an I-frame rate of 10, and a BER of 5×10^{-4}. The resilience was chosen to minimize the quality degradation that would result if the rate were reduced excessively. The PSNR using this resilience at this error rate is 17.4 dB. This is 2 dB higher than just using temporal resilience, and 1.4 dB higher than just using spatial resilience (based on an interpolated data point). In addition, at this BER, the combination of both moderate temporal and spatial resilience outperforms stronger spatial or temporal resilience added separately. These results indicate that additional improvements can be expected by combining temporal and spatial resilience in the transcoder.

5.4. Perceptual Observations

In this section we discuss some observations gathered during informal perceptual viewing, and correlate the changes in PSNR with the perceptibility of the degradation. For example, we would like to compare the perceived degradation due to bit errors and the more gradual degradation due to rate reduction.

First, when the BER increases significantly, e.g. from 10^{-5} to 10^{-4}, and from 10^{-4} to 5×10^{-4}, the perceived quality visibly degrades for a given spatial and temporal resilience. In each of these cases, the corresponding change in PSNR is about 4-5 dB.

Second, for a given BER, the change in spatial and temporal resilience is perceptually noticeable, but not to the same extent as discussed above when changing BER. This agrees with the data that show differences in PSNR are less than 3 dB for changes in either slice length or

I-frame rate for a given BER. Depending on the simulation run, the improvements caused by added resilience may or may not be visible. Indeed, the standard deviation (s.d.) of the PSNR is large enough that the PSNR for different degrees of spatial resilience overlap. However, in general the s.d. for temporal resilience is smaller than the s.d. for spatial resilience.

Last, for a low BER of 10^{-5}, PSNR decreases up to 1.7 dB as the spatial or temporal resilience increases. In this case, reducing the bit-rate to accommodate the increased resilience degrades the quality more than the increased resilience improves the quality. However, this gradual degradation in quality is just barely noticeable for this level of resilience.

Overall, increased resilience has a visible effect on the perceptual quality of the video, especially at BERs greater than 10^{-4}.

6. Summary and Conclusions

We applied our transcoder to video coded using H.263. Our baseline for comparison is the quality using a slice length equal to the number of macroblocks in one row of the image and a typical H.263 bitstream with an I-frame at the beginning. Bit error rates considered ranged from 10^{-5} to 10^{-3}. In this paper, we did not explore the ability of the transcoder to introduce channel coding. Instead, we assumed a given amount of channel coding for all experiments.

We developed an analytical model for the spatial resilience that captures the video quality (as measured by MBs lost to errors) as a function of slice length and BER. The simulation results indicate similar non-linearities to those predicted by our model. We are developing a Markov model to characterize the improvements of temporal resilience, as well as spatial resilience. These models, along with the simulations, will enable an understanding of the tradeoffs among spatial resilience, temporal resilience, and channel coding.

Separately, spatial and temporal localization each result in an improvement of 2-3 dB across the range of resilience. At a low BER, e.g. 10^{-5}, increasing the resilience using a transcoder incurs some small quality degradation because of the rate reduction required to make room for the increased resilience. However, at somewhat higher BER, this degradation is less than the improvement in quality. Therefore, the amount of resilience added by the transcoder should match the prevailing channel conditions to obtain the best overall quality.

To effectively match the added resilience to the prevailing channel conditions, we will incorporate measurements of the wireless channel into our system. A model of how the errors will affect the video will also enable us to adjust the resilience to improve video quality in the presence of errors while not degrading the quality during benign conditions.

7. References

[1] Chang, S.-F. and D. G. Messershmitt, "Manipulation and Compositing of MC-DCT Compressed Video." IEEE JSAC, Vol. 13, No. 1, January 1995, pp. 1-11.

[2] Eleftheriadis, A. and D. Anastassiou, "Meeting Arbitrary QoS Constraints Using Dynamic Rate Shaping of Compressed Digital Video," Proceedings of the 5th International Workshop on Network and Operating System Support for Digital Audio and Video, Durham, NH, April 1995.

[3] Draft ITU-T Recommendation H.263, Video Coding for Low Bitrate Communication, Sept. 1997.

[4] Keesman, G., R. Hellinghuizen, F. Hoeksema, G. Heidman, "Transcoding of MPEG Bitstreams." Signal Processing: Image Comm., 8 (1996), pp 481-500.

[5] Liao, J. Y. and J. D. Villasenor, "Adaptive Intra Update for Video Coding over Noisy Channels." IEEE ICIP, September 1996.

[6] Morrison, D. G., M. E. Nilsson, and M. Ghanbari, "Reduction of Bit-rate of Compressed Video while in its Coded Form." Proceedings of the Sixth International Workshop on Packet Video, September 1994.

[7] Ortega, A. and M. Khansari, "Rate Control for Video Coding over Variable Bit Rate Channels with Applications to Wireless Transmission." Proceedings IEEE ICIP, October 1995, pp. 23-26.

[8] Safranek, R. J., C. R. Kalmanek, and R. Garg, "Methods for Matching Compressed Video to ATM Networks" Proceedings IEEE ICIP, October 1995, Vol 1, pp. 13-16.

[9] ITU-T, SG15/WP15/1, LBC-95-033, Telnor R&D, "An error resilience method based on back channel signaling and FEC", Jan 1996.

[10] University of British Columbia, Copyright © 1997.

[11] Yong, M., Q.-F. Zhu, and V. Eyuboglu, VBR Transport of CBR-encoded Video over ATM Networks. Proceedings of the Sixth International Workshop on Packet Video, September 1994.