

Adaptive Intra Update for Video Coding over Noisy Channels

Judy Y. Liao and John D. Villasenor

Electrical Engineering Department
University of California, Los Angeles
405 Hilgard Avenue, Los Angeles, CA 90095-1594, U.S.A.
E-mail: jliao@icsl.ucla.edu, villa@icsl.ucla.edu

ABSTRACT

An algorithm for increasing error robustness in an H.263 video coding system in the absence of a feedback channel is proposed. For each 16 x 16 macroblock, the encoder accumulates a metric representing the vulnerability to channel errors. As each new frame is encoded, the accumulated metric for each block is examined, and those blocks deemed to have an unacceptably high metric are sent using intra as opposed to inter coding. This approach is fully compatible with H.263 and involves a negligible increase in encoder complexity and no change in the decoder complexity. Simulations performed using an H.263 bitstream corrupted by channel errors demonstrate a significant improvement over non-adaptive intra update strategies.

1. INTRODUCTION

Wireless transmission environments add the additional challenges of noisy channels to the already difficult problem of enabling efficient narrowband video communications. Because of the growing importance of mobile communications, wireless image and video transmission have begun to receive significant attention in both the research community and in industry standardization efforts. On the research side, a number of techniques have been proposed which combine well-designed source coding approaches with efficient error protection schemes [1-3]. In addition, the International Telecommunications Union is now studying modifications to the existing H.324 multimedia terminal standard to support communications over error-prone channels. These mobile terminals, referred to as H.324M terminals, employ the same underlying video and audio codecs (e.g. H.263 and G.723.1) as H.324, but include adaptation and multiplexing layers that provide error detection and correction and fixed-length packets with built-in synchronization information.

It is clear that the most error-robust image communication algorithms are those in which complete flexibility is allowed in the choice of image decomposition technique, method of interframe prediction, and design of channel error protection. However, for at least next few years the

more constrained set of solutions that retain compatibility with existing and emerging image and video coding standards are likely to find wider application.

We describe a method for obtaining improved picture quality using an H.263 bitstream transmitted over an error-prone channel. We utilize the flexibility in H.263 to choose for each 16 x 16-pixel block whether intra or inter transmission is used. This block-specific intra coding permits image refreshing to be performed in a distributed manner as opposed to with a single intra coded frame. In many H.263 implementations, a "forced update" is applied in which blocks are intra updated in raster order. For example, in a QCIF image composed of 9 rows of 11 16 x 16 blocks, an update cycle based on 3 intra updates/frame takes 33 frames for a complete refresh, beginning with the left 3 blocks of the first row and terminating 32 frames later in the last row. More recently, schemes have been proposed that utilize feedback information from the receiver to select blocks for intra coding [4]. In these algorithms, the receiver identifies the locations of errors either using a CRC check or some other means, and communicates back to the transmitter information which allows the error to be corrected in a future image frame.

This paper presents a scheme for adaptive intra update in which a feedback channel is not required. Inevitably, this falls short of the performance of a system in which information about specific transmission errors is made available to the transmitter. However, algorithms in this class are of interest for applications such as broadcast where feedback information from multiple receivers is difficult to incorporate. In addition, these algorithms involve lower signaling complexity than those involving a feedback channel, and are therefore attractive from a system cost perspective as well.

2. ALGORITHM FRAMEWORK

The adaptive update algorithm uses the fact that different portions of an encoded video bitstream have different sensitivities to channel errors. In the absence of a feedback channel it is impossible for the encoder to observe and correct for specific realizations of channel error events.

In: *Proceedings of the International Conference on Image Processing, vol. III*

October, 1996

However, if the encoder behavior is modified to counter expected errors, an improved video quality will result.

The error sensitivity of a particular set of bits is a function of several attributes. First, bits that code more important information such as low-frequency DCT or wavelet coefficients will result in greater degradation in image quality if they are corrupted. The second attribute is the location of the bits with respect to synchronization indicators embedded in the bitstream. These indicators have the effect of partitioning the bitstream into segments, with bits that occur immediately after synchronization indicators less likely to be in error than bits occurring near the end of a segment. The positional dependence can be reduced if error concealment techniques are used at the decoder. The basic idea of the algorithm is that error sensitivities can be used to construct an array of accumulated metrics representing the likelihood that an error has corrupted the corresponding location in the pixel plane. These metrics can then be used to drive decisions on how the encoding is performed. For practical reasons, the sensitivity metrics should be accumulated using a resolution consistent with the granularity of the coding algorithm.

We have applied these concepts to an H.263 bitstream. H.263 is based on the well-known combination of block motion compensation in combination with DCT coding of the prediction residual. Each frame in H.263 begins with 22-bit picture start code which serves as a synchronization indicator for the frame. Frames are divided into groups of blocks (GOB), each of which contains one or more rows of 16 x 16 pixel macroblocks. GOBs have their own 22-bit start code and represent the finest level of synchronization available in the bitstream. Within each GOB, motion vector and prediction residual information is contained on a block by block basis.

3. SENSITIVITY METRICS

Let $x(n,k,t)$ denote the k th macroblock of the n th GOB in frame number t . The sensitivity of the macroblock to channel errors can be denoted by $S(n,k,t)$. After coding, let the first bit representing information from each macroblock be located $x(n,k,t)_{\text{start}}$ bits after the nearest previous synchronization marker. Assuming that the decoder is correctly synchronized at the start of each GOB, there are several possible outcomes for a given macroblock in response to transmission errors. If an error occurs before bit $x(n,k,t)_{\text{start}}$, a decoder implementing concealment will typically detect the error, abort decoding of the GOB, and replace macroblock $x(n,k,t)$ with the corresponding macroblock $x(n,k,t-1)$ from the previous frame. This type of error occurs with probability $P_1 = 1 - (1 - P_e)^{x(n,k,t)_{\text{start}}}$, where

P_e is the probability of error assuming a binary symmetric channel with random errors. The contribution to the sensitivity $S(n,k,t)$ is set equal to $P_1 S(n,k,t-1)$, the scaled sensitivity of the corresponding macroblock from the previous frame.

If no bit errors occur prior to $x(n,k,t)_{\text{start}}$ but an error occurs in one or more of the $x(n,k+1,t)_{\text{start}} - x(n,k,t)_{\text{start}}$ bits representing macroblock $x(n,k,t)$ then all bits after the error will be improperly decoded. An error of this type occurs with probability $P_2 = (1 - P_e)^{x(n,k,t)_{\text{start}}} - (1 - P_e)^{x(n,k+1,t)_{\text{start}}}$, and results in a mean absolute difference (MAD) of C , where C is a constant that depends on the output of the decoder given random input. We have used a value of 128 for C based on experimental observations. It is also helpful within this class of errors to account for the case where the macroblock contains exactly 1 bit (identifying the macroblock as "uncoded"), which if corrupted would cause a very large perceptual error at the decoder.

The final (and usually most likely) probability to consider is $P_3 = (1 - P_e)^{x(n,k+1,t)_{\text{start}}}$, which describes the case where no bit errors occur in the GOB before or during the transmission of macroblock $x(n,k,t)$. Even in this case, however, it is necessary to consider propagation of errors from macroblocks in the previous frame that are used in the motion compensation of $x(n,k,t)$. This is accomplished by using a linear combination the sensitivity metrics from the 1, 2, or 4 macroblocks from frame $t-1$ used to predict $x(n,k,t)$.

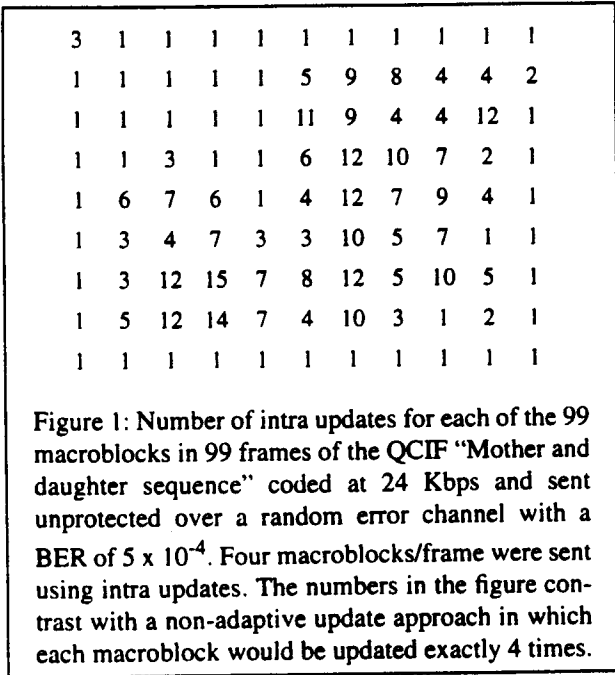
To sum up, $S(n,k,t)$ is constructed as an expected MAD based on the three cases of (i) errors in the GOB before the start of the macroblock, (ii) errors within (but not before) the macroblock, and (iii) no errors before or within the macroblock. These cases have probabilities P_1 , P_2 , and P_3 respectively.

4. EXPERIMENTAL RESULTS

To verify that the approach described above constitutes a reasonable framework for characterizing sensitivity, we ran a series of experiments in which an input video sequence was coded using H.263, corrupted by channel errors, and then decoded. The H.263 coder was a modified version of version 1.5 of the Telenor H.263 simulator [5] in which each frame contained five GOBs. In each experiment the received image was decoded after application of a different realization of the error pattern. As expected, when the results for many trials were averaged there was a correlation between the sensitivity metric $S(n,k,t)$ and the square error (MSE) for that macroblock at the receiver.

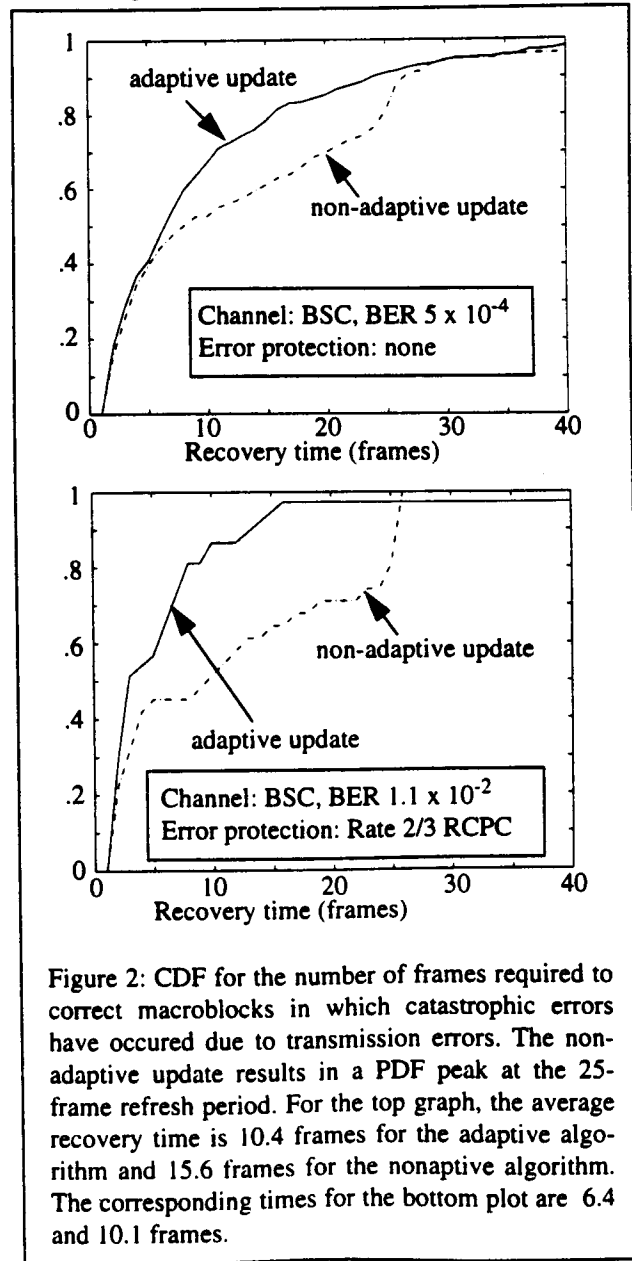
Scatter plots showed that the relationship was approximately linear though there were a significant number of macroblocks for which the fit was poor.

There are many ways in which the sensitivity metric S can be used in adaptive intra update decisions. The key advantage of an adaptive algorithm over a raster based approach is that the bandwidth used for refresh can be intelligently directed to the locations in the image frame where it is likely to be needed. Since intra coding of a predetermined number of macroblocks is desirable from a buffer management standpoint, one obvious solution is to identify the macroblocks with the highest accumulated sensitivity values, transmit them using intra coding, and reset S to zero. A disadvantage of this approach is that errors occurring in a macroblock without a high S value can remain uncorrected. This can be resolved if the "budget" of intra blocks per frame is shared between a raster-based update and an adaptive update, thereby ensuring that all blocks are periodically refreshed.



The effect on the intra update pattern when the adaptive algorithm is used is significant. Figure 1 shows the number of times each of the 99 macroblocks in the "Mother and daughter" sequence was updated when the algorithm was applied over the course of 99 frames. This is a QCIF (176 by 144) sequence that was coded at a rate of 24 Kbits/sec. No error protection was used, and the channel was a BSC with a BER of 5×10^{-4} . In each frame the three macroblocks with the highest sensitivity metrics were intra updated and their metrics were then reinitialized to zero. In addition, one macroblock per frame was

updated using a raster-based forced update strategy. The contrast between the update distribution in Figure 1 and the distribution that would be experienced in a pure raster based forced update with 4 blocks/frame (in which each block would be refreshed 4 times during the 99 frames) is very strong. It is clear that the updates have been concentrated in the regions of the image where there is the most activity (the heads of the mother and daughter), and where there is the greatest likelihood of encountering an error.



Another evaluation can be performed by examining the number of frames required to refresh a macroblock that has been "catastrophically" corrupted. When viewing a video decoded from an error-corrupted bitstream, it is

clear that some errors cause macroblocks to be very obviously corrupted, with intensity and color values that are completely inconsistent with those of nearby blocks. These catastrophically corrupted blocks are very disturbing visually and should be corrected as rapidly as possible. Figure 2 shows the cumulative distribution function for the number of frames required to correct catastrophically corrupted macroblocks for the non-adaptive and adaptive update cases. The adaptive update implementation used was the same one described above. The non-adaptive update was raster-based with 4 contiguous intra macroblocks per frame. The video sequence and coding rate are the same as for Figure 1. Figure 2 contains the cumulative data from 15 trials of 99 frames each. The top plot in the figure shows the results of transmitting the H.263 bitstream over an unprotected BSC with a BER of 5×10^{-4} . The bottom plot shows the results when the H.263 bitstream is protected with a rate 2/3 RCPC code and then transmitted over a BSC with a BER of 1.1×10^{-2} .

For both the adaptive and nonadaptive intra update algorithms, many errors get corrected within a few frames of their appearance. This result is expected for the adaptive case, and shows that intra updates are being directed to the appropriate places. It is also expected for the non-adaptive case, though for very different reasons. Macroblocks occurring in the same GOB and after the 4 intra blocks in a non-adaptive algorithm are very vulnerable to error because of the high number of bits used for intra coding. These blocks are farther from the GOB header and will often be corrupted due to loss of sync in the intra coding, but then will be quickly repaired within a few frames as the raster updating proceeds.

The CDFs in Figure 2 show that for the 5×10^{-4} BSC (top plot in figure) the adaptive algorithm recovers from 70% of errors within 11 frames, versus 20 frames to achieve the same recovery rate for the non-adaptive algorithm. The mean recovery time is 10.4 frames for the adaptive algorithm and 15.6 frames for the nonadaptive algorithm. In the bottom plot the average recovery time is 6.4 frames for the adaptive algorithm and 10.1 frames for the non-adaptive algorithm.

5. CONCLUSIONS

We have introduced an H.263-compatible approach to adaptive updating of intra blocks that utilizes only the information available to the encoder in a system without a feedback channel. A metric is constructed allowing the sensitivity of each macroblock in the image to be quantified and accumulated over a series of video frames. At each frame, those macroblocks with the greatest accumu-

lated error sensitivity are intra updated. A small number of blocks are also updated using a raster based forced update scheme to ensure that blocks with low sensitivity metrics are eventually refreshed. Experiments performed at low bit rates show that the adaptive update strategy significantly reduces the length of time required to correct highly visible decoding errors arising from transmission errors.

This work demonstrates the utility of using intelligence at the encoder in the decision regarding which macroblocks to select for intra updating. It is applicable to wireless multimedia systems in general, and specifically to broadcast environments where feedback from multiple receivers is difficult to incorporate. Although the algorithm here was designed for an H.263 coder, these principles could be applied to other implementations of H.263 (with different GOB sizes or concealment techniques) or to other wireless video coding algorithms.

6. ACKNOWLEDGEMENTS

This work was supported by ARPA/ITO under contract J-FBI-93-112. The authors thank Feng Chen of UCLA for useful discussions.

7. REFERENCES

- [1] R. Stedman, H. Gharavi, L. Hanzo, and R. Steele, "Transmission of Subband-Coded Images via Mobile Channels", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 3, No. 1, pp. 15-25, Feb. 1993.
- [2] B. Belzer, J. Villasenor, and B. Girod, Joint source channel coding of images with trellis coded quantization and convolutional codes", *Proc. 1995 IEEE Intl. Conf. on Image Proc.*, Wash. D.C., vol. II, pp. 85-88, Oct. 1995.
- [3] M. J. Ruf and J. W. Modestino, "Rate-Distortion Performance for Joint Source and Channel Coding of Images," *Proc. 1995 IEEE Intl. Conf. on Image Proc.*, Wash. D.C., vol. II, pp. 77-80, 1995.
- [4] N. Faerber, E. Steinbach, B. Girod, "Robust H.263 Compatible Video Transmission Over Wireless Channels," *Proceedings of the Intl. Picture Coding Symposium*, Melbourne, vol. 2, pp. 575-578, March 1996.
- [5] H.263 code available at ftp site bonde.nta.no in /pub/tmn/software