Perceptual Quality Feedback Based Progressive Frame-level Refreshing For Robust Video Communication

Liang Cheng and Magda El Zarki
School of ICS, UC, Irvine
Email: lcheng61@ics.uci.edu, elzarki@uci.edu

Abstract

In this paper, we propose an encoder side error resilience technique, namely the progressive group of picture (PGOP), which is well suited for rate controlled real-time video transmission over wireless channels. The PGOP scheme can be seen as an alternative approach to the group of picture (GOP) that uses periodic intra-frames (I-frames) for low bit-rate channels. This is because: (1) the PGOP scheme can completely stop error propagation in a timely manner; (2) the PGOP scheme avoids the bit rate unevenness problem of I-frames by a progressive intra-updating method. We further apply the PGOP scheme in a novel video perceptual quality feedback framework, and design an on-line algorithm both to help shape the bit rate and to explore the maximum recovery capability of PGOP. The performance of our scheme is demonstrated in a wireless LAN channel. It is to be noted that the PGOP can additionally provide the random access functionality and the technique is applicable to all inter frame based video standards, e.g., MPEG-4.

1 Introduction

Wireless networks experience high bit error rates due to long and short fades, shadowing and environmental noise, none of which are conducive to the transmission of time sensitive highly compressed data. Channel coding, such as forward error correction (FEC) and automatic repeat request (ARQ), can help protect and recover the data [1]. These schemes, however, add overhead and delay to the transmission and cannot guarantee 100% recovery. The consequence of a packet loss is exacerbated in the case of predictive video coding schemes because the prediction loop, especially motion compensation, propagates errors.

The most intuitive and effective method to stop these propagating errors caused by the imperfections of the network is to insert intra-frames (I-frame). Each I-frame is usually followed by a sequence of predictive frames (P-frames) and bidirectional-frames (B-frames), e.g., IBBPBBPBBPBB is an example of a 12 frame group of picture (GOP) structure. One GOP is an independent decodable entity. The propagation of errors terminates at the beginning of each GOP with the I-frame. For low-bandwidth (especially wireless) video transmission (e.g. $\leq$ 128kbit/s) environments, the use of I-frames is, however, highly undesirable: I-frames incur buffer overflow and higher delay because they have a large number of bits due to the lack of predictive coding. Therefore, in most encoders designed for wireless use, only one I-frame is used at the beginning of the sequence, followed by all P-frames. Unless we can guarantee that no errors will occur, some form of data refreshing has to take place continually to halt the propagation of errors. Over the past few years, researchers have proposed a variety of approaches to increase the robustness of low bit-rate video communications [2]. Selectively forcing the intra coding of a number of macroblocks (henceforth referred to as forced intra MBs, updated MBs, or refreshed MBs) in the P-frames is well recognized as an effective mechanism to mitigate the propagation effect of interframe prediction [3]-[6]. These proposed schemes fall into two categories: heuristic non-rate-distortion (RD) model based schemes [3] and RD model based schemes [4]-[6].

RD based selection methods (designed for error prone environments) calculate the end-to-end distortion, under the following assumptions: (1) the encoder receives the packet loss rate [4] and/or the addresses of the lost MBs from the decoder [6]; and (2) the encoder knows the error concealment algorithm applied at the decoder. Given this information, these methods can then optimally select the inter/intra mode for each MB and the appropriate quantization parameter to optimize a quality performance measure, e.g., mean square error (MSE) or peak signal to noise ratio (PSNR). The drawback of the RD based methods, however, stems primarily from the assumptions that they make. First, any feedback channel will experience varying delays, errors and losses, this will impact the accuracy of the information that the encoder receives and uses to make decisions. Moreover, the feedback cannot be cumulative: the received feedback cannot reflect the content of prior lost feedback. Secondly, the tight coupling between the encoder and the techniques applied at the decoder (i.e., error concealment, post-processing filters, etc.) converts the mapping relationship between the encoder and the decoder from a “one-to-many” to a “one-to-one”, which limits the flexibility of the encoder and the versatility of the decoder. But most importantly, the subjective quality of the quantization distortion and the distortion caused by the packet loss is different even though they both may result in an equal value for the objective quality performance measurement. This is illustrated in Fig. 1. We notice that a loss of a MB (Fig. 1-a) and a quantization distortion (Fig. 1-b) both result in the same PSNR value. We can conclude therefore that an optimal encoding scheme based on MSE as a performance measure is by no means optimal from a subjective quality perspective. We believe that the elimination of irregular distortion caused by channel losses should be taken care of before any other optimizations are applied. In addition, neither non-RD schemes nor RD based schemes can provide a refresh pattern to completely eliminate the propagation effect of the existing errors in a timely manner. Therefore, all of these conventional schemes don't support random access and it may take a long time to completely recover the whole frame.


(a) transm. loss (b) quant. loss (c) original

Figure 1: The 3rd frame of *Foreman.* $PSNR_a = PSNR_b = 30.3dB$

The main contribution of our work is a method that is able to completely refresh errors existing in the prediction loop in a progressive systematic manner, which we refer to as the progressive group of picture (PGOP). The PGOP method uses a tracing pixel-map table to preventively trace all of the possible errors that may propagate from un-updated MBs to updated MBs. We demonstrate in this paper that the PGOP is able to control the frame-level error recovery time and provide random access functionality. We further apply PGOP in a feedback-enabled network and explore the maximum refreshing capability of PGOP. In order to avoid the inaccurate network characterization of packet loss rate [4] and the delicate "error tracing" mechanisms [6], we propose to use decoder side reconstruction quality features [7][8] as feedback, thereby enabling the encoder to monitor the in-service quality degradation at the decoder side.

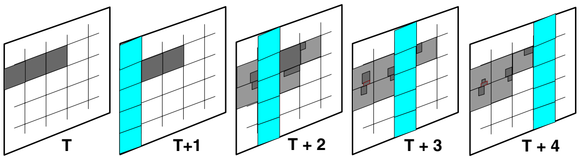
The rest of the paper is divided as follows. In section 2, we describe the PGOP algorithm. In section 3, we discuss the PGOP in the quality feedback network. In section 4, we present the simulation results to demonstrate the performance of our scheme. In section 5, we conclude and outline future work.

2 Frame level progressive refresh

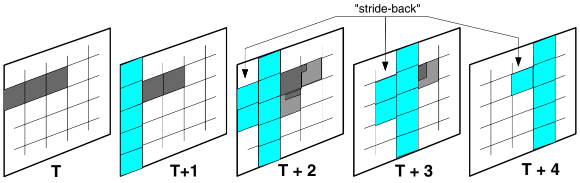
In order to maximally scatter the intra-MBs and help the rate control mechanism (which usually changes the quantization scale of each MB in a scan order) allocate the bit budget more uniformly, we refresh intra-MBs on a column-by-column basis from left to right. We denote MBs that have just been intra-updated as the "clean" area, MBs that have not been intra-updated as the "dirty" area, and MBs that are affected by the error propagation from "dirty" area to "clean" area as "unclean" area. As the refresh proceeds from left to right, the "clean" (left of refresh column) area grows while the "dirty" (right of refresh column) part shrinks. We note that errors may propagate across the column being refreshed from what we call the "dirty" area of the frame to the "clean" area (we refer to this as "crossing errors"). What was a "clean" MB may become "unclean" if it uses part of the "dirty" area as its reference for motion estimation. In addition, these now "unclean" MBs in the "clean" area may potentially affect a number of their neighboring MBs (up to 8 MBs if the maximal motion searching range is 16 pixels) in the "clean" area. To solve this problem, we propose to augment the refresh process to trap these error propagations by refreshing the affected MBs. We call this *stride back*.

Fig. 2 illustrates the comparison between the heuristic column-based refresh and the proposed refresh scheme with *stride back*. Before the refresh, we assume some MBs are in error and only those MBs are in motion. Fig. 2-b shows how our refresh scheme can refresh all of the MBs (including error MBs) by adapting to possible error propagations, which are normally ignored by heuristic column refreshing (i.e., Fig. 2-a). At time T + 1, we refresh the first column. We then mark the first column as "clean", and the rest of the columns remain "dirty". At time T + 2, some MBs in the first column can become "unclean" because they use MBs from the "dirty" part as their reference. Then, at time T + 2, we not only refresh the second column, but also those "unclean" MBs in the first column. This is the *stride back* function of our scheme. Note that the *stride back* function only affects those MBs that are adjacent to the refresh column (as we are using a refresh interval of 1). We continuously "clean" MBs until we come to the last column, say at time T + m (assuming m column refreshes). We refer to the m frames from time T + 1 to time T + m as a progressive group of picture (PGOP). By time T + m, all errors have been completely corrected that existed in the frame at time T. Note that any new errors that affect the frame to the left of the refresh column will not be corrected during this period unless they coincide with an "unclean" MB.

The period of PGOP is determined by two parameters: 1) refresh interval (i.e., $RefrInt$: the frame in-



(a) intuitive column refresh



(b) error propagation adapted refreshing with *stride back*

☐ : Unaffected MB

■ : Damaged MB/ Affected pixles

▨ : Affected MB

🟦 : Refreshing MB

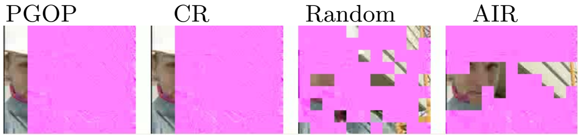
Figure 2: the proposed frame level refresh scheme

terval between frames with force-updated MBs) and 2) refresh column number (i.e., $RefrColNum$: columns regularly refreshed in one frame). Thus, we notice that the period of PGOP is constant and can be described as Eq. 1.

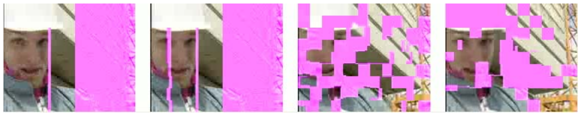
$$PGOPperiod = \lceil \frac{TotalNumCol \times RefrInt}{RefrColNum} \rceil. \qquad (1)$$

Note $TotalNumCol$ denotes the total number of MB columns in one frame, which is decided by the picture format (e.g., for QCIF, $TotalNumCol = 11$).

Our implementation of PGOP is based on the MPEG-4 baseline encoder [9]. In the case that $RefrInt$ is greater than 1, we setup a bit-map table to keep the status of every pixel, because for some frames there is no refreshing while the error propagation is continuing. For one MB (which has 256 pixels), we need 32 bytes. Each bit represents the status of one pixel, i.e., if that pixel is affected or not. Those MBs that have at least 1 pixel marked "dirty" or "un-clean" will be marked for refreshing. For a QCIF video, we need in total 3168 bytes (i.e., 32 * 99 = 3168) extra memory at the encoder. When half-pixel motion estimation is used, some extra pixel interpolation operations need to be done. In the simulation throughout this paper, we use half-pixel motion estimation. Basically, the added complexity of those bit operations is trivial. When $RefrInt$ is 1, no extra memory is needed because all propagating errors can be promptly eliminated.



(a) the 275th frame



(b) the 277th frame

(c) the 280th frame

Figure 3: Comparison of PGOP and conventional refresh schemes

Fig. 3 shows the refresh comparison between PGOP ($RefrColNum = 2$, $RefrInt = 1$), column based refresh without *stride back* (CR), random intra update (Random), and adaptive intra refresh (AIR). Note AIR updates the specified number of MBs that have higher difference from the corresponding MBs in the previous frame. For both random update and AIR schemes, 24 MBs are intra-coded per frame. We assume the whole 274th frame of sequence Foreman is lost without any concealing. We can see the *stride back* function enables PGOP to eliminate the "crossing error", which affects neighboring MBs for the CR scheme. By the 280th frame, the PGOP completely removes all of the residual errors. Thus, the total period of PGOP illustrated in Fig 3 is 6 frames, which, from another perspective, can be grouped to provide a random access point. Using the similar number of intra-MBs, none of other refresh schemes are able to achieve the goal of frame-level error elimination.

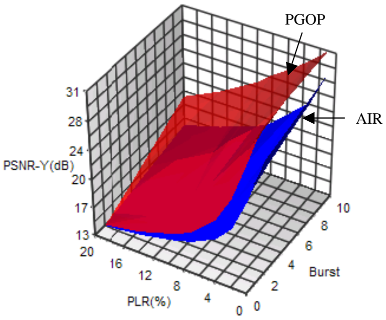


Figure 4: Comparison of PGOP and AIR under different error environment, Bitrate: 64kbps; Pakcet length: 300 bits; Sequence: *Foreman*; Video format: QCIF.

In Fig. 4, we test the error resilience capability of PGOP under different error environments (i.e., different packet loss rates and packet loss burst lengths) and compare it to that of AIR [9]. Note we use the Gilbert-

Eilliot packet erasure model [6] to generate the loss pattern. The average number of intra-MBs per frame is equal for both PGOP ($RefrInt = 1$, $RefrColNum = 1$) and AIR (12 intra-MBs per frame). The result shows PGOP achieves higher PSNR than AIR under the same error situation, and, especially, the difference in quality performance between PGOP and AIR increases as the channel environment becomes worse (i.e., packet loss rate and the burst length increases).

Of course, we cannot accurately locate the lost MBs. Therefore, when applying PGOP, if some MBs are in motion but are not in error, the regular column-refresh and the *stride back* might also update those MBs, which is obviously unnecessary. The technique that will be introduced in next section can help to decrease the needless intra-coding by informing the encoder the occurrence of channel loss.

3 Feedback enabled progressive frame refreshing

In this section, we incorporate PGOP into a novel perceptual quality feedback framework. The major advantage of the quality feedback over the conventional channel status feedback (e.g., packet loss rate) is that the quality feedback reflects the distortion caused by the channel loss in a cumulative way. This can effectively solve the drawbacks (normally caused by the feedback delay and loss) that are inevitable in channel-layer feedback based system, because the quality feedback doesn't rely on the current channel status (which might be not "current" when the channel status is received by the encoder). Instead, the received quality feedback shows the quality distortion jointly determined by all the losses before it, even if the feeback is received late or some of its previous feedbacks (if any) are lost.

The Institute for Telecommunications Sciences (ITS) in [7] developed a spatial-temporal distortion metrics. Instead of using the pixel-comparison, the ITS model based quality assessment algorithm calculates the quality features of processed spatial-temporal (S-T) regions. Thus, it is suitable for in-service digital video quality monitoring. In our implementation, we extract the quality features of the decoded video, send them back to the encoder, and compare them to the features of the encoder side reconstructed video. Once the encoder realizes any inconsistencies, which indicates the occurrence of channel loss, the encoder will send one PGOP to quickly remove the errors jointly caused by all previous channel losses and propagation effects thereof. For error free periods, nevertheless, the highest possible coding efficiency can be preserved as no MBs will be force-updated.

The quality feature feedback rate is extraordinarily low. For example, for quarter-common-interchangeable-format (QCIF) video, the feedback rate is only 15.47 kbit/s with a 6-frame feedback interval. In addition, the complexity of the quality feature extraction is highly scalable because the feedback interval is adjustable and the quality feature can be spatially sampled. And thus, the feedback rate can be further lowered.

There is a tradeoff between the error recovery capability (i.e., PGOP period) and the bit rate fluctuation. On the one hand, it is desirable to set the PGOP period as short as possible. I-frame is an extreme instantiation: I-frame has the maximum error recovery capability because it can stop error propagation within one frame. On the other hand, we could lower the bit rate fluctuation if we increase the length of PGOP, i.e., decrease the number of intra-MBs of each frame. When the PGOP is sporadically inserted into the compressed stream, the bit rate fluctuation may happen due to the mismatch of the quantization scale and the percentage of intra-MBs in the first frame of the PGOP. The percentage of the intra-MBs in the first frame of the PGOP is usually larger than its previous frames, which are regular P-frames. The rate control mechanism (which is independent of the PGOP design in this paper) is not able to increase the quantization scale promptly to compensate for the higher number of bits generated by the force-updated MBs. After the first frame of the PGOP, the rate control mechanism can adapt to the large intra-MBs percentage of rest frames (which may have slightly more intra-MBs than the first frame due to *stride back* function). The bit rate will converge quickly to the desired value.

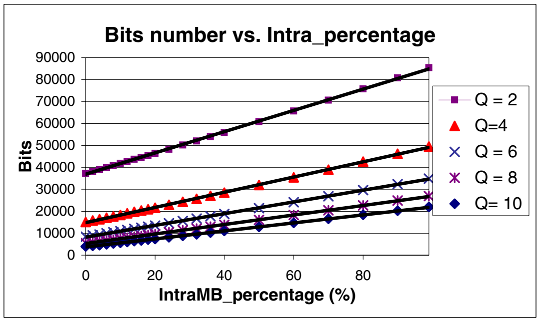


Figure 5: Linear relation between average bit number and intra-MB's percentage for the total 400 frames of *Foreman*.

Therefore, we determine the number of intra-MBs (or say intra-MB columns) of each frame in the PGOP, which can maximally expedite the frame-level refresh period and keep the output bit rate fluctuation within a designated range. From empirical studies we conducted on several video sequences (i.e., *Foreman*, *Suzie*, and *MotherDaughter*), we know that the relationship between the percentage of the forced intra MBs and the bit number per frame is linear, as shown in Fig.5. Of course a significant scene change or high motion may cause this relation to be non-linear because those

Table 1: Parameters definition.

$\bar{Q}_{pre}$	Average quantization scale of previous frame
R	Bit rate of the video sequence
F	Frame rate of the video sequence
C	Total number of MB-columns in one frame
α	Percentage that the bit number can exceed the average bit allocation
β	Percentage of intra-MBs for the first frame of PGOP
Col_{pgop}	Maximum number of intra-columns for the first frame of PGOP

forced intra-MBs may overlap with the intra-MBs that have already been determined by the coding loop. As a matter of fact, the overlapping of some intra-MBs reduces the bit rate fluctuation.

We propose an algorithm to minimize the impact of PGOP on the bit rate fluctuation. Before we describe our algorithm, we define parameters in Tab. 1.

Thus, the linear relation between the percentage of intra-MBs and the bit number is described as:

$$Bits = a \times \beta + b. \tag{2}$$

Where b represents the number of bits for a predictive frame with zero intra-MB percentage. Then $(a + b)$ represents the number of bits for a predictive frame with 100% intra-MB percentage (i.e., $\beta = 1$).

From Fig. 5, we also notice that the number of bits in a frame with a certain percentage of intra-MBs relies on the quantization scale. Instead of using rate-quantization (R-Q) models, we design an online algorithm to determine the values of a and b by pre-coding the current frame once with 0% intra-MBs and once with 100% intra-MBs. We use $\bar{Q}_{pre}$ for all the MBs in the current frame. Note that a MB based rate control mechanism will increase the quantization scale if it is aware of a sudden increase in bits, which makes our assumption of using the for every MB fairly conservative.

The algorithm is as follows:

(1) The last quality feedback shows that channel loss has occurred and one PGOP is requested.

(2) Get $\bar{Q}_{pre}$ and apply it to every MB of the current frame, i.e., first frame of PGOP.

(3) Encode the current frame in the regular predictive mode and assign the encoded number of bits to parameter b.

(4) Encode the current frame with intra mode. Assign the encoded number of bits to $(a + b)$ to obtain the linear relationship denoted by Eq. 2.

(5) Calculate the maximum number of bits (including the fluctuation range) that can be used in the current frame, which is

$$Bits = \frac{R}{F} \times (1 + \alpha). \tag{3}$$

(6) Calculate the maximum intra-MB percentage that can be applied in the current frame by using Eq. 2 and Eq. 3:

$$\tfrac{R}{F} \times (1 + \alpha) = a \times \beta + b,$$

from which we get

$$\beta = \frac{(1 + \alpha) \times (R/F) - b}{a}. \tag{4}$$

(7) Calculate the number of intra-columns for the PGOP.

$$Col_{pgop} = \lfloor C \times \beta \rfloor . \tag{5}$$

Note: $\lfloor \cdot \rfloor$ denotes rounding to the lowest integer. If Col_{pgop} is negative, set it to zero.

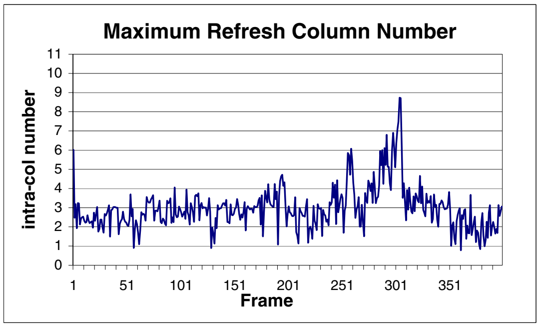


Figure 6: Simulation result of Eq 5. Bitrate: 64kbps; Sequence: *Foreman*; Video format: QCIF;.$\alpha = 1$; $F = 10$.

Fig. 6 shows the explored maximal number of intra columns. It is observed that Col_{pgop} for the motion intensive part (frame 250 to frame 320) is high. That means, given the same bit number fluctuation range, we can apply intra-modes to more MBs in the motion intensive frames than in the frames with less motion. Therefore, our scheme automatically gives more protection to the scene with intensive motion, which usually cannot be concealed well at the decoder.

Of course, the proposed algorithm increases the encoding complexity as the encoder must determine the value of a and b; the first frame of the PGOP is encoded three times. However, a lot of coding routines (such as motion estimation, packetization, I/O, etc.) do not need to be repeated. Hence, our scheme only needs moderate additional complexity at the encoder when a perceptual quality drop is detected and reported.

4 Performance

In Fig. 7, we demonstrate the performance of feedback-based PGOP for a wireless LAN packet loss

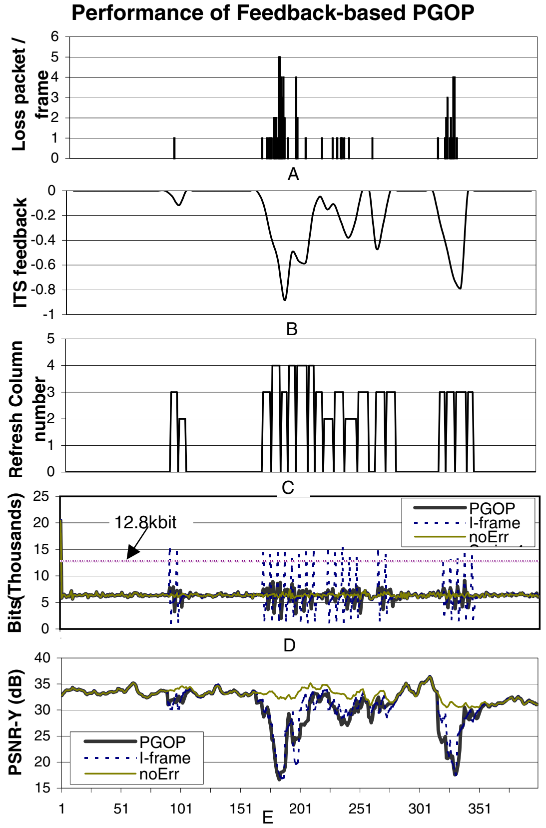


Figure 7: Performance of feedback-based PGOP scheme in the wireless LAN environment. Bit rate: 64 kbps; packet size: 200 bytes; error concealment: direct copy from corresponding MBs in the previous frame; feedback interval: 6 frames; $\alpha = 1$.

pattern (Fig. 7-A), which reflects the pedestrian movement and moderate traffic congestion around the ICS School building at the University of California, Irvine. Fig. 7-B shows that the encoder is able to determine the correct positions to insert PGOPs by receiving ITS quality feedback. In Fig. 7-C, we notice that the number of intra-columns of PGOP is different and adapts to the video content (i.e., motion intensity) and error locations. In Fig. 7-D, the bit number fluctuation is within the range (i.e. for $\alpha = 1$, the upper bound of the fluctuation range is 12.8kbits/frame if the average bit rate is 6.4kbits/frame). If we insert I-frames instead, we observe a higher bit rate fluctuation, which may incur buffer overflow and thereby frame dropping. In Fig. 7-E, we see that the PGOP can effectively and promptly recover the decoded video from errors without influencing the decoded video quality in the error free period, which is comparable to the direct I-frame insertion.

5 Conclusion

In this paper, we first propose an encoder side error resilient scheme named PGOP, which can progressively and completely recover from channel loss and its propagation on a frame level. We then incorporate PGOP into a perceptual quality feedback framework, discuss its effect to the bit rate fluctuation, and evaluate its performance. The proposed methods are independent from any other encoder/decoder side control mechanisms (i.e. rate control, channel coding, etc.). Further optimization, however, is possible if these control mechanisms are taken into consideration.

6 Acknowledgement

The authors would like to appreciate the help from Mr. Haining Liu for providing packet loss patterns.

References

[1] H. Liu and M. El Zarki, "Performance of H.263 Video Transmission over Wireless Channels Using Hybrid ARQ," *IEEE J. Select. Areas Commun.*, vol. 15, no. 9, 12/97.

[2] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974-997, 05/98.

[3] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, CA, pp. III545-548, 05/92.

[4] R. Zhang, S. L. Regunathan and K. Rose, "Video Coding with Optimal Inter/Intra Mode Switching for Packet Loss Resilience," *IEEE J. Select. Areas Commun.*, pp. 966-976, vol. 18, no. 6, 06/00.

[5] G. Cote, S. Shirani, and F. Kossentini, "Optimal Mode Selection and Synchronization for Robust Video Communications over Error-Prone Networks," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 952-965, 06/00.

[6] B. Girod and N. Farber, "Feedback-Based error control for mobile video transmission," *Proceedings of the IEEE, special issue on video for mobile multimedia*, vol. 97, no. 10, pp. 1707-1723, 10/99.

[7] S. Wolf, and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system", *Proc. of SPIE Int. Symp. on Voice, Video, and Data Commun.*, Boston, MA, 09/99.

[8] L. Cheng and M. El Zarki, "An adaptive error resilient video encoder," *Visual Commun. and Image Proc.*,Switzerland, 07/03.

[9] ISO/IEC 14496-5: Information technology - Coding of audio-visual objects Part 5: Reference software, 2001/Amd 1: 2002.