Telecommunication Standardisation Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-511
July, 1993

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG93/528
July 1993

Source: M. Biggar (Telecom Australia)
CHAIR, Ad-hoc Group on ATM Cell Loss and Error Resilience

Title: Report of Ad-hoc Group on ATM Cell Loss and Error Resilience

Status: Information and Proposal

Establishment and Mandate

The Ad-hoc Group on ATM Cell Loss and Error Resilience was established at the Sydney meeting to study means by which the MPEG-2 video syntax can provide for robustness in the presence of cell loss on ATM networks, and other errored transmission or storage conditions.

Meetings: No meetings were held.

Membership:

Mike Biggar (chair)	Telecom Australia
John Arnold	ADFA
Dimitris Anastassiou	Columbia U.
Gisle Bjontegaard	Norwegian Telecom
Peter Borgwardt	Tektronix
TiHao Chiang	Columbia University
Stuart Dunstan	Siemens Ltd.
Gerard Eude	CNET
Cesar Gonzales	IBM
Bernard Hammer	Siemens AG
Barry Haskell	AT&T
Carsten Herpel	Thomson
Jae-Kyoon Kim	KAIST
Takuyo Kogure	Matsushita
Arian Koster	PTT Research
Matthew Leditschke	Telecom Australia
Jean-Jacques Lhuillier	NA Philips
Sandy MacInnis	Kaleida
Gerard Madec	CCETT
Ken McCann	NTL
King Ngan	Monash University
Gilles Nocture	LEP
Satoshi Nogaki	NEC
Sakae Okubo	NTT
Ian Parke	BT Labs
Olivier Poncin	RTT Belgacom
Amy Reibman	A.T.&T.
Daniel Reininger	David Sarnoff
Kiyoshi Sakai	Fujitsu
Tristan Savatier	TCE
Gunther Schamel	HHI
Thomas Sikora	Monash University
Kenji Sugiyama	JVC R&D Center
Ali Tabatabai	Tektronix
Roel ter Horst	PTT Research
Fouad Tobagi	Starlight Networks
Metin Uz	David Sarnoff
Jan van der Meer	Philips
Christel Verreth	Telia Research
Jennifer Vollbrecht	Hughes Aircraft
Yoichi Yagasaki	Sony
Takeshi Yukitake	Matsushita
Joel Zdepski	David Sarnoff

Experiments & Results

Experiments on Cell Loss Resilience remaining after the Sydney meeting covered AC-leak and layered coding using Data Partitioning, but no experimental results were distributed or discussed within the Ad-hoc Group.

Discussions

Topics discussed by email covered:

- The need for better characterisation of real-world error characteristics.

It was recognised that expected values of error rates or cell loss would be beneficial in identifying the suitability of particular resilience techniques in different environments. However, given that the only variables are the average loss/error rate and the burst or temporal characteristics, it was expected that further knowledge of actual numbers would not influence the design of alternative error resilience strategies.

- Error resilience from layered coding with non-prioritised transport or storage.

There was some speculation about how a gain in error resilience could be achieved in this situation. The suggested mechanism is that there is a gain that actually arises from the shorter data unit lengths that result from generation of two bitstreams. The topic has not been very widely discussed and may not be fully understood yet, so no description is included in the draft text offered for inclusion in the Working Draft.

- Proposed text on error concealment for the WD.

WD text on error resilience has been prepared for previous meetings, but has not yet been included. The TSS SG15 ATM Experts Group considered the draft at its Melbourne meeting in April, and recommended amendments. The draft has been reorganised, reworked and expanded, and is offered again for inclusion at the New York meeting.

Topics for New York meeting

Particular issues in video error resilience that should be included in considerations by MPEG at the New York meeting are:

- Experimental results as described in TM5;
- Further improvements to the draft text;
- Particular attention should be given to consistency between Video and Systems groups in dealing with, or protecting in the event of, errors and losses. This is particularly true for multi-layer schemes reliant on the transport of multiple bitstreams.

Proposal

It is proposed that the following Annex be added to the MPEG Video WD at the New York meeting.

Annex - Text offered for Working Draft

D.3 Error Resilience

The coded video bitstream generated by this Recommendation | International Standard may be carried by different transport systems including packet systems. The Asynchronous Transfer Mode (ATM) of B-ISDN is an example of a packet system. ATM uses short, fixed length packets, called cells, consisting of a 5 byte header containing routing information, and a user payload of 48 bytes. The nature of errors on ATM is such that some cells may be lost, and the user payload of some cells may contain bit errors. Depending on AAL (ATM Adaptation Layer) functionality, indications of lost cells and cells containing bit errors may be available.

As an indication of the impact of cell loss in an ATM environment, the following table summarises the average interval between cell losses for a range of CLR and service bit rates. (A cell payload must be assumed for this. Allowing 1 byte/cell for AAL functions leaves 376 bits = 47 bytes). Note, however, that this summary ignores cell loss bursts and other shorter term temporal statistics.

	Average interval time of error			
CLR	5 Mb/s	10 Mb/s	50 Mb/s	100 Mb/s
10^{-2}	$7.52*10^{-3}$ s	$3.76*10^{-3}$ s	$7.52*10^{-4}$ s	$3.76*10^{-4}$ s
10^{-3}	$7.52*10^{-2}$ s	$3.76*10^{-2}$ s	$7.52*10^{-3}$ s	$3.76*10^{-3}$ s
10^{-4}	$7.52*10^{-1}$ s	$3.76*10^{-1}$ s	$7.52*10^{-2}$ s	$3.76*10^{-2}$ s
10^{-5}	7.52 s	3.76 s	$7.52*10^{-1}$ s	$3.76*10^{-1}$ s
10^{-6}	1.25 m	37.6 s	7.52 s	3.76 s
10^{-7}	12.5 m	6.27 m	1.25 m	37.6 s
10^{-8}	2.09 h	1.04 h	12.5 m	6.27 m

Bit Error Ratios (BERs) corresponding to the above mean times between errors can be calculated easily for the case of isolated bit errors. The BER that would cause the same incidence rate of errors is found by dividing by the cell payload size. ie. BER = CLR/376.

The following techniques of minimising the impact of lost cells and other error/loss effects are provided for reference, and indicate example methods of using the various tools available in this Recommendation | International Standard to provide good performance in the presence of those errors. Note that the techniques described may be applicable in the cases of packets of other sizes (eg. LANs or certain storage media) or video data with uncorrected errors of different characteristics, in addition to cell loss. It may be appropriate to treat a known erasure (uncorrected bit error(s) known to exist somewhere in a data block) as a lost data block, since the impact of bit errors cannot be predicted. However, this should be a decoder option.

The error resilience techniques are summarised in three categories, covering the restriction of the influence of a loss or error in both space (within a picture) and time (from picture to picture), and the methods of concealing the error once it has occurred.

D.3.1. Concealment

Concealment techniques hide the effect of losses/errors once they have occurred. Some concealment methods can be implemented using any encoded bitstream, while others are reliant on the encoder to structure the data or provide additional information to enable enhanced performance.

D.3.1.1 Temporal Predictive.

A decoder can provide concealment of the errors by estimating the lost data from spatio-temporally adjacent data. The decoder uses information which has been successfully received to make an informed guess at what should be displayed, under the assumption that the picture characteristics are fairly similar across adjacent blocks (in both the spatial and temporal dimensions). In the temporal case, this means estimation of errored or lost data from nearby fields or frames.

D.3.1.1.1 Substitution from previous frame

The simplest possible approach is to replace a lost macroblock with the macroblock in the same location in the previous picture. This approach is suitable for relatively static picture areas but block displacement is noticeable for moving areas.

Note that "previous picture" must be interpreted with care due to the use of bidirectional prediction and a difference between picture decoding order and picture display order. When a MB is lost in a P- or I-picture, it can be concealed by copying the data corresponding to the same MB in the previous P-picture or I-picture. This ensures that the picture is complete before it is used for further prediction. Lost MBs in B-pictures can be substituted from the last displayed picture, of any type.

D.3.1.1.2 Motion compensated

The concealment from neighbouring pictures can be improved by estimating the motion vectors for the lost MB, based on the motion vectors of neighbouring MBs (provided these are not also lost). This improves the concealment in moving picture areas, but there is an obvious problem with errors in intra coded MBs, because there are ordinarily no motion vectors. Encoder assistance to get around this problem is discussed in Section D.3.1.1.3.

Sophisticated motion vector estimation might involve storage of adjacent MB motion vectors from above and below the lost MB, for predictions both forward and backward (for B-pictures) in time. The motion vectors from above and below (if available) could then be averaged

Less complex decoders could use, for example, only forward prediction and/or only the motion vector from the MB above the lost MB. This would save on storage and interpolation.

D.3.1.1.3 Use of Intra MVs

The motion compensated concealment technique could not ordinarily be applied when the above and below MBs are Intra-coded, since there is no motion vector associated with Intra-coded MBs. In particular, in I-pictures, this type of concealment would not be possible with the normal calculation and use of motion vectors.

The encoding process can be extended to include motion vectors for intra MBs. Of course, the motion vector and coded information for a particular macroblock must be transmitted separately so that the motion vector is still available in the event that the image data is lost.

When "concealment_motion_vectors" = 1, motion vectors are transmitted with Intra MBs, allowing improved concealment performance of the decoders. The concealment motion vector associated with an Intra-coded MB is intended to be used only for concealment (if necessary) of the MB located immediately below the Intra-coded MB.

For simplification, concealment motion vectors associated with Intra-coded MBs are always forward, and are considered as frame motion-vector in Frame-pictures and field motion-vector in Field-pictures.

Therefore, encoders that choose to generate concealment motion vectors should transmit, for a given Intra-coded MB, the frame- or field-motion vector that should be used to conceal (i.e. to predict, with forward frame- or field-based prediction respectively) the MB located immediately below the Intra-coded MB.

Concealment motion vectors are intended for primarily for I- and P-pictures, but the syntax allows their use in B-pictures. Concealment in B-picture is not critical, since B-picture are not used as predictors, and it may be a waste to transmit concealment motion vectors in B-pictures.

Concealment motion vectors transmitted with Intra MBs located in the bottom row of a picture cannot be used for concealment. However, if "concealment_motion_vectors" = 1, those concealment motion vectors must be transmitted. Encoders can use the (0, 0) motion vector to minimise the coding overhead.

When concealment motion vectors are used, it is a good idea that one slice be equal to one row of MBs (or smaller), so that concealment can be limited to less than one row of MBs when a slice, or part of a slice, is lost.

Note that, when "concealment_motion_vectors"=1, Prediction Motion Vectors (PMVs) are NOT reset when an Intra MB is transmitted. Ordinarily, an Intra MB would reset the PMVs.

D.3.1.2 Spatial predictive

The generation of predicted, concealment MBs is also possible by interpolation from neighbouring MBs within the one picture. This is best suited to areas of high motion, where temporal prediction is not successful, or as an alternative means of concealment for Intra MBs when concealment motion vectors (Section D.3.1.1.3) are not available.

There are several possible approaches to spatial interpolation, and it could be carried out in the spatial or DCT domain, but normally it is only feasible and useful to predict the broad features of a lost MB, such as the DC coefficient and perhaps the lowest AC coefficients. Spatial prediction of fine detail (high frequencies) is likely to be unsuccessful and is of little value in fast-moving pictures anyway.

D.3.1.3 Layered Coding

It is possible to assist the concealment process further by arranging the coded video information such that the most important information is most likely to be received. The loss of the less important information can then be more effectively concealed. This approach is most effective if used with a transmission medium or storage device with different priority levels (such as priority-controlled cell-based transmission in the B-ISDN, or where different error protection or correction is provided on different channels). Such techniques include:

The components produced by the coding process can be placed in a hierarchy of importance according to the effect of loss on the reconstructed image. By indicating the importance of bitstream components and treating the individual components with due importance, superior error concealment performance may be possible.

Strategies available for producing hierarchically ordered bitstreams, or layers, include

data partitioning - the coded macroblock data is partitioned into multiple channels such that lower order channels contain address and control information and lower order DCT coefficients, while upper channels contain high frequency DCT coefficients. The channels are multiplexed at the System level.

frequency scalable pyramid - similar to data partitioning but lower order DCT coefficients are included within every channel. These lower order coefficients are a refinement of the quantisation error of the layer below.

spatial scalable pyramid - all DCT coefficients are a refinement of the quantisation error of the layer below.

These strategies produce channels which, when added progressively, produce increasing quality of the reconstructed signal. While some of these source coding techniques may result in a bit rate increase compared to the system without layering, the performance of the layered systems, when subjected to channel errors, may be greater.

The hierarchically ordered channels are handled with due quality, such that some cost function is minimised. This cost should be lower than that required for the system without layering. The bitstream components may be treated differently at one or more of the following locations:

- coder - different channel coding might be used
- channel - the channel may be able to provide different cell loss probabilities or error characteristics to the different bitstream components.
- decoder - error concealment could be performed differently within each bitstream

The costs respectively are the amount of redundancy introduced, the channel quality, and the decoder hardware complexity.

D.3.1.3.1 Use of Data Partitioning

A two layer data partitioning scheme, and the use of two channels with different error performance, are used here to illustrate the error concealment capability of layering.

At the coder the value of the PBP pointer may be

fixed - this implements a subband scheme

variable - the PBP pointer may be different for each slice such that the distribution of bits between the two channels is constant. The distribution may be different for I, P, and B frames.

Data partition 0 contains the address and control information and the low frequency DCT coefficients, while data partition 1 contains the high frequency DCT coefficients.

Two independent channels are required. Channel errors are distributed so that data partition 1 receives most errors.

It is assumed that at the decoder channel errors can be detected, so that actions can be taken to prevent errored data from being displayed. For data partition 1 errored data is simply not displayed. For data partition 0 a suitably high quality channel, or decoder concealment actions are required.

D.3.1.3.2 Use of Frequency Scalable Coding

Frequency Scalability can combine the advantages and functionality of a simple layered implementation for resolution and SNR scaling with error robustness features. In particular the Frequency Scalability syntax can be used to provide two bitstreams suitable for transmission in B-ISDN or other transmission or storage systems with unbalanced error characteristics. One bitstream containing the most important DCT-coefficients may then be allocated to a high priority virtual channel. Providing limited scalability features (but not the full capability offered by the scalability extension) would be an advantage that is a by-product of providing significant error resilience in this manner.

The layering is based on coefficient scanning in the DCT domain in a structured form, which allows the transmission of particular DCT-coefficients in specific layers with appropriate network priority if necessary. In a simple 2 layer pyramid scheme every 8x8 DCT-domain block in the encoder prediction error loop is divided into two DCT domain partitions. One partition is provided for the lower 4x4 coefficients transmitted in the lower layer and a second partition can contain information about all 8x8 DCT coefficients to be transmitted in the higher layer. The selected lower 4x4 coefficients are quantised using a quantiser QP4 and are next scanned and transmitted in the lower layer. In a second step the already coded coefficients from the lower 4x4 partition are reconstructed and subtracted from the original DCT coefficients. They are again quantised (QP8) and scanned for transmission together with the remaining 48 coefficients. At the decoder the coefficients contained in both layers are simply individually reconstructed (using inverse quantisation IQP4 and IQP8) and added before being treated the same as in a single layer MPEG2 Core decoder.

Only additional blocks for quantisation, inverse quantisation, summation and subtraction of lower layer 4x4 coefficients are required at the encoder. The decoder requires one additional inverse quantiser and one summation element.

For error robustness the following functionality provided by the Frequency Scalability syntax is emphasised:

- Dynamic allocation of coefficients into priority channels. Note that not all of the lower 4x4 coefficients have to be transmitted in the lower layer. Depending on bitrate constraints or visual perceptibility, the lower 4x4 DCT-coefficients can be allocated to either the high or low priority channel. In the extreme, if none of the lower coefficients is transmitted in the lower 4x4 layer, the coefficients are re-scanned for transmission in the higher layer. In this case the video is coded similar to a single layer implementation.
- Flexible allocation of bitrate between channels

The pyramid structure with quantisers QP4 and QP8 and re-scanning of lower layer coefficients in the higher layers enables the flexible allocation of bitrates between layers. Frequency scalability provides a "scalable side information" extension for transmission of the lower layer at low target bitrates.

D.3.1.3.3 Use of Spatial Scalable Coding

[Text required]

D.3.1.3.4 Layered coding in non-prioritised channels

[Text required]

D.3.2. Spatial localisation.

Spatial localisation encompasses those methods aimed at minimising the extent to which errors propagate within a picture, by providing early resynchronisation of the elements in the bitstream that are coded differentially between macroblocks.

Isolated bit errors may be detected through invalid codewords and so a decoder designer may choose to allow an errored sequence to be decoded. However, the effect on the picture is difficult to predict (legal, but incorrect, codewords could be generated) and it may be preferable to control the error through concealment of the entire affected slice(s) even when only one bit is known to be in error somewhere in a block of data.

When long consecutive errors occur (eg packet or cell loss), virtually the only option is to discard data until the next resynchronisation point is located (a start code at the next slice or frame header). By providing more resynchronisation points, the area of the screen affected by a cell loss can be reduced, in turn reducing the demands on the concealment techniques, and overall producing a better quality picture. Spatial localisation of errors is therefore dependent on controlling the slice size since this is the smallest coded unit with resynchronisation points (start codes).

D.3.2.1 Small slices

The most basic method for achieving spatial localisation of errors it to reduce the (fixed) number of macroblocks in a slice. The increased frequency of resynchronisation points will reduce the affected picture area in the event of a loss. It is effective in any transport or storage media, and in any profile since the slice structure is always present in MPEG coded video.

The method results in a small loss of coding efficiency due to the increase of overhead information. The loss is about 3% for 11 Macroblocks per slice and 12% for 4 Macroblocks per slice based on Rec. 601 picture format at 4 Mb/s, (percentages calculated relative to a system using 44 Macroblocks, or one picture width, per slice). The efficiency loss results in degradation of picture quality up to about 1 dB with 4 Macroblocks per slice and 0.2 dB with 11 Macroblocks per slice without errors at 4 Mb/s. However, the method performs approximately 1 to 5 dB better at CLR = 10^{-2}, depending on the concealment method used (simple MB replacement or motion compensated concealment).

From the view point of psychovisual picture quality, the performance of this method is generally dependent on the relative size of slice size and picture. Therefore, the slice size should be decided by considering the picture size (in macroblocks) and the trade-off between coding efficiency and visual degradation due to errors.

D.3.2.2 Adaptive slice size

There is a significant variation in the number of bits required to code a picture slice, depending on the coding mode, picture activity, etc. If slices contain only a few macroblocks, it will be possible that one cell could contain several slices. Offering multiple resynchronisation points in the same cell serves no purpose. Another problem with the simplistic short slice approach is that, because no account is taken of the cell structure, the first valid cell after a loss could contain most of the information for a slice, but it is unusable because the start code was lost.

An improvement over the small slice method may be to use adaptive slice sizes. As the encoder is producing the bitstream, it keeps track of the data contents within ATM cells. The start of a slice is placed at the first opportunity in every cell (or in every second, third, ...). This approach can achieve about the same spatial localisation of errors as small, but fixed size, slices, but with a greater efficiency.

Note, however, that this method ONLY gives an advantage for cell or packet based transmission, or where error detection occurs over a large block of data. The frequent resynchronisation points of small slice localisation are only wasteful if they are all lost in the event of an error. If isolated bit errors affect just one slice anyway, then there is no advantage in adapting the slice size.

Furthermore, the adaptive slice size technique requires an intimate connection between encoder and packetiser, to allow a new slice for a new packet or cell. As such, it may not be appropriate for some applications (eg. stored video intended to be distributed by multiple means).

D.3.3. Temporal Localisation

Temporal localisation encompasses those methods aimed at minimising the extent to which errors propagate from picture to picture in the temporal sequence, by providing early resynchronisation of pictures that are coded differentially. An obvious way to do this is to make use of intra mode coding.

D.3.3.1 Intra Pictures

Use of intra pictures is an efficient method for obtaining one form of temporal localisation. By use of intra pictures a single error will not stay in the decoded picture longer than (N + M -1) pictures if every Nth picture is coded intra and (M-1) B pictures are displayed before each I picture.

While the intra pictures, normally used as "anchors" for synchronising the video decoding part way through a sequence, are useful for temporal localisation, care should be taken in adding extra intra pictures (ie. reducing N) for error resilience. Intra pictures are expensive in terms of the number of bits produced and, as a result, are more likely to be affected by cell losses themselves.

D.3.3.2 Intra Slices

To avoid the additional delay caused by intra pictures, some applications requiring low delay may want to update the picture by coding only parts of the picture intra. This may provide the same kind of error resilience as intra pictures. As an example assume that a constant number of slices per picture from top to bottom are intra coded so that the whole picture is updated every P pictures. Three aspects of this kind of updating should be kept in mind:

- While an errored portion of the scene will ordinarily be erased within P pictures (with an average duration of about P/2), it is possible that motion compensation will allow the disturbance to bypass the intra refresh and it may persist as long as 2P pictures.
- To ensure that errors are not propagating into the updated region of the picture, restrictions could be put on motion vectors, limiting the vertical vector components to ensure that predictions are not made from the "oldest" parts of the picture.
- The visual effect of clearing errors can be similar to a windscreen wiper clearing water. This *windscreen wiper* effect can become noticeable in some cases in the unerrored sequence, unless the rate control mechanism ensures that the quality of the intra slice is close to that of the surrounding inter macroblocks.

D.3.4. Summary

The following table summarises the above error resilience techniques, with a guide to their applicability.

Category	Technique	Profile/Applicability
Concealment	Temporal predictive - substitution from previous picture	Any. Most suited to static pictures.
	Temporal predictive - Motion compensated	Any. Choice of sophistication in motion vector estimation.
	Temporal predictive - using concealment MVs	Any, but calculation of Intra MVs is an encoder option.
	Spatial predictive	Any. Not suitable for static, complex pictures.
	Layered coding	Freq. scalability, Spatial scalability, Data partitioning. Suitable for very high error/loss conditions.
Spatial Localisation	Small Slices	Any
	Adaptive slice sizes	Any, but requires knowledge of transmission characteristics when packet size is decided.
Temporal Localisation	Intra pictures	Any
	Intra slices	Any, but errors may persist longer than for Intra picture method.

It is not possible to provide a concise indication of error resilience performance, because assessments must necessarily be subjective and application dependent, and so should be taken as nothing more than a guide. It is also true that several different approaches to error resilience are likely to be used in combination. However, to provide some guidance to performance, the following descriptions are provided. They are the results of cell loss experiments, looking only at cell-based transmission of video information, coded at 4 Mbit/s according to the Main Profile, Main Level specification.

A simple MB substitution from a previous frame combined with the small-slice method (11 MBs per slice) will provide adequate picture quality for most sequences in the presence of rather low error rates of around $CLR=10^{-5}$.

Including sophisticated motion compensated concealment (with full spatial and temporal interpolation of motion vectors for lost MBs, and concealing losses in I, P and B pictures) provides adequate picture quality at $CLR=10^{-3}$.

Operation in environments with greater loss may require use of one of the layered coding methods. With adequate protection of the high priority channel, these schemes can provide adequate performance in the face of CLRs as high as 10^{-2} or even 10^{-1}.

- END -