A Cell-Loss Concealment Technique for MPEG-2 Coded Video

Jian Zhang, Member, IEEE, John F. Arnold, Senior Member, IEEE, and Michael R. Frater, Member, IEEE

Abstract—Audio-visual and other multimedia services are seen as important sources of traffic for future telecommunication networks, including wireless networks. A major drawback with some wireless networks is that they introduce a significant number of transmission errors into the digital bitstream. For video, such errors can have the effect of degrading the quality of service to the point where it is unusable. In this paper, we introduce a technique that allows for the concealment of the impact of these errors. Our work is based on MPEG-2 encoded video transmitted over a wireless network whose data structures are similar to those of asynchronous transfer mode (ATM) networks. Our simulations include the impact of the MPEG-2 systems layer and cover cell-loss rates up to 5%. This is substantially higher than those that have been discussed in the literature up to this time. We demonstrate that our new approach can significantly increase received video quality, but at the cost of a considerable computational overhead. We then extend our technique to allow for higher computational efficiency and demonstrate that a significant quality improvement is still possible.

Index Terms—Author: please supply index terms. E-mail: keywords@ieee.org for information.

I. INTRODUCTION

With the continuing trend toward the provision of mobile communications services, recent interest has been directed toward the transmission of real time video over wireless and other error-prone communications networks. With the possibility of cell-loss rates as high as 1%, the development of techniques to minimize the visual degradation caused by cell loss is clearly of considerable importance. Error resilience can be conveniently divided into four parts: error detection, resynchronization, data recovery, and concealment. While all of these parts are important, it is probably true to say that the use of good error concealment, the topic of this paper, will lead to the greatest improvement in the subjective quality.

A number of authors have proposed techniques that can aid in the error resilience of coded video [2]–[12]. In this paper, we briefly review techniques that have been proposed to achieve error concealment in compressed digital video bitstreams transmitted over cell-based telecommunications networks, such as asynchronous transfer mode (ATM) networks. A new concealment technique is then introduced. We verify this new technique by simulation running at the Multiplexing Layer so that the impact of the MPEG-2 systems layer is included. As we have shown elsewhere [12], this has a significant impact on the simulation results. Cell-loss probabilities up to 5% were applied to the well-known video sequences Flower Garden and Bus. These errored bitstreams were then used to measure the performance of our proposed concealment technique, as well as for comparison with other published techniques. We demonstrate that our technique significantly increases decoded video quality.

Error-resilience strategies based on concealment can operate with existing and future standards without requiring modifications to the syntax of the coded video. Examples include Lam et al. [17], Horst [21], and this paper. Strategies based on improved error detection, resynchronization and data recovery tend to require implementation of specific features in the video-coding syntax, and are therefore usually not compatible with existing standards. Examples of previous work in each of these categories include: error detection [2], [20], resynchronization [7], [18], [20] and data recovery [1], [2], [4], [6], [10]. As can be seen from these examples, many authors propose hybrid techniques to maximize performance.

Other approaches to error resilience are based on protecting certain parts of the bitstream from loss. These approaches are known to provide good performance at very high cell-loss probabilities, but can only be used in networks where an error-free channel is available to carry the base layer. This may be available in broadband packet-switched networks but would not usually be available in, for example, wireless applications. Such approaches are often based on scalability [19], [22] or data partitioning [23]. In each case, the base layer is transmitted in the protected channel and all loss occurs in the enhancement layer.

While Aravind et al. [22] present results using an MPEG-compatible scheme, Wen and Chung-Lin [19] combine the tools available in MPEG-2 with interleaving to minimize loss of adjacent blocks. Like the technique proposed here, Wen and Chung-Lin [19] use motion-compensated concealment using motion vectors derived from a motion search carried out in the decoder. The important differences between this and our technique are that we do not rely on any syntax modifications to the video bitstream nor on any part of the video bitstream being protected from error. Also, the computational complexity of our technique is easily scalable and we require only one inverse DCT operation per block in the decoder.

The novel aspects of this paper are:

1) our error-concealment technique is able to improve both objective and subjective video quality without requiring any change to the video compression algorithm (i.e., it is completely compatible with all existing standards);
2) the cell-loss probabilities that are examined in this paper range up to 5%—the upper end of this range is much
higher than that normally considered in video error resilience work, but is potentially very important for future applications such as wireless ATM networks and other mobile telecommunications systems.

The paper is organized as follows. In Section II, descriptions of relevant aspects of the MPEG-2 Video and Systems standards are briefly presented. In Section III, several error-concealment techniques are described including those suggested in the MPEG-2 standard, other published algorithms and our new concealment algorithm. Details of the simulation experiments performed are set out in Section IV, while the results are discussed in Section V. Further results, obtained by combining concealment techniques with three different resynchronization strategies, are outlined in Section VI, emphasising that good performance can be obtained using MPEG-2 at very high cell-loss rates. Conclusions are drawn in Section VII.

II. OVERVIEW OF THE MPEG-2 STANDARD

A. Video Compression

For a standard size television picture (720 × 576 pixels) and frame rate (25 Hz), MPEG 2 is designed to provide distribution quality television at bitrates between 4 and 9 Mbit/s depending on scene content. Each picture in a video sequence is coded in one of three modes. These are Intra (I), Predictive (P) and Bidirectional (B) [13]. The different picture types are usually used in a regular group of pictures structure. The one used in this work is the sequence IBBPBBPBBPBBI...

Coding within a picture is based on blocks, each of which consists of 64 luminance or chrominance pixels in an 8 × 8 pixel square. Luminance blocks are combined in groups of four, which, when combined with the associated chrominance information for this region of the picture, form macroblocks, which are of size 16 × 16 pixels. In P and B pictures, the prediction mode used can change on a macroblock by macroblock basis. Adjacent macroblocks are grouped into a slice. A picture consists of a number of slices preceding a picture header.

Similarly, a slice consists of a number of macroblocks preceded by a slice header. Each macroblock also begins with a header, which includes information on the macroblock location in the picture and motion vectors for use in the motion-compensated prediction if required. In the first macroblock of each slice, the macroblock address and motion vector are coded absolutely. In each remaining macroblock in a slice, these parameters are coded differentially with respect to the corresponding values in the macroblock immediately before it.

B. MPEG-2 Systems Layer

For many audio-visual applications, it is necessary to transmit simultaneously streams of both audio and video data on a single channel. For example, in videoconferencing applications there would usually be at least one video and one audio channel. The MPEG-2 systems layer [14] allows multiple streams of audio and video data to be combined to produce a single output stream. Packetization is carried out in two steps.

1) Each source bitstream to be transmitted is broken up into packets, known as packetized elementary stream (PES) packets. These packets are of variable length. For video, there would usually be one packet per slice.

2) PES packets are broken into transport stream (TS) packets, each of length 188 bytes. These TS packets are then time-multiplexed onto the output channel. TS packet headers include information that allows the decoder to channel the received data to the correct decoder (e.g., audio or video).

The systems layer decoder's ability to correctly decode one TS packet is not in any way affected by errors in the previous packet. It can also be seen, however, that any error in the TS packet header that corrupts the source stream identification will result in the loss of a whole transport packet, even though the remainder of the data in that packet may be received correctly. Almost twice as many cells are not available to the video decoder when the effect of the systems layer is taken into account [12]. In the experiments described here, data is packed into ATM cells using AAL-1 as described in [12].

III. METHODS FOR CONCEALMENT OF CELL LOSS

The MPEG-2 standard suggests three methods for enhancing the error resilience of coded video information. These are temporal localization, spatial localization and error concealment. The major contribution of this paper lies in the area of error concealment and so in this section we review concealment techniques suggested in the MPEG-2 standard and in the technical literature. We also introduce our new concealment approach.

A. Error Concealment in the MPEG-2 Standard

These techniques attempt to conceal an error once it occurs by taking into account the remaining spatial and temporal correlation in the decoded video sequence. In areas of the picture that do not change very much with time, it is effective to conceal the effect of cell loss by temporal replacement, i.e., by using information from the corresponding position in a previous decoded picture. Naturally, this approach is not very effective in high-motion areas. In this situation, spatial interpolation, where missing parts of the picture are interpolated from decoded information in macroblocks surrounding the lost macroblocks, tends to be more effective. Spatial interpolation tends to work well in low detail areas of a picture but is of little use in areas containing significant detail.

Motion-compensated concealment, which combines both temporal replacement and motion compensation, can be used to improve the effectiveness of concealment. The technique works by exploiting the fact that there is generally high correlation between nearby motion vectors in a picture. Motion vectors for macroblocks above or below a macroblock missing due to cell loss can be used to predict the motion vectors of the lost macroblock. These motion vectors are then used to find a block in the previous decoded picture which will hopefully provide a good estimate of the lost information. However, this approach is not able to conceal errors for a lost macroblock which is surrounded by intra-coded macroblocks. To avoid this, the MPEG-2 standard [13] allows the encoding process to be optionally extended to include motion vectors for intra-coded macroblocks. Of course, the motion vector and the coded
information for a macroblock should be transmitted separately (e.g., in different ATM packets) so that the motion vector is still available in the event that the other macroblock data is lost.

In our experiments, when both the motion vectors in the macroblock directly above and below the lost macroblock are known, using only the motion vectors in the macroblock directly above and using the interpolated vectors of the two led to indistinguishable performance. We therefore used the above motion=vector approach in our simulation.

B. The Boundary Matching Algorithm (BMA)

This algorithm [17] exploits the fact that adjacent pixels in a video picture exhibit high spatial correlation. It takes the lines of pixels above, below, and to the left of the lost macroblock in the current picture and uses them to surround each candidate block from the previous decoded picture. It then calculates the total squared difference between these three lines and the corresponding three lines on the edge of a 16 × 16 block of data within a previous decoded picture. This is illustrated in Fig. 1. The BMA estimates the lost motion vector as the one in which the squared difference between the surrounding lines (from the current decoded picture) and the block (from the previous decoded picture) is a minimum. Referring to Fig. 1, this means we minimize the total squared difference calculated by summing the following three squared differences:

1) the squared difference between the pixels above the block and the pixels on the top line of the block (i.e., region A in Fig. 1);
2) the squared difference between the pixels to the left of the block and the pixels on the left edge of the block (i.e., region B in Fig. 1);
3) the squared difference between the pixels below the block and the pixels on the bottom line of the block (i.e., region C in Fig. 1).

The search method employed to estimate the lost motion vector could be a full search over some area in the previous picture. Alternatively, the search process can be greatly speeded up if only a small number of candidate motion vectors are considered. These might include:

1) motion vectors for the same macroblock in the previous picture;
2) motion vectors associated with available neighboring macroblocks;
3) median of the motion vectors of available neighboring macroblocks;
4) average motion vectors of the available neighboring macroblocks;
5) zero motion vectors.

As described, the algorithm has two forms. When motion compensation is employed, there are two types of data that need to be transmitted in a macroblock: namely, the motion vectors and the coded displaced picture difference. Initially, only the loss of the motion vectors was considered with the coded displaced picture difference assumed to be received correctly. In addition, the case where both the motion vector and the coded displaced picture difference were lost was also considered. Only this latter case is relevant in our study, since in an MPEG-2 video bitstream, it is certain that if the motion vector is lost, then the coded DCT coefficients in that macroblock will also be lost since resynchronization cannot occur until the next macroblock at the earliest (and then only if the next macroblock is proceeded by a slice header).

This technique has some significant limitations. In the first place, using only the three boundary lines to match the entire 16 × 16 block is not sufficient in many cases. Furthermore, very often all three of these lines are not available for matching when cell loss occurs. The BMA allows backward operation (i.e., from the first correctly received macroblock backward to predict lost macroblocks). However, if all the remaining macroblocks in a row of macroblocks are lost, then this does not help. If the macroblock directly above or directly below is also lost, then the performance of the technique is further degraded.

C. Decoder Motion-Vector Estimation (DMVE) Algorithm

We now introduce our new algorithm which, like BMA, aims to accurately estimate the motion vectors of any lost macroblocks using correctly received information at the decoder. While BMA uses spatial correlation to estimate the motion vectors, the DMVE algorithm primarily exploits temporal correlation in the estimation process. As we explain below, a process similar to the motion estimation performed at the encoder is used to compute the missing motion vectors.

When cell loss occurs, several lines (two to eight) of information around any lost macroblocks are taken. This includes information in the macroblock above the lost macroblock (even if this macroblock is itself a concealed macroblock), the macroblock below the lost macroblock (if received correctly), and the macroblock to the left of the lost macroblock (even if this macroblock is itself a concealed macroblock). In addition, we include pixels from the above-left (even if this macroblock is itself a concealed macroblock) and below-left macroblocks (if received correctly) to complete the encirclement of the lost macroblock. If we assume that only two surrounding lines are used...
Fig. 2. Matching areas employed in DMVE algorithm.

and that all the required macroblocks in the current picture are available, then the pixels used are as shown in Fig. 2.

The algorithm then performs a full search within the previous picture for the best match to the lines of decoded pixels that surround the lost macroblock from the current picture. The macroblock of data which is surrounded by the lines which best match the lines from the current picture is assumed to be the best match to the lost macroblock. In our experiments, we used a search area of $+16$ pixels both horizontally and vertically. However, this is a decoder option and can be used to trade performance for lower computational complexity. Motion-vector estimation is performed at half pixel accuracy since the filtering associated with half pixel accuracy prediction tends to smooth any blocking effects at macroblock boundaries. In order to speed up processing, a search at single pixel accuracy was first performed, and then the best match within $+0.5$ pixels horizontally and vertically surrounding that point was chosen for concealment.

Let us now consider the advantages of this method when compared to the motion-compensated concealment scheme proposed within the MPEG-2 standard. For the approach proposed in MPEG-2, we have assumed that the motion vectors associated with the macroblock directly above the lost macroblock should be used for motion-compensated concealment. Due to the strong spatial correlation within typical motion-vector fields, this is a reasonable assumption. However, noticeable artefacts appear when this assumption is incorrect. We found that this could be a particular problem at the edges of low complexity regions within the picture. For example, in the Flower Garden sequence, we found that macroblocks from the roof of a house in the scene background could be relocated into the flat blue sky with obvious subjective impairment. Using several lines of pixels that surround the entire lost macroblock as the basis of motion esti-

IV. EXPERIMENTAL PROCEDURE

Two well-known video test sequences used during the development of MPEG-2 (Flower Garden and Bus) were used for our experiments. Each video sequence was coded using MPEG-2 compatible software, with an output bit rate of 4 Mbit/s. Concealment motion vectors were used in I pictures and these motion vectors was used for concealment if available. If the concealment motion vectors was not available due to the loss of the macroblock above the current (lost) macroblock, then the concealment technique being studied was employed. Random cell loss was used in our simulation. In all the results reported below, we assume one slice per row of macroblocks (i.e., 44 macroblocks per slice). Two sets of results are shown based on cell-loss probabilities of 1% and 5%. For each error condition, we studied each of the following techniques to conceal the effect of cell loss:

- block replacement (i.e., replace the lost macroblock with the corresponding macroblock in the prediction picture);
- above motion-vector scheme (i.e., motion-compensated concealment using the motion vector of the macroblock directly above the lost macroblock);
- BMA algorithm;
- DMVE algorithm with optional candidate search (eight-line search);
- DMVE algorithm (two-line search);
- standard DMVE algorithm (eight-line search);
- DMVE algorithm with bi-directional prediction (eight-line search).

The order shown provides a gradation from low to high computational complexity.

So as to achieve statistically significant results, each cell-loss experiment was repeated with 25 different cell-loss patterns, each based on a different random number seed. It should be pointed out that since the coded bitstream in each case is iden-
tical and the position of lost cells for a given seed is also identical, the difference in decoded service quality is only a function of the performance of the concealment algorithm.

We choose random rather than bursty cell loss (as was used during the development of the MPEG-2 standard), since this represents the more severe test. When any cell loss occurs, the remainder of the received bitstream is lost until the next resynchronization point (in our case, a slice header) is received. If several cells are lost consecutively, it is still quite possible that resynchronization will be achieved at the same point as if a single cell loss had occurred. It is therefore cell-loss events, rather than total cells lost, which is most important when studying the impact of cell loss. Random cell-loss represents close to the worst-case number of cell-loss events for a given cell-loss probability since consecutive cell-loss is rare.

The decoded video quality is significantly reduced if either an I picture or a P picture is completely lost due to the loss of the picture header information as result of a cell loss. For example, the loss of a single I picture can result in a PSNR drop of up to 1 dB in the quality of a decoded sequence. The feature of the MPEG-2 Transport Layer permitting repeated transmission of packets was used in all simulations to overcome this problem.

As well as considering the quality of the decoded errored bitstream (measured using PSNR), we also attempted to quantify the computational complexity of each method. We did this by measuring the decoding time required for each technique and comparing it to the decoding time for a nonerrored sequence. Thus, a CPU time of 2.0 indicates that the decoding time for the errored sequence was twice the time required to decode the nonerrored sequence.

V. EXPERIMENTAL RESULTS

The experimental results obtained when cell loss was introduced are shown in Tables I and II. The error-free PSNR’s were 28.38 dB for Flower Garden and 30.71 dB for Bus.

At high cell-loss rates (1%) using only block replacement results in a significant loss of about 6.5 dB–7.0 dB in decoded video quality. Using motion-compensated concealment based on the motion vector in the macroblock directly above the lost macroblock recovers approximately 3 dB of this loss. The BMA performs slightly better in the case of the Flower Garden sequence and somewhat worse in the case of the Bus sequence. The DMV algorithm further improves the performance from 0.4–0.8 dB in the case of DMVE-OCS. More than 1 dB of improvement is achieved when an eight-line search is utilized together with bi-directional search in the case of B pictures. When this last technique is used, decoded video quality is reduced by around only 2 dB, compared to the nonerrored case.

Computational complexity is, of course, an important issue. The results show that DMVE, when using the fast DMVE-OCS approach, has comparable computational complexity to the simpler approaches, and thus, its benefits can be gained with little overhead. The more computationally complex forms of DMVE do result in further improvements in decoded video quality, but at a computational overhead of a factor of two (DMVE with two-line search), five (DMVE with eight-line search) or six (DMVE with eight-line search and bi-directional search in B pictures).

At very high cell-loss rates (5%), the drop in decoded video quality using block replacement is as high as 12 dB. Using motion-compensated concealment based on the motion vector in the macroblock directly above the lost macroblock or the BMA recovers around 1.5 dB of this loss and the various forms of DMVE achieving a further gain of around 1.5 dB. Computational complexity rapidly rises as more complex forms of DMVE are employed. Even DMVE-OCS introduces a 60%–70% increase in the computation requirement, which is comparable to BMA, but considerably higher than the simpler motion-compensated concealment based on the motion vector in the macroblock directly above the lost macroblock.

The increase in computational complexity as loss probability is increased reflects the increased number of macroblocks requiring concealment. The use of more regular resynchronization can reduce the number of macroblocks requiring concealment, and hence, can reduce the amount of extra computation required.

All decoded sequences were subjectively viewed with viewers agreeing that the subjective improvement achieved by DMVE was even more than might have been expected from the PSNR values achieved. Figs. 3 and 4 show pictures from the two sequences used in our experiments for motion-compensated concealment based on the motion vector in the macroblock directly above the lost macroblock is used and DMV with eight-line search and bi-directional search for B pictures is employed. These figures indicate that the subjective improvement is indeed significant.

In summary, we have shown that simple application of the error-concealment techniques defined within the MPEG-2 standard can improve the quality of the decoded video significantly when cell loss occurs. However, decoded video quality can be further improved in high cell-loss environments by the use of the DMVE. This approach can be implemented in a decoder that
is completely compliant with the MPEG-2 video standard. Further, this technique can be combined with other error resilient coding techniques, such as temporal localization, spatial localization, or macroblock-resynchronization [12], [18], to further improve decoded video quality.

VI. COMBINED RESYNCHRONISATION AND CONCEALMENT

We performed some experiments to quantify the improvement in decoded service quality when the DMVE approach was combined with improved resynchronization approaches. Four methods of resynchronization after cell loss were considered namely 44 macroblocks per slice (as used in previous experiments), 11 macroblocks per slice, adaptive slice sizes (where the number of bits per slice is held approximately constant, and thus, the number of macroblocks per slice varies) and macroblock resynchronization [18]. The first three of these are MPEG-2 compliant, while the last is not. Results are shown in Table III for concealment using the motion vector of the macroblock above the lost macroblock and in Table IV for the DMVE approach. The Flower Garden sequence coded at 4 Mbit/s was again used.

Using the improved resynchronization approaches reduced the no-loss quality due to the extra overhead bits required to provide the resynchronization. However, the improvement in decoded service quality after cell loss is dramatic, with around a 5-dB improvement between the poorest resynchronization approach (44 macroblocks per slice) and the best (macroblock resynchronization) in both cases for a cell-loss rate of 5%. Note also that the adaptive slice approach (which is MPEG-2 compliant) performs almost as well as macroblock resynchronization. Finally, DMVE continues to perform substantially better than the above motion-vector scheme (by around 1.5 dB) even...
when improved resynchronization is employed. In the case of both the adaptive slice approach and macrorblock resynchronization, the PSNR decrease is just above 2 dB at a cell-loss rate of 5%.

VII. CONCLUSION

This paper has reviewed several algorithms that can be used to conceal the effect of cell loss in the transmission of MPEG-2 coded video. We have shown that the DMVE algorithm can significantly increase decoded video quality even at high cell-loss rates (1% or more). Recognizing that the costs of decoders needs to be minimized in many applications, we have developed a number of different implementations of the DMVE algorithm which allow a trade-off to be made between extra computational complexity required at the decoder and decoded service quality. For all but the very highest cell-loss rates studied, DMVE-OCS seems to provide reasonable decoded service quality with low computational overhead. When combined with good resynchronization, DMVE can provide good quality decoded services, even at cell-loss probabilities of 5%.

REFERENCES


Since 1997, he has been with the Motorola Australia Research Center, Sydney, Australia, where he is a Senior Engineer. His research interests include data compression, image processing, and video coding for wireless communications.

John F. Arnold (S’77–M’85–SM’96) received the B.E. and M.Eng.Sc. degrees from the University of Melbourne, Australia, in 1976 and 1979, respectively, and the Ph.D. degree from the University of New South Wales, Canberra, Australia, in 1984. Since 1978, he has been with the School of Electrical Engineering, University of New South Wales, initially at the Royal Military College, Duntroon, and more recently at the Australian Defence Force Academy, Canberra, Australia. He is currently a Professor and Head of the School of Electrical Engineering. His research interests lie in the fields of video coding, error resilience of compressed digital video, and coding of remotely sensed data.

Dr. Frater is currently an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.