ERROR CONCEALMENT TECHNIQUES FOR ENCODED VIDEO STREAMS

Paul Salama, Ness B. Shroff, Edward J. Coyle, and Edward J. Delp

Purdue University
School of Electrical Engineering
Computer Vision and Image Processing Laboratory
West Lafayette, Indiana 47907-1285

ABSTRACT

In this paper we describe two error-recovery approaches for MPEG encoded video over ATM networks. The first approach aims at reconstructing each lost pixel by spatial interpolation from the nearest undamaged pixels. The second approach recovers lost macroblocks by minimizing intersample variations within each block and across its boundaries. Moreover, a new technique for packing ATM cells with compressed data is also proposed.

1. INTRODUCTION

Broadband networks are expected to support a variety of exciting applications involving high-resolution video and images. Asynchronous Transfer Mode (ATM) is expected to be the target communication protocol for these networks. To effectively transmit video traffic over these networks we need to study the issues involved in packetizing encoded video sequences. In particular it is important to study the effect of ATM cell loss, and develop post processing techniques that can be used for error concealment.

Issues such as the influence of packetization defects on picture quality, suitability of current video encoding techniques for ATM networks, and the benefits of bandwidth flexibility offered by ATM techniques have been examined [1, 2]. Error-concealment approaches in [3] have assumed that both encoding and decoding occur simultaneously with the decoder communicating to the encoder the location of damaged picture blocks. Many of these techniques are not realistic for real-time applications since they require retransmission of ATM cells. Prioritization approaches to ATM cell loss concealment have been proposed [4, 5, 6]. Techniques involving interleaving data have also been proposed [7, 8], along with post processing techniques for error concealment [9, 10, 7, 11, 12, 13]. In all of the above techniques there has been no mention of how the loss of macroblocks is detected. Furthermore, there been no reference as to how bits pertinent to one macroblock are distinguished from those belonging to another macroblock in the event that cell loss has occurred. Our approach allows for the detection of macroblock loss and the recognition of the location of the lost macroblocks. The decoder then does not have to decide whether a group of bits belongs to a damaged macroblock or to an undamaged one. This reduces the complexity of the decoding process.

This work was partially supported by the National Science Foundation, under Grant No. CDR-8803017 to the Engineering Research Center for Collaborative Manufacturing and a grant from the AT&T Foundation. Address all correspondence to E. J. Delp, ace@ecn.purdue.edu or (317) 494-1740.

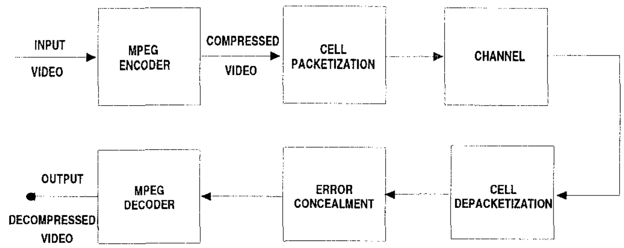

Figure 1: Block diagram of the packing/error concealment scheme

In this paper we describe two approaches to lost signal restoration in encoded MPEG video sequences. The inherent assumption is that an error in transmission has resulted in complete loss of some of the macroblocks. This arises when cells carrying bits from more than one macroblock are discarded due to network congestion. The first approach aims at reconstructing each lost pixel by spatial interpolation from the closest undamaged pixels. The second approach recovers lost macroblocks by minimizing intersample variations within each block and across its boundaries.

We also propose a new technique for packing ATM cells with compressed data. The idea is to pack high priority cells with group of pictures (GOP), Picture and Slice header information and pack integral numbers of macroblocks into single cells.

Figure 1 shows our approach to video packetization using ATM. The cell depacketization operation also provides information as to which macroblocks are missing. This information is passed to the error concealment algorithm which attempts to conceal the missing blocks.

2. CELL PACKING

A study of MPEG encoded video sequences reveals that it is possible to pack more than one macroblock into a single ATM cell. In fact, in the case of the *salesman* sequence, the average number of bits per macroblock is 75 with a minimum of 6, a maximum of 623, and a standard deviation

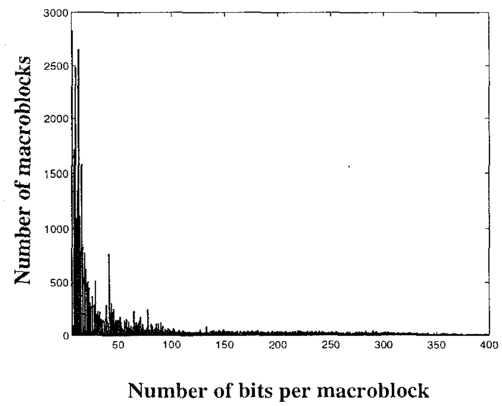

Figure 2: Macroblock size histogram of the *salesman* sequence

of 103. Figure 2 is a plot of the *salesman* sequence macroblock histogram. Our approach to packing cells with a user payload of 48 bytes has been to, first, separately pack GOP, Picture and Slice data into a high priority cell and, secondly, to pack successive cells with an integral number of macroblocks. The aim from such a scheme is two fold. The first is to guarantee the safe arrival of synchronizing information present in the GOP, Picture and Slice headers, and the second is to localize the loss of macroblocks within a frame while maintaining the ease of decoding the correctly received macroblocks. This is crucial to MPEG streams since a loss of a few bits can perturb the decoding process, resulting in the loss of whole frames.

3. ERROR CONCEALMENT

The goal of error concealment is to estimate missing macroblocks in the MPEG data that were caused by dropped ATM cells. The use of spatial, temporal, and picture quality concepts are exploited.

Let D denote a macroblock of NxN samples that has been lost during the process of transmission, $x_{i,j}$ the decoded value of the sample at the i^{th} row and j^{th} column of D and $\hat{x}_{i,j}$ the reconstructed value. Below we discuss two methods for reconstructing lost macroblocks.

3.1. Interpolation

For this method, every pixel in D is reconstructed by spatially averaging the values of its four closest neighbors as shown in Figure 3.

$$\begin{aligned} x_{i,j} &= \lambda[\mu_1 x_{i,-1} + (1-\mu_1)x_{i,N}] + \\ &\quad (1-\lambda)[(1-\mu_2)x_{-1,j} + \mu_2 x_{N,j}] \end{aligned} \quad (1)$$

The weighting coefficients μ_1 and $1-\mu_1$ are used to weigh the contributions from the pixels on either side of the lost pixel, and μ_2 and $1-\mu_2$ are used to weigh the contributions from those above and below the lost pixel. The coefficient μ_1 is a function of the distances between the lost

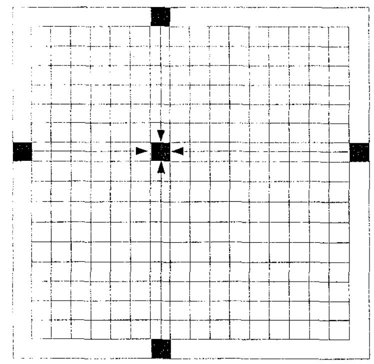

Figure 3: The four closest pixels used in spatial averaging

pixel and its closest neighbors lying on the same row, while μ_2 depends on the distances between the lost pixel and its closest neighbors on the same column. The contributions from the macroblocks on either side are weighted by λ, and those from the ones above and below are weighted by $1-\lambda$. The advantage of this method is that it is simple, fast, and results in good reconstruction. We have also investigated the use of more complicated approaches that use *a-priori* models of the image structure, in particular the use of regularization techniques [14], B-Splines [15],and wavelet approaches [16].

An alternative approach has been proposed to estimate missing edges in each block from edges in the surrounding blocks [12]. For each direction of an estimated edge, a version of the lost block is reconstructed by performing a number of one dimensional interpolations carried out along several lines parallel to the edge. A final version of the missing block is obtained by merging all previously attained versions together.

3.2. Optimal Iterative Reconstruction

The second proposed method aims at reconstructing the lost macroblocks by minimizing a cost function. This cost function, f, is the sum of the weighted square differences between each lost pixel value and its neighbors [17, 9], which include pixels from surrounding undamaged blocks shown in the dark region in Figure 4. Thus,

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}\sum_{(i,j)\epsilon D} [\omega^w_{i,j}(\hat{x}_{i,j} - \hat{x}_{i,j-1})^2 \\ &\quad +\omega^e_{i,j}(\hat{x}_{i,j} - \hat{x}_{i,j+1})^2 \\ &\quad +\omega^n_{i,j}(\hat{x}_{i,j} - \hat{x}_{i-1,j})^2 \\ &\quad +\omega^s_{i,j}(\hat{x}_{i,j} - \hat{x}_{i+1,j})^2] \end{aligned} \quad (2)$$

where $\mathbf{x}$ is a vector of length N^2 composed of the samples of D arranged in lexicographic order, and $\omega^w_{i,j}$, $\omega^e_{i,j}$, $\omega^n_{i,j}$ and $\omega^s_{i,j}$ are the weighting coefficients for the pixel values west, east, north, and south, respectively, of the current pixel. Equation 2 can be written as

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{x}^T\mathbf{b} + c \quad (3)$$

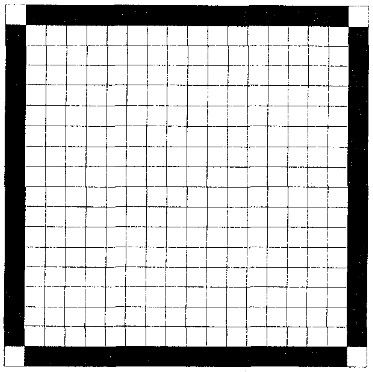

Figure 4: Boundary information used in the iterative reconstruction process

where

$$\mathbf{b} = \sum_{k=1}^{4} [\mathbf{S}_k - \mathbf{Q}_k^T \mathbf{S}_k] \mathbf{b}_k \tag{4}$$

$$\mathbf{Q} = \sum_{k=1}^{4} [\mathbf{S}_k - \mathbf{Q}_k - \mathbf{Q}_k^T + \mathbf{Q}_k^T \mathbf{Q}_k] \tag{5}$$

$$c = \sum_{k=1}^{4} \mathbf{b}_k^T \mathbf{S}_k \mathbf{b}_k. \tag{6}$$

Here $\mathbf{S}_k$, $k = 1, 2, 3, 4$, are diagonal matrices with entries $\omega_{i,j}$ for the four directions respectively. The matrices $\mathbf{Q}_k$ are upper and lower diagonal matrices with zeros along the diagonals and satisfy

$$\mathbf{Q}_2^T = \mathbf{Q}_1 \tag{7}$$

and

$$\mathbf{Q}_4^T = \mathbf{Q}_3 \tag{8}$$

The vectors $\mathbf{b}_k$, $k = 1, 2, 3, 4$, contain zeros and values from the four borders respectively.

It can be shown that, when all weighting coefficients are set to one, $\mathbf{Q}$ is positive definite, which is necessary for attaining the optimal solution. In this case the optimal solution is

$$\mathbf{x}_{opt} = \mathbf{Q}^{-1} \mathbf{b}. \tag{9}$$

The above equation can be iteratively solved. For instance, the following equations depict the l^{th} iteration for $\mathbf{x}$ in the event that the steepest descent algorithm is used to find $\mathbf{x}$.

$$\mathbf{x}_l = \mathbf{x}_{l-1} - \frac{\mathbf{g}_{l-1}^T \mathbf{g}_{l-1}}{\mathbf{g}_{l-1}^T \mathbf{Q} \mathbf{g}_{l-1}} \mathbf{g}_{l-1} \tag{10}$$

where

$$\mathbf{g}_l = \mathbf{Q} \mathbf{x}_l - \mathbf{b} \tag{11}$$

is the l^{th} iteration of the gradient vector of f.

3.3. Reconstruction with Motion Compensation

Since MPEG sequences contain predicted and bidirectionally interpolated pictures, it is necessary to consider reconstruction of macroblocks lost within such frames. Our approach has been to average the motion vectors of surrounding macroblocks [3, 4] and use the average vector to retrieve a version of the lost macroblock. This retrieved macroblock is then averaged with another version obtained via spatial interpolation.

4. RESULTS

Figure 5a is a frame from the *salesman* sequence, Figure 5b is the same frame that is missing ten percent of its macroblocks. Figure 5c shows the reconstructed version using linear interpolation and Figure 5d shows the difference between the original and reconstructed version. It can be observed that the reconstruction algorithm performed well except that it smeared edges. The advantage to such a technique is its simplicity and speed.

Figure 6a is a motion compensated frame from the same sequence. Figure 6b shows, with very little motion artifacts, a reconstructed version using interpolation only. Figure 6c shows a reconstructed version by means of motion estimation and replacement of lost macroblocks by those indicated in the previous frame, while Figure 6d shows the average of Figure 6b and Figure 6c. The reconstructed versions are acceptable with very little motion artifacts.

More information describing our proposed techniques, the video sequences used in our study along with a postscript version of our paper, is available via anonymous ftp to `skynet.ecn.purdue.edu` in the directory `/pub/dist/delp/icip95-errcon`.

5. REFERENCES

[1] W. Verbiest, L. Pinnoo, and B. Voeten, "The impact of the ATM concept on video coding," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1623–1632, December 1988.

[2] F. Kishino, K. Matanabe, Y. Hayashi, and H. Yasuda, "Variable bit rate coding of video signals for ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 5, pp. 801–806, June 1989.

[3] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE Journal on Selected Areas in Communication*, vol. 7, no. 5, pp. 807–814, June 1989.

[4] M. Ghanbari and V. Seferidis, "Cell-loss concealment in ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, pp. 238–247, June 1993.

[5] M. Ghanbari and C. Hughes, "Packing coded video signals into ATM cells," *IEEE/ACM Transactions on Networking*, vol. 1, no. 5, pp. 505–508, October 1993.

[6] P. Pancha and M. E. Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Communications Magazine*, vol. 32, no. 5, pp. 54–66, May 1994.

[7] Q. Zhu, Y. Wang, and L. Shaw, "Coding and cell loss recovery in DCT based packet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 3, pp. 248–258, June 1993.

[8] A. S. Tom, C. L. Yeh, and F. Chu, "Packet video for cell loss protection using deinterleaving and scrambling," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 1991, Toronto, Canada, pp. 2857–2860.

[9] Y. Wang, Q. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Transactions on Communications*, vol. 41, no. 10, pp. 1544–1551, October 1993.

[10] Y. Wang and Q. Zhu, "Signal loss recovery in DCT-based image and video codecs," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, November 1991, Boston, Massachusetts, pp. 667–678.

[11] H. Sun and J. Zdepski, "Adaptive error concealment algorithm for MPEG compressed video," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, November 1992, Boston, Massachusetts, pp. 814–824.

[12] W. Kwok and H. Sun, "Multidirectional interpolation for spatial error concealment," *IEEE Transactions on Consumer Electronics*, vol. 3, no. 39, pp. 455–460, August 1993.

[13] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 470–477, April 1995.

[14] R. L. Stevenson, "Reduction of coding artifacts in transform image coding," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, April 1993, Minneapolis, Minnesota, pp. V401–V403.

[15] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I–theory," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–833, February 1993.

[16] M. Unser, "Efficient dyadic wavelet transformation of images using interpolation filters," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, April 1993, Minneapolis, Minnesota, pp. V149–V152.

[17] W. Grimson, "An implementation of a computational theory of visual surface interpolation," *Computer Vision, Graphics, Image Processing*, vol. 22, no. 1, pp. 39–69, April 1983.

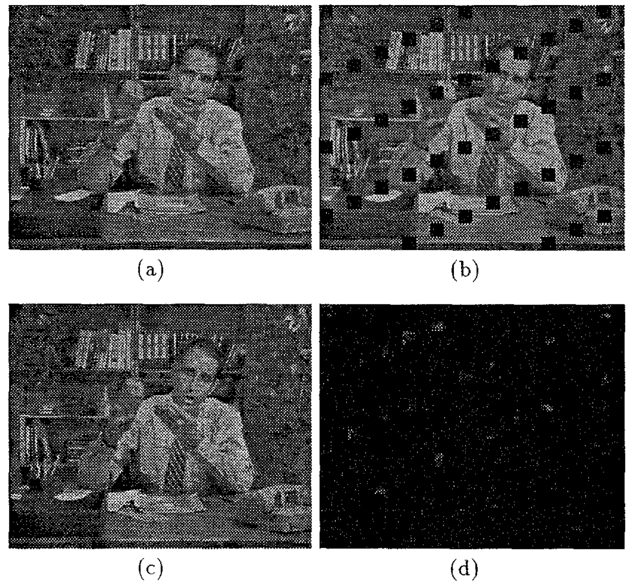

Figure 5: (a): Original image from the sequence. (b): Image with ten percent of its macroblocks missing. (c): Image reconstructed via spatial interpolation. (d): Difference image of (a) and (c).

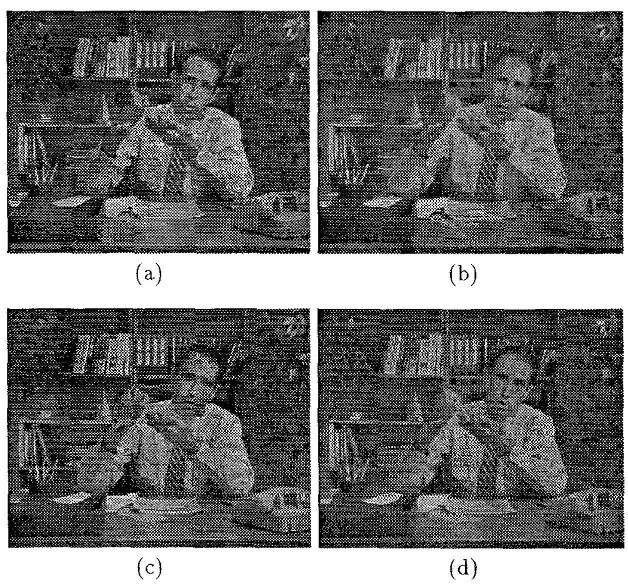

Figure 6: (a): Predicted frame from the sequence. (b): Predicted frame reconstructed using spatial interpolation. Ten percent of the macroblocks were missing. (c): Predicted frame reconstructed using temporal replacement. (d): Average image of (b) and (c).