# Error-Propagation Prevention Technique for Real-Time Video Transmission over ATM Networks

Jong-Tzy Wang and Pao-Chi Chang

***Abstract*—Video sequences compressed by the current video-compression standards—such as MPEG-1/2 and H.261/H.263, which include motion compensation and variable-length coding—are very sensitive to channel disturbances. There exist many error-concealment techniques that can improve the video quality substantially. However, they generally do not prevent or terminate the error propagation. The forced intraupdate technique was proposed for H.245 recently. In this paper, we present an efficient temporal error-propagation prevention method that implements the forced intraupdate concept implicitly. Once an error is detected at the decoder, the starting address of the erroneous macroblocks is sent back to the encoder. The encoder marks the possible damaged area, and the encoding process continues as normal except that the motion estimation will probably not refer to the erroneous part. Thus, the impact of the cell loss is limited to the erroneous slice only, and the damage from the error propagation is greatly reduced. Simulation results of MPEG-2 coding over asynchronous transfer mode (ATM) networks show that the error concealment with feedback can effectively isolate the error and reduce the damage to give satisfactory performance even when the cell-loss rate is higher than 1%. With the delay analysis of ATM networks, we also show that in most cases, the encoder has adequate time to get the feedback information before processing the next I- or P-frames.**

***Index Terms*—Asynchronous transfer mode (ATM), error concealment, error prevention, error propagation, quality of service, video coding.**

## I. INTRODUCTION

THE MOST commonly used video-compression techniques are based on MPEG-1/2 [1]–[3] or H.26x—i.e., H.261 or H.263 [4], [5]—where motion compensation, discrete cosine transform (DCT)-based coding, variable-length coding (VLC), and adaptive quantization are performed. These coding techniques are very sensitive to channel disturbances. A single error may cause error propagation in both spatial and temporal domains. Both MPEG and H.26x streams are constructed by layered structures. In MPEG, a group-of-pictures (GOP) layer generally includes one intrapicture (I-frame), several predicted pictures (P-frames), and several bidirectional predictive pictures (B-frames). The H.26x coding streams may consist of one I-frame and many P-frames or PB-frames. Each picture is further divided into a number of slices, or group of blocks (GOB) in H.26x, of contiguous macroblocks (MB's). A slice is the minimum element that is equipped with a start-code that provides resynchronization and thus avoids the spatial error propagation.

Many concealment techniques have been developed to reduce the damage from errors. Wada [6] proposed a selective recovery method that replaces the lost areas with pixel blocks of the previous frame shifted by the average motion vector of the neighboring blocks. Wang *et al.* [7], [8] developed a cell-loss concealment method that recovers the damaged regions by dc and low-frequency coefficients for DCT-based video coding in asynchronous transfer mode (ATM) networks. Aigh and Fazel [9] presented an error-concealment technique that interpolates each pixel of the whole lost MB with the adjacent pixels of the four neighboring MB's. All of these methods can partially recover the lost parts and improve the overall quality. However, none of them can totally prevent the error from propagation.

To reduce the error-propagation effect in the spatial domain, we can isolate the dc level or reduce the slice size. Fernández *et al.* [10] presented an early resynchronization technique with dc recovery at slices, where the dc recovery technique is only applied to intracoded pictures.

The forced intraupdate technique has been proposed for H.245 [11], [12]. In this paper, we present a temporal error-propagation prevention technique partially based on this idea. When the decoder detects an error in the I- or P-pictures, it sends the starting address of the damaged MB to the encoder. The encoder marks the possible damaged area and the normal encoding process continues. By this operation, the motion compensation of the following frames will probably not refer to the damaged area of the current frame, and the error propagation in the temporal domain is then terminated by this feedback mechanism. Note that only a small amount of information is needed for feedback. This is significantly different from the commonly used automatic retransmission mechanism, which requires the retransmittal of whole lost data. This approach can be applied to either MPEG or H.26x video coding over networks with feedback channels and limited delay. In the rest of this paper, we use MPEG/H.261 over ATM networks as examples to discuss the operations and performance.

This approach is designed for real-time encoding and decoding conditions, such as live video transmission. In our experiments, we apply it to a real-time video system over an ATM testbed with quality-of-service (QoS) guarantees. We observe that the round-trip delay, which includes the video delivery to the decoder and the feedback to the encoder, in most cases, is less than several video frames in time. The quick

Manuscript received November 22, 1997; revised December 17, 1998. This work was supported by the National Science Council, Taiwan, R.O.C., under Grant NSC-86-2213-E-008-017. This paper was recommended by Associate Editor J. Villasenor.

The authors are with the Department of Electrical Engineering, National Central University, Chung-Li, Taiwan 320 R.O.C.

Publisher Item Identifier S 1051-8215(99)02955-9.

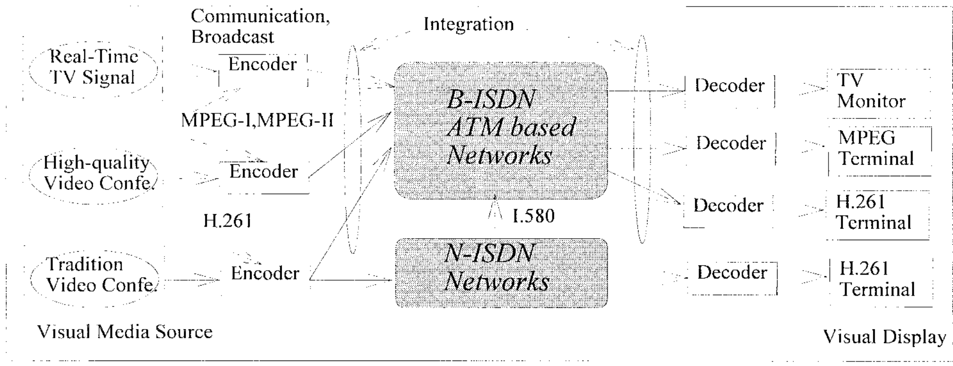



Fig. 1. Configuration of real-time video systems on N-ISDN and B-ISDN.

response ensures that the encoder has adequate time to take the prevention action to avoid referring to the damaged areas in the next P-frame encoding. Together with the error-concealment techniques, the error-propagation damage is greatly reduced.

This paper is organized as follows. Section II contains a brief description of the end-to-end real-time video system that has been used throughout this work. Section III presents the video error-propagation effect, the general error-concealment techniques, and the error-propagation prevention method we propose. We describe the simulation environment and show the simulation results in Sections IV and V, respectively. The delay analysis is described in Section VI. Last, conclusions are given in Section VII.

## II. Overview of the Real-Time Video System

The capacity of broadband-ISDN (B-ISDN) makes high-quality real-time video applications possible. ATM networks have been proposed to provide a unified transport structure for B-ISDN. As B-ISDN is intended to support a wide variety of services, it is necessary to cluster these services into classes. A primary class of B-ISDN services will be visual communication services for distributed real-time multimedia applications and integrated media services, supported by the broadcasting industry and communications industry. The real-time multimedia applications include videotelephony, videoconferencing, distance learning, digital TV broadcasting, intracompany TV, and road-traffic monitoring.

The delivery of video signals over ATM networks can exploit the high speed and QoS advantages that ATM networks support. However, the video delivery needs special care because of its complex bandwidth requirements and its high sensitivity to cell losses.

ATM networks provide many opportunities for new and evolved services. Although traditional videotelephony and videoconferencing coded by H.261 or H.263 attempts to work at low transmission rates, B-ISDN is capable of delivering high-quality video sequences coded by MPEG-1/MPEG-2 with its high bandwidth capacity. ATM networks also provide the possibility of variable bit-rate (VBR) transmission for real-time services. Fig. 1 shows the configuration of real-time video applications on narrowband-ISDN (N-ISDN) and B-ISDN. N-ISDN delivers H.26x video, while B-ISDN delivers MPEG and H.26x. I.580 is the ITU-T recommendation that covers the general arrangements for internetworking between B-ISDN and 64 Kbit/s-based N-ISDN [13], [14].

To provide a real-time video-transmission service, the round-trip delay time must be shorter than a few hundred milliseconds, e.g., 300 ms. In practice, the delay should be designed to be much less than this value for satisfactory interactive responses.

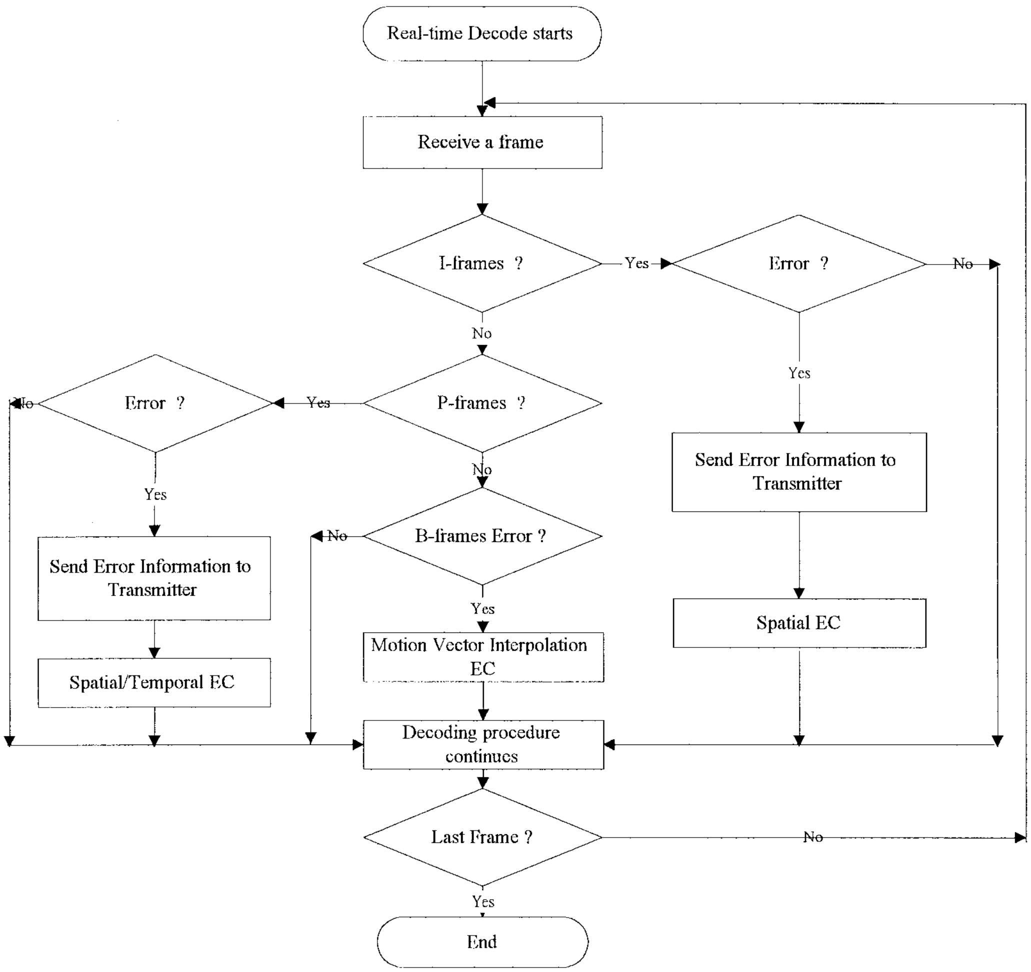
## III. Error-Propagation Prevention

A cell loss may be detected by communication protocols—e.g., a wrong cell will be indicated in the transport header—or entropy decoding—e.g., if some codeword cannot be mapped into the table of the variable-length decoder [15], [16]. In this section, we discuss the damage from the error-propagation effect, the general error-concealment techniques, the concept of error propagation prevention, and the combined error prevention and concealment procedure performed at the decoder and the encoder.

### A. Error-Propagation Effect

An MPEG GOP consists of three types of frames: I-frames, P-frames, and B-frames. Error propagation exists in both the spatial and the temporal domains. The error-propagation effect in the temporal domain is usually more serious and more difficult to terminate. The I-picture is intracoded and will be a reference frame for the following pictures in a GOP. Thus, the error effect propagates to all the succeeding pictures within a GOP until the I-frame of the next GOP. Since an I-frame has no dependency on any other frames, it serves as an effective synchronization point that terminates error propagation. Thus, any error propagation is terminated in a GOP, e.g., 12–15 frames in MPEG. This is, however, already unacceptably long for a high-quality real-time video application.

The P-picture is a forward motion-compensated prediction picture with a reference to the previous I-frame or P-frame. The P-frame itself again will be refereed by the motion-compensation operations of following pictures in a GOP. Thus, the error in P-pictures will also propagate to all the following pictures in a GOP. On the other hand, the B-picture is a motion-compensated interpolation picture. The error in the decoded B-frames does not propagate in the temporal domain because it is never used as a reference picture by other frames.




(a)

Fig. 2. (a) Error prevention and concealment procedure operated at the receiver.

Usually, I-frames and P-frames generate higher rates than B-frames and thus are more likely to have cell losses in a network. In addition, B-frames do not have the temporal error-propagation problem. Hence we only focus on I- and P-pictures in this paper regarding the error prevention and concealment.

### *B. Error-Concealment Techniques*

Error concealment is commonly used to reduce damage from the propagation of errors. We use the combined spatial and temporal error-concealment techniques, which shows significant improvement with moderate complexity [17]. The spatial error concealment is proposed for the damaged area of I-frames, in which no motion estimation exists. It interpolates each pixel of the whole damaged MB with the adjacent pixels of the four neighboring MB's [9], [17]. The interpolation for each pixel is based on the following equation:

$$\begin{aligned} mb(i,\,k) = {} & \frac{1}{d_L + d_R + d_T + d_B} \\ & \cdot [d_R mb_L(i,\,2N) \\ & \quad + d_L mb_R(i,\,1) + d_B mb_T(2N,\,k) \\ & \quad + d_T mb_B(1,\,k)] \qquad i,\,k = 1,\cdots,2N \end{aligned} \tag{1}$$

where $mb$ is the damage MB, $mb_L, mb_R, mb_T, mb_B$ is the respective neighboring MB, and $d_L, d_R, d_T, d_B$ is the distance from the respective pixel of the MB to the damage pixel. If any neighbors are also lost, the remaining correct neighbors are used for interpolation. An alternative is to apply block interleaving before the encoding so that all neighbors are more likely to be correct, and the formula can be applied without modifications [17].

We apply the combined spatial/temporal error concealment with motion vector interpolation to the P-frames and B-frames [9], [10]. The spatial error concealment is used for intracoded MB's, and the temporal error concealment with motion vector interpolation is used for intercoded MB's.

### *C. Error-Propagation Prevention Algorithm*

The reason that the errors propagate in the temporal domain is that the encoder and decoder are out of synchronization. The encoder does not have information on the errors that occurred in the decoding bitstream, and thus continues its normal encoding work. The decoder, even with error concealment, may not reach its original state and hence performs in a suboptimal way. For instance, if the frame buffers in the encoder and the decoder have different contents due to transmission errors, the


(b)

Fig. 2. *(Continued.)* (b) Error prevention and concealment procedure operated at the transmitter.

MB's pointed to by the motion vector at the decoder may contain incorrect data; hence the reconstructed image quality is deteriorated.

In this paper, we develop an efficient method of temporal error-propagation prevention with feedback assuming that a low-rate reverse channel is available. When the receiver detects an error in the I- or P-frames, it acknowledges the transmitter the starting address of the damaged MB's via the reverse channel. The transmitter, having received the notice, simply fills in all the pixels in the rest of the erroneous slice with an infrequently occurring value, e.g., zero ("00"), and continues the encoding process. With the "00"-marked MB's, the following P- or B-pictures will be unlikely to have motion estimation referred to these MB's because of the large differences in pixel values. Thus, further error propagation is prevented. For instance, if a slice in an I-frame is damaged, the MB's in the following pictures that originally referred to this slice may refer to a different area with low distortion or change the encoding mode to the intramode for P-pictures or to the backward prediction mode for B-pictures. Thus, the reconstructed image quality can be maintained. The overhead for the feedback operation is low. Only the ID of the lost MB, which includes the frame ID and the MB location in a frame, needs to be acknowledged to the transmitter. The transmitter needs neither to reencode the damaged frame nor to retransmit the damaged frame. This is significantly different from the acknowledgment-repeat-request (ARQ) scheme often used in data communications. Since the motion estimation may still be applied to error-affected MB's in this scheme, it has an advantage over the forced-intraupdate scheme [11], [12], in which the bit rate may increase substantially at high error rates due to the forced intramode encoding.

### *D. Error Prevention and Concealment Procedures*

The error-prevention concept and error-concealment techniques are combined into two procedures performed at the receiver and the transmitter, shown in Fig. 2(a) and (b), respectively.

*1) Procedure Operated at the Receiver:* When the receiver gets a frame ready for decoding, it performs error detection for each type of frame. The cell loss is detected by the cyclic redundancy check, or the errors in VLC, or the synchronization words in the receiver. If the damaged picture is an I- or P-picture, the starting addresses of the damaged MB's in a slice are sent back to the transmitter. Depending on the type of damaged picture, a suitable error-concealment technique is applied to estimate the damaged blocks. A spatial error-concealment technique is chosen for I-pictures, while temporal or motion-compensated error-concealment techniques are used for P- and B-pictures. Finally, the normal decoding procedure continues.

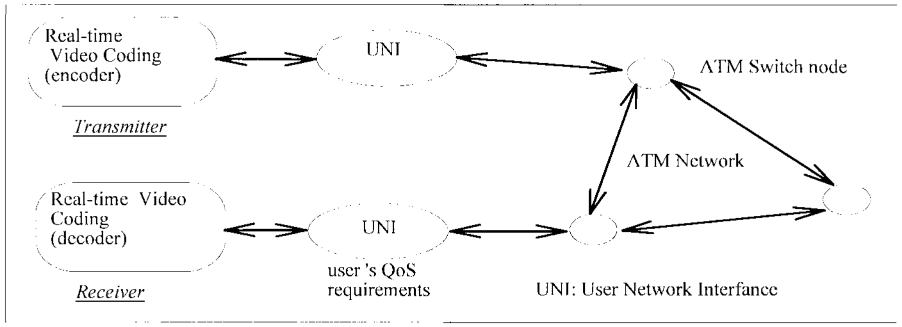



Fig. 3. Architecture of a real-time video system over an ATM testbed.

*2) Procedure Operated at the Transmitter:* The error-prevention action is only applied to the encoding of P and B-pictures. When the acknowledgment of error information with the starting address of the damaged MB's is received, the offset between the damaged frame and the current encoding frame is first calculated. If the offset is within a P-frame period, defined as the time duration between two adjacent P-frames, the encoder marks the corresponding error slice with "00" in the reference-frame buffer, which is used as reference in motion estimation. If the end-to-end delay is relatively long such that the offset is larger than a P-frame period, the "00"-marked area in the reference-frame buffer needs to be expanded to avoid being referred in the motion estimation of later frames. One of the expansion methods for H.263 in the mobile environment was addressed by Steinbach [12]. For a low delay network such as ATM, however, a simple expansion may be adequate. A simple and safe expansion in each direction is the motion-estimation search range for each additional P-frame-period delay, e.g., a 16-pixel expansion in each direction for a ($-16$ to $+15$) motion-estimation search.

The normal encoding procedure continues after the error-mark operations. Since the difference in pixel values may be large, the motion estimation in the following pictures will be unlikely to refer to the marked area because the encoding type of an MB may also be changed to avoid the references to the damaged area. For P-frames, intramode, which does not perform motion estimation, may be applied to replace the intermode. For B-frames, the motion estimation of one side prediction, e.g., backward prediction (BP), may be applied to replace the other side prediction, e.g., forward prediction (FP), or interpolated prediction (IP) so that the damage area is not referred. If, unfortunately, the corresponding areas in both P-frames that enclose the B-frame are marked, the encoding mode can be changed to the intramode to improve the decoded image quality, or simply ignore this case since the B-frame errors will not propagate temporally anyway. With this method, the impact of errors is confined in the damaged picture only.

## IV. The Simulation Environment

We have incorporated our algorithm in a real-time video system over a campus ATM testbed, shown in Fig. 3. An MPEG coder based on TM5 is used in the simulations [2], [18]. Fore SBA-200 ATM cards are used as the user–network interface (UNI) to the ATM network. The ATM adaptation layers (AAL) it supports are AAL5 and AAL3/4 segmentation and reassembly. FORE ASX-200 ATM switches delivering switching capacity of 2.5 Gbps are used as ATM nodes. All connections are running on 155-Mbps Oc-3c fibers. The video coding and ATM networking functions are running under the Unix operating system. At the transmitter side, the connection management module responds to the receiver requests. The receiver is responsible for requesting and receiving data from the transmitter and putting data into shared memory. ATM connections are established on Fore native ATM application programming interface (API).

The system provides QoS control to fit the requirements of different services. A user-level service class is used to represent the user's desired level of service. We discuss QoS parameters, which are related to video coding in the ATM and AAL [19], [20], and are used in the simulation setup and the delay analysis in later sections. The QoS parameters include the following:

1) *end-to-end delay:* the total delay experienced by a frame in traveling from transmitter to the receiver;
2) *frame loss ratio:* the percentage of lost frames during transmission between the transmitter and the receiver;

TABLE I
PSNR Reductions of "No EC & EP" (No Error Concealment and Error Prevention), "EC1" (Simple Spatial Replacement), "EC2" (Spatial Interpolation EC), and "ECP" (Error Concealment and Error Prevention) for Y, U, and V Components of I-Frames with Cell-Loss Rate $10^{-2}$

| PSNR reductions in dB of 'Garden' | | | | |
|---|---|---|---|---|
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 11.94 | 4.20 | 2.48 | 0.48 |
| U | 12.99 | 0.81 | 0.43 | 0.10 |
| V | 15.85 | 0.75 | 0.28 | 0.06 |
| PSNR reductions in dB of 'Table Tennis' | | | | |
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 10.83 | 2.67 | 2.26 | 0.32 |
| U | 19.6 | 0.24 | 0.13 | 0.02 |
| V | 21.67 | 4.03 | 2.04 | 0.22 |
| PSNR reduction in dB of 'Football' | | | | |
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 7.79 | 2.78 | 1.05 | 0.5 |
| U | 10.86 | 2.34 | 0.55 | 0.07 |
| V | 15.13 | 1.26 | 0.32 | 0.06 |

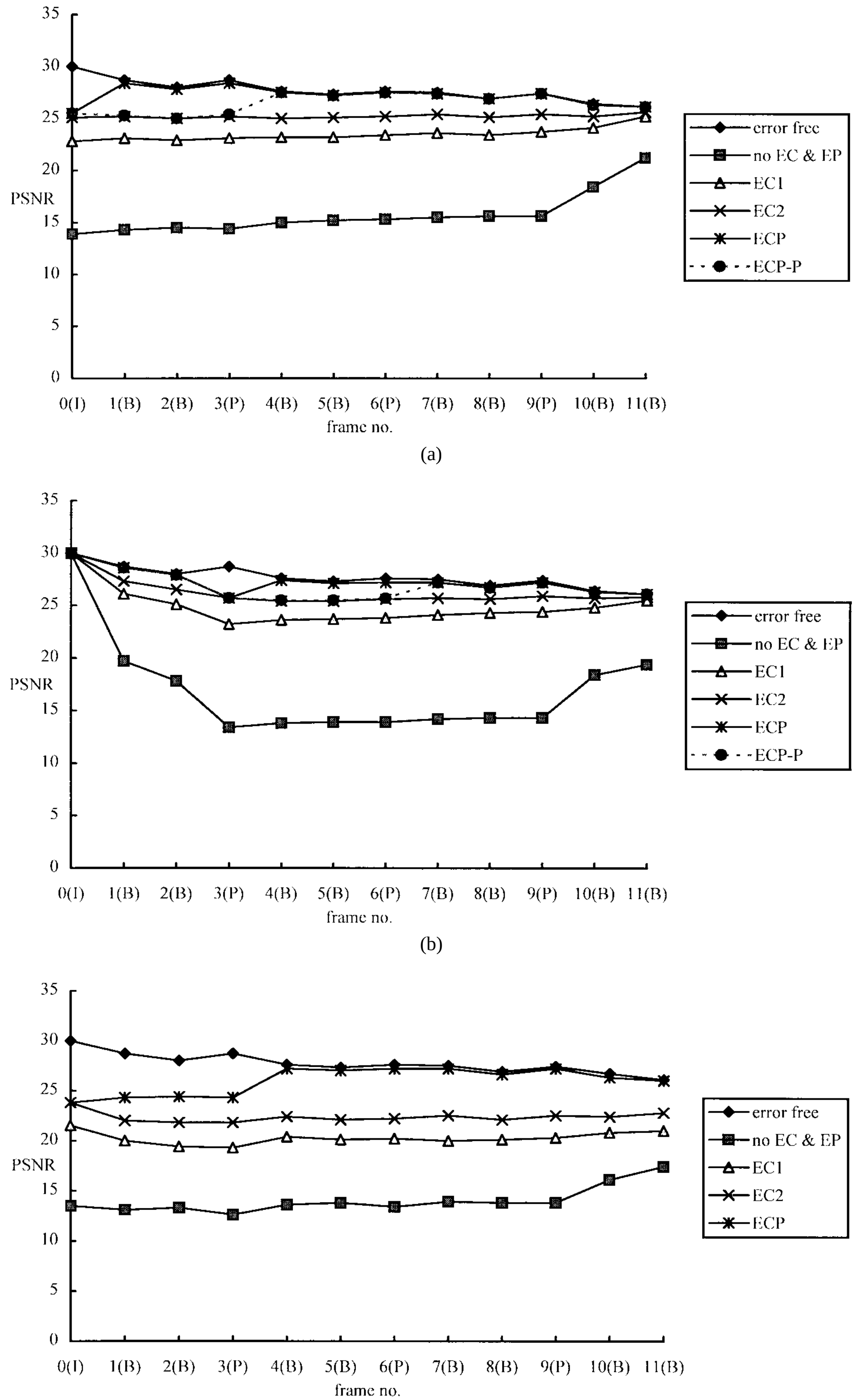
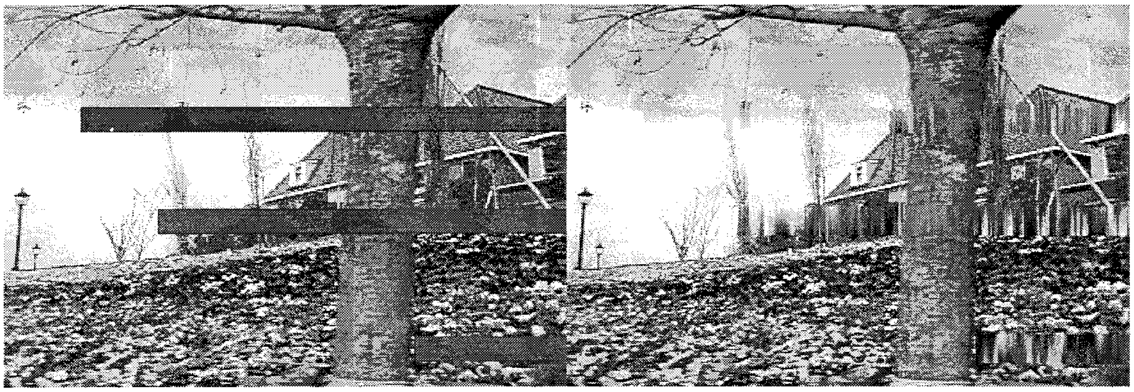



Fig. 4. PSNR's of the video sequence *Garden* with different techniques: (a) cell loss at I-frames, (b) cell loss at first P-frames (frame 3), and (c) random cell loss with cell-loss rate $10^{-2}$.

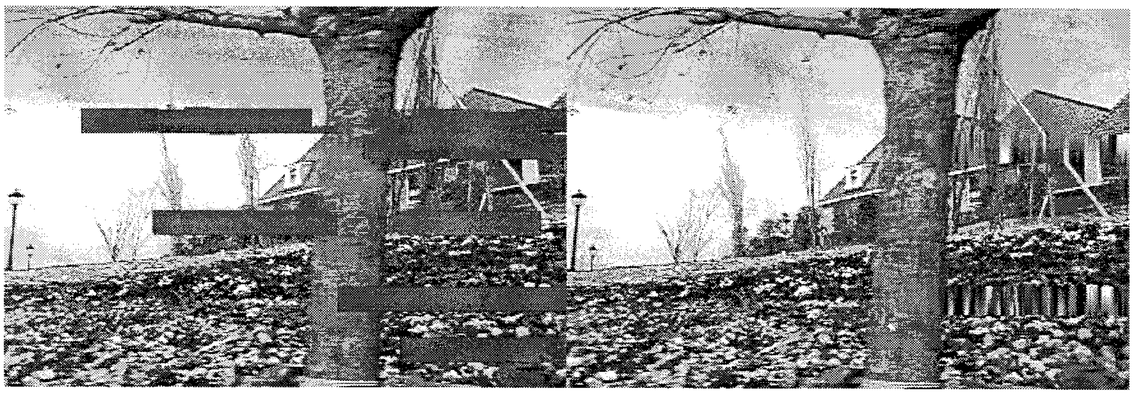
3) *peak bandwidth:* the maximum (burst) rate at which the transmitter produces data, measured in kilobits per second;
4) *mean bandwidth:* the average bandwidth expected over the lifetime of the connection, also measured in kilobits per second.

## V. Simulation Results

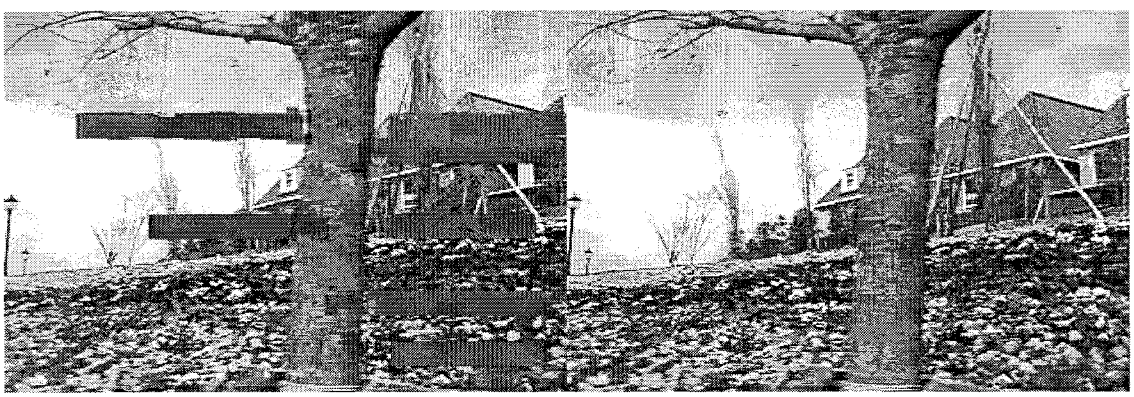
The simulations are carried out under the conditions that the cell-loss rate is controlled under 7% and the peak signal-to-noise ratio (PSNR) between the reconstructed and original images is used as an objective image-quality measure. The video sequences, including *Table Tennis, Football, Flower,*



(a) (d)



(b) (e)



(c) (f)

Fig. 5. Effect of error-propagation prevention and error concealment for *Garden* sequence. (a) Frame 0 (I-frame) with cell loss. (b) Frame 3 (P-frame) with cell loss and error propagation. (c) Frame 5 (B-frame) with error propagation. (d) Frame 0 (I-frame) with spatial EC. (e) Frame 3 (P-frame) with ECP. (f) Frame 5 (B-frame) with ECP.

*Sales Man*, and *Miss America*, with CIF sequence format (24 fps, $352 \times 240$ pels, $4:2:0$ chrominance format, 15 slices per picture), are MPEG coded at 1.5 Mbps, 12 pictures per GOP, one slice per MB row. The feedback channel is assumed to be error free in the simulations.

### *A. Performance of Error-Propagation Prevention and Concealment*

To examine the impact of errors in different types of frames, we first control the cell loss to occur in certain frames. The average PSNR reductions with various concealment and prevention techniques at the cell-loss rate $10^{-2}$ are given in Table I for I-frame errors and in Table II for the first P-frame (frame 3) errors. Different concealment and prevention/concealment techniques are compared. We observe that both the simple spatial replacement (EC1), which copies from the above MB, and the spatial interpolation error concealment (EC2), shown in (1), can reduce the error damage substantially. However, the degradation is still significant compared with the error-free case. Only the error concealment and prevention (ECP) technique we proposed yields less than 0.8 dB degradation, which is the best performance among all tested techniques.

TABLE II
PSNR REDUCTION OF "NO EC & EP" (NO ERROR CONCEALMENT AND ERROR PREVENTION), "EC1" (SIMPLE SPATIAL REPLACEMENT), "EC2" (SPATIAL INTERPOLATION EC), AND "ECP" (ERROR CONCEALMENT AND ERROR PREVENTION) FOR Y, U, AND V COMPONENTS OF THE FIRST P-FRAMES IN GOP'S WITH CELL-LOSS RATE $10^{-2}$

| | PSNR reduction in dB of 'Garden' | | | |
|---|---|---|---|---|
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 10.76 | 2.8 | 1.47 | 0.4 |
| U | 11.86 | 0.15 | 0.05 | 0.03 |
| V | 12.78 | 0.26 | 0.12 | 0.06 |
| | PSNR reduction in dB of 'Table Tennis' | | | |
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 9.79 | 1.85 | 1.55 | 0.45 |
| U | 17.35 | 0.13 | 0.05 | 0.04 |
| V | 20.19 | 2.74 | 0.21 | 0.19 |
| | PSNR reduction in dB of 'Football' | | | |
| | no EC & EP | EC1 | EC2 | ECP |
| Y | 8.48 | 2.16 | 1.81 | 0.8 |
| U | 10.77 | 1.85 | 0.36 | 0.09 |
| V | 14.89 | 0.82 | 0.16 | 0.05 |

Fig. 4 shows the PSNR's of the individual frames of GOP's in the *Garden* sequence with different concealment and prevention techniques. In Fig. 4(a), the errors are controlled to occur at the first frame (I-frame). The two concealment techniques "EC1" and "EC2" give good results. However, the error still propagates to the whole GOP with concealment only. Our proposed ECP technique not only improves the quality of the current frame but also prevents the error from propagation to all following frames. Note that if, unfortunately, the round-trip delay time is too long to acknowledge the encoder before processing the next P-frame, ECP can still be performed to prevent the error from further propagation, shown as the ECP-P line (partial ECP) in Fig. 4(a). The case of P-frame errors is shown in Fig. 4(b), in which errors occur at the third frame (P-frame). ECP still has the best performance. Fig. 4(c) demonstrates the case of random cell loss with a loss rate of $2 * 10^{-2}$. In this specific example, two- or three-cell losses occur in frames 0(I), 1(B), 2(B), 3(P), and one-cell losses occur in 6(P), 9(P). In other frames, even though no cell losses occur, the quality is degraded; except for the ECP technique, which terminates the error propagation effectively.

Fig. 5 shows the effect of error prevention and error concealment for the above specific case. Fig. 5(a) is an example of video frame 0(I), which is contaminated by three errors in the MPEG bitstream. Spatial propagation of errors is manifested as the loss of a slice. Fig. 5(b) shows two new errors and the errors propagated from previous frame 0(I). Although frame 5(B) has no error, serious error propagation occurs, shown as in Fig. 5(c). Fig. 5(d) is the I-frame with spatial interpolation error concealment (EC) based on (1). Fig. 5(e) and (f) shows the corresponding P- and B-frames, respectively, with error propagation prevention technique ECP, which yields significantly better quality.

To evaluate the performance of the ECP method at various cell-loss rates, we compute the average PSNR of the video sequence *Football*. A sufficiently long test sequence is formed by concatenating the original sequence with itself by 100 times. Fig. 6 shows the PSNR at the cell-loss ratio 0.1% $\sim$ 7% for both the random and burst cases. The average burst length

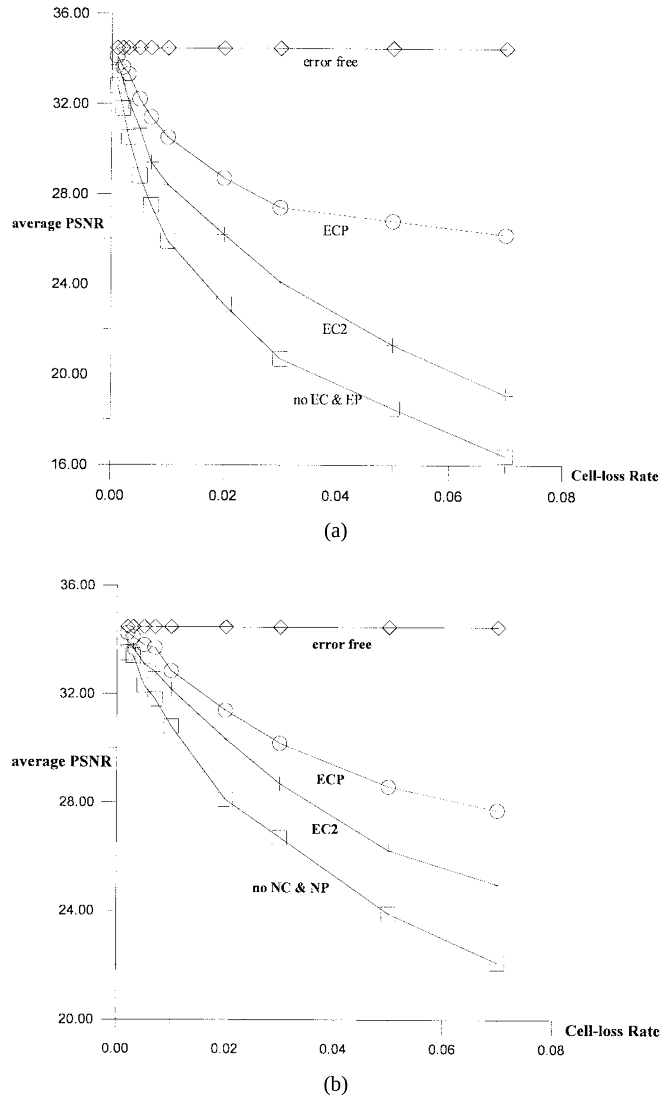



Fig. 6. Average PSNR of the video sequence *Football* with different techniques: (a) random cell loss and (b) burst cell loss at the rate 0.1% $\sim$ 7%.

is set to 4.68 cells in this simulation. The random case yields worse picture quality than the burst case. The ECP technique outperforms other concealment techniques at all ranges of tested cell-loss rates. In the simulations, if errors occur in the picture header, the current decoding picture is ignored or, equivalently, replaced by the previous picture stored in the frame buffer.

### *B. Bit-Rate Comparison*

Since the ECP algorithm changes the encoding from intermode to intramode for some MB's, it is a concern that the bit rate will increase. With the embedded rate-control mechanism of the TM5 MPEG encoder [2], the output rate can be maintained almost constant. However, the effect of the inter/intramode change is reflected to the PSNR's, for which we have shown that the ECP method has significant improvements. Nevertheless, Table III gives the comparisons of bit rates of each frame with the ECP algorithm to the error-

TABLE III
BIT RATES IN Kbps OF EACH FRAME OF A GOP FOR ECP TECHNIQUES, TESTED FOR THREE VIDEO SEQUENCES

| (a). 'Garden' sequence | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0(I) | 1(B) | 2(B) | 3(P) | 4(B) | 5(B) | 6(P) | 7(B) | 8(B) | 9(P) | 10(B) | 11(B) | increase |
| error free | 180.2 | 38.3 | 33.1 | 84.5 | 31.8 | 26.8 | 80.5 | 26.2 | 24.1 | 76.0 | 25.5 | 21.3 | |
| ECP(I) | 180.2 | 38.7 | 32.8 | 85.8 | 31.7 | 26.1 | 80.4 | 25.9 | 23.8 | 75.9 | 25.2 | 21.8 | 0.087% |
| ECP(p) | 180.2 | 38.5 | 34.1 | 84.5 | 32.2 | 26.8 | 80.8 | 25.9 | 23.7 | 74.1 | 25.4 | 21.5 | 0.022% |
| (b).'Football' sequence | | | | | | | | | | | | | |
| | 0(I) | 1(B) | 2(B) | 3(P) | 4(B) | 5(B) | 6(P) | 7(B) | 8(B) | 9(P) | 10(B) | 11(B) | increase |
| error free | 169.0 | 44.1 | 38.6 | 72.5 | 39.2 | 35.2 | 70.7 | 36.8 | 35.5 | 66.6 | 36.5 | 38.2 | |
| ECP(I) | 169.0 | 44.8 | 38.6 | 72.9 | 38.6 | 35.1 | 70.5 | 36.9 | 35.4 | 66.4 | 36.3 | 38.2 | 0.022% |
| ECP(p) | 169.0 | 44.3 | 39.1 | 72.5 | 39.2 | 35.1 | 70.8 | 36.3 | 35.7 | 65.7 | 36.6 | 38.3 | 0.005% |
| (c).'Table Tennis' sequence | | | | | | | | | | | | | |
| | 0(I) | 1(B) | 2(B) | 3(P) | 4(B) | 5(B) | 6(P) | 7(B) | 8(B) | 9(P) | 10(B) | 11(B) | increase |
| error free | 168.9 | 29.0 | 36.2 | 76.4 | 29.4 | 27.5 | 82.7 | 30.4 | 35.4 | 92.9 | 28.1 | 20.3 | |
| ECP(I) | 168.9 | 30.3 | 35.7 | 77.9 | 29.4 | 26.9 | 81.6 | 30.3 | 35.3 | 92.6 | 28.1 | 19.9 | 0.096% |
| ECP(p) | 168.9 | 29.4 | 36.3 | 76.4 | 31.0 | 27.9 | 83.3 | 30.4 | 34.4 | 90.0 | 27.7 | 20.1 | 0.038% |

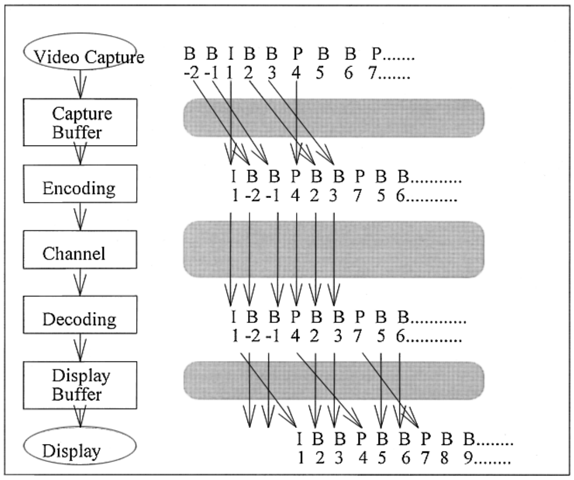



Fig. 7. Frame resequence in MPEG(12, 3) coding system.

free case. Two cases, ECP(I) (error occurs at the I-frame) and ECP(P) (error occurs at the P-frame) are tested. For the target bit rate of 1.5 Mb/s, the average bit rate increasing by ECP is only 0.005% $\sim$ 0.096%, which is almost negligible.

## VI. END-TO-END DELAY-TIME ANALYSIS

In this section, we discuss how much delay budget we can get from a video over an ATM system for feedback and the estimation of end-to-end delay in a real-time video system over ATM networks.

### A. Round-Trip Delay Budget

In order to have effective error-prevention operation, the encoder needs to receive the feedback information on the lost MB's before processing the next P-frame. This maximum round-trip delay time is therefore referred to as the delay budget, which is the upper bound of the delay used in the design of the video over ATM systems.

In H.261–like video coding, since the encoding of each frame will refer to the previous frame, the delay budget is a frame period, e.g., 1/12 or 1/15 s. In MPEG coding, however, the delay budget could be larger in terms of the number of frames because the encoding of B-frames does not need the feedback of error information.

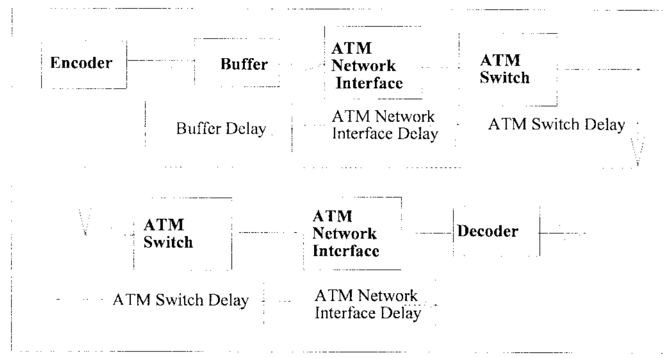



Fig. 8. End-to-end delay model in video-over-ATM networks.

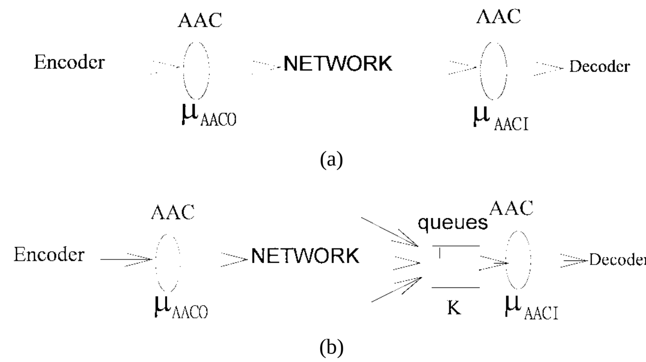



Fig. 9. Model of ATM network interface. (a) Point to point and (b) multipoint to point.

TABLE IV
END-TO-END DELAY PARAMETERS

| | |
|---|---|
| $d_{buffer}$ | Encoder Buffer Delay |
| $\beta_o$ | ATM Network Interface Delay (output) |
| $P$ | Optical fiber Propagation Delay |
| $d_{switch}$ | ATM Switches Delay |
| $\beta_i$ | ATM Network Interface Delay (input) |
| $q$ | Number of ATM Switch |

Since the encoding of each B-frame needs to refer to both adjacent I- and P-frames, the encoding of a B-frame must be performed after the encoding of the next P-frame. Fig. 7 shows the example of an MPEG ($N = 12$, $M = 3$) coding system, where $N$ is the number of frames in a GOP and $M$ is the distance between I–P or P–P frames. Only the relative positions of the video frames are shown. The delays generated in the encoder, the network, and the decoder will be discussed in the next section. Since the encoding sequence order is reversed for P–B-frames, each B-frame must be delayed for at least $M$ frames for encoding. I- and P-frames, however, do not need delay, which can be encoded immediately after capture. The decoding sequence is exactly the same as the encoding sequence in which the B- and P-frames are with reversed order. Finally, the display buffer is used to resequence the frames to the original order for correct display. For any error occurring in I- or P-frames, there exists $M$ frames end-to-end delay budget available for information feedback.

### B. Delay-Time Analysis

We give primitive estimates of the delay occurring in each part of a video-over-ATM system. Depending on the actual

TABLE V
ATM QoS PARAMETERS AND BUFFER DELAY

| class | mean source rate ($r_{MS}$) | peak source rate ($r_{PS}$) | mean transmission rate ($r_{MT}$) | frame rate ($f$) | worst case buffer requirement ($B$) | worst case buffer delay ($d_{worst\text{-}case}$) |
|---|---|---|---|---|---|---|
| MPEG | 1.5 Mbps | 3.5 Mbps | 1.5 Mbps | 24 fps | 83.3 Kbits | 55.56 ms |
| H.261 | 0.75 Mbps | 1.29 Mbps | 0.75 Mbps | 12 fps | 45 Kbits | 60 ms |

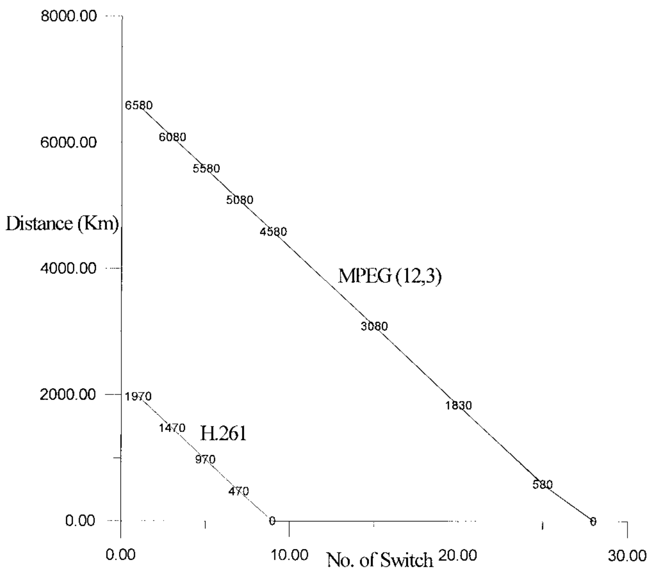



Fig. 10. The allowable size of an ATM network that supports the error-prevention scheme.

implementation, extra delay may be needed in some part. Fig. 8 shows the delay model of an end-to-end connection. To have an approximately zero cell-loss probability in the end-to-end connection, the end-to-end delay can be calculated by

$$D_e = d_{\text{buffer}} + \beta_o + P + q \times d_{\text{switch}} + \beta_i \quad (2)$$

and the round-trip delay can be approximated by

$$D_e = d_{\text{buffer}} + 2 \times (\beta_o + P + q \times d_{\text{switch}} + \beta_i) \quad (3)$$

where delay parameters are listed in Table IV.

*1) ATM Switch Delay:* The Fore System ASX series ATM switch used in this project has first in, first out queues. The upper bound of the delay for a 155-Mbps link is 1.402 ms for the ASX-100 ATM switch and 1.25 ms for ASX-200 ATM switch [21].

*2) ATM Network Interface Delay:* The transmission delay of an SBA-200 Sbus ATM adapter card can be obtained in a similar manner as that of the switch [22] (Fig. 9). There are two cases related to the video system in which we are interested.

*a) Point-to-point case:* This case supports point-to-point videoconferencing. The output and input adapter-card delays are approximately $\beta_o = n/\mu_{\text{AACO}}$ and $\beta_i = n/\mu_{\text{AACI}}$, respectively, where $n = 424$ bits for ATM cell length and $\mu_{\text{AACO}}$ and $\mu_{\text{AACI}}$ are adapter card throughputs. If $\mu_{\text{AACO}} = \mu_{\text{AACI}} = 1.5$ Mbps, then $\beta_o = \beta_i = 0.283$ ms, and the end-to-end adapter-card delay is $d_{\text{AAC}} = \beta_o + \beta_i = 0.566$ ms.

*b) Multipoint-to-point case:* This case supports point-to-multipoint videoconferencing. The worst case of end-to-end adapter-card delay without cell loss is calculated as $d_{\text{AAC}} = \beta_o + \beta_i = n/\mu_{\text{AACO}_i} + nK/\mu_{\text{AACI}}$, where $K$ is queue size.

*3) Propagation Delay:* The optical-fiber propagation delay is approximated as $T_{\max} \cong T_{\min} = Ln_1/C = 5.0\,\mu$s km at multimode graded index fibers or single-mode fibers for ATM networks, where $L$ is the length of a fiber along which the axial ray travels, $n_1$ is the refractive index of the core, and $C$ is the velocity of light in vacuum.

*4) Encoder Buffer Delay:* The video-coding output is naturally VBR, which may not be perfectly matched to the network throughput. Thus, an encoder buffer is needed to temporarily store the video data for transmission. The network throughput is expressed by the QoS parameters, including peak transmission rate, mean transmission rate, and average burst length. The buffer requirement and the buffer delay can be derived from the relationship between the source rate and the transmission rate. Since in current video codecs the high-rate frames, e.g., I-frames, are enclosed by low-rate frames, e.g., B-frames, the buffer requirement of the encoder is calculated as the maximum buffer space that is enough for a full frame.

Depending on the maximum burst length specified by QoS, the buffer requirement may be derived differently. In the worst case, if the maximum burst length is specified as close to zero, the buffer length $B$ should be larger than the difference between the peak source rate $r_{\text{PS}}$ and the mean transmission rate $r_{\text{MT}}$ to avoid cell loss, i.e.,

$$B = \frac{r_{\text{PS}} - r_{\text{MT}}}{f} \quad \text{in the unit of Kbits/frame} \quad (4)$$

where $f$ is the frame rate.

The worst case buffer delay $d_{\text{worst-case}}$ happens for the burst source, i.e., the encoded data of a frame enter the buffer instantaneously. Its delay can be calculated by

$$d_{\text{worst-case}} = \frac{B}{\mu} \quad (5)$$

where $\mu$ is the maximum throughput, which is between the peak transmission rate and the mean transmission rate.

Two examples of user service classes and buffer requirements with different video-coding schemes are shown in Table V. The mean source is given by the QoS requirement for a service class. The peak source rate is measured from several video sequences. The mean transmission rate is set to be equal to the mean source rate, which is the worst case in terms of the buffer requirement. The frame rate of H.261 is set to be one-half of the MPEG frame rate for the tradeoff between the frame rate and the video quality. Then the worst case buffer requirement and delay are calculated by (4) and (5), respectively.

### *C. Delay-Budget Allocation*

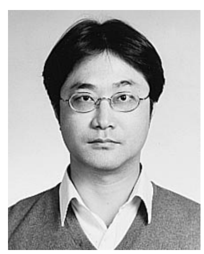
We have discussed the delay budget allowed for feedback and the delay in each part of ATM networks. Based on the above analysis, we figure out the maximum size of ATM networks that support the error prevention by feedback scheme. Since the delay of a connection imposed in a network is affected by the total length of the links $d$ and the number of switches $s$, we express the network size by the combination of $d$ and $s$. Fig. 10 shows the maximum size of an ATM network that can support the error-prevention scheme for two typical video-coding systems. For an MPEG ($N = 12, M = 3$) coding system, the maximum network size can span over 4000 Km with ten ATM hubs. For an H.261 application, the delay budget is relatively small. Nevertheless, it can still span over 1000 Km, which covers a large part of video applications.

## VII. Conclusion

We have proposed an efficient temporal error-propagation prevention method in which only the starting address of the lost MB's is feedback from the decoder to the encoder and the impact of the cell loss can be limited to the damaged slice only. The impact from the error propagation is thus greatly reduced. Experimental results show that the error-concealment-only schemes cannot fully eliminate the error-propagation effect. On the other hand, the error concealment with feedback can effectively isolate the error and conceal the damage to give satisfactory performance even when the cell-loss rate is as high as $10^{-2}$. For systems with long round-trip delays, the prevention scheme can still terminate the error propagation at the following P-frames and improve the video quality.

## References

[1] "Coded representation of picture and audio information," ISO/IEC/JTC1/SC29/WG11 MPEG 93/457, Apr. 1993.
[2] "Test model 5," ISO/IEC/JTC1/SC29/WG11 MPEG 93/457, Draft Version 1, Apr. 1993.
[3] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, pp. 47–58, Apr. 1991.
[4] M. Ghanbari, "An adapted H.261 two-layer video codec for ATM networks," *IEEE Trans. Commun.,* vol. 40, pp. 1481–1490, Sept. 1992.
[5] "Video coding for low bitrate communication," Draft ITU-T Recommendation H.263, May 1996.
[6] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE J. Select. Areas Commun.,* vol. 7, pp. 807–814, June 1989.
[7] Y. Wang, Q. F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.,* vol. 41, pp. 1544–1551, Oct. 1993.
[8] Q. F. Zhu, Y. Wang, and L. Shaw, "Coding and cell-loss recovery in DCT-based packet video," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 3, pp. 248–258, June 1993.
[9] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec," in *Proc. IEEE INFOCOM'95,* Seattle, WA, June 1995, pp. 1778–1783.
[10] C. L. Ferńandez, A. Basso, and J. P. Hubaux, "Error concealment and early resynchronization techniques for MPEG-2 video streams damaged by transmission over ATM networks," in *Proc. SPIE Digital Video Compression: Algorithms and Technologies,* San Jose, CA, Feb. 1996, vol. 2668, pp. 372–383.
[11] "Control protocol for multimedia communication: Draft ITU-T recommendation H.245," May 1996.
[12] E. Steinbach, N. Farber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 7, pp. 872–881, Dec. 1997.
[13] S. Okubo, S. Dunstan, G. Morrion, M. Nilsson, H. Rodha, D. L. Skran, and G. Thom, "ITU-T standardization of audiovisual communication systems in ATM and LAN environments," *IEEE J. Select. Areas Commun.,* vol. 15, pp. 965-982, Aug. 1997.
[14] T. Kinoshita, T. Nakahashi, and M. Maruyama, "Variable bit-rate HDTV codec with ATM-cell-loss compensation," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 3, pp. 230–237, June 1993.
[15] H. Ohta and T. Kitami "A technique to detect and compensate consecutive cell loss in ATM networks," in *Proc. IEEE INFOCOM'91,* Bal Harbour, FL, Apr. 1991, vol. 2, pp. 781–790.
[16] F. C. Jeng and S. H. Lee "Concealment of bit error and cell loss in inter-frame coded video transmission," in *Proc. IEEE Int. Conf. Communications'91,* Denver, CO, June 1991, vol. 1, pp. 496–500.
[17] P. C. Chang and M. C. Chien, "Interleaving and error concealment for MPEG video over ATM networks," *SPIE Photon.,* pp. 271–282, Nov. 1996.
[18] C. Gao and J. S. Meditch, "An efficient hierarchical VBR video codec over ATM networks," in *Proc. IEEE GLOBECOM'95,* Singapore, Nov. 1995, vol. 3, pp. 1910–1914.
[19] A. Sahai, K. Tseng, and W. Wang, "A QoS-controlled distributed interactive multimedia system on ATM networks," in *Proc. IEEE GLOBECOM'95,* Singapore, Nov. 1995, vol. 1, pp. 188–192.
[20] J. M. Hyman, A. A. Lazar, and G. Pacifici, "Real-time scheduling with quality of service constraints," *IEEE J. Select. Areas Commun.,* vol. 9, pp. 1052–1063, Sept. 1991.
[21] Fore Systems Inc., "ForeRunner ASX-200 ATM switch user's manual," MANU0013-Rev.D, Warrendale, PA, May 1995.
[22] ——, "ForeRunner SBA-100/-200 ATM Sbus adapter user's manual," MANU0002-Rev.E, Warrendale, PA, May 1995.



**Jong-Tzy Wang** was born in Taiwan, R.O.C., on May 24, 1959. He received the B.S.E.E. and M.S.E.E. degrees from National Taiwan Institute of Technology, Taiwan, in 1987 and 1991, respectively. He is currently pursuing the Ph.D. degree at National Central University, Taiwan.

In 1991, he joined the Department of Electronics Engineering at Jiing Wen College, Taiwan. His current interests are in the areas of image/video coding, error prevention/concealment technique, and ATM network applications.

**Pao-Chi Chang** received the B.S. and M.S. degrees from National Chiao Tung University, Taiwan, R.O.C., in 1977 and 1979, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 1986, all in electrical engineering.

From 1986 to 1993, he was a Research Staff Member of the Department of Communications, IBM T. J. Watson Research Center, Hawthorne, NY. At Watson, his work centered on high-speed switching systems, efficient network design algorithms, network management, and multimedia teleconferencing. In 1993, he joined the Faculty of National Central University, Taiwan. His current interests are in the areas of speech/image coding, video conferencing, and video coding over high-speed networks and wireless communications.