ERROR RESILIENCE SUPPORTING BI-DIRECTIONAL FRAME RECOVERY FOR VIDEO STREAMING

Chun-Ming Huang, Kai-Chao Yang, and Jia-Shung Wang

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan 30043

ABSTRACT

In this paper, we propose a novel coding dependency among video frames which supports efficient error resilience without adding any redundancy. Our approach is based upon reorganizing the regular GOP structure. The proposed scheme can effectively recover a lost frame and prevent error propagation phenomenon. For a single frame loss, we guarantee that both of its next and previous frames still can be successfully decoded, thus we have sufficient temporal and spatial information to reconstruct the damaged frame. Our new coding structure even can recover the successive lost frames. The experimental results showed that the video quality has a graceful degradation when the loss rate increases rapidly. Comparing with the conventional GOP, PSNR values can improve from 0.5 dB to 3 dB.

1. INTRODUCTION

For the popularity of applications of streaming videos such as video-on-demand (VOD), video conferencing and 3G phones [1], the protection of compressed video streams has become an important concern. Recently, the increasing number of mobile users has spurred the research on video transmission over wireless networks. Wireless streaming environments present more challenges in multimedia delivery due to the error-prone, bandwidth-limited and time-varying characteristics.

Transmission of streaming video is very sensitive to delay and loss of information and cannot easily make use of retransmission. Several error resilience approaches [3-6] have been proposed to minimize the transmission loss with a certain degree of redundancy. Recently, the standard of H.263 [7] suggests a mode of the *Reference Picture Selection* (RPS), which relies upon a feedback channel to efficiently stop error propagation. The RPS mode allows the encoder to select one of several previously decoded frames as a reference picture for prediction. In addition to error resilient coding schemes in H.263, *Multiple-Description Coding* (MDC) [2][8] generates several bit-streams of the same source signal and transmits them over separate channels. The reconstructed signal is acceptable with receiving any channel. On the other hand, another scheme to recover the damaged block by interpolating the combination of motion vectors is proposed in [9].

Due to the nature of inter-frame coding, once all packets of a frame are lost during transmission, the only available temporal information for recovery of the damaged frame is its previous frame, since all frames behind it will also crash. In this paper, our definition of error resilience is addressed as: Once an inter-frame is damaged, both of its neighboring frames still can be compensated through some other frames using the pre-stored prediction information. The damaged inter-frame can be precisely reconstructed from these two neighboring frames by any existing interpolation method.

In the remaining part of this article, Section 2 describes the nature of the conventional GOP. Our structure and error concealment scheme are illustrated in Section 3. And then Section 4 demonstrates the performance of the proposed structure. Section 5 describes some advanced applications for our structure. Finally, several conclusions are given in Section 6.

2. TEMPORAL DEPENDANCY AMONG FRAMES

In the conventional coding systems, a video clip consists of many groups of pictures (GOPs), and three types of frames are defined in each GOP. I-frames are encoded and decoded by themselves. P-frames have to be encoded and decoded by referring to the previous I- or P-frame. B-frames are similar to P-frames, but the motion compensation is done by referring to the previous I- or P-frame, the next I- or P-frame, or an interpolation between them.

An example of the predictive dependencies among frames is shown in Figure 1(a). Typically, the first frame in a GOP is I-frame, followed by a series of P-frames. And B-frames are inserted between them. In this paper,

---

*This research is supported in part by the National Science Council, Taiwan, R.O.C., under Contract NSC 93-2752-E-007-004-PAE and NSC 92-2213-E-007-027.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

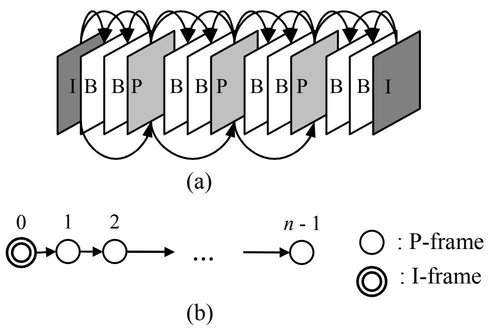

Figure 1. (a) Conventional GOP structure. (b) The dependencies between I- and P-frames in a GOP.

without loss of generality, we focus the analysis on the cases that the bit-stream contains I- and P-frames only, which has been adopted by the H.263 standard, as illustrated in Figure 1(b). The results can be easily extended to the general I-P-B-frame structures.

The above temporal dependencies are unfavorable to recovering the damaged inter-frames. Packet loss in the current frame causes error propagation in the successive frames. This situation will become even more critical in the low bit-rate streaming video, where the whole frame might be packetized into few packets. And a loss of one packet or two can cause crash of the whole frame such that all the successive frames will also fail to be decoded.

Suppose a GOP is of size 15. We calculate that once an inter-frame is lost or damaged, how many frames in the GOP might also be damaged. For the conventional GOP, the average number of frames influenced by the damaged inter-frame will be

$$\frac{0+1+...+13}{14}=6.5.$$

To recover the damaged frame, the most available information of temporal domain is the previous undamaged frame. Almost all error resilient mechanisms apply this character, but the quality level of the reconstructed frame is not acceptable, especially when frames lose continuously.

If both neighbors of the damaged frame can be decoded correctly, we will have both side of temporal information to recover this frame. By the nature that most motions are smooth in a quite small time interval, the damaged frame can be reconstructed through the interpolation of these two neighbors.

3. DOUBLE-BINARY TREE STRUCTURE

Now we present a novel predictive dependency among frames in a GOP, called *double-binary tree structure*, which guarantee that loss of an inter-frame will not affect the decoding process of both of its neighbors. Suppose the size of the GOP is n. For simplicity, $(n-1)$ is set to the power of 2. The following steps are given to establish the double-binary tree structure:

Step1. Choose frame $\lfloor (n\text{-}1)/2 \rfloor$ as I-frame.

Step2. For all P-frames with odd number, each pair of forward and backward prediction is performed striding across half of remaining continuous odd P-frames starting from I-frame until all odd P-frames are predicted.

Step3. For remaining P-frames with even number, each pair of forward and backward prediction is performed striding across half of remaining continuous even P-frames starting from I-frame until all even P-frames are predicted.

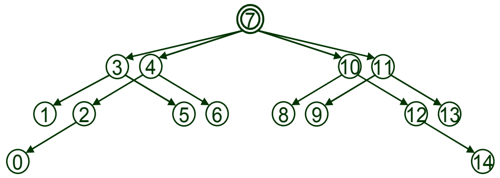

Figure 2. Double-binary tree structure -- Coding dependencies in a GOP.

An example with $n = 15$ is given in Figure 2. In this figure, the 8th frame (frame 7) is selected as I-frame, and two binary trees consisting of odd frames and even frames refer to I-frame, respectively. We choose the center frame to be I-frame instead of using the first frame, since this frame on average takes the minimum distance to any other inter-frame in the same GOP. Comparing with the conventional GOP, the average number of frames influenced by a damaged inter-frame is merely

$$\frac{0\times 8+1\times 2+2\times 2+3\times 2}{14}=0.86.$$

It is apparent that any two successive P-frames (in playback order) are not located in the same sub-tree, so a frame still can be compensated even though both of its neighbors are lost. In other words, a damaged frame can be recovered from both of its neighbors because loss of one frame will not affect the decoding process of it neighbors.

Errors in one P-frame propagate to at most three more frames in our example. However, the same error in the conventional GOP might propagate to at most 13 frames. Moreover, these damaged frames in the proposed structure are not continuous in playback order, so the viewer can hardly tell from them while seeing movies.

In video streaming applications, we transmit the double-binary tree in depth-first search (DFS) order, which guarantees that any errors in an inter-frame will never propagate onto its neighbors.

In the condition of single or burst frame-loss, an error concealment scheme similar to [9] is utilized. By assuming the motion is smooth, the nearby motion or pixel information is used to approximate the lost motion. For example, if frame 3 in Figure 2 cannot be decoded successfully, it can be replaced by any of the following approaches:

1) (1/2) × ($P_2 + P_4$) (the interpolation scheme, used to replace P_3);
2) (1/2) × $V_{4\rightarrow 2}$ (used to replace $V_{4\rightarrow 3}$);
3) (4/3) × $V_{7\rightarrow 4}$ (used to replace $V_{7\rightarrow 3}$);
4) $-V_{3\rightarrow 1}$ (used to replace $V_{1\rightarrow 3}$).

Here $V_{i\rightarrow j}$ means the motion vector from frame i to frame j, and P_i means the pixel values of frame i. By applying our error concealment mechanism, even if all frames in a sub-tree of the root are lost, we can reconstruct them from the other sub-tree containing all of their neighbors. We conclude the discussions as the following theorems.

Theorem 1 (single frame damaged). *The double-binary tree structure with DFS streaming order guarantees that any P-frame error does not propagate onto its neighboring frames in the playback order.*

We have known that loss of a frame will infect decoding of its descendants. Since any two successive P-frames are not located at the same sub-tree, and an error will never propagate to frames in other sub-trees, the above theorem can be simply proved. Therefore, the damaged P-frame can be reconstructed from its two neighboring frames by any existing interpolation method.

Theorem 2 (successive frames damaged). *For the left (right) sub-trees of I-frame, if burst frame-loss occurs not across other sub-trees, all lost frames can be recovered by interpolating from their neighbors in the playback order.*

As mentioned before, any two successive frames in double-binary tree are not located in the same sub-tree, and frames are sent in DFS order. This represents that frames are transmitted and decoded in the interleaving order, and the burst error often takes place within a sub-tree. In this condition, even though all frames of some sub-tree are lost, the remaining alternating frames still can be decoded from the other sub-tree, and then the damaged frames also can be recovered from those undamaged frames. In contrast, burst frame-loss in the conventional GOP causes that some succeeding damaged frames can be only recovered from the previous damaged frames such that the outcomes are worse and worse.

4. EXPERIMENTAL RESULTS

We modified the GOP structure in "XVID" [11] codec supporting MPEG-4 and H.263 encoding/decoding process to measure the efficiency of our structure. In our simulations, the GOP size and frame-rate is set as 15 and 30 fps, respectively.

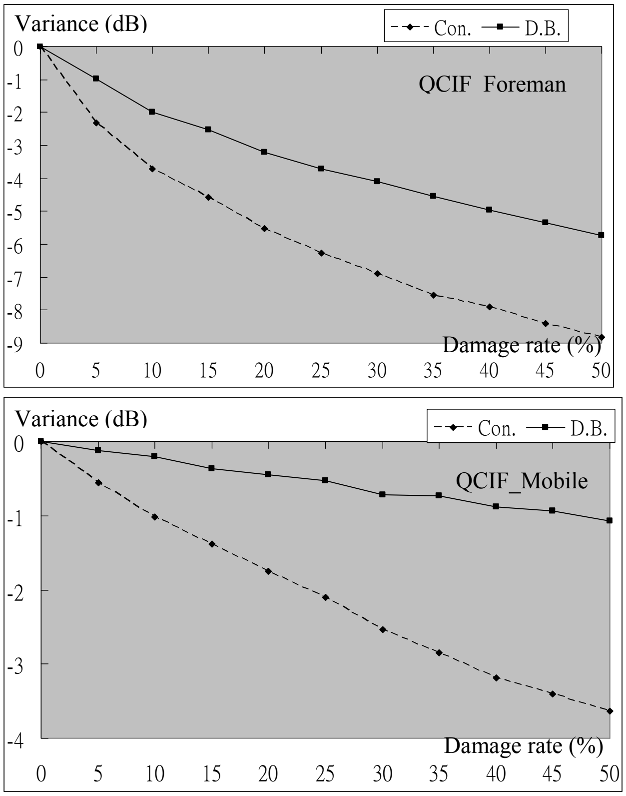

Figure 3. Comparisons of error concealment for "Foreman" and "Mobile" in QCIF format.

Figure 3 gives the variance of PSNR at different damage rate (% of damaged blocks in a single frame). Without loss of fairness and simplicity, we use the easiest mechanism, where the damaged block is recovered from the corresponding block in the neighboring undamaged frame. In the conventional GOP structure, one damaged block is directly pasted from the previous frame; while in the double-binary tree, we apply the average of the corresponding blocks in both the previous frame and the next frame instead of the damaged block.

To compare the difference among these schemes, the variation of the PSNR value between the specified damage rate and zero damage rates is represented with the vertical axis in Figure 3. Instead of rapid reduction of the quality level, we have a graceful degradation when the damage rate grows up. PSNR improves from 0.5 dB to 3 dB.

The most important question should be noticed is the degradation of quality level due to the long predictive distance in the double-binary tree GOP structure. TABLE I has shown that the average quality is similar to the conventional GOP structure in most cases, even better for some sequences.

TABLE I. Average of PSNR and bit-rate comparison between the conventional GOP structure and the double-binary tree structure for 150 frames. 'C': conventional GOP. 'D': double-binary tree GOP.

		C	D
Foreman (CIF)	PSNR (dB)	35.0	34.9
	Bit-rate (Mb/s)	0.50	0.50
Mobile (CIF)	PSNR (dB)	29	**29.8**
	Bit-rate (Mb/s)	1.79	**1.68**
Stefan (CIF)	PSNR (dB)	32.3	31.8
	Bit-rate (Mb/s)	1.32	1.32
Foreman (QCIF)	PSNR (dB)	32.9	**33.5**
	Bit-rate (Mb/s)	0.15	0.15
Mobile (QCIF)	PSNR (dB)	28.1	**29.1**
	Bit-rate (Mb/s)	0.47	**0.45**
Salesman (QCIF)	PSNR (dB)	34.9	**35.1**
	Bit-rate (Mb/s)	0.11	0.11

5. ADVANCED APPLICATIONS ON DOUBLE-BINARY TREE

In addition to error resilience, double-binary tree GOP can be extended to some advanced applications such as layered coding or multiple description coding techniques. In the application of layered coding, either frames at higher levels or frames with odd numbers in the double-binary tree could be chosen as the base layer, and the remainder could be enhancement layers. Similarly, in MDC, we could transmit odd frames and even frames separately over two different channel.

Under the above two applications, the received frames can be successfully compensated. Remaining frames are reconstructed by our error concealment scheme to get a basic quality level. When we have enough bandwidth to receive all data, the video clip gets the best quality.

6. CONCLUSIONS

In this paper, we proposed an error resilient mechanism for streaming video with effective dependent relations among frames. Double-binary tree reorganizes the regular predictive dependency to construct a new dependent relation. With our effective temporal dependencies among frames, the quality of the reconstructed damaged frame can be more precisely, and the error propagation phenomenon is significantly reduced. Besides, we do not need any feedback or redundant information. The proposed structure also can be used along with most of existing error resilient mechanisms to strengthen the protection of data. In addition to error resilience, we have successfully proposed another structured-GOP solution to support high efficiency and low cost VCR operations on VOD systems [10].

7. REFERENCE

[1] R. T. Derryberry, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman, "*Transmit diversity in **3G** CDMA systems*," *IEEE Communications Magazine*, vol. 40, April 2002.

[2] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Prentice Hall, 2002.

[3] S. Fukunaga, T. Nakai, and H. Inoue, "Error-resilient video coding by dynamic replacing of reference pictures," *GLOBECOM'96*, Nov. 1996.

[4] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified interframe coding system for limited reference picture memories," in *Proc. Picture Coding Symp.*, Berlin, Germany, Sept. 1997, pp. 743–748.

[5] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, Vol. 18, pp. 966–976, June 2000.

[6] R. Talluri, "Error-resilient video coding in ISO MPEG-4 standard," *IEEE Commun. Mag.*, Vol. 36, No. 6, pp. 112–119, July 1998.

[7] S. Wenger, "Video redundancy coding in H.263+," in *Proc. AVSPN 97*, Aberdeen, U.K., 1997.

[8] J. Apostolopoulos, T. Wong, W. Tan, and S. Wee, "On multiple description streaming with content delivery networks," *IEEE Infocom*, July 2002.

[9] Tamer Shanableh and Mohammed Ghanbari, "The Importance of the Bi-Directionally Predicted Pictures in Video Streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 3, March 2001.

[10] Kai-Chao Yang, Chun-Ming Huang, and Jia-Shung Wang, "Restructuring GOP algorithm to Reduce Video Server Load on VCR Functionality," in *International Conference on Parallel Processing 2003 (ICPP'03)*, 2003, pp.107-114.

[11] http://www.xvid.org