Content-based retransmission for video streaming system with error concealment

Yi Zhao[a], Stanley C. Ahalt[a] and Jianyu Dong[b]

[a] Department of Electrical and Computer Engineering
The Ohio State University
Columbus, OH 43210
zhaoy, sca@ee.eng.ohio-state.edu

[b] Department of Electrical and Computer Engineering
California State University, Los Angeles
Los Angeles, CA 90032
jdong2@calstatela.edu

ABSTRACT

In order to provide good QoS for video streaming in error-prone environments, effective error control methods are essential. Current error control methods can be classified into two categories: 1) Transport layer approaches such as FEC and retransmission; 2) Application layer approaches such as error resilience coding and error concealment. By far, most existing research is aimed towards optimizing one of the above approaches to reduce the impact of transmission errors. However, there is usually more than one error control method in a real video streaming system. In this case, how to optimize the system performance becomes more complicated, and is not standardized. This paper presents the research effort to joint-optimize the effects of two error control methods, retransmission and error concealment, in wavelet-based video streaming system. The major difficulty of the joint-optimization is that the two methods are mutually dependent; the system cannot be optimized by improving each error control method independently. To tackle this problem, a new content index, namely "reconstruction distortion", is defined to quantify both the packet content and its importance in error concealment. Based on the defined content index, a content-based retransmission approach is developed to select the best packet-sending scheme to maximize the quality of the received video under the given error concealment method. Experiments results demonstrate the effectiveness of the proposed method.

Keywords: Video streaming, error concealment, retransmission, content index, joint-optimization

1. INTRODUCTION

In recent years, video streaming applications on the Internet have become increasingly popular. Due to the huge volume of the video data and the limited bandwidth of the Internet users, compression is essential in streaming applications. While reducing the data volume efficiently, compression makes the video packets vulnerable to bit errors and packet losses. Therefore, error control techniques are critical to provide a good quality to the streamed video. Due to the popularity of the DCT-based video such as MPEG and H.261/H.263, many corresponding error control method have been proposed.[1–4] In,[1] Gallant et al proposed unequal protection method based on the property of layered coding; adaptive intra/inter mode switching proved to be efficient to increase the error resilience for DCT-based video.[3, 4] Although effective, these approaches cannot be applied to the streaming application of 3-D wavelet video due to the difference between the two compression methods.

Our research focuses on the quality streaming of 3-D wavelet video. Although not so widely used, 3-D wavelet video has some advantages such as great scalability and error resilience ability,[5] which are desirable in streaming applications. Since no motion compensation is used in 3-D wavelet based compression, the video bitstream can be put into independent packets.[6, 7] Therefore, loss of one packet will not propagate the damage to other packets. Nevertheless, to improve the quality of 3-D video streaming, more active error control method is needed. In our previous research, the concept of content-based retransmission was proposed.[8] Based on delay-constrained retransmission,[9, 10] the proposed method jointly considers the characteristics of the 3-D wavelet video, the packets' contents, and the channel conditions to optimize the transmission schedule. Although proved to be effective by experiments, one limitation of this approach is that the optimization is conducted assuming there is only one error control method, retransmission, in the streaming system. However, this assumption is not always true. When two or more error control methods co-exist in one system, how to optimize the performance becomes more complicated.

As an extension to the previous research, this paper presents the joint-optimization of two error control methods: retransmission and error concealment in the wavelet-based video streaming system. The major difficulty of the joint-optimization is that the two methods are mutually dependent. Thus the system cannot be optimized by improving each error control method independently. To tackle this problem, a new content index, namely "reconstruction distortion", is defined to quantify both the packet content and its importance in error concealment. Based on the defined content index, the previously proposed content-based retransmission approach is revised to select the best packet-sending scheme to maximize the quality of the received video under the given error concealment method.

The rest of the paper is organized as follows: Section 2 reviews the 3-D wavelet video streaming system with error concealment, and formulates the joint-optimization problem for retransmission and error concealment. Section 3 presents the proposed content-based retransmission approach while taking into account the effect of error concealment. Experimental results are presented in Section 4 to evaluate the performance of the proposed approach. Section 5 concludes the paper.

2. PRELIMINARIES

2.1. 3-D wavelet compression with error concealment

Unlike the MPEG/H.263 video codec, 3-D wavelet compression uses wavelet transform to remove both the spatial and temporal correlation. Based on different coding methods used after the wavelet transform, the 3-D wavelet compression can be classified into two categories: 1) subband wavelet coding; 2) Zerotree wavelet coding. Subband wavelet coding uses scalar quantization in each subband first, followed by run-length and entropy coding; the latter employs successive quantization, followed by tree-based coding. Since the bitstream structure produced by zerotree wavelet coding leads to more robust packetization,[7, 11] it is more desirable in streaming applications. Therefore, our research emphasizes on optimizing the streaming quality of zerotree wavelet video.

2.1.1. 3-D zerotree wavelet coding

The overall structure of 3-D zerotree wavelet codec is depicted in Figure 1. First, 3-D wavelet transform is applied to a group of frames (GOF). Secondly, an optimal tree structure is formed based on the subband correlation.[12] Then, successive quantization and coding will be applied to the coefficients on the tree structure. The structure of the output bitstream is determined by the way quantization and coding is conducted, which leads to different packetization schemes.

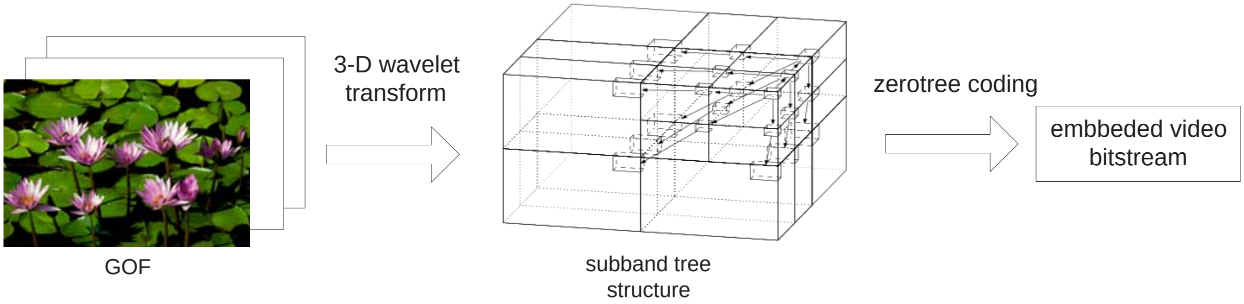
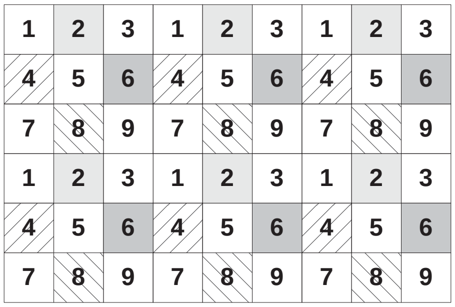


Figure 1. Overall structure of 3-D zerotree wavelet codec.

2.1.2. Robust packetization

Since the bitstream will be delivered in a packet-switch network, a good packetization method is essential to prevent error propagation. In 1990, the concept of Application Level Framing (ALF)[13] was proposed, which requires the packet boundary be conformed with the frame boundary of the application data unit (ADU). Thus the loss of one packet will not affect the decoding of other received packets.

In reality, it might be impossible to conform with ALF since the bitstream output from certain codec has very strong dependency. Fortunately, Packetized Zerotree Wavelet (PZW)[7] revised the original zerotree coding method to allow each tree be coded and packetized independently. To further enhance the error resilience of the video packet, the trees are spatially interleaved during packetization. One interleave pattern is illustrated in Figure 2, where the blocks labeled with the same number are grouped together and put in one packet. Therefore, losing one packet causes neither error propagation, nor aggregate loss of adjacent trees.



1	2	3	1	2	3	1	2	3
4	5	6	4	5	6	4	5	6
7	8	9	7	8	9	7	8	9
1	2	3	1	2	3	1	2	3
4	5	6	4	5	6	4	5	6
7	8	9	7	8	9	7	8	9

Figure 2. Illustration of spatial interleave.

2.2. Error concealment for 3-D wavelet video

Error concealment method for 3-D wavelet video is very different from the that for DCT-based video. DCT-based video used frame-based compression that introduce strong dependency among adjacent frames. Therefore, lost frame can predicted using the previous frame. In 3-D wavelet video, a tree corresponds to a 3-D local region. If a tree is lost, it can be predicted using the neighboring trees. Hence, to facilitate the error concealment for 3-D wavelet video, spatial interleaving at the tree level is needed to reduce the probability of losing adjacent trees.

The procedure of the tree-based error concealment can be illustrated by Figure 3. The lost coefficient of lowest LL band is approximated by interpolating the four neighbors, while the lost coefficients of the LH and HL subbands after the last level of wavelet decomposition are predicted by linear interpolation from the two nearest neighbor coefficients in that direction that the low-pass filter is applied.

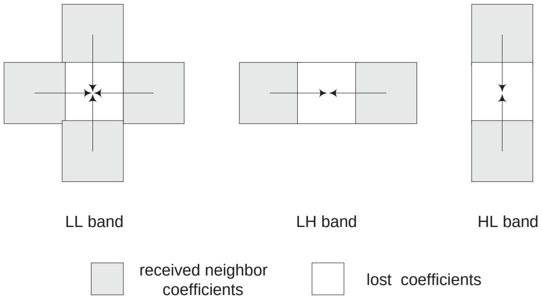


Figure 3. Illustration of error concealment method for 3-D wavelet video.

3. CONTENT-BASED RETRANSMISSION UNDER ERROR CONCEALMENT

3.1. Basics of content-based retransmission

The basic idea of the content-based retransmission proposed in our previous research is choosing a transmission scheme based on the packet's content to minimize the distortion caused by packet loss. When bandwidth is not sufficient to send all the packets (including those requested to be retransmitted), packets with less significant content may be sacrificed to ensure the transmission of more important packets.

To quantitatively measure the significance of the packet content, a content index called *loss-distortion* was defined[8]:

$$\Delta D_i = D_i - D_i^* + \sum (1 - p_{lost}) \cdot \Delta D_{dependent}, \tag{1}$$

where D_i is the distortion of the video reconstructed from all packets except both packet i and its dependent packets[*], D_i^* is the distortion of the video reconstructed from all packets except only the dependents of packet i, p_{lost} and $\Delta D_{dependent}$ are the loss probability and loss-distortion of packet i's direct dependent packet, respectively.

Since our previous research focused on subband wavelet coding where there exists dependency among different subbands, the definition in the equation (1) considers the loss impact of the dependent packets. For zerotree wavelet video with robust packetization, the video packets are independent decodable. Hence, the definition can be simplified. Assuming that the set of coefficient trees contained in packet i is denoted by Ω_i, and the wavelet coefficient is denoted by $c_{x,y,z}, x, y, z \in \Omega_i$, then the loss distortion can be approximated as:

$$\Delta D_i = \sum_{x,y,x \in \Omega_i} c_{x,y,z}^2 .$$

3.2. Problem formulation for joint-optimization

The advantage of the content-based retransmission is that it utilizes limited network resources to transmit the most important packets. Since the importance of the packet is measured by the square sum of wavelet coefficients (which approximates the energy contained in the packet), the foreground or the regions containing larger motion is more likely to be transmitted successfully than the background or smooth, slow moving regions. This effect is generally desirable since human perception is more sensitive to the foreground. However, it may cause problem when error concealment is used as the decoder.

Consider the following scenario. Assuming the loss behavior of the network is random, then the effect of packet loss is usually spread across the whole video frame. Content-based retransmission sacrifices the packets with small loss distortion to ensure the transmission of more significant ones. In another words, the loss pattern is changed. Content-based retransmission causes more severe loss in non-significant packets. If massive loss happens to the non-significant packets in the same local area, the performance of error concealment will be seriously degraded.

Apparently, if the content-based retransmission operates independently with the error concealment, it is impossible to achieve the best visual quality at the receiver. When two error control methods co-exist, both of their effects should be taken into account to optimize the system performance. Assume that the average data transmission rate is constrained by R, and the delay constraint imposed on the i^{th} group of frame (GOF) in the application is $T_D(i), i = 1, 2, ...$ The objective of the joint-optimization is to find a policy π which chooses a subset of the packets to transmit such that the quality of the received video is maximized under the rate and delay constraints:

$$\begin{cases} r_s \leq R \\ \sum_{j \in \Omega_T(\pi(i))} B_j \leq r_s \cdot T_D(i) \\ \pi = \arg\min_{\pi_k} \left(\sum \left(D\left(P_{lost}\left(\pi_k\right)\right) - \Delta d\left(P_{lost}\right) \right) \right) \end{cases}, \tag{2}$$

where r_s is the average data sending rate, and B_j is the size of the j^{th} packet selected to transmit in the i^{th} GOF under policy π. $D\left(P_{lost}\right)$ and $\Delta d\left(P_{lost}\right)$ are the distortion caused by packet loss and the change of distortion caused by error concealment of the lost packet, respectively.

---

[*] If a packet is lost, all its dependent packets become useless in reconstruction.

3.3. Near-optimal retransmission with error concealment

Based on the formulation, the key to achieve the best error control performance is to incorporate the effect of error concealment as well as the importance of the packet content when selecting the transmission scheme. In this subsection, we first propose an improved content index, *reconstruction distortion*, which reflects the impact of packet loss on the overall video quality when error concealment is presented, and then describe a real-time decision-making system to optimize the retransmission.

3.3.1. Definition of reconstruction distortion

Reconstruction distortion is a quantitative measurement of the quality reduction caused by the packet loss in a streaming system with error concealment method described in Section 2. Assuming that the video packets are independently decodable, the *reconstruction distortion* of packet i, ΔRD_i, is defined as:

$$\Delta RD_i = \Delta D_i - (1 - p_{lost})^k \, \Delta d_i, \tag{3}$$

where ΔD_i is the loss distortion of the lost packet i, p_{lost} is the loss probability in the channel, k is the number of packets containing necessary information for error concealment, and Δd_i is the distortion reduction achieved by error concealment for packet i.

The value of reconstruction distortion depends on three factors. The first factor is the loss distortion of the packet which reflects the importance of the packet content. The second is the effect of error concealment; a packet that is easy to be predicted by other received ones usually has small reconstruction distortion although its may contain complicated content. The third factor is the channel loss rate. Based on (3), reconstruction distortion increases with channel loss rate, since successful error concealment relies on the recipient of the neighbor coefficients. As a result, higher channel loss rate leads to worse performance of error concealment.

3.3.2. Near-optimal transmission system

Based on the proposed new content index, to find the best transmission scheme is to select a subset of packets to send under the rate and delay constraints. As described in,[8] this is an NP problem. The optimal solution can be obtained by exhaustive search, but the computation complexity is too high for real-time application. To meet the delay constraint, the fast-decision making algorithm presented in[8] is employed to determine a near-optimal transmission scheme. In the fast-decision making algorithm the sending order of the packets is determined by $\Delta RD_i/B_i$, where B_i is the size of the packet in bits. $\Delta RD_i/B_i$ represents reconstruction distortion per bit. Apparently, packet with larger $\Delta RD_i/B_i$ is more important in transmission with rate constraint. To further speed up the computation, the reconstruction distortion is generated for each coefficient tree during compression. To do so, error concealment is needed at the sender to obtain Δd.

4. PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed joint-optimization method, a lot of experiments have been conducted. In the experiments, the delivered video bitstream is generated by the 3-D zerotree wavelet codec described in Section 2. Since the goal of the proposed approach is to achieve the best video quality, PSNR of the received video is used to evaluate the performance. The experiments are conducted under two different channel conditions: a channel with low packet loss rate (5%), and a channel with high packet loss rate (20%). Here, the packet loss is independent. To rule out the random factor, the presented results is the average results of many measurements.

Table 1 and 2 compare the average PSNR of the received video sequences, “Football” and “Miss American”, under different transmission schemes:

- Scheme 1: no retransmission, no error concealment (no ARQ and EC)
- Scheme 2: no retransmission, with error concealment (EC only)
- Scheme 3: Content-based retransmission, no error concealment (CARQ only)

Table 1. Performance comparison of different transmission schemes for Football sequence.

loss rate	5%	20%
no ARQ and EC	26.6747	22.8966
EC only	27.2369	23.8313
CARQ only	28.9686	27.5476
CARQ with EC	29.0189	27.4116
EC-based CARQ with EC	29.0296	27.5750

Table 2. Performance comparison of different transmission schemes for Miss American sequence.

loss rate	5%	20%
no ARQ and EC	30.9525	25.4084
EC only	34.8813	30.5595
CARQ only	36.7257	33.3450
CARQ with EC	36.2912	33.3895
EC-based CARQ with EC	37.5792	34.3327

- Scheme 4: Content-based retransmission, error concealment, applied independently (CARQ with EC)
- Scheme 5: Joint-optimization of content-based retransmission with error concealment (EC-based CARQ with EC)

Figures 4 and 5 depict the frame-by-frame PSNR of the video sequence "Miss American" delivered by the above five different transmission schemes under different channel conditions, respectively. In order to indicate the best achievable video quality, PSNR of the reconstructed video without packet loss is also displayed in the figure. From the simulation results, we can see that the proposed joint-optimization of the content-based retransmission with error concealment can achieve the best video quality. When the packet loss rate is 20%, joint considering the content-based retransmission and error concealment can improve the average video quality by nearly 1 dB compared to applying the two error control methods independently. An interesting observation is that in the football sequence, the video quality achieved by applying content-based retransmission and error concealment independently is worse than that achieved by content-based retransmission only. This is because the transmission scheme achieved solely based on packet content sometimes compromises the effect of error concealment by dropping packets containing adjacent trees.

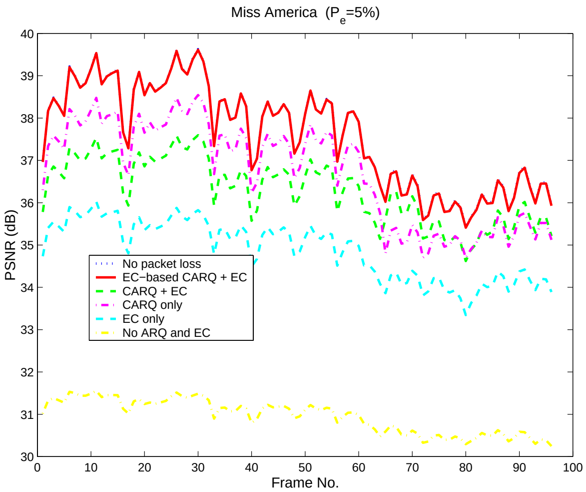


Figure 4. Frame-by-frame PSNR comparison at packet loss rate 5%.

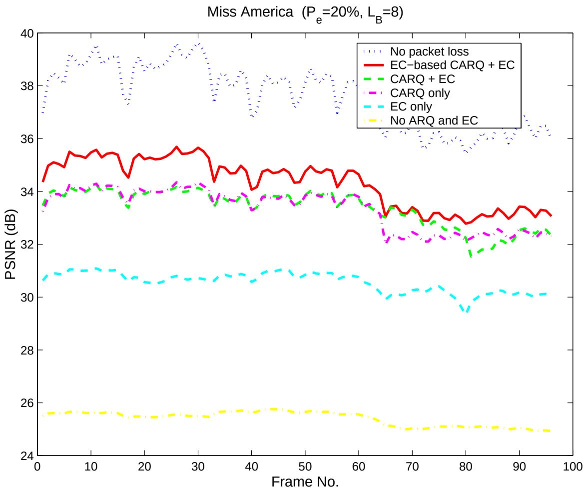


Figure 5. Frame-by-frame PSNR comparison at packet loss rate 20%.

5. CONCLUSIONS

In this paper, we presented the joint-optimization of content-based retransmission with error concealment, with the goal of maximizing the streaming quality of 3-D zerotree wavelet video in lossy channels. Considering both the significance of the content in each video packet and the effect of error concealment at the decoder, the previously proposed content index is revised to reflect the impact of packet loss more accurately. In order to select the best transmission scheme in real-time fashion, the content index of each packet is pre-calculated and stored at the sender. Experimental results showed that the proposed approach is very effective. In our future work, we will investigate how to update the content index of the packet on the runtime based on the changing channel condition and transmission status of other packet.

ACKNOWLEDGMENT

This publication was partially funded through support provided by DoD HPCMP PET activities through Mississippi State University under the terms of Agreement No. # GS04T01BFC0060. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the DoD or Mississippi State University.

REFERENCES

1. M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Transactions on Circuits and Systems for Video Technology* **11**, pp. 357–372, Mar. 2001.
2. F. Wu, Z. Li, and Y. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology* **11**, pp. 332–344, Mar. 2001.
3. J. Liao and J. Villasenor, "Adaptive intra block update for robust transmission of H.263," *IEEE Transactions on Circuits and Systems for Video Technology* **10**, pp. 30–35, Feb. 2000.
4. R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications* **18**, pp. 966–976, June 2000.
5. S. D. Servetto and K. Nahrstedt, "Broadcast quality video over IP," *IEEE Transactions on Multimedia* **3**, pp. 162–173, March 2001.
6. X. Wu, S. Cheng, and Z. Xiong, "On packetization of embedded multimedia bitstreams," *IEEE Transactions on Multimedia* **3**, pp. 132–140, Mar. 2001.
7. J. K. Rogers and P. C. Cosman, "Wavelet zerotree image compression with packetizatio," *IEEE Signal Processing Letters* **5**, pp. 105–107, May 1998.
8. J. Dong and Y. F. Zheng, "Content-based retransmission for 3-D wavelet video on the Internet," in *Proc. IEEE Int. Conference on Information Technology, Coding and Computation*, pp. 452–457, Apr. 2002.
9. B. J. Dempsey, *Retransmission-based error control for continuous media traffic in packet-switched networks.* PhD dissertation, University of Virginia, 1994.
10. B. J. Dempsey, J. Liebeherr, and A. C. Weaver, "On retransmission-based error control for continuous media traffic in packet-switched networks," *Computer Networks and ISDN System* **28**, pp. 719–736, Mar. 1996.

11. W. T. Tan and A. Zakhor, “Real-time Internet video using error resilient scalable compression and TCP-friendly transport protocol,” *IEEE Transactions on Multimedia* **1**, pp. 172–186, June 1999.
12. C. He, J. Dong, and Y. F. Zheng, “Studies on tree-structure for 3-D wavelet video coding,” *IEEE Trans. on Circuits and Systems on Video Technology* **13**, pp. 961–972, October 2003.
13. D. Clark and D. Tennenhouse, “Architectural considerations for a new generation of network protocols,” in *Proc. ACM Sigcomm 90*, 1990.