

PROXY-BASED SNR SCALABLE ERROR TRACKING FOR REAL-TIME VIDEO TRANSMISSION OVER WIRELESS BROADBAND

Hui-Ya Li and Wei-Kuo Chiang

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION ENGINEERING, NATIONAL
CHUNG CHENG UNIVERSITY, CHIAYI, TAIWAN 621, R.O.C.
{alyh93, wkchiang}@cs.ccu.edu.tw

ABSTRACT

In this paper, we propose a proxy-based SNR scalable error tracking framework for real-time video transmission where the server is wired connected to Internet and the client is connected to Internet through wireless broadband networks. We assume that all errors (packet losses) result from wireless links, and wired links are assumed to be error-free. The client sends back NACKs with the information about base layer lost packets to the proxy via a feedback channel. Once the NACK is received, the proxy uses the motion data and the side information received from the streaming server to perform error tracking. We compare our approach to the original proxy-based error tracking scheme without scalability support at the same bitrate and bit error rate. Experimental results show that the proposed method can effectively improve performance.

1. INTRODUCTION

The demand for real-time video streaming services is steadily on the increase. In order to enable more efficient use of transmission and storage resources, there are several video compression standards developed to compress video sequences into bitstreams. These video encoding standards achieve a high compression ratio by removing redundancy both in the temporal and spatial domains, while making the compressed data more sensitive to transmission error. Decoding of erroneous or incomplete video bitstreams may lead to serious video quality degradation, which not only affects the quality of current frame, but also results in error propagation to following frames because of the use of the motion-compensated prediction technique in video coding standards. For achieving error robust transmission, especially for transmitting over unreliable wireless networks, an error resilient transmission scheme [1],[2] is essential.

In these error resilient techniques, Error Tracking (ET) [3],[4] allows clients to report any corrupted blocks due to transmission errors back to the server via a feedback channel. According to the negative acknowledgment

(NACK), the server can track spatio-temporal propagated errors to determine the INTRA update allocation in the current encoding frame. This INTRA refreshment terminates the error propagation effect. Due to its relatively low additional memory requirement and computation complexity [5], the ET algorithm is suitable for real-time application.

The performance of ET is closely related to round-trip time between the server and the client. Since the wired link can be viewed as almost error-free, a proxy-based approach is proposed in [6], which performs ET on the proxy to reduce the error propagation and the round-trip delay. In [7] an additional side information stream is sent to the proxy so as to reduce the computational complexity on the proxy. The proxy first extracts the coding mode and the motion vector information of contaminated macro-blocks (MBs), and then uses the side information to backfill these corrupted image areas with INTRA-MBs. Therefore, the incoming bitstream does not have to be fully decoded and affected MBs not to be transcoded anymore.

The side information stream transmitted in [7] is the INTRA mode of the reconstructed original bitstream. Although this transmission reduces the computational complexity on the proxy, it has several disadvantages. First, transmitting large INTRA frames requires large bandwidth on the wired network and may lead to high congestion, which makes it difficult to claim error-free on the wired side. Second, it's hard to guarantee that the original stream and the side information can be simultaneously in flight; which will create difficulties in backfilling affected MBs. Third, a large memory requirement proxy is needed.

We deal with these problems by proposing a proxy-based SNR scalable error tracking (PSET) scheme. The proposed PSET partitions the original video stream into layers (one base layer with one or more enhancement layers) without increasing bitrate, and simply performs ET on the base layer stream using the INTRA mode of the reconstructed base layer bitstream as the side information.

The remainder of this paper is organized as follows. How and what improvements are made by PSET is described in Section 2. In Section 3, our proposed PSET framework for error robust real-time video transmission is

illustrated. Section 4 presents experimental results that show improvements achieved by our proposal compared to the original proxy-based ET. Finally the conclusion is drawn in Section 5.

2. PARTITIONING THE VIDEO STREAM WITH THE SCALABLE EXTENSION OF H.264/AVC

We use the scalable extension of H.264/AVC video codec [8],[9],[10], which has been chosen to be the starting point of the new standardization activity on Scalable Video Coding of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG). The scalable extension of H.264/AVC uses the Fine Grain SNR Scalability (FGS) to support the SNR scalability feature. FGS is a way to generate the refinement data to improve the visual quality of the reconstructed video bitstream. As in Figure 1, the base layer is the H.264/AVC compliant stream whose coefficients are quantized with a coarse quantization step size. The enhancement layer encodes successive refinements of transform coefficients, starting with the minimum quality provided by the base layer. The refinement data is directly coded in the transform coefficient domain and corresponds to a bisection of the quantization step size. Furthermore, each enhancement layer stream can be truncated at any arbitrary point (progressive refinement).

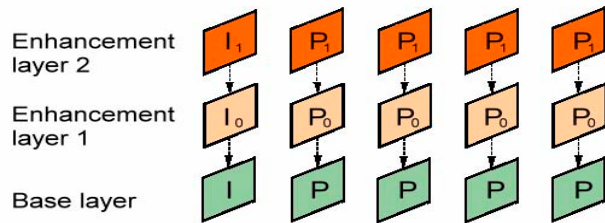


Figure 1- Example for the FGS SNR scalability of video coding.

Since there is no temporal correlation within the same enhancement layer, lost enhancement layer packets will not cause any drift problem. We can thus ignore the loss of the enhancement layer packet and only perform ET on the base layer, which means the packet size of the original stream and the side information are both decreased. This not only decreases the internet congestion probability, but also offers following advantages.

- Lower packet loss rate in the wireless environment. The main cause for packet loss in the wireless channel is the high BER (bit error rate). Since the PER (packet error rate) is proportional to the packet size [11],[12], smaller packet size results in lower packet error rate under the uniform BER model.
- Prevent large INTRA-refresh loss. When perform ET to do INTRA update (to be

discussed in Section 3), some INTER-MBs have to be replaced by INTRA-MBs, which inevitably increase the packet size and/or the packet number, and eventually make the packet loss rate increases since the ET needs to be triggered more times.

- Scalable. Clients have different decoding capabilities depending on their characteristics and transmission environments. Therefore, it is highly desirable to have a video coding technology which can be adapted to various channels with different bandwidth and bandwidth variation. With this scalable extension, the client still can ensure a lowest acceptable visual quality even in a poor transmission environment since the base layer has an error resilience scheme to avoid error propagation.

3. PROXY-BASED SNR SCALABLE ERROR TRACKING

An example for an error robust transmission using the proxy-based SNR scalable error tracking is illustrated in Figure 2.

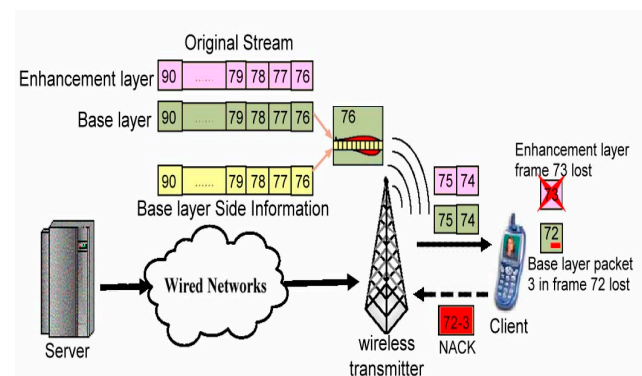


Figure 2- A proxy-based SNR scalable error tracking scenario.

Assume that the server is wired connected to Internet and the client is connected to Internet through wireless broadband networks. When the client asks for a real-time video streaming service, it sends the server a request which indicates the quality requirement (layers on demand), for example two layers as shown in Figure 2. The server will then transmit three kinds of video streams to the wireless transmitter (the access point of a WLAN or the base station of a 3G mobile), as shown in Figure 2, the base layer (P-frame), the enhancement layer (FGS progressive refinement), and the side information (INTRA encoded P-frames of the base layer bitstream).

As shown in Figure 2, the round-trip delay between the access point and the client is 3 video frames. In this example, the client first detects the base layer packet 3 of

frame 72 is lost and sends a NACK to the access point. When the access point receives the NACK which indicates an error occurred at packet 3 of frame 72 while the next frame to be sent is 76, in contrast to the ET algorithm used in [7], the access point first uses the fast error tracking algorithm proposed in [5] to backward trace the four corners of every single MB in frame 76 along with the corresponding motion vectors and the frame distance 4 to see if any corner of an MB points to the erroneous area of frame 72. Based upon the illustration, several MBs in frame 76 are then determined to be contaminated MBs. Then the access point extracts the coding mode and the motion vector information of contaminated MBs in frame 76, and uses the side information of frame 76 to backfill these contaminated MBs with INTRA-MBs. Finally, the access point sends the refresh packet to the client.

After reporting the NACK about frame 72, the client then detects that the enhancement layer packet of frame 73 is also lost. Since the enhancement layer packet includes the quantization error between the base layer and the original subband picture, the packet loss can not cause further damage; the client ignores the loss and keeps decoding.

Please note that the side information is assumed to be sent every frame in this scenario. This assumption is not necessary. A subset of INTRA frames or a subset of INTRA slices can be transmitted if the transmission is good over the wireless side.

In summary, the two major assumptions made are the following. We assume that the client can send feedback about base layer lost packets to the access point, and the access point can perform ET to do INTRA-refresh on the next waiting frame. The second assumption is that all errors come from wireless links, and the wired link can be viewed as error-free.

4. EXPERIMENTAL RESULTS

The video sequence Foreman is tested in the experiment. It is in CIF (352×288) format, frame rate at 30 Hz, and 300 frames in length. The scalable H.264/AVC extension reference software version JSVM 2.0 [13],[14],[15] is adopted as the video codec.

We analyze the use of PSET with one base layer and one enhancement layer with three different base layer quantization parameters of 33, 35, and 37 compared to proxy-based ET with quantization parameter of 29 [ref] at the same bitrate. The side information is sent every frame, and its quantization parameter is the same as used in the base layer. Table 1 shows the bitrate and PSNR for test streams in an error-free environment, and the overhead of the side information is specified. The coding structure is pure IPPPP-sequence without B-frames. The FGS layer QP corresponds to a decrease of the base layer QP by 6. The FGS layer bitrate shown in Table 1 is the cumulative bitrate which includes the base layer and the enhancement layer.

In our simulation, we first assume that all errors come from wireless links, and the wired link can be viewed as error-free. The wireless link bit error rate is set to 2×10^{-5} , and a packet is lost as long as any bit is flipped. Figure 3 shows the curve of the relationship between the packet size and the packet error rate.

One frame is placed into one packet if its compressed bitstream is smaller than the MTU (Maximum Transfer Unit) packet size. The MTU packet size we use is 1400 bytes. The round-trip delay between the access point and the client is set to be 2 frames, which corresponds to a round-trip time of 200ms. For every lost base layer packet we perform one ET. If one FGS packet gets lost, the complete frame is marked as lost. In case of base layer packet loss, the corresponding position of the last frame is repeated in order to allow a reasonable PSNR calculation.

		Proxy-based ET	Proxy-based SNR scalable ET			
			QP=29	QP=33	QP=35	QP=37
Base layer (H.264)	Bitrate (kBit/s)	414	209	154	118	
	PSNR (dB)	36.04	33.40	32.13	30.99	
FGS layer	Bitrate (kBit/s)	/	414	414	414	
	PSNR (dB)	/	33.82	32.79	31.92	
Side Info.	Bitrate (kBit/s)	1811	1102	842	666	

Table 1- The test streams in an error-free environment with 4 different QP, CIF format, 30 fps.

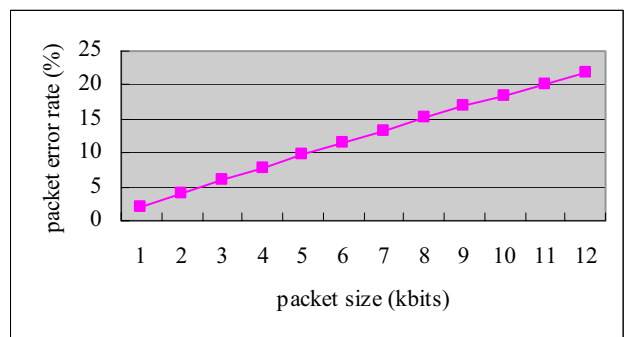


Figure 3- The relationship between the packet size and the PER. BER is 2×10^{-5} .

The PSNR results for test streams are shown in Table 2. The packet loss results in a decrease of 7.38dB for original proxy-based ET, 2.14dB for PSET using QP of 33, 1.4dB for PSET using QP of 35, and 1.03dB for PSET using QP of 37 respectively. The proxy-based ET loses quality significantly due to the high PER and high INTRA-refresh loss resulting from the larger packet size. In contrast,

the PSET protects the most important part (the base layer) and reduces the INTRA-refresh loss by decreasing the base layer packet size. Moreover, the quantization error included in the enhancement layer packet can possibly overcome the channel loss to further enhance the quality.

	Proxy-based ET	Proxy-based SNR scalable ET		
	QP=29	QP=33	QP=35	QP=37
PSNR(dB)	28.66	31.68	31.39	30.89

Table 2- The PSNR of test streams in a BER= 2×10^{-5} environment with 4 different QP.

Comparing results in PSET of using different QP, we notice that the PSNR differences between the error-free and the error-prone environments are smaller and the PSNR values of the pure base layer in the error-free environment are closer with increasing QP. Otherwise, increasing QP will cause serious destruction to the enhancement layer. In order to provide an acceptable quality, we should not decrease the packet size of the base layer without any strong justification; even we can further decrease the packet more to get a lower PER. As indicated in Table 1, we find that the bitrate percentages of the base layer of all layers are 50.48% for QP of 33, 37.20% for QP of 35, and 28.5% for QP of 37 respectively. To have a better visual quality than the original proxy-base ET in an error-prone environment, we suggest that the base layer bitrate percentage of all layers should not be lower than 28%.

With a suitable bitrate allocation of every layer, our system can be easily extended to multiple enhancement layers without loss of generality.

5. CONCLUSION

Proxy-based error tracking with additional side information can significantly reduce the computational complexity on the proxy. But the large side information can cause high congestion and synchronization problem. In our proposal, the original stream is partitioned into several layers using the SNR scalable extension of H.264/AVC. The server simply sends the side information of base layer packets to the access point, which only performs error tracking whenever there is a base layer packet loss. Therefore, our approach not only reduces the overhead of the side information but also the packet error rate in the wireless link. Performance evaluations show that our approach works better than the original proxy-based ET in the error-prone environment in spite of a slight sacrifice of quality when using FGS technique to encode the video sequence. Moreover, our approach can adapt to variable channel bandwidth and clients with different decoding capabilities.

6. REFERENCES

- [1] Y. Wang, S. Wenger, J.T. Wen, A.K. Katsaggelos, "Review of Error Resilient Coding Techniques for real-time Video Communications," *IEEE Signal Proc. Magazine*, vol. 17, no. 4, pp. 61-82, Jul. 2000.
- [2] Y. Wang and Q. Zhu, "Error control and concealment for video communications: A review," *Proceedings of IEEE Special Issue on Multimedia Signal Processing*, pp. 974-977, May 1998.
- [3] W. Wada, "Selective recovery of video packet losses using error concealment," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 807-814, June 1989.
- [4] E. Steinbach, N. Färber, 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.
- [5] P.C. Chang, T.H. Lee, "Precise and Fast Error Tracking for Error-Resilient Transmission of H.263 Video," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 15, no. 4, Jun. 2000.
- [6] T.C. Wang, H.C. Fang, L.G. Chen, "Low Delay and Error Robust Wireless Video Transmission for Video Communications," 2002 IEEE International Conference on Multimedia and Expo (ICME 2002), Lausanne, Switzerland, Aug. 2002.
- [7] Wei Tu, Eckehard Steinbach, "Proxy-based Error Tracking for H.264 Based Real-time Video Transmission in Mobile Environments," *IEEE International Conference on Multimedia & Expo (ICME 2004)*, Taipei, Taiwan, June 2004.
- [8] H. Schwarz, D. Marpe, T. Schierl, T. Wiegand, "Combined Scalability Support for the scalable Extensions of H.264/AVC," *ICME*, July 2005.
- [9] H. Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, T. Wiegand, "Technical Description of the HHI proposal for SVC CE1," *ISO/IEC JTC1/SC29/WG11, M11244*, Oct. 2004.
- [10] H. Schwarz, D. Marpe, T. Wiegand, "MCTF and Scalability Extension of H.264/AVC," *Proc. Of PCS, San Francisco, CA, USA, Dec.2004*.
- [11] David Eckhardt and Peter Steenkiste, "Improving Wireless LAN Performance via Adaptive Local Error Control," *Sixth IEEE International conference on Network Protocols (ICNP'98)*, Austin, October, 1998.
- [12] Dan Duchamp and Neil F. Reynolds, "Measured Performance of a Wireless LAN," 17th conference on Local Computer Networks, *IEEE 1992*, pp. 494-499.
- [13] J.Reichel, H.Schwarz, M. Wien (eds.), "Scalable Video Coding-Working Draft 2," *Joint Video Team (JVT)*, Doc. JVT-O201, Busan, Korea, April, 2005.
- [14] J.Reichel, H.Schwarz, M. Wien (eds.), "Joint Scalable Video Model JSVM-2," *Joint Video Team (JVT)*, Doc. JVT-O202, Busan, Korea, April, 2005.
- [15] ITUT and ISO/IEC JTC 1, "JSVM 2 Software", *JVT-O203*, Apr, 2005.