A Dynamic Transmission Method for Scalable Video with Unequal Error Protection

Xi Zheng and Yonglin Xue

Abstract. In this paper, we propose a dynamic transmission method for scalable video with unequal error protection (UEP). Forward error correction (FEC) is used in the UEP scheme. FEC codes, such as Reed-Solomon (RS) codes, are allocated to different layers of scalable video according to their importance. The UEP scheme is applied on each group of pictures (GOP) with taking account of the data size fluctuation of GOPs. The proposed method dynamically adjusts the number of packets according to the source data size of each GOP in order to achieve similar channel code rates for different GOPs. Meanwhile, the buffer management is also considered to limit transmission delay on the sender side. Simulation results show that the proposed method can make significant improvement of average PSNR of reconstructed video in a wide range of channel coding rate and packet loss rate compared with the conventional UEP scheme.

Keywords: dynamic transmission • scalable video • group of pictures • unequal error protection

1 Introduction

With the rapid popularization and development of video applications, video coding and transmission technology has attracted growing research interests. Meanwhile, heterogeneous network conditions, terminal capability and requirements make it a big challenge to provide reliable and effective video service. On one hand, the high efficiently compressed video data is very sensitive to transmission errors caused by network congestion, channel fading, etc. On the other hand, video

X. Zheng (✉)
Tsinghua National Laboratory of Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China
e-mail: zhengx06@mails.tsinghua.edu.cn

Y. Xue
Tsinghua National Laboratory of Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China

applications are used on various terminals with differences in computing capability, screen size and so on.

Scalable video coding (SVC) provides a solution to the problem mentioned above. SVC allows transmission and decoding of partial bit streams. Videos with different temporal or spatial resolutions or fidelity can be reconstructed from partial bit streams at different channel rates. Consequently, it provides flexibility to adapt video resources to variable network conditions and terminal capabilities [1].

Although scalable video can adapt to heterogeneous application scenarios, it still remains a problem to improve the efficiency and robustness on bandwidth-constrained transmission. Scalable video usually contains video layers with different contributes to the video quality. Considering the unequal importance of scalable video layers, the unequal error protection (UEP) scheme is naturally adopted. One kind of UEP scheme is efficiently distributing FEC codes such as RS codes to scalable video layers [2, 3]. The importance of different layers is measured. For example, in [4], unequal amounts of protection were allocated by jointly considering the temporal dependency of frames in a GOP and the quality dependency in each frame. In [5], a performance metric namely layer-weighted expected zone of error propagation (LW-EZEP) was used for quantifying the error propagation effect from packet loss in scalable video layers. An adaptive and unequal cross-layer FEC mechanism was applied for scalable video transmission over WLANs in [6] according to network channel conditions and the importance of different scalable video layers. By applying UEP on each GOP, scalable video transmission can be more effective over packet-lossy networks.

Generally both the packet size and the number of packets for each GOP are fixed while the source data sizes of GOPs differ, sometimes a lot. The data size fluctuation leads to unequal error protection for GOPs which is not reasonable and appropriate. In this paper, we take the data size fluctuation into consideration. The proposed method mainly aims to achieve similar error protection for different GOPs.

2 UEP Scheme for Scalable Video

Scalable video coding extension of the H.264/AVC standard [1] is the most recent standard for SVC. Temporal, spatial and quality scalability can be supported by a single bit stream. In this paper, we consider the case of combined temporal and quality scalability based on SVC extension of the H.264/AVC.

Temporal scalability is supported by hierarchical prediction structures. In the temporal layer sets $\{T_0,...,T_k\}$, the temporal base layer pictures which form the set T_0 are coded as I or P-pictures, while the temporal enhancement layer pictures which form the set T_k ($k > 0$) are coded as hierarchical B or P-pictures. According to the prediction relationship among all pictures, the temporal base layer can be decoded independently while the temporal enhancement layer can only be decoded on the basis of that all layers with a less temporal layer identifier. Similarly, quality layers consist of a base layer and one or more enhancement layers and quality enhancement layers can improve the quality gradually. Therefore different temporal and quality layers have unequal importance. Obviously the base layer is more important than all enhancement layers and the enhancement layers with less identifier are more important than the ones with larger identifier, respectively among temporal or quality scalable layers. Thus, it is necessary to apply the UEP scheme to scalable video transmission over packet-lossy networks.

Generally, the UEP scheme is applied on a GOP unit. FEC is utilized for every scalable video layer of different importance. A GOP is encoded into some temporal and quality scalable video layers called as scalable units (SU), as shown in Fig. 1. The number of maximum temporal level is T and the number of maximum quality level is Q. The video data with temporal layer identifier i and quality layer identifier j is defined as $SU(i, j)$. The number of source data bytes for $SU(i, j)$ is $R(i, j)$. M is the packet size and N represents the number of packets respectively. The length of FEC codes for $SU(i, j)$ is denoted as $k_{i,j}$ and consequently $N-k_{i,j}$ is the length of $SU(i, j)$. Therefore, the height of each $SU(i, j)$ denoted as $h_{i,j}$ can be calculated as $h(i, j) = \lceil R_{i,j} / (N-k_{i,j}) \rceil$. The problem is to find the best FEC codes assignment for each $SU(i, j)$ with the bandwidth limitation from ($M \times N$). Ha proposed a layer-weighted UEP scheme by using a performance metric namely layer-weighted expected zone of error propagation (LW-EZEP) [5] which can quantify the error propagation effect on decoded video quality degradation from packet loss in scalable video layers. Thus, FEC codes can be efficiently assigned aiming to minimize the overall distortion of reconstructed video.

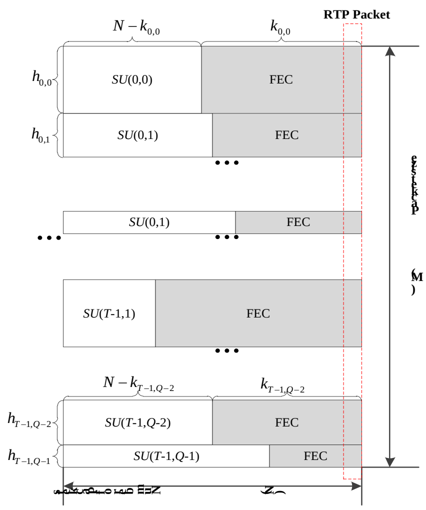

Fig. 1 UEP scheme for SVC

3 Dynamic Transmission Method for Scalable Video

In the UEP scheme proposed in [5], both the packet size and the number of packets for each GOP are fixed with the constrained bandwidth. Therefore, given a video sequence, the FEC code rate is limited. The average FEC code rate R can be calculated as

$$R = R_b / B \tag{1}$$

where R_b represents the bit rate of coded video data and B represents the bandwidth. According to (1), the limited average code rate also means the constrained bandwidth. Therefore, given a target code rate R and a fixed packet size M, the average number of packets N for a GOP can be calculated as

$$N = \left\lfloor \frac{R_b \cdot G}{R \cdot R_f \cdot M} \right\rfloor \tag{1}$$

where G represents the GOP size and R_f represents the frame rate of video sequence.

According to the JSVM (Joint Scalable Video Model) software for SVC, rate control can only be applied on the base layer coding. The numbers of source bytes of GOPs are different from each other. With both the packet size and the number of packets fixed, the number of total bytes of FEC coded GOP is constant. Thus, the data size fluctuation leads to unequal FEC code rate for each GOP. To be specific, the GOPs with more source bytes obtain less FEC codes, hence having weaker error resilience capability.

Considering the drawback of conventional UEP scheme analyzed above, we propose a dynamic transmission method for scalable video. The number of packets of each GOP can be dynamically adjusted to achieve close FEC code rates for every GOP. Therefore, different GOPs obtain similar error recovery capabilities, which means a fair distribution of channel resources between GOPs.

The constant FEC code rate means that the GOPs with more source bytes will generate more packets for transmission, thus costing more time. When the time delay accumulates, it may become unacceptable in a continuous transmission. Thus it is improper to transmit video data without delay limitation in video services such as video conversation. Therefore, in this paper, the buffer management is considered to limit the number of coded packets. Given a fixed packet size M and a target code rate R, the average number of packets for a GOP is defined as N and can be derived from the encoding parameter GOP size, the average bit rate and the frame rate of the coded bit stream according to (2). For convenience, one packet size M is defined as the size of a buffer unit and applying FEC on a GOP is called as a UEP operation in this paper.

The circular buffer is considered and the buffer management mechanism is illustrated in Fig. 2. It works as follows:

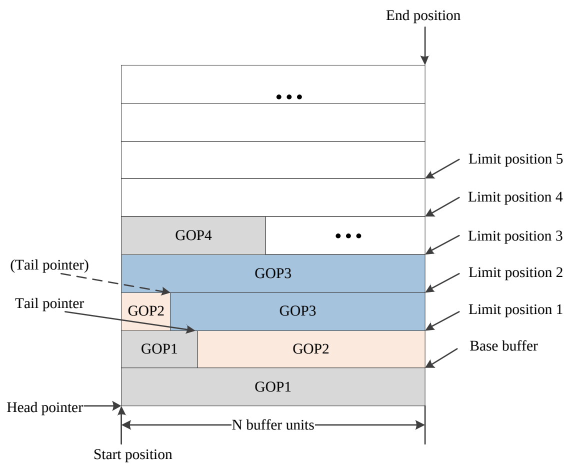

Fig. 2 Buffer management mechanism

(1) The circular buffer is initialized. The buffer size is indicated by the start position and the end position of the buffer. A part of buffer from the start position is assigned as the base buffer. In Fig. 2, the base buffer has N buffer units. At the start, both the head pointer and the tail pointer are at the start position, and the limit position is in the end of the base buffer.

(2) Before each UEP operation, the limit position adds by N buffer units.

(3) Take UEP operations. The maximum size of coded data of each GOP is limited to the limit position. Thus, the number of packets for the kth GOP can be chosen as

$$N_k = \begin{cases} \lfloor S_k/(R \cdot M) \rfloor, & \text{if } \lfloor S_k/(R \cdot M) \rfloor \le (PL_k - PT_k)/M \\ (PL_k - PT_k)/M, & \text{otherwise} \end{cases} \tag{1}$$

where S_k is the number of source bytes of the kth GOP, while PL_k and PT_k represent the limit position and the tail pointer position for the kth GOP.

(4) After a UEP operation finished, the tail pointer shifts to the corresponding bound of the coded data.

(5) The head pointer shifts as the coded packets have been transmitted at a constant rate. The tail pointer and the limit position should not exceed the head pointer.

As shown in Fig. 2, the maximum size of coded data for the 1st GOP is limited to limitation position 1. Some packets, the number of which is between N and 2N, are generated after the UEP operation, thus filling the base buffer without exceeding the limit position 1. The tail pointer shifts to the corresponding position. Then the UEP operation for the 2nd GOP is taken. Similarly, the maximum size of the 2nd GOP is limited to limit position 2 and the tail pointer shifts to the position indicated as the dashed arrow. When taking the UEP operation for the 3rd GOP, the target code rate cannot be achieved because of the limit position 3, so the coded data fill the buffer to limit position 3. Hence, the code rate for 3rd GOP is lower than the target code rate. According to the buffer management mechanism analyzed above and also as shown in Fig. 2, the transmission delay of every GOP is restricted in $2N$ packets transmission time.

4 Simulation Results

JSVM9.19.12 [7] is used to test the proposed dynamic transmission method for scalable video. The combined temporal and quality scalability is applied. Four types of QCIF video sequence 'City, 'Crew', 'Foreman', 'Soccer' are coded to scalable bit stream. The GOP size is set to 8 so that the number of maximum temporal level is 4. Every frame is coded to a quality base layer and three quality enhancement layers. That means T = 4 and Q = 4 respectively.

The packet loss channel is simulated with the two-state Markov model [8]. Different packet loss rates (PLR) and average burst lengths of packet losses are chosen to simulate different channel condition as shown in Table 1. The packet size M is set as 200, and Table 2 shows and the number of average packets N for tested video sequences with the given code rate. In addition, the base buffer is assigned as zero for all tests so that there is no extra delay on the sender side when the proposed dynamic transmission method is adopted.

First, the proposed method is compared to the conventional UEP scheme in terms of the FEC code rate for each GOP. Fig. 3 shows the comparison result of the 'Foreman' sequence where UEP-Ha represents the conventional UEP method proposed in [5]. The FEC code rate ranges from 0.32 to 0.94 with fixed number of

Table 1 Channel condition

Channel condition index	PLR	Average burst length
1	1%	2
2	5%	3
3	10%	4
4	15%	4.5
5	20%	5

Table 2 Simulation setup parameters of UEP

Sequence	Number of packets (N)		
	Code rate 0.60	Code rate 0.65	Code rate 0.70
City	134	124	115
Crew	200	184	171
Foreman	173	159	148
Soccer	217	200	186

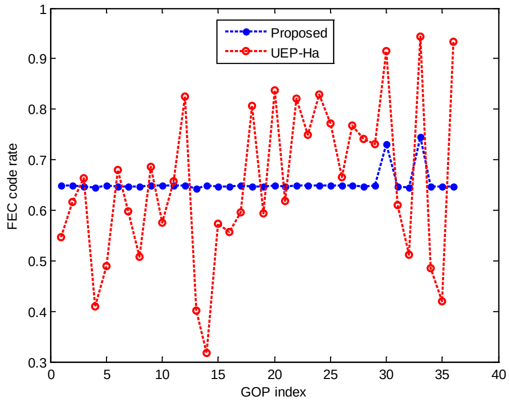

Fig. 3 FEC code rate as a function of GOP index of 'Foreman' when the average code rate is 0.65

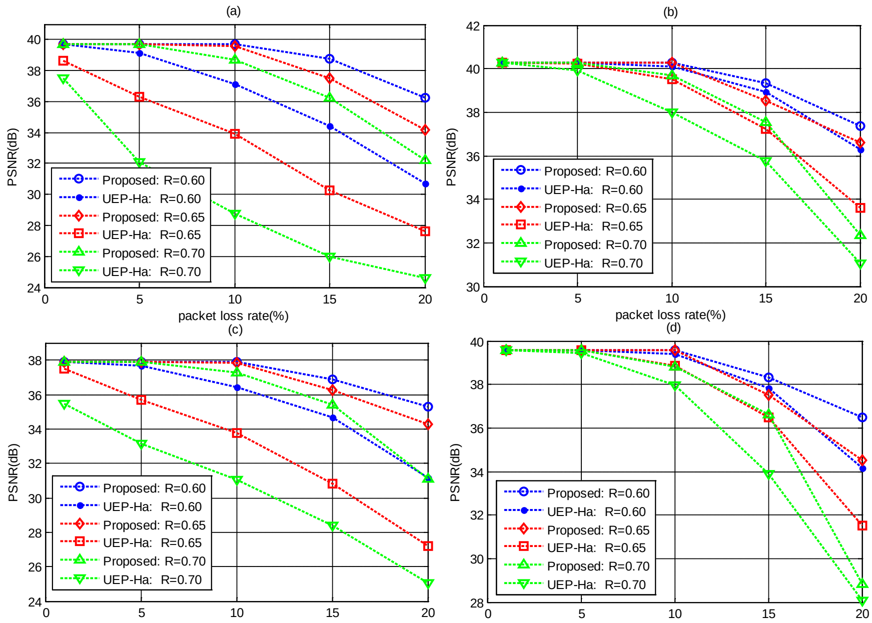

Fig. 4 Average PSNR performance on four types of video sequence: (a) City (b) Crew (c) Foreman (d) Soccer

packets, while the proposed method leads to a stable code rate ranging from 0.64 to 0.65 except the 30th and 33rd GOP. The results of other situations are almost similar. Next we compare the average PSNR performance of the proposed method and the conventional UEP scheme. Each video sequence over different channel condition and at different average channel code rate was tested for 100 times. The simulation results are shown in Fig. 4. The proposed method achieves a significant improvement of 2.54dB in average PSNR in comparison with the conventional UEP scheme in a wide range of packet loss rate and channel code rate.

5 Conclusions

In this paper, we propose a dynamic transmission method for scalable video based on UEP. By adjusting the number of packets for each coded GOP, the proposed method ensures all GOPs of a video sequence have similar error recovery capabilities. Meanwhile, the transmission delay on the sender side is strictly limited by the buffer management so that the method can be used in some video applications. The proposed method was tested in a wide range of channel coding rates and packet loss rates for different types of video sequence. The simulation results show a significant improvement of average PSNR compared with the conventional UEP scheme.

Acknowledgments This work was supported by Tsinghua University Initiative Scientific Research Program 2010THZ0 and National Basic Research Program of China (973 Program) 2012CB316000.

References

1. Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the scalable video coding extension of the H. 264/AVC standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(9), 1103-1120.
2. Maani, E., & Katsaggelos, A. K. (2010). Unequal error protection for robust streaming of scalable video over packet lossy networks. *Circuits and Systems for Video Technology, IEEE Transactions on*, 20(3), 407-416..
3. Naghdinezhad, A., Hashemi, M. R., & Fatemi, O. (2007, June). A novel adaptive unequal error protection method for scalable video over wireless networks. In *Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on* (pp. 1-6). IEEE.
4. Fang, T., & Chau, L. P. (2006). GOP-based channel rate allocation using genetic algorithm for scalable video streaming over error-prone networks. *Image Processing, IEEE Transactions on*, 15(6), 1323-1330..
5. Ha, H., & Yim, C. (2008). Layer-weighted unequal error protection for scalable video coding extension of H. 264/AVC. *Consumer Electronics, IEEE Transactions on*, 54(2), 736-744..
6. In, H. P., & Kang, S. S. (2010). An Adaptive and Unequal Cross-layer Forward Error Correction Mechanism for Scalable Video Transmission over WLANs. In *2010 Second International Conference on Communication Systems, Networks and Applications (ICCSNA 2010)*.
7. Schwarz, H., Hinz, T., & Suehring, K. (2011, March). JSVM (Joint Scalable Video Model) software manual 9.19.12.
8. Majumda, A., Sachs, D. G., Kozintsev, I. V., Ramchandran, K., & Yeung, M. M. (2002). Multicast and unicast real-time video streaming over wireless LANs. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(6), 524-534.