Dynamic FEC-Distortion Optimization for H.264 Scalable Video Streaming

Wei-Chung Wen and Hsu-Feng Hsiao
Dept. of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
{wcwen, hillhsiao}@cs.nctu.edu.tw

Jen-Yu Yu
Information and Comm. Research Labs
Industrial Technology Research Institute
Hsinchu, Taiwan
KevinYu@itri.org.tw

Abstract—**Forward error correction codes have been shown to be a feasible solution either in application layer or in link layer to fulfill the need of Quality of Service for multimedia streaming over the fluctuant channels. In this paper, we propose FEC-distortion optimization algorithms to efficiently utilize the bandwidth for better video quality. The optimization criterions are based on the unequal error protection by taking account of the error drifting problems from both temporal motion compensation and inter-layer prediction of H.264/MPEG-4 AVC scalable video coding. Also, it can adapt to the content-dependent quality contribution of each video frame in a video layer. Lightweight error-concealment is also incorporated with the proposed algorithms for better H.264 SVC streaming. For some applications where either computation might be the bottleneck or the upper bound of non-decodable probability of each video layer is specified, alternative bandwidth allocation algorithm is provided with the trade-off of slight quality degradation.**

Keywords—FEC optimization;H.264;scalable video coding;unequal error protection

Topic area—multimedia communication.

I. INTRODUCTION

Personal, home, or handheld entertainment systems, such as DVB-H [1] and IPTV which is under construction to be a standard by ITU-T, have been an emerging research and industrial emphasis due to the great progress of the network communications and joint multimedia/channel coding technologies. It is rather challenging to fulfill the needs for Quality of Service and Quality of Experience requirements in the mobile environments of such entertainment systems that might suffer from dynamic channel fluctuation.

Besides Automatic Repeat reQuest (ARQ) which possibly suffers from the intolerable end-to-end packet delay and exacerbated jitter, forward error correction codes have been shown to be a feasible solution. In DVB-H, Multi-Protocol Encapsulated Forward Error Correction (MPE-FEC) is used by interleaving the information packets and the protection packets from Reed-Solomon code to deal with the burst error. The error protection strength in MPE-FEC is not really content-dependent. Besides Reed-Solomon code, rateless erasure codes (also known as fountain code [2]), such as raptor code [3], provide virtually infinite protection symbols and the modified version of such code has been recently adopted in 3GPP [4]. However, unlike Reed-Solomon error erasure code which shows maximum distance separable property, fountain codes generally have less coding efficiency.

In [5], Tan et al. proposed layered FEC for sub-band coded scalable video multicast using equation-based rate control while adaptive FEC is adopted to recover the lost packets so that the distortion function can be minimized with the optimized subscription of video and FEC layers, under an assumption that different frames in a video layer shall have the same distortion.

In [6], an adaptive FEC scheme as part of the reliable layered multimedia streaming over either unicast or multicast was proposed. The main objective of the FEC scheme is to maximize the streaming throughput while maintaining an upper bound of the error rate for each scalable video layer that FEC fails to decode. However, the upper bounds are preset without further explanation.

The impact of packet loss and FEC overhead on scalable bit-plane coded video in best-effort networks is analyzed in [7] and similar optimization algorithm was proposed to allocate the bandwidth resource to FEC and video data, respectively.

In this paper, we propose *FEC-Distortion* optimization algorithms that take account of the error drifting problems from both temporal motion compensation and inter-layer prediction of H.264/MPEG-4 AVC scalable video coding, as well as the content-dependent visual quality contribution of each video frame in a video layer to achieve better quality of service with the same resource. In case of occasional packet error that is not recoverable by the FEC scheme, lightweight error-concealment is also incorporated with the proposed algorithms for better quality of reconstructed video.

The rest of this paper is organized as follows. In Section II we modify the FEC optimization algorithm in [5] to be used with H.264 scalable video coding in a non-FEC-layer fashion. We present the dynamic FEC-distortion optimization algorithm in Section III and discuss the error-bounded optimization algorithm in Section IV, followed by the simulation results and concluding remarks in Section V and Section VI, respectively.

II. FLAT FEC-DISTORTION OPTIMIZATION

In [5], Tan et al. proposed layered FEC algorithm for sub-band coded scalable video multicast using equation-based rate control such that packet loss is one of the parameters to regulate the sending rate while adaptive FEC is adopted to recover the lost packets so that the distortion can be minimized with optimized subscription S^* as described in (1) and (2), under an assumption that different frames in a video layer shall have the same distortion measure.

$$S^* = \underset{s \in M, R(s) \le B}{\arg\min} D(s,p), \tag{1}$$

$$D(s,p) \approx \sum_{i=0}^{L-1} p_i \cdot D_i, \tag{2}$$

where M is a set of possible subscriptions of video and FEC layers that fit into the available bandwidth B. p is the average packet loss rate. $D(s, p)$ is the distortion function while p_i is the decodable probability of only the accumulated i video layers, D_i is the associated distortion, and L is the total encoded video layers. The packet losses are assumed to be independent and identically distributed across all the packets and the relationship between p_i and p of this Bernoulli error model is shown below.

$$p_i = \begin{cases} q_{i+1} \prod_{k=1}^{i} (1 - q_k) & ,0 \le i < L \\ \prod_{k=1}^{L} (1 - q_k) & ,i = L \end{cases} \tag{3}$$

$$q_i = p \sum_{w=0}^{K-1} \binom{M_i + K - 1}{w} (1-p)^w p^{M+K-w-1} \tag{4}$$

where q_i in (3) stands for the probability that layer i can not be recovered. M_i is the number of protection symbols in a FEC session for layer i.

Instead of the sub-band scalable video coding with layered structure on both video and FEC data in [5], our proposed FEC optimization algorithms are based on the H.264/MPEG-4 AVC scalable extension, which is an amendment to the H.264/MPEG-4 AVC standard and it is scheduled to be finalized in 2007. The base layer of a Scalable Video Coding (SVC) bit-stream is usually coded in compliance with H.264 while new scalable tools are added for supporting spatial, SNR, and temporal scalability [8]. For each Group of Pictures (GOP) of a scalable video layer, we apply Reed-Solomon erasure code [9] to form an *(n, k)* code which has *k* symbols of the video layer data and the amount of *n-k* protection symbols. It will take a few FEC coding sessions if the data rate of a video layer in the same GOP is high. Sequence Parameter Set Network Abstraction Layer (NAL) units and Picture Parameter Set NAL units [10] have essential header information in order to decode the video properly and they are assigned strongest error correction code $(n=256)$, as compared to the other NAL data units. We modify (1) and (2) to accommodate H.264 SVC and define PSNR function *PSNR(s, p)* to be maximized as shown in (5) and (6).

$$S^* = \underset{s \in M, R(s) \le B}{\arg\max} PSNR(s,p), \tag{5}$$

$$PSNR(s,p) = \sum_{i=0}^{L-1} p_i \cdot PSNR_{ave,i}, \tag{6}$$

where p_i is the decodable probability of the error erasure codes for only the accumulated i video layers and $PSNR_{ave,i}$ is the corresponding average PSNR, respectively. Each subscription *s* from the subscription set M means a vector $(n_1, n_2, ..., n_N)$. N is the number of the transmitted video layers $(N \le L)$ and each vector element n_i means the output symbol number of the FEC erasure code for the i^{th} video layer as the *n* in the *(n, k)* code. In addition, if the packet loss distribution is modeled by the Gilbert/Elliot's 2-state Markov chain [11], which is usually adopted to describe fading channel, the relationship between p_i and p in [6] is used in (6). The modified optimization algorithm is designated as the Flat FEC-Distortion Optimization (FFDO) algorithm.

III. DYNAMIC FEC-DISTORTION OPTIMIZATION

FFDO is based on the assumption that different frames in the same video layer exhibit constant distortion. However, this is usually not the case for the real H.264 SVC videos. The distortion (or PSNR) depends on the content of each video frame as well as the quantization parameter and mode decision used in each block. Due to the error propagation effect resulting from not only the prediction coding across the video layers but also the temporal motion compensation coding in each individual video layer, the distortion caused by different frame of a video layer can also vary. As a result, the global optimal bit allocation of H.264 SVC and FEC shall be found over all the possible bit allocation and packet loss combinations.

We further propose the Dynamic FEC-Distortion Optimization (DFDO) algorithm to perform the optimization not only across the video layers but also within each video layer. Since the PSNR variation of different pictures within a video layer is smaller than that across the video layers, the DFDO algorithm first uses FFDO to decide the number of video layers N and also the total amount of protection packets per GOP for each video layer to subscribe. Then the algorithm finds the distribution pattern S_n^* of those protection packets among all the FEC sessions in each video layer n $(1 \le n \le N)$ to remove the constant distortion assumption within the same video layer. The criterion of this search within a video layer is based on the FEC-distortion optimization of the video layer as shown in (7) and (8).

$$S_n^* = \underset{s \in m}{\arg\max}\, psnr_n(s,p), \tag{7}$$

$$psnr_n(s,p) = \sum_{i=1}^{\text{GOP size}} (p^1{}_i \cdot psnr^1{}_{n,i} + p^2{}_i \cdot psnr^2{}_{n,i} + p^3{}_i \cdot psnr^3{}_{n,i}), \tag{8}$$

where $psnr_n(s,p)$ is defined as the PSNR summation of the accumulated video layers up to layer n among all the video frames in each GOP. m is the set of all the possible FEC distribution patterns over all the FEC sessions in the n^{th} video layer. $p^1{}_i$ is the decodable probability of i^{th} picture of that GOP and all its reference pictures in the current video layer. $psnr^1{}_{n,i}$ is the PSNR of picture i in the same layer. $p^2{}_i$ is the non-decodable probability of i^{th} picture regardless the successful decoding of its reference pictures in the current video layer. In this case, the implemented error concealment method is to reuse the reconstructed i^{th} picture of the $n\text{-}1^{th}$ layer. $psnr^2{}_{n,i}$ is the PSNR of picture i in the same layer and thus it is equal to $psnr^2{}_{n-1,i}$. $p^3{}_i$ is the decodable probability of i^{th} picture but not all of its reference pictures in the current video layer. $psnr^3{}_{n,i}$ is the PSNR of picture i based on the residual video frame and the reconstructed reference pictures with the same error concealment technique mentioned above.

The three probabilities (p^1_i, p^2_i, p^3_i) can be determined by (9), (10), and (11), respectively.

$$p^1_i = \prod_{r \in R_A} (1 - q_r), \tag{9}$$

$$p^2_i = 1 - \prod_{r \in R_B} (1 - q_r), \tag{10}$$

$$p^3_i = 1 - p^1_i - p^2_i, \tag{11}$$

where R_A is the set of all the FEC sessions involved for picture i in the current video layer and all the reference pictures in the same layer. R_B is the set of all the FEC sessions involved for picture i in the current video layer. r is for each and every FEC session in the set. q_r is the decodable probability of FEC session r.

If the packet losses are assumed to be independent and identically distributed with packet loss rate p across all the packets, q_r of FEC session (n, k) is shown in (12).

$$q_r = \sum_{i=0}^{n-1} \binom{n}{i} (1-p)^i p^{n-i}. \tag{12}$$

On the other hand, if the packet loss distribution is modeled by Gilbert/Elliot's 2-state Markov chain, the decodable probability of an FEC session can be found in [6].

IV. ERROR-BOUNDED FEC ALLOCATION

Both the flat and dynamic FEC-Distortion optimization algorithms compare all the possible video layers and the FEC allocation combinations for each GOP, which might require considerable computation effort. In [6], an adaptive FEC scheme for reliable layered multimedia streaming was proposed. The main objective of the FEC scheme is to maximize the streaming throughput while maintaining an upper bound of the error rate of each video layer that FEC cannot decode. Inspired by this concept, we determine the upper bounds of the non-decodable probability of an FEC session for each layer and use those upper bounds to calculate the protection strength for each video layer from the base layer of H.264 SVC to each enhancement layer until all the available bandwidth is consumed as mentioned in [6]. If there is unused bandwidth after all the video layers are included and they all satisfy the upper bounds of the decoding error probability, we distribute the remaining bandwidth as additional error protection equally among all the layers.

To derive the upper bound ε_i for the i^{th} layer, we draw the Q-PSNR graph for the i^{th} video layer, where Q stands for the non-decodable probability of an FEC session of this video layer while all the lower video layers can be decoded . The upper bound ε_i is defined as the Q value with the steepest slope on the Q-PSNR graph within the range of first 20% of the higher PSNR values, so that the upper bound will have reasonable high video quality and it introduces most PSNR increase by the same Q decrease.

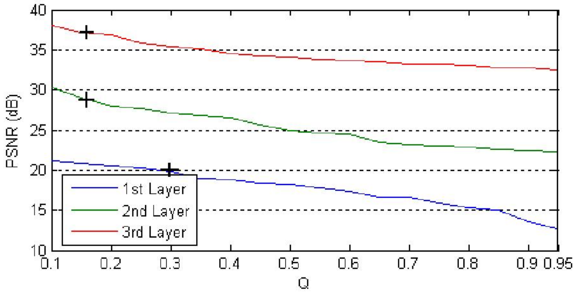

Figure 1. Q-PSNR graph.

Fig. 1 is the Q-PSNR with three H.264 SVC layers on the video sequence mobile at 4CIF resolution. In the case of base layer, the range of the 20% of the higher PSNR values is from 19.45 to 21.14 dB, which is roughly corresponding to Q values from 0.1 to 0.3. Within this range, the steepest slop occurs at Q = 0.3 and it is selected as the upper bound ε_1 for the base layer. Similarly, ε_2 and ε_3 can be found in the same way.

V. SIMULATIONS

We perform simulations for both the flat and dynamic FEC-distortion optimization algorithms (noted as FFDO and DFDO, respectively) as well as the error-bounded allocation algorithm (Error-Bounded). As a comparison, we also show the PSNR performance of equally-distributed FEC scheme (Uniform Distribution) among all the video layers.

The video sequence is mobile at 30 fps and the video resolution is 4CIF. The H.264 SVC encoder and decoder are based on the Joint Scalable Video Model (JSVM) reference software and the error concealment technique described in Section III is applied to all the algorithms. For the spatial scalability, the PSNR is calculated after the picture is up-sampled back to its raw video resolution (4CIF in this case). Some of the encoding parameters for each scalable video layer are listed in Table I and the GOP size is 16. The available bandwidth over time is shown in Fig. 2.

TABLE I. ENCODING SETTINGS FOR THE VIDEO

Layer	Resolution	QP	Bitrate kbps
1	QCIF	30	298.80
2	CIF	34	872.49
3	4CIF	26	3610.09

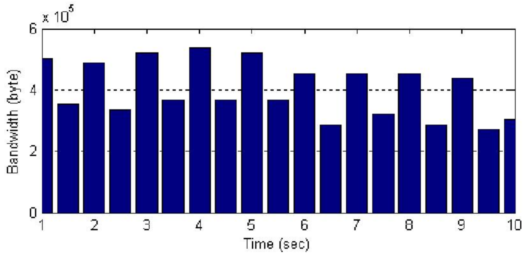

Figure 2. Available bandwidth over time.

First we consider the packet losses to be independent and identically distributed with packet loss rate p=0.25 across all the packets. The primitive results in terms of the average PSNR of four algorithms are shown in Table II. It clearly shows the importance of the unequal error protection provided in DFDO, FFDO, and Error-Bounded, when compared to the equal error protection scheme. The DFDO is always better than the FFDO even though the PSNR increase is not significant. This can be due to the error concealment technique which eliminates some of the distortion caused by the error propagation. Fig. 3 shows the PSNR performance.

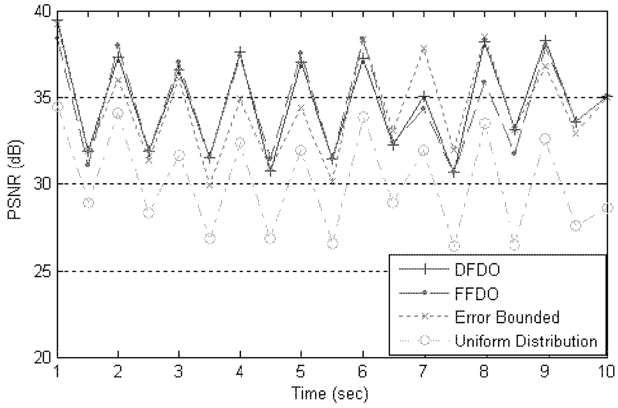

Figure 3. PSNR performance comparison.

TABLE II. THE AVERAGE PSNR

Algorithm	Average PSNR (dB)
DFDO	34.70
FFDO	34.53
Error-Bounded	34.36
Uniform Distribution	30.15

Secondly, we use Gilbert/Elliot's 2-state Markov chain to model the packet loss behavior in a fading channel and the transition probabilities are chosen so that the average packet lost probability is also 0.25 and the available bandwidth profile is kept the same. The corresponding results in terms of average PSNR for four algorithms are shown in Table III.

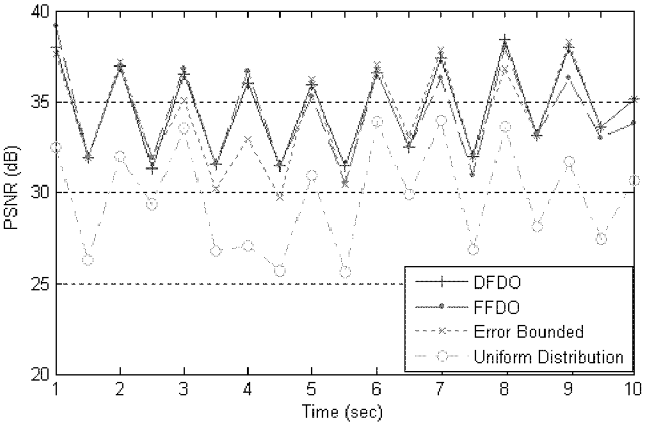

Figure 4. PSNR performance comparison.

TABLE III.THE AVERAGE PSNR

Algorithm	Average PSNR (dB)
DFDO	34.65
FFDO	34.41
Error-Bounded	34.23
Uniform Distribution	29.78

The available bandwidth over time is shown in Fig. 4. The simulation results are very similar to those in Fig. 3. It confirms that if we distinguish the distortion difference with greater details, we can perform the unequal error protection better.

VI. CONCLUSIONS

In this paper, the *FEC-Distortion* optimization algorithms are proposed. The algorithms take account of the error drifting problems from both temporal motion compensation and inter-layer prediction of H.264/MPEG-4 AVC scalable video coding, as well as the content-dependent visual quality contribution of each video frame in each video layer to achieve better quality of service with the same resource. In case of occasional packet error that is not recoverable by the FEC scheme, lightweight error-concealment is also incorporated with the proposed algorithms for better H.264 SVC streaming. For some applications where either computation might be the bottleneck or the upper bound of error probability for each video layer is required, alternative bandwidth allocation algorithm is provided with the trade-off of slight quality degradation.

REFERENCES

[1] ETSI, "Digital Video Broadcasting (DVB): Transmission systems for handheld terminals," *ETSI standard*, EN 302 304 V1.1.1, 2004.

[2] J. Byers, M. Luby, and M. Mitzenmacher, "A Digital Fountain Approach to Asynchronous Reliable Multicast," *IEEE Journal on SelectedAreas in Communications*, 20(8), pp. 1528-1540, October 2002.

[3] M. Luby et al., "Raptor Codes for Reliable Download Delivery in Wireless Broadcast Systems," *IEEE CCNC*, Las Vegas, NV, Jan. 2006.

[4] 3GPP TS 26.346 V6.4.0, "Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs," Mar. 2006.

[5] W.-T. Tan, A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Transactions on Circuits and Systems for Video Technology*, March 2001.

[6] H.-F. Hsiao, A. Chindapol, J. A. Ritcey, and J.-N. Hwang, "Adaptive FEC Scheme for Layered Multimedia Streaming over Wired/Wireless Channels," *Workshop on Multimedia Signal Processing, IEEE*, pp. 1-4, Oct. 2005.

[7] S.-R. Kang and D. Loguinov, "Modeling Best-Effort and FEC Streaming of Scalable Video in Lossy Network Channels," *IEEE/ACM Trans. on Networking*, Feb. 2007.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable H.264/MPEG4-AVC Extension," *International Conference on Image Processing, IEEE*, Oct., 2006.

[9] L. Rizzo, "Effective Erasure Codes for Reliable Computer Communication Protocols", *Computer Communication Review*, 27(2):24-36, April 1997.

[10] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services," version 3, 2005.

[11] E. Elliot, "Estimates of error rates for codes on burst-noise channels," *Bell System Technique Journal*, 1963.