# Towards Robust UHD Video Streaming Systems Using Scalable High Efficiency Video Coding

Eun-Seok Ryu[§], Yeongil Ryu[§], Hyun-Joon Roh[§], Joongheon Kim[†], Bok-Gi Lee[§]

[§]Department of Computer Engineering, Gachon University, Rep. of Korea

[†]Platform Engineering Group, Intel Corporation, Santa Clara, California, USA

Emails: esryu@gachon.ac.kr[§], wlrmwlrm99@gc.gachon.ac.kr[§], ggyo@gc.gachon.ac.kr[§], joonghek@usc.edu[†], bglee@gachon.ac.kr[§]

*Abstract*—With a new video coding standard high efficiency video coding (HEVC), the ultra high definition (UHD) TV service with robust video streaming technology is emerging in the TV industry. This paper addresses the system architecture for the UHD video streaming and proposes two main ideas: (i) picture prioritization method, (ii) error concealment mode signaling (ECMS). In the experiments using HEVC reference model conducted, the proposed picture prioritization method shows the gains in video quality from 2.2 to 7.5 dB in Y-PSNR, and the error concealment mode signaling gains from 0.2 to 2.5 dB in Y-PSNR, with corresponding subjective improvements.

## I. INTRODUCTION

Among the emerging technologies in the TV industry, ultra high definition TV with robust video streaming is one of the most important technologies. Korea's cable TV companies opened the world's first UHD (4K resolution, 3840x2160) channel in April 2014, and Japan started the UHD service around 11 months later. Korea plans to activate UHD market with PyeongChang 2018 Winter Olympics, and Japan also has a plan to open 8K (7680x4320) UHD TV service in time for the Tokyo 2020 Olympic games.

To provide these UHD TV services over error prone networks, robust video streaming technologies including video packet error protection and error concealment (EC) are essential. However, the current video coding standard works of HEVC and SHVC (scalable HEVC) are only focusing on the video compression without careful consideration of video transmission. Besides, the MPEG-H part 1 system standard, MPEG media transport (MMT), that is considering the transmission issue also does not have any syntax and semantics for the picture priority in the same temporal level of hierarchical-B-structure and EC at all. Regarding the EC technology, it is very difficult to find the best EC mode among multiple EC methods provided by decoder without original pictures. This is the limitation of the EC method that only works at a video decoder side. Thus, this paper proposes two main ideas; (i) a new picture prioritization method in the hierarchical B structure of HEVC, and (ii) a new EC mode signaling method that signals best EC mode(s) which is calculated and determined at an encoder side to a decoder.

## II. BACKGROUND

### A. Video Coding Standards: HEVC, SVC, and SHVC

After successfully standardizing H.264/AVC (Advanced Video Coding) [1], ISO/IEC MPEG and ITU-T VCEG have been jointly developing next generation video standard called HEVC. This new standard targets next-generation HDTV displays and IPTV services, addressing the concern of error resilient streaming in HEVC-based IPTV. As shown in Table I, comparing to H.264/AVC, HEVC includes new features such as extended prediction block sizes (up to 64x64), large transform block sizes (up to 32x32), tile and slice picture segmentations for loss resilience and parallelism, sample adaptive offset (SAO), and so on [2].

TABLE I: AVC and HEVC

| AVC | HEVC |
|---|---|
| Picture ¿ Slices ¿ MB (Max 16x16) | Picture ¿ Slices ¿ CTUs (Max 64x64) |
| Macroblock / Block | CU / PU / TU (Coding/Prediction/Transform Unit) |
| - Transform 4x4 / 8x8 | - Transform 4x4   32/32 (DCT+DST for Intra) |
| Intrapicture prediction | Intrapicture prediction |
| - Up to 9 directional modes | - Angular prediction 33 directional+2 modes |
| Variable Block Size | Asymmetric Motion Partitioning (AMP) |
| Motion copy mode | Motion copy mode |
| - Direct mode | - Merge Mode / |
| | Advanced Motion Vector Prediction(AMVP) |
| | - Transmit MVD |
| Deblocking filter only | Deblocking + Sample Adaptive Offset (SAO) |
| CAVLC / CABAC | CABAC |
| Slices (FMO) | Tiles / Wavefront |

H.264 scalable video coding (SVC) is an H.264 AVC scalable extension that combines spatial, temporal, and quality scalabilities simultaneously. That is, the SVC can support multiple resolutions, frame rates, and video qualities within a single bitstream because the bitstream consists of multiple layers.

In the layered structure of SVC encoder. The original high quality video input is spatially down-sampled for multiple layers, and each layer encodes the input video with interlayer prediction. Because of the layered feature, the SVC has several advantages. First, by supporting many clients with a single video content file (bitstream) the SVC enables video service providers to reduce overall network bandwidth (BW) [3], disk storage for video contents, and computational complexity for transcoding. Second, the SVC is applicable to many unequal error protection (UEP) methods using the priorities of each layer [4]. For example, a base layer (BL) can be provided

at a level of error protection higher than that of the other enhancement layers (ELs) because the decoder cannot reconstruct a video sequence without the BL [5] [6], suggesting a higher priority for it. Third, SVC can support the diverse screen sizes and resolutions of user devices, and also diverse network BWs. The SHVC standard is designed to have low complexity for ELs by adding the reconstructed BL picture to the reference picture lists in EL [7] [8]. In addition, SHVC uses multiple loops decoding to make a decoder chipset simple while the SVC uses single loop decoding. SHVC also provides a standard scalability by supporting AVC with BL and HEVC with EL. Thus, UHD TV services that supporting legacy HD TVs as well as simple bitstream level rate control (layer switching) need the SHVC.

Because the bitstream level layer switching and UEP with Raptor codes [9] [10] [11] are explained in our previous work in detail [4], this paper focuses on the picture prioritization and EC to provide robust UHD video streaming over error-prone networks.

### B. Video Streaming System with Picture Prioritization and Error Concealment Methods

In video compression and transmission, picture prioritization is of utmost importance for the role it plays in UEP, picture dropping for bandwidth adaptation, as well as quantization parameter (QP) control for enhanced video quality, to name a few. There have been many studies to prioritize individual video pictures and slices with precision and reliability. Layer information of video packets is widely used. For example, in the encoded bitstream of H.264 SVC, the reconstruction pictures of BL is used to decode the pictures of the ELs, and the video packet of BL must be processed with the highest priority and transmitted with greater reliability or lower packet loss rates. Otherwise, losing a single BL packet could result in severe error propagation in both layers. Fig. 1 shows four different methods of picture prioritization based on picture characteristics. (a) Use picture type information which is related to temporal reference dependency for picture prioritization; (b) Use temporal level information in hierarchical B structure, and higher layer will not be referenced by lower layer; (c) Use location information of slice groups (SGs) (SG-level prioritization); and (d) Use the layer information of the SVC. In most cases, I-pictures, pictures in low temporal level, slice group of region of interest (ROI), and pictures in BL of the SVC have higher priority than the others. Regarding the ROI, the flexible macroblock ordering (FMO) method in H.264 or the tile method in HEVC could be used. Picture prioritization can also be used for QoS handling in video streaming. In Fig. 2, the video encoder (or QoS component in the server) determines the priority of each picture ($P_n$) where $n$ is a picture number. Picture priorities are then used for several QoS purposes: (i) dropping less important pictures in the transmitter or scheduler of the server for bandwidth adaptation; (ii) allocating more important pictures to more stable channels (or antennas) in multi-channel networks or MIMO; (iii) protecting more important pictures with larger overhead of forward error correction (FEC) code in application or physical layers; (iv) scheduling more important pictures first in application or MAC layers; and (v) differentiating services in the media aware network element (MANE), edge server, or home gateway.
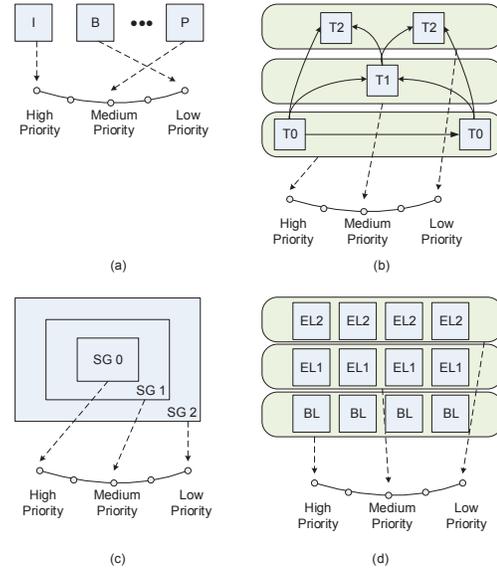


Fig. 1: Examples of picture prioritization methods.
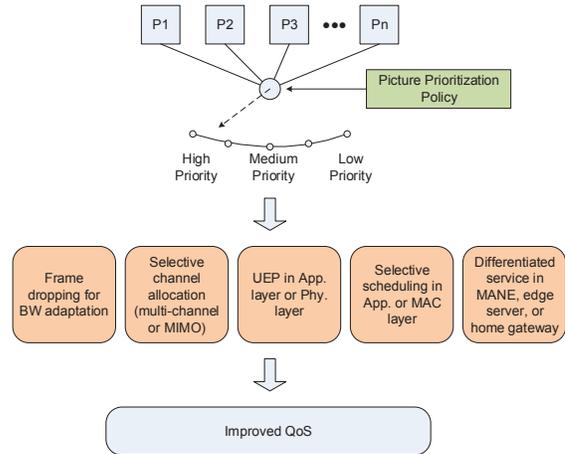


Fig. 2: Examples of QoS handling with picture priority.

Among them, Fig. 3 shows two use cases in detail. Once the encoder has decided the priority of a picture, the UEP and/or transmission scheduler can use the priority in both robust streaming and QoS handling. Fig. 3a applies different FEC overheads to pictures according to picture priority ($PP_n$), and Fig. 3b allocates pictures to different prioritized queues according to picture priority, the high priority queue has higher throughput. Therefore, picture priority is essential for optimal QoS handling in video streaming and communication applications. Other standardization working groups such as MMT and IETF H.264 over RTP consider picture priority at the system level, which can enhance video server (scheduler) and
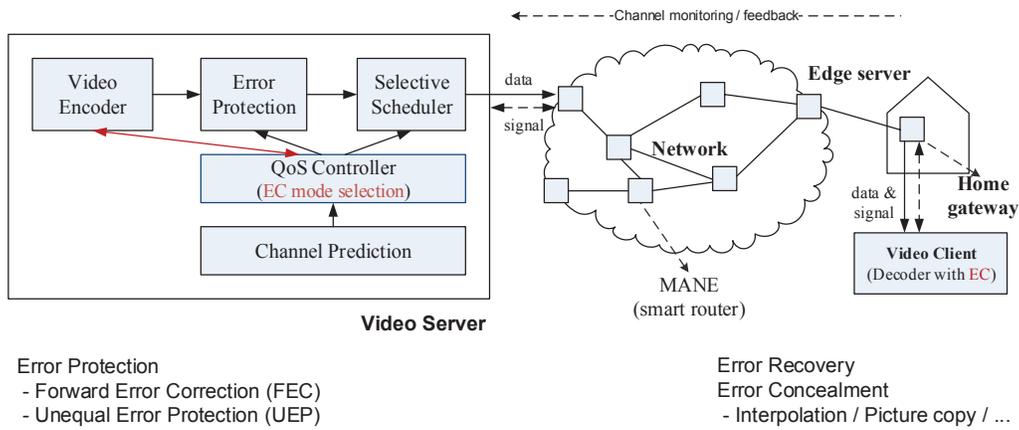
Fig. 4: General architecture of video streaming system.
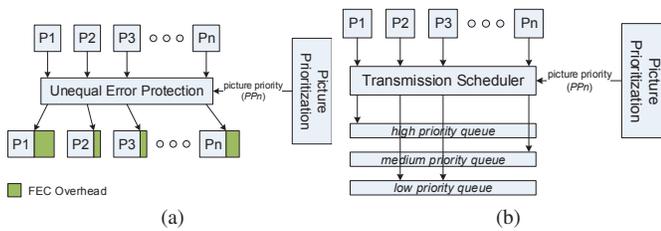


Fig. 3: Two use cases of the picture prioritization method: (a) UEP and (b) transmission scheduler.

MANE (smart router) for QoS improvement by differentiating among packets with various priorities when congestion occurs in networks.

## III. ROBUST 4K UHD VIDEO STREAMING SYSTEM: ERROR RESILIENT VIDEO STREAMING

The general architecture of the video streaming system showing the picture priority can be used effectively in the error protection module, selective scheduler, MANE, edge server, and home gateway as shown in Fig. 4. The video server consists of multiple modules such as video encoder, error protection, selective scheduler, and quality of service (QoS) controller for streaming. The video client includes an EC module. From a network point of view, the video packet could be transmitted over error-prone network. Thus, the transmission has to consider the packet loss condition that could happen in wireless connection by signal interference or by dropping packets for congestion control. The network may use the methods such as automatic repeat request (ARQ) and FEC to recover the packets from the network error, but extra transmission delay and jitter may occur unpredictably. Because of the undesirable delay and jitter, the cross-layer optimization is avoiding to use the retransmission (e.g. ARQ) and error protection (e.g. FEC) in the link and physical layers, instead, the technologies such as video content-aware error protection (e.g. UEP) and EC methods are preferred in the application layer. Consequently, the video server and client need to provide

error resilient streaming methods and EC methods along with flow control and congestion control technologies. In Fig. 4, the server and client exchange control messages (signal) to control the QoS metrics, and the signaling effort may enhance the overall video quality significantly.

### A. Picture Prioritization and Unequal Error Protection

This section explains the picture prioritization and UEP method with our previous researches [12] [13] [14]. Fig. 5 denotes the current uniform prioritization method (applies same priority to pictures in the same temporal level of hierarchical B structure) with four dyadic stages in temporal domain. Although the temporal levels could tell the priority of a picture, the current HEVC standard provides no additional methods for assigning priorities to pictures at the same temporal level. Fig. 5 shows the random access (RA) setting in the common test condition of HEVC, with picture order count (POC) 2 and 6 having a same priority. However, the uniform prioritization at the same temporal level presents a problem when the importance of two (or more) POCs in each group of picture (GOP) varies according to both the reference picture set (RPS) and the size of the reference picture lists (RPL). In order to illustrate the problem, this paper uses two pictures at the same temporal level as an example, and defines $Pos.A$ as pictures with a POC equal to $2 + N \times$GOP, and $Pos.B$ as pictures with a POC equal to $6 + N \times$GOP, where GOP is 8 and $N$ represents the number of GOP(s).
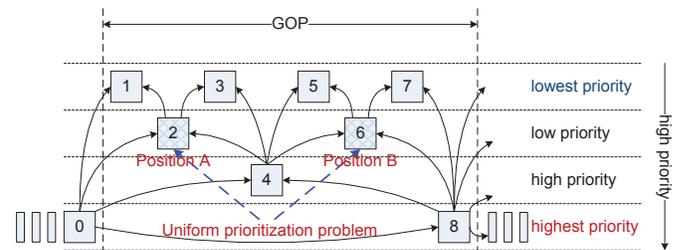


Fig. 5: Current uniform prioritization in hierarchical B pictures.

Fig. 6 shows corresponding rate-distortion (RD) curves, indicating that the picture in $Pos.B$ is more important than the picture in $Pos.A$; red curve is from original HEVC reference software (HM) 6.1 EC, and black curve is from modified HM 6.1 EC. The average BD-rate differences between the blue and black curves were 23.4% ($Kimono$) and 20.2% ($ParkScene$) when test sequences were encoded with the same TID (all TID is 0); error propagation effect. When TIDs from 0 to 3 were used according to their temporal levels, the average differences were 9.9% ($Kimono$) and 10.2% ($ParkScene$) respectively. The PSNR degradation caused by dropping a picture per intraperiod (=32 in this example) in $Pos.A$ was less than the PSNR degradation caused by dropping pictures in $Pos.B$ indicating that pictures even in the same temporal level in hierarchical B pictures should have different priorities in accordance with their prediction information.
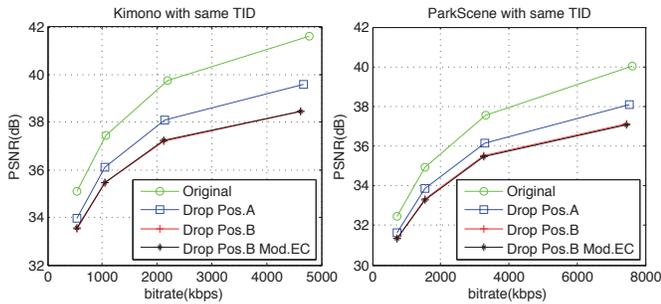


Fig. 6: RD curves for dropped packets in $Pos.A$ and $Pos.B$.

To solve the uniform picture prioritization problem, this paper proposes an implicit picture prioritization method that the encoder assigns priorities to pictures according to the RPS and the size of RPL of the encoding option without any additional delays as shown in Fig. 7a. If a POC number is observed more often in the RPL, then the corresponding picture will earn a higher priority; this is because the number of observations implies the opportunity of being referenced in motion estimation. In case of a POC consists of multiple slices, the priority of the POC is assigned to those slices. The proposed picture prioritization method was combined with the implemented FEC code, Raptor codes, to show its UEP usage and performance gain. Each picture was encoded with a NAL packet and was protected with selected FEC redundancies. For example, when combined UEP, the proposed picture prioritization method protected pictures in $Pos.A$ with 28% FEC redundancies (medium-low priority), and protected pictures in $Pos.B$ with 32% FEC redundancies (medium-high priority), In contrast, when uniform UEP is used, pictures at $Pos.A$ and at $Pos.B$ were both protected with 30% FEC redundancies (medium priority). The other redundancies were as these: highest = 44%, high = 37%, low = 24%. Because the hierarchical B pictures with GOP 8 has 4 temporal levels, pictures in the lowest temporal level (e.g. POC 0 and 8) were protected with the highest priority, picture in temporal level 1 (e.g. POC 4) was protected with high priority, and pictures in the highest temporal level (e.g. POC 1, 3, 5, 7) were protected

with low priority.
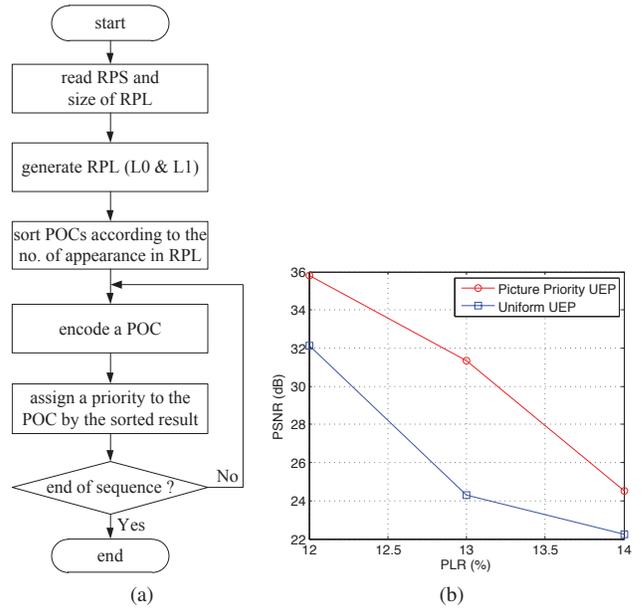Fig. 7b shows the gain of proposed picture prioritization method (from 2.2 dB to 7.5 dB in PSNR).



Fig. 7: (a) Algorithm of the proposed picture prioritization method, (b) PSNR comparison results (sequence: $ParkScene$, QP27).

## B. Error Concealment with Mode Signaling

This section explains the EC mode signaling method with our previous researches [15] [16]. The EC method is important for a scalable video coding transmission system over error-prone network [17]. Fig. 8 shows an example of scalable coding with two layers (BL and EL, following number represents POC), where the picture EL2 in EL is lost. In the example, decoder can copy one of EL0, EL4, or BL2 to conceal the lost EL2 as a simple EC method (picture copy). Because EL2 could be referenced by EL1, EL3, and EL6, losing EL2 can cause error propagation in EL1, EL3, EL5, EL6, and EL7 (marked with red wave). Thus, applying the best EC method can improve not only the quality of the lost picture EL2, but also the quality of the other pictures such as EL1, EL3, EL5, EL6, and EL7 that are affected due to error propagation. The proposed EC mode signaling method works as the diagram in Fig. 9.

Though general video decoder supports some EC methods [18], it is difficult to find the best EC method among the supported EC methods at the decoder side without original pictures. The proposed EC mode signaling method enables the video encoder to (i) simulate various EC methods on a damaged picture, (ii) determine the best EC method that provides minimal disparity between an original image and a reconstructed image, and (iii) signal the best EC mode to the video decoder at the client. Fig. 10 shows the algorithm in detail, and the term $dependent layer$ means higher EL of current layer..
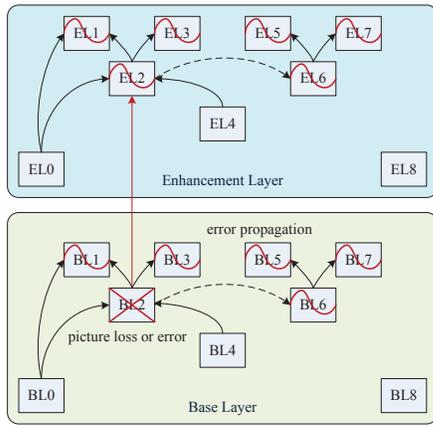
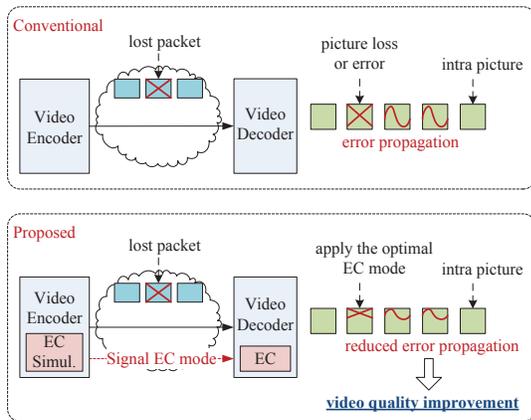Fig. 8: The effect of the error propagation in scalable video coding.



Fig. 9: EC mode signaling to reduce the error propagation (upper: conventional method, lower: proposed method).

At the beginning, the decoder sets EC mode to dealt EC mode 0, and it means the decoder copies previous reference picture if the picture is lost. In assumption, this study does not consider the picture loss of first intra-picture (I picture) of BL. Normally, in video streaming systems, the first intra-picture is guaranteed to transmit by using retransmission and FEC. There are some ways to determine the picture loss in the decoder side. If the whole one picture is lost during transmission, the picture number POC cannot be continued, and the decoder easily knows the lost picture number. If the partial one picture is lost during transmission, the decoder internally faces decoding errors in coding unit level, and its error handler lets the decoder conceal the damage by using the signaled EC mode. If the error handler just fills the damaged coding unit (in H.264, macro-block) with average Y/U/V values of neighboring coding units, severe blocky artifacts can be observed. If there is no EC method in a decoder, the decoder normally crashes with the packet loss.

Regarding the complexity of the proposed EC mode signaling method, it does not increase the computational complexity at encoder side. Comparing the disparities between the original
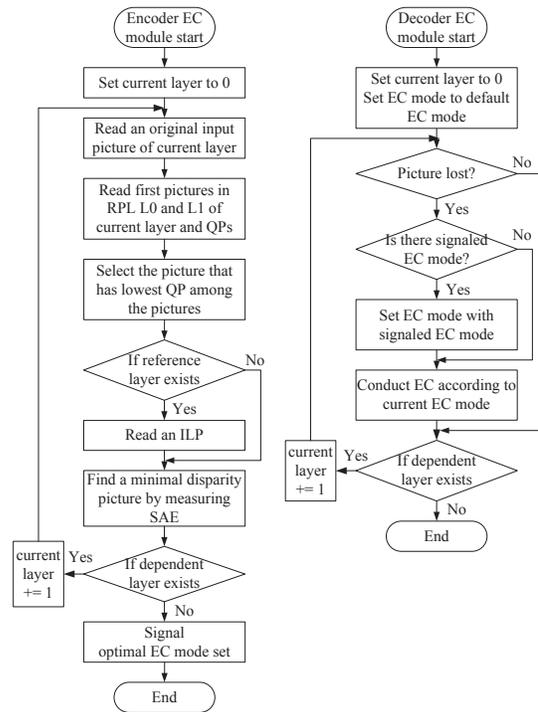


Fig. 10: EC mode signaling algorithms of an encoder (left) and a decoder (right).

picture and the reconstructed reference pictures is already included in the default encoding processes such as a motion estimation. Thus, the proposed method simply compares the disparities only and signals the best EC mode that has minimal disparities. In addition, the bitrate increasing for the EC mode signaling is very little because the method just sends a few bits per one picture. For example, if the encoder and decoder support 4 types of EC methods, the method needs only 2 bits to indicate the best EC mode. To verify the benefit of the proposed method, this study implements the optimal EC mode determination module in SHVC reference software (SHM) encoder and decoder version 2.0 [19]. For the fair performance comparison, this study implements some simple EC methods that use the picture copy method with multiple options as follows.

(a) EC0 (EC mode 0): picture copy from previous reference picture (1st picture in RPL0),

(b) EC1 (EC mode 1): picture copy from next reference picture (1st picture in RPL1),

(c) EC2 (EC mode 2): picture copy from the upsampled BL picture (picture in inter-layer picture buffer),

(d) EC3 (EC mode 3): picture copy from the reference picture that has lower QP (among 1st pictures in RPL0 and RPL1),

(e) EC4 (EC mode 4): (proposed method) signaling best EC mode.

The test bitstream (sequence named $Han$ provided by Vidyo Inc.) consists of two layers (BL and EL) with three types of scalability, and their resolutions of the $spatial\ 2X$ are 1080p and 540p. For $spatial\ 1.5X$, 1080p and 720p sequences are

used, and SNR (quality scalability) uses same 1080p sequences with different QPs.

TABLE II: Average PSNR gain between EC modes for referenced pictures.

| Scalability | QP<br>BL 38 | Avg. Y-PSNR gain (dB) | | |
|---|---|---|---|---|
| | | EC4 - EC0 | EC4 - EC1 | EC4 - EC3 |
| Spatial 2X | EL 32 | 1.83 | 1.98 | 1.79 |
| | EL 33 | 1.77 | 1.91 | 1.74 |
| | EL 34 | 1.77 | 1.74 | 1.72 |
| Spatial 1.5X | EL 32 | 1.91 | 2.1 | 0.2 |
| | EL 33 | 1.91 | 1.89 | 1.87 |
| | EL 34 | 1.86 | 1.89 | 1.82 |
| SNR | EL 26 | 2.38 | 2.5 | 2.34 |
| | EL 28 | 2.29 | 2.45 | 2.26 |
| | EL 30 | 2.22 | 2.33 | 2.18 |

TABLE III: PSNR gain between EC4 and EC2.

| | Scalability | QP | | Y-PSNR gain (dB) |
|---|---|---|---|---|
| | | BL | EL | in GOP (POC 65-72) |
| EC 4 - EC 2 | Spatial 2X | 38 | 32 | 1.03 |
| | | | 33 | 0.83 |
| | | | 34 | 0.81 |
| | Spatial 1.5X | 38 | 32 | 0.38 |
| | | | 33 | 0.35 |
| | | | 34 | 0.27 |
| | SNR | 38 | 26 | 0.37 |
| | | | 28 | 0.34 |
| | | | 30 | 0.2 |



Fig. 11: Reconstructed image comparison with enlarged noticeable sections: (upper) EC mode 2, (lower) EC mode 4 (proposed method).

Table II and Table III show the experimental results. The gains in video quality vary from 0.2dB to 2.5 dB in PSNR. Fig. 11 shows their constructed images (picture number 68) by EC mode 2 and EC mode 4. As shown in the figure, there are several noticeable visual differences; the area of hairs, plants, face, arm, and edge-line of table. Thus, the advantage of the proposed EC mode signaling method is verified.

## IV. CONCLUSION

This paper addresses the system architecture for the UHD video streaming and proposes two main ideas: (i) picture prioritization method, (ii) error concealment mode signaling (ECMS). In the experiments using HEVC reference model conducted, the proposed picture prioritization method shows the gains in video quality from 2.2 to 7.5 dB in Y-PSNR, and the error concealment mode signaling gains from 0.2 to 2.5 dB in Y-PSNR, with corresponding subjective improvements.

## REFERENCES

[1] G. J. Sullivan, P. N. Topiwala, and A. Luthra, "The h.264/avc advanced video coding standard: overview and introduction to the fidelity range extensions," pp. 454–474, 2004.

[2] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[3] M. G. Martini, M. Mazzotti, C. Lamy-Bergot, J. Huusko, and P. Amon, "Content adaptive network aware joint optimization of wireless video transmission," *Communications Magazine, IEEE*, vol. 45, no. 1, pp. 84 –90, Jan. 2007.

[4] E.-S. Ryu and N. Jayant, "Home gateway for three-screen tv using h.264 svc and raptor fec," *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 4, pp. 1652–1660, November 2011.

[5] C. Hellge, T. Schierl, and T. Wiegand, "Receiver driven layered multicast with layer-aware forward error correction," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, oct. 2008, pp. 2304 –2307.

[6] T. Schierl, H. Schwarz, D. Marpe, and T. Wiegand, "Wireless broadcasting using the scalable extension of h. 264/avc," *Multimedia and Expo, IEEE International Conference on*, vol. 0, pp. 884–887, 2005.

[7] Y. Ye, G. W. McClellan, Y. He, X. Xiu, Y. He, J. Dong, C. Bal, and E. Ryu, "Codec architecture for multiple layer video coding," in *U.S. Patent Application 13/937,645*.

[8] Y. Ye and P. Andrivon, "The scalable extensions of hevc for ultra-high-definition video delivery," *MultiMedia, IEEE*, vol. 21, no. 3, pp. 58–64, July 2014.

[9] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[10] M. Luby, "Lt codes," in *Proc. 43rd Annual IEEE Symp. Foundations of Computer Science*, 2002, pp. 271–280.

[11] M. Sayit and G. Seckin, "Scalable video with raptor for wireless multicast networks," in *Proc. Packet Video 2007*, 2007, pp. 336–341.

[12] E.-S. Ryu, Y. Ye, Y. He, and Y. He, "Frame prioritization method based on prediction information," in *JCTVC-J0063 for Stockholm meeting on HEVC standard*, July 2012.

[13] E. Ryu, Y. Ye, Y. He, and Y. He, "Frame prioritization based on prediction information," in *U.S. Patent Application 13/931,362*.

[14] E.-S. Ryu, "Prediction-based picture prioritisation method for hierarchical b-structure of high efficiency video coding," *Electronics Letters*, vol. 49, no. 20, pp. 1268–1270, 2013.

[15] E.-S. Ryu, Y. He, Y. Ye, and Y. He, "On error concealment mode signaling," in *ISO/IEC JTC1/SC29/WG11 MPEG2013/m31189 for Geneva meeting*, October 2013.

[16] E.-S. Ryu and J. Kim, "Error concealment mode signaling for robust mobile video transmission," *AEU-International Journal of Electronics and Communications*, vol. 69, no. 7, pp. 1070–1073, 2015.

[17] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Joint media-channel aware unequal error protection for wireless scalable video streaming," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 2008*, 2008, pp. 1129–1132.

[18] Y. Guo, Y. Chen, Y.-K. Wang, H. Li, M. M. Hannuksela, and M. Gabbouj, "Error resilient coding and error concealment in scalable video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 19, no. 6, pp. 781–795, 2009.

[19] J. Chen, J. Boyce, Y. Ye, and M. M. Hannuksela, "Shvc test model 2 (shm 2)," in *JCTVC-M1007, 13th Joint Collaborative Team on Video Coding (JCT-VC) Meeting, Incheon, South Korea*, 2013.