Scalable Extension of HEVC for Flexible High-Quality Digital Video Content Services

Hahyun Lee, Jung Won Kang, Jinho Lee, Jin Soo Choi, Jinwoong Kim, and Donggyu Sim

This paper describes the scalable extension of High Efficiency Video Coding (HEVC) to provide flexible high-quality digital video content services. The proposed scalable codec is designed on multi-loop decoding architecture to support inter-layer sample prediction and inter-layer motion parameter prediction. Inter-layer sample prediction is enabled by inserting the reconstructed picture of the reference layer (RL) into the decoded picture buffer of the enhancement layer (EL). To reduce the motion parameter redundancies between layers, the motion parameter of the RL is used as one of the candidates in merge mode and motion vector prediction in the EL. The proposed scalable extension can support scalabilities with minimum changes to the HEVC and provide average Bjøntegaard delta bitrate gains of about 24% for spatial scalability and of about 21% for SNR scalability compared to simulcast coding with HEVC.

Keywords: HEVC, H.264/SVC, Scalable Video Coding.

I. Introduction

Recently, a large amount of digital video content has been distributed over the Internet. In contrast to traditional TV broadcasting, the receiving devices are characterized by widely varying properties. With the explosive growth of mobile devices, people regularly use such devices as smartphones, tablets, and notebooks for sharing and browsing video contents. Mobile devices typically have a lower screen resolution as well as lower computing capabilities and battery power. Furthermore, video applications such as YouTube and Hulu use the Internet and mobile networks. These are characterized by adaptive resource sharing, which can bring about unreliable connection qualities.

To provide each user with a video quality that fits the capabilities of its receiving devices and its network connection, multiple coded streams of the same video content with different spatial resolutions and qualities have to be generated. In these video content consuming environments, which are characterized by receiving devices with heterogeneous properties and unreliable connections, Scalable Video Coding (SVC) is an attractive solution since it enables a single bitstream to simultaneously serve various devices with different display resolutions and qualities. During the past several decades, SVC has been an active area of research and standardization. Such video coding standards as H.262|MPEG-2, H.263, and MPEG-4 Part 2 already support most important scalability tools. However, owing to the significant loss in coding efficiency as well as a large increase in decoder complexity, the scalable profiles of these standards have rarely been used. The SVC extension of H.264/AVC includes an approach, referred to as single-loop decoding, to balance the coding efficiency and decoder complexity more than in prior

Manuscript received Mar. 28, 2013; revised Oct. 21, 2013: accepted Nov. 11, 2013.

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

Hahyun Lee (phone: +82 42 860 6138, hanilee@etri.re.kr), Jung Won Kang (jungwon@etri.re.kr), Jinho Lee (jinosoul@etri.re.kr), Jin Soo Choi (jschoi@etri.re.kr), and Jinwoong Kim (jwkim@etri.re.kr) are with the Broadcasting & Telecommunications Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Donggyu Sim (dgsim@kw.ac.kr) is with the Department of Computer Engineering, Kwangwoon University, Seoul, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.13.2013.0040

standards. Although the SVC extension of H.264/AVC is used in some video conferencing applications, this design was basically not successful from an industry adoption perspective. One reason for this is that there are no advantages in terms of coding efficiency compared to the alternatives of simulcast and transcoding, but also additional implementation costs and an increased decoding complexity are incurred.

Nowadays, with the ever increasing demand for higher quality video, ultra high definition (UHD), which provides a four to 16 times higher resolution than HD, is being investigated as a new video content format. Although UHD video is an attractive option, and UHDTV will be launched in the near future, UHD video contents will not completely replace HD contents because of increased data rates and backward compatibility with legacy devices. In particular, backward compatibility with legacy devices can be supported using scalable coding. UHD contents can be encoded as the enhancement layer (EL) of HD contents, so legacy devices capable of decoding HD contents can be used continuously, while new devices can decode both UHD and HD contents and the UHD content can be displayed. Recently, the Joint Collaborative Team on Video Coding (JCT-VC) of both ITU-T SG16 WP3 Video Coding Experts Group (VCEG) and ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) developed the High Efficiency Video Coding (HEVC) standard for targeting UHD and HD video contents. The aim of HEVC was to improve coding efficiency by about two times that of H.264/AVC. The first edition of the HEVC standard was finalized in January 2013. As the next standardization step, JCT-VC issued a Call for Proposals (CfP) on Scalable Video Coding Extensions for HEVC in July 2012.

The proposed scalable extension of HEVC was first described in response to the CfP on the scalable extension of HEVC [1]. The remainder of this paper is organized as follows. In section II, the background on SVC and HEVC is briefly reviewed. The proposed scalable extension of HEVC is described in section III. In section IV, a performance analysis is provided to evaluate the coding efficiency of the proposed scalable extension. Finally, we provide some concluding remarks in section V.

II. Background

In this section, we review the SVC extension of H.264/AVC and briefly describe the basic concepts of HEVC.

1. SVC Extension of H.264/AVC

The latest SVC standard, the SVC extension of H.264/AVC, supports spatial scalability, temporal scalability, and quality scalability. To support spatial scalability, the H.264/AVC scalable extension follows the conventional approach of multi-layer coding [2]. In each spatial layer, motion compensated prediction and intra prediction are employed as in H.264/AVC [3]. However, to improve the coding efficiency, additional inter-layer prediction tools are employed. There are three inter-layer prediction tools, namely, inter-layer intra prediction, inter-layer motion prediction, and inter-layer residual prediction. First, inter-layer intra prediction uses the reconstructed samples of the reference layer (RL) for the prediction signals of the EL. Second, inter-layer motion prediction utilizes the motion information of the RL to reduce the motion redundancy between layers. In this case, the partitioning data, reference indexes, and motion vectors of the EL block are derived from the corresponding block of the RL. Finally, inter-layer residual prediction is used to further reduce the residual data of the EL. Inter-layer intra prediction can only be used for the EL block whose corresponding block of the RL is intra-coded, whereas inter-layer motion prediction and inter-layer residual prediction are utilized for the EL block whose corresponding block of the RL is inter-coded. The temporal scalability is provided by hierarchical prediction structures. For quality scalability, two different possibilities are provided: coarse grain scalability and medium grain scalability [4].

2. High Efficiency Video Coding

Similar to previously established video coding standards, such as H.262|MPEG-2 and H.264/AVC, HEVC has a hybrid video coding structure. As in H.264/AVC, HEVC consists of inter prediction, intra prediction, 2D transformation, entropy coding, and in-loop filters. One fundamental difference between HEVC and H.264/AVC is that HEVC uses a quadtree coding structure [5]. In HEVC, a slice is partitioned into multiple coding tree units (CTUs), which is similar to the concept of a macroblock in H.264/AVC. All CTUs are allowed to have a size between 8×8 and 64×64, and each CTU is subdivided into smaller coding units (CUs) according to a quadtree structure. A CU is either intra-coded or inter-coded and is subdivided into one, two, or four prediction units (PU) according to the PU splitting type. HEVC defines two splitting types for intra-coded CUs and eight splitting types for inter-coded CUs [6]. When the PU is coded in intra mode, the previously decoded samples of the adjacent blocks are used to predict the current PU samples. For the intra prediction, HEVC supports 33 directional, planar, and DC prediction modes [7]. When the PU is coded in inter mode, the samples of the previously decoded pictures stored in the decoded picture buffer (DPB) are used to predict the current PU samples. A quarter-sample precision is used for the motion vectors, and

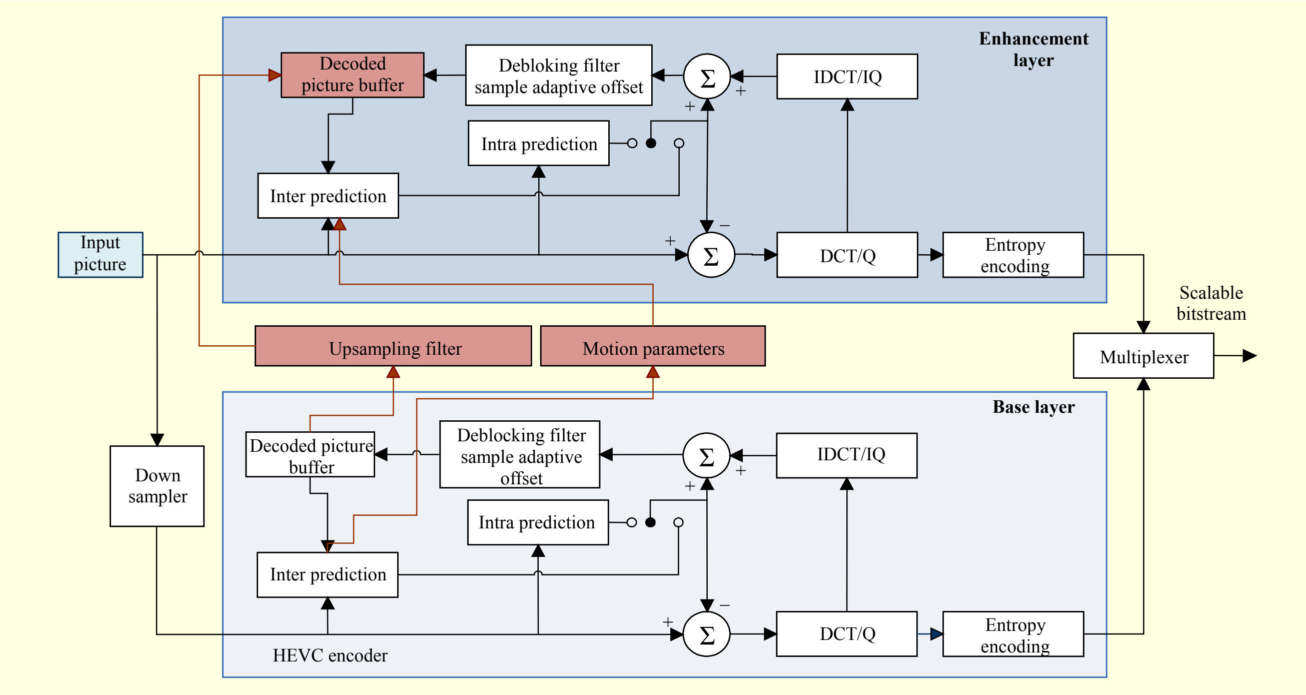


Fig. 1. Encoder block diagram of proposed scalable extension of HEVC.

7-tap or 8-tap filters are used for the interpolation of the fractional-sample positions. For motion vector signaling, HEVC includes a merge mode to derive motion information from spatially or temporally neighboring blocks [8]. HEVC also includes an advanced motion vector prediction (AMVP) to derive several of the most probable candidates from spatially or temporally neighboring blocks. A context-adaptive binary arithmetic coding is used for entropy coding. This has undergone several improvements to improve its throughput speed and compression performance, and to reduce its context memory requirements [9]. The in-loop filtering process includes a sample-adaptive offset as well as deblocking filtering [10], [11].

III. Proposed Scalable Extension of HEVC

Scalable video codec can be designed on either single-loop decoding architecture or multi-loop decoding architecture. In single-loop decoding architecture, which the SVC extension of H.264/AVC is based on, motion compensation (MC) is only enacted once in a target layer. Since inter-coded blocks in the RLs are not reconstructed, it has the advantage that the DPB size and memory bandwidth for MC remain the same on the decoder regardless of the number of layers. However, it requires that constrained intra prediction has to be used in all RLs to avoid MC in the RL. Also, it requires more complicated inter-layer prediction schemes, such as residual prediction, to achieve comparable coding efficiency to that of the multi-loop decoding architecture. In the multi-loop decoding architecture, MC is done in every RL, which is needed to reconstruct the target layer. Both inter-coded blocks and intra-coded blocks are reconstructed in all RLs, and the reconstructed samples of the RLs can be used as additional predicted samples for the EL. Although the multi-loop decoding architecture increases the DPB size and memory bandwidth for MC on the decoder side depending on the number of layers, it is known that the coding efficiency is better than that of the single-loop decoding architecture. It also has the advantage that the multi-view scalability can easily be supported at the same time since the scalable codec based on multi-loop decoding architecture can display any view of the multi-view configuration as view scalability. For these reasons, we propose a scalable extension of HEVC based on multi-loop decoding architecture that employs inter-layer sample prediction and motion parameter prediction.

Figures 1 and 2 illustrate the simplified block diagrams of the encoder and decoder for the proposed scalable extension for HEVC, respectively. The base layer (BL), which is also referred to as the RL, can be encoded or decoded with HEVC coding tools. For the EL, the same concepts as HEVC and additional inter-layer prediction schemes represented with red boxes in Figs. 1 and 2 are incorporated to improve the coding efficiency relative to simulcast coding. The proposed approach has an upsampler and inter-layer prediction tools consisting of inter-layer sample prediction and inter-layer motion parameter prediction. Although, in this paper, it is assumed that two layers

http://dx.doi.org/10.4218/etrij.13.2013.0040

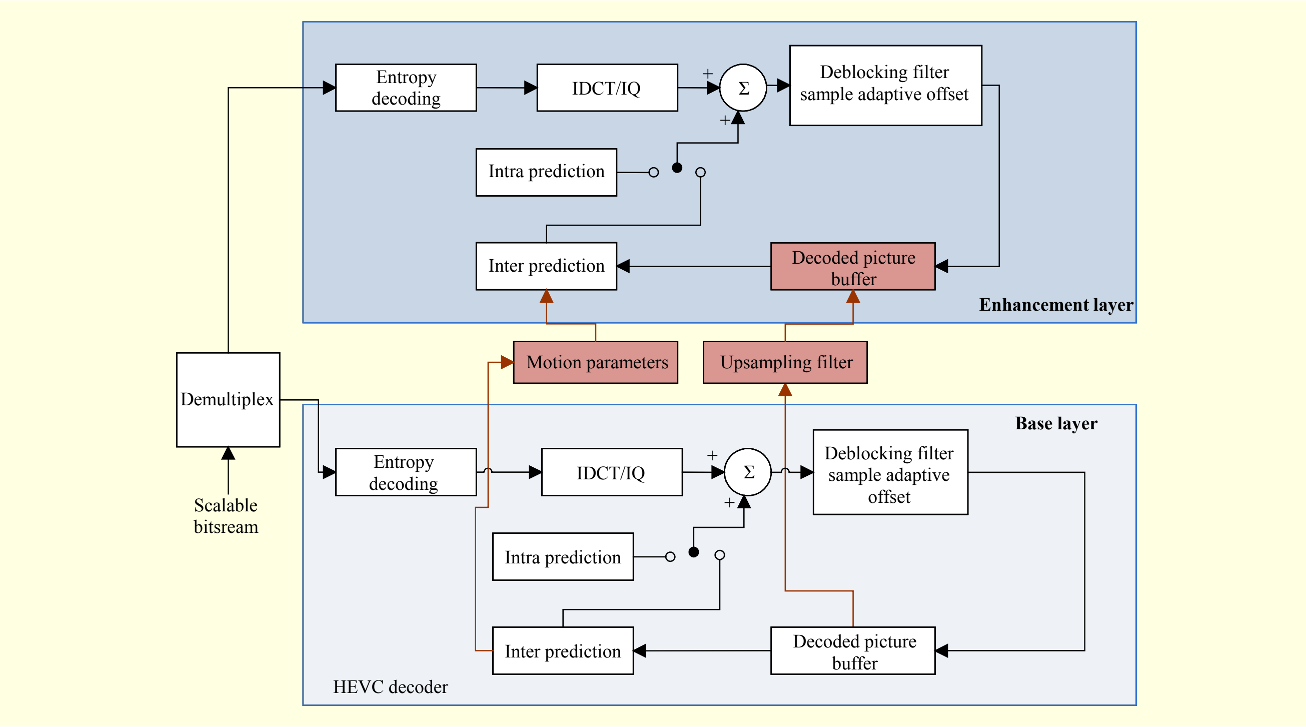


Fig. 2. Decoder block diagram of proposed scalable extension of HEVC.

are coded in a single bitstream, the proposed approach can be extended to support multiple layers without changes.

1. Upsampling Filter

As the proposed scalable extension design is based on multi-loop decoding architecture, the reconstructed picture of the BL is upsampled when the spatial resolution of the BL is different from that of the EL. The proposed approach employs the interpolation filter based on the discrete cosine transform (DCT-IF) to upsample the reconstructed picture of the BL. Table 1 shows the upsampling filter coefficients for the luma components. In the case of a spatial resolution factor of 2, filter coefficients defined in HEVC are used for the upsampling of luma components. In the case of a spatial resolution factor of 1.5, an 8-tap DCT-IF is designed for the upsampling of the luma components. Table 2 shows the upsampling filter coefficients for the chroma components. The filter coefficients defined in HEVC and newly designed filter coefficients are used for the upsampling of the chroma components when the spatial resolution factor is 2 and 1.5, respectively.

Table 1. Upsampling filter coefficients for luma components.

Phase	Filter coefficients (8-tap)
1/3	–1, 4, –11,52, 26, –8, 3, –1
1/2	–1, 4, –11, 40, 40, –11, 4, –1
2/3	–1, 3, –8, 26, 52, –11, 4, –1

Table 2. Upsampling filter coefficients for chroma components.

Phase	Filter coefficients (4-tap)
1/3	–5, 50, 22, –3
1/2	–4, 36, 36, –4
2/3	–3, 22, 50, –5

2. Inter-layer Sample Prediction

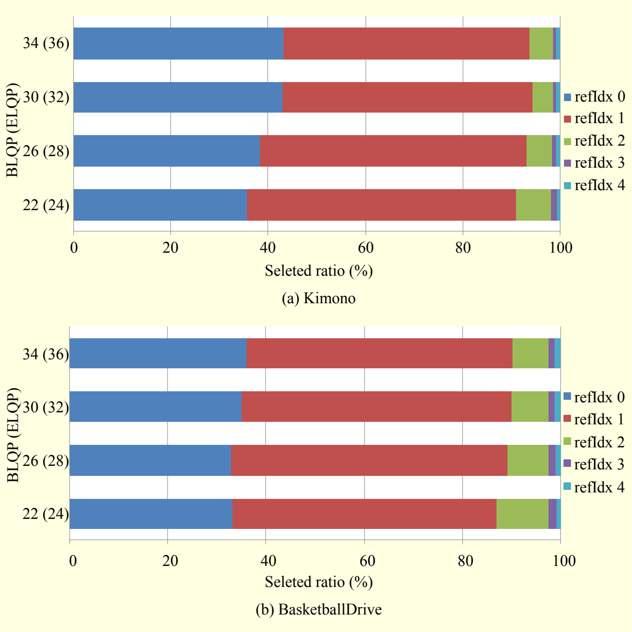
In the proposed scalable extension of HEVC, inter-layer sample prediction is enabled entirely by means of motion estimation (ME) [12] and MC in the EL. For the inter-layer sample prediction, a picture of the BL is entirely reconstructed, upsampled (if necessary), and inserted into the reference picture lists for the picture currently being processed in the EL. This reconstructed picture of the BL, whose spatial resolution is the same as that of the picture in the EL, is defined as an inter-layer reference (ILR) picture. After that, the ILR picture is used as an additional reference picture during the inter prediction process to predict the picture currently being processed in the EL. This means that reference indexes specifying the reference picture are used as a method for the signaling of inter-layer sample prediction without introducing a new prediction mode, such as IntraBL, which is used in the

H.264/SVC. In detail, the process for the reference picture list construction of the EL consists of the following.

Step 1. An ILR picture is generated by reconstructing (and upsampling, if necessary) the corresponding picture of the BL.

Step 2. According to the reference picture sets in the slice segment header, the reference picture lists for the temporal reference picture are initialized and constructed.

Step 3. The ILR picture is inserted into the reference picture lists for the picture currently being processed in the EL. When the picture currently being processed in the EL is a random access picture (RAP), the ILR picture is assigned to the reference index of "0" (refIdx 0), which is the first position in the reference picture lists. Since a non-RAP can use multiple reference pictures, the position of the ILR picture might influence the coding efficiency. Figure 3 shows the selected percentage of ILR and temporal reference pictures in the EL when the ILR picture is assigned to refIdx 0. In this test, the BL quantization parameter (BLQP) is set as 22, 26, 30, 34 in the random access configuration, respectively. As the ILR picture has the same picture order count as that of the corresponding picture of the EL, it is assumed that the ILR picture might correlate more to the EL picture than to the temporal reference picture in terms of temporal distance. With this assumption, the ILR picture is assigned to refIdx 0 and temporal reference pictures are assigned to a reference index from 1 up to 4 (refIdx 1 to refIdx 4) regarding the temporal distance to the current picture. It is observed that the most selected index is refIdx 1 and the second most selected index is refIdx 0 in the range of low and high quality. Regarding entropy coding, the most selected index should be assigned to the shortest index; when the EL picture currently being processed is a non-RAP, the ILR picture is assigned to refIdx 1 instead of refIdx 0. If a reference picture exists, the reference index of the reference picture whose reference index was originally greater than "1" is increased by "1." After the reference picture lists are constructed, the inter prediction process specified in the HEVC shall be performed.



(a) Kimono

(b) BasketballDrive

Fig. 3. Selected ratio of reference picture index when ILR picture is assigned to refIdx 0 (ELQP=BLQP+2).

Table 3. Example of reference picture list of EB_2 picture.

refIdx	LIST0	LIST1
0	EP_0	EB_1
1	**ILR_{B2}**	**ILR_{B2}**
2	EB_1	EP_0

It should be noted that since the ILR picture is included in the reference picture lists for the picture currently being processed in the EL and used as an additional reference picture, each picture in the EL is encoded as an inter-coded picture. Even the first picture in the EL, which should have originally been coded as an I-picture, is coded as a P-picture, using the ILR picture as a reference. In addition, allowing the ILR picture to be used as a reference for the picture currently being processed in the EL makes the following prediction processes in the EL. First, the picture block for the picture currently being processed in the EL can be predicted from not only the co-located region of the ILR picture but also from the other region of the ILR picture with motion vectors. Second, the picture currently being processed in the EL can be bi-predicted from the temporal reference picture and ILR picture. This is defined as combined inter-layer prediction. Therefore, the prediction types used in the proposed scalable extension approach can be conceptually categorized into intra prediction, inter prediction, inter-layer prediction, and combined inter-layer prediction. Figure 4 illustrates an example of the prediction structures in the proposed scalable extension of HEVC. An RAP, EP_0, whose BL is coded as an I-picture can be coded through intra prediction or inter-layer prediction, as depicted in Fig. 4(a). For the inter-layer prediction, the ILR picture can be inserted into the reference picture LIST0. The EP_0 picture can be predicted from the ILR picture. EB_2, whose BL is coded as a B-picture, can be coded through intra prediction, inter prediction, inter-layer prediction, or combined inter-layer prediction. For the inter-layer prediction, the ILR picture is assigned to refIdx 1 in both the reference picture LIST0 and reference picture LIST1 in the EL, as shown in Table 3. Therefore, the EB_2 picture can be uni-predicted or bi-predicted from only the EP_0 picture, only the EB_1 picture, or only the ILR picture, or can be bi-predicted

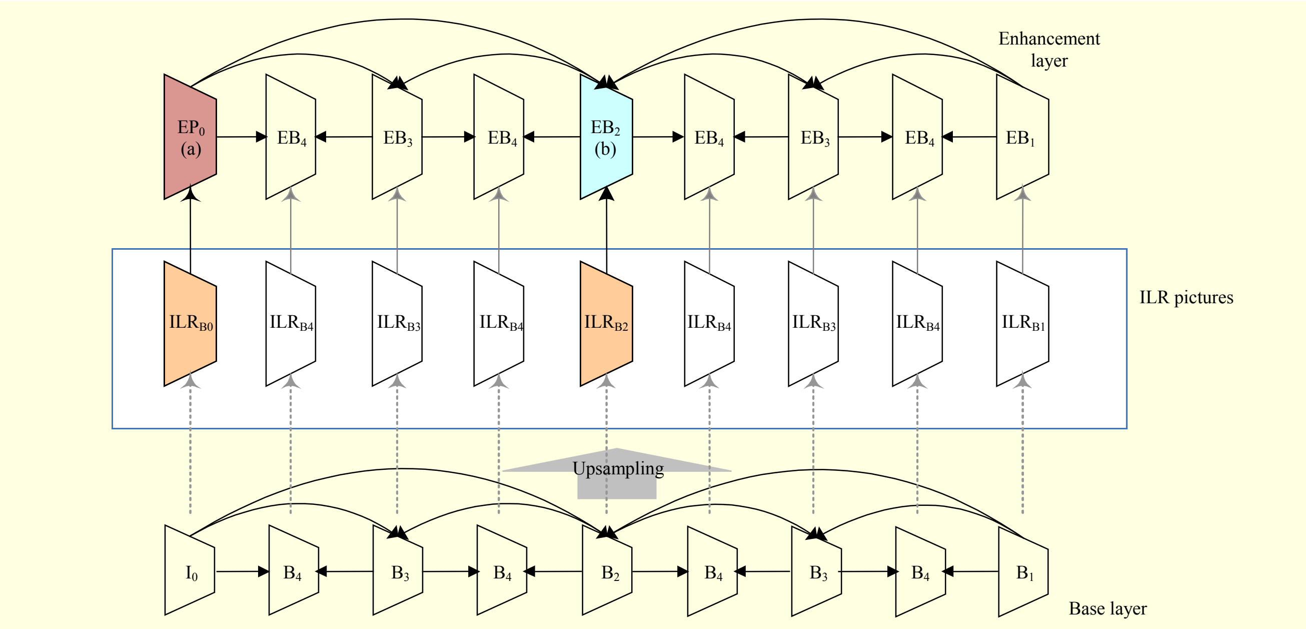


Fig. 4. Example of prediction structure with inter-layer prediction for proposed scalable extension of HEVC.

from any combination of the EP$_0$, EB$_1$, or ILR picture, as described in Fig. 4(b).

3. Inter-layer Motion Parameter Prediction

In general, the motion parameters in the EL, such as reference indexes and motion vectors, are likely to be similar to the corresponding parameters of the co-located PU in the BL. To reduce the motion parameter redundancy between layers, the motion parameters of the co-located PU in the BL can be assigned as one of the candidates for both merge and motion vector prediction in the EL. While H.264/SVC requires an explicit signaling to indicate whether the motion parameter of the BL or the spatial neighboring motion is used as a predictor, no additional signaling is required in the proposed method because the motion parameter of the BL is used as an additional candidate in the merge and motion vector prediction. When the motion vector derived from the co-located PU in the BL is used as a merge or motion vector prediction candidate, it is scaled as follows.

$$mvELx = mvBLx \times ScaleX,$$

$$mvELy = mvBLy \times ScaleY,$$

where *ScaleX* is the ratio between the EL picture width and BL picture width, *ScaleY* is the ratio between the EL picture height and BL picture height, (*mvBLx, mvBLy*) is the motion vector of the co-located PU in the BL, and (*mvELx, mvELy*) is the scaled motion vector.

In HEVC, a motion compression scheme is employed to reduce the buffer size for motion storage. As the motion parameters of BL are compressed into a 16×16 unit after encoding and decoding a picture, the compressed motion parameters can make the inter-layer motion parameter prediction inefficient. To enhance the prediction efficiency, we use the uncompressed motion information of the BL for inter-layer motion prediction, which means we postpone the motion compression of the BL after encoding and decoding a picture in the EL in the access unit.

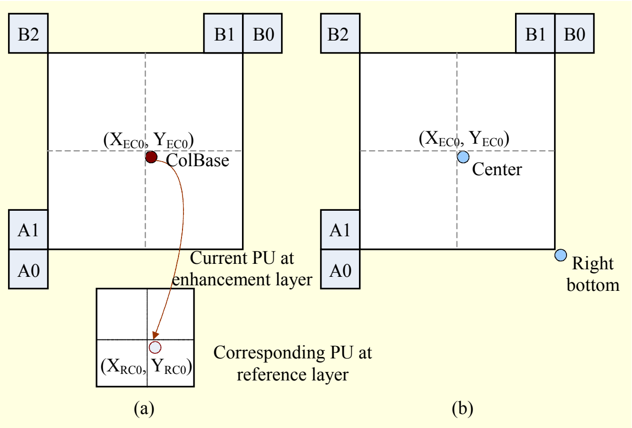


Fig. 5. Derivation of merge and motion vector prediction candidate list using spatial, temporal, and inter-layer candidates.

The process of the merge candidate list construction of the EL is modified as follows. As there will be a high correlation between the motion parameters of the co-located PU in the BL (ColBase) and that of the current PU in the EL, the motion parameters of the ColBase are placed in the first candidate list. In the derivation of the merge candidates, up to four merge candidates are selected according to the order {ColBase, A$_1$, B$_1$,

B_0, A_0, B_2} described in Fig. 5(a). The temporal merge candidate is then selected between two candidate positions, as described in Fig. 5(b). In the derivation of the temporal merge candidate, when the reference index of the temporal merge candidate refers to the ILR picture or when the co-located reference picture is an ILR picture, the temporal merge candidate is considered to be unavailable. The remaining parts of the construction process of the candidate list are the same as in HEVC [13]. Also, the construction process of the motion vector predictor candidates in the EL is modified. In the derivation of the motion vector predictor candidates, only two candidates are selected according to the order of {ColBase, {A_0, A_1}, {B_0, B_1, B_2}, {right bottom position, center position}}, as described in Fig. 5. When the reference picture index of the current PU indicates an ILR picture, the neighboring PUs that have a reference picture index indicating a temporal reference picture are considered unavailable. Likewise, when the reference picture index of the current PU indicates a temporal reference picture, the neighboring PUs that have a reference picture index indicating an ILR picture are considered unavailable. The remaining parts of the construction process of the candidate list are the same as that of HEVC [13].

IV. Experiment Results

The proposed scalable extension of HEVC is implemented on HM6.1 reference software. To evaluate the performance of the proposed solution, the coding efficiency of the proposed scalable coding with two layers is compared to that of the simulcast, in which the BL and the EL are coded independently with HEVC, and to that of the single-layer coding, in which the EL is coded with HEVC. For the experiments, the test conditions specified in [14] are followed. From these test conditions, two scalabilities, namely, spatial scalability and quality scalability, also referred to as SNR scalability, are applied. For spatial scalability, two layers are coded with a resolution ratio of 1.5 and 2, respectively. In these cases, the BLQP is set to 22, 26, 30, 34. For the SNR scalability, two layers are coded with a resolution ratio of 1, and the BLQP is set to 26, 30, 34, 38. For each BLQP, two different ELQPs are tested. For spatial scalability, the ELQPs are set to BLQP+0 and BLQP+2. For SNR scalability, the ELQPs are set to BLQP-6 and BLQP-4. Test video sequences according to these configurations are summarized in Table 4. Class A sequences are used for the spatial scalability with a resolution ratio of 2 and SNR scalability. Class B sequences are used for all spatial and SNR scalabilities. For downsampled video sequences, a downsampling filter with a 0.9π cut-off frequency is used [15]. The simulcast anchor bitstreams are constructed using HM6.1 reference software with the main profile.

Table 4. Test video sequences.

Class	Sequence	Encoded frames	Frame rate	BL resolution	EL resolution
A	Traffic	150	30	1920×1024 3840×2048	3840×2048 3840×2048
	PeopleOnStreet	150	30	1920×1080 3840×2160	3840×2160 3840×2160
B	Kimono	240	24	960×540 1280×720 1920×1080	1920×1080 1920×1080 1920×1080
	ParkScene	240	24	960×540 1280×720 1920×1080	1920×1080 1920×1080 1920×1080
	Cactus	500	50	960×540 1280×720 1920×1080	1920×1080 1920×1080 1920×1080
	BasketballDrive	500	50	960×540 1280×720 1920×1080	1920×1080 1920×1080 1920×1080
	BQTerrace	600	60	960×540 1280×720 1920×1080	1920×1080 1920×1080 1920×1080

Table 5. Average BD-bitrate gain (%) compared to simulcast coding when inter-layer texture prediction is enabled.

Sequence	AI-2x	AI-1.5x	RA-2x	RA-1.5x	RA-SNR
Traffic	–24.3	-	–13.2	-	–12.8
PeopleOnStreet	–27.5	-	–19.8	-	–23.3
Kimono	–29.7	–38.9	–20.2	–32.7	–19.6
ParkScene	–22.5	–34.6	–13.5	–25.5	–18.8
Cactus	–19.8	–32.6	–12.0	–23.7	–15.4
BasketballDrive	–16.1	–28.0	–12.9	–25.7	–17.3
BQTerrace	–12.6	–25.1	–5.3	–14.2	–13.2
Class A average	–25.9	-	–16.5	-	–18.1
Class B average	–20.1	–31.8	–12.8	–24.4	–16.9
Average	–21.8	–31.8	–13.9	–24.4	–17.2

As the first evaluation, we investigate the coding efficiency of the inter-layer sample prediction scheme without inter-layer motion parameter prediction. Table 5 summarizes the coding efficiency of the proposed approach with respect to the simulcast coding for the spatial and SNR scalability. The results shown in Table 5 are averaged over the two tested ELQPs for each sequence. For spatial scalability, an average luma Bjøntegaard delta bitrate (BD-bitrate) savings of 21.8%, 31.8%, 13.9%, 24.4% is achieved for All Intra (AI)-2x, AI-1.5x, Random Access (RA)-2x, and RA-1.5x, respectively. In the

Table 6. Average BD-rate savings (%) compared to simulcast coding with both inter-layer prediction and inter-layer motion parameter prediction.

Sequence	Spatial scalability												SNR scalability		
	AI-2x			AI-1.5x			RA-2x			RA-1.5x			RA-SNR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Traffic	−24.3	−22.6	−23.2	-	-	-	−15.2	−10.6	−9.5	-	-	-	−20.4	−17.3	−14.9
PeopleOnStreet	−27.5	−21.6	−22.8	-	-	-	−22.2	−0.8	−2.6	-	-	-	−25.4	−8.7	−9.7
Kimono	−29.7	−25.5	−27.5	−38.9	−38.0	−39.6	−23.5	−15.2	−13.3	−34.8	−29.3	−28.0	−22.7	−16.3	−14.8
ParkScene	−22.5	−20.5	−21.9	−34.6	−33.9	−36.0	−14.8	−9.4	−9.9	−27.0	−21.9	−22.3	−20.6	−16.0	−16.0
Cactus	−19.8	−19.5	−18.1	−32.6	−33.5	−33.2	−14.9	−8.7	−4.9	−26.7	−22.4	−19.1	−19.6	−14.6	−7.5
BasketballDrive	−16.1	−11.9	−11.3	−28.0	−28.5	−26.9	−16.3	−4.3	−4.3	−28.4	−20.8	−19.1	−21.1	−9.8	−9.9
BQTerrace	−12.6	−10.4	−8.7	−25.1	−25.1	−25.6	−6.4	3.9	9.3	−15.8	−2.1	1.2	−15.7	11.0	24.0
Class A average	−25.9	−22.1	−23.0	-	-	-	−18.7	−5.7	−6.0	-	-	-	−22.9	−13.0	−12.3
Class B average	−20.1	−17.6	−17.5	−31.8	−31.8	−32.3	−15.2	−6.7	−4.6	−26.5	−19.3	−17.5	−19.9	−9.2	−4.9
Average	−21.8	−18.9	−19.1	−31.8	−31.8	−32.3	−16.2	−6.4	−5.0	−26.5	−19.3	−17.5	−20.8	−10.3	−7.0

SNR scalability, an average luma BD-bitrate savings of 17.2% is achieved. It can be observed that the BD-bitrate savings is consistent over all test configurations. As the proposed inter-layer sample prediction operates at the picture level in the EL, the CU level or PU level operations in the EL for HEVC remain the same. These make advantages of achieving significant coding gains and supporting scalabilities with minimum changes relative to the single-layer HEVC architecture. Table 6 shows the coding efficiency of the proposed HEVC scalable extension when both inter-layer sample prediction and inter-layer motion parameter prediction are enabled. For inter-layer motion parameter prediction, the merge and AMVP candidate list construction processes are modified as explained in section III. In the spatial scalability, an average luma BD-bitrate savings of 21.8%, 31.8%, 16.2%, and 26.5% is achieved for AI-2x, AI-1.5x, RA-2x, and RA-1.5x, respectively. In the SNR scalability, an average luma BD-bitrate savings of 20.8% is achieved. Compared to the results shown in Table 5, it is clear that using the inter-layer motion parameter prediction can provide a higher coding efficiency improvement. Compared to using only inter-layer sample prediction, an additional 2.3%, 2.1%, and 3.6% can be provided for the RA-2x, RA-1.5x, and RA-SNR scalabilities, respectively.

Table 7. Average BD-bitrate overhead (%) and runtime ratio of proposed scalable extension compared to single-layer coding.

Sequence	BD-bitrate (Y)		Time ratio		BD-bitrate (Y)		Time ratio		BD-bitrate (Y)	Time ratio	
	AI-2x	AI-1.5x	Enc.	Dec.	RA-2x	RA-1.5x	Enc.	Dec.	RA-SNR	Enc.	Dec.
Traffic	12.9	-	2.5	1.4	22.4	-	1.6	1.8	17.3	2.2	2.2
PeopleOnStreet	9.8	-	2.5	1.4	18.5	-	1.4	1.6	13.1	2.2	2.1
Kimono	6.6	4.0	2.1	1.5	14.3	9.7	1.6	1.7	15.3	2.3	2.2
ParkScene	9.4	6.5	2.4	1.5	19.1	17.9	1.6	1.9	16.1	2.2	2.2
Cactus	15.2	10.7	2.4	1.5	21.6	19.5	1.6	1.9	16.6	2.1	2.2
BasketballDrive	19.7	16.3	2.3	1.5	19.9	16.1	1.5	1.8	14.9	2.0	2.1
BQTerrace	17.2	15.2	2.6	1.6	19.0	20.8	1.6	1.9	8.9	2.1	2.2
Class A average	11.3	-	2.5	1.4	20.4	-	1.5	1.7	15.2	2.2	2.2
Class B average	13.6	10.5	2.4	1.5	18.8	28.0	1.6	1.8	14.4	2.1	2.2
Average	13.0	10.5	2.4	1.5	19.2	16.8	1.6	1.8	14.6	2.2	2.2

Table 7 shows the BD-bitrate overhead and encoding and decoding runtime ratio of the proposed solution compared to single-layer coding. As the results show, the average BD-bitrate overhead relative to single-layer coding is 13.0%, 10.5%, 19.2%, 16.8%, and 14.6% for the AI-2x, AI-1.5x, RA-2x, RA-1.5x, and RA-SNR scalabilities, respectively. These values indicate that it is possible to service both the UHD and HD contents in the single bitstream with about 15% bitrate overhead compared to the UHD single bitstream. The average encoding runtime ratio of the proposed solution is 2.4 times for the AI spatial scalability, 1.6 times for the RA spatial scalability, and 2.2 times for the RA SNR scalability relative to single-layer coding. The reason for having a higher runtime ratio in the AI case is that an ME/MC process in the EL is required. The average decoding runtime ratio of the proposed solution is

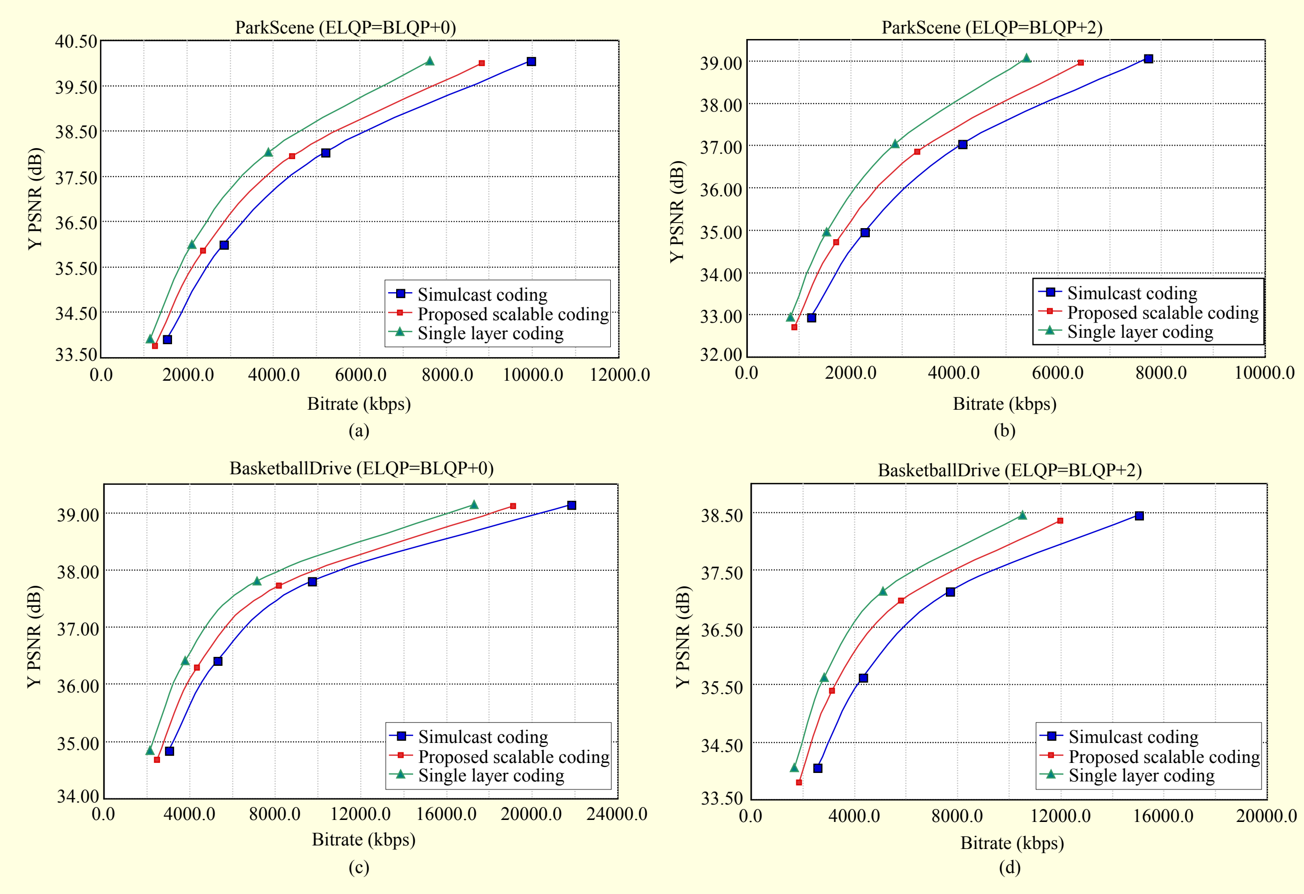


Fig. 6. RD-curves for RA spatial scalability with resolution ratio of 2.

1.5 times for the AI spatial scalability, 1.8 times for the RA spatial scalability, and 2.2 times for the RA SNR scalability relative to single-layer coding. Figure 6 shows the selected rate-distortion (RD) curves for RA spatial scalability with a resolution ratio of 2. The effectiveness of the proposed approach generally improves throughout the whole range of bitrates, as can be seen in Fig. 6. The selected RD-curves for RA spatial scalability with a resolution ratio of 1.5 are shown in Fig. 7. The effectiveness of the proposed approach generally improves throughout the whole range of bitrates. Furthermore, it can be seen that the effectiveness of the proposed scalable extension approach for a resolution ratio of 1.5 is higher than that for a resolution ratio of 2. In fact, it is certainly inherent to a better correlation between layers at a resolution ratio of 1.5 than at a resolution ratio of 2. Also, since the proposed solution is based on multi-loop decoding architecture, it makes more use of inter-layer correlation at a resolution ratio of 1.5.

V. Conclusion

In this paper, we described the scalable extension approach of HEVC for delivering high-quality digital video contents. The proposed scalable extension is designed on multi-loop decoding architecture, which fully reconstructs the picture of all layers with MCs for each layer. To support the scalabilities, inter-layer sample prediction is enabled by inserting the reconstructed RL picture into the DPB of the EL picture. Additionally, to reduce the inter-layer motion parameter redundancy, the motion parameters of the co-located PU in the RL is assigned as one of the candidates for both merge mode and motion vector prediction in the EL. Compared to simulcast coding, average BD-bitrate savings of about 24% and 21% are achieved for spatial and SNR scalability, respectively. Compared to HEVC single-layer coding, the average overhead is about 15% for spatial scalability and 15% for SNR scalability. The experiment results show that the proposed approach has the advantage of significant coding gains and supporting scalabilities with a simple extension of HEVC. We can thereby conclude that the proposed HEVC scalable extension is appropriate for the simultaneous delivery of different formats of high-quality digital video contents, such as UHD and HD, in the single bitstreams to support the endpoint devices with different capabilities. In addition, it can be easily extended to support multi-view scalability and coding standard

http://dx.doi.org/10.4218/etrij.13.2013.0040

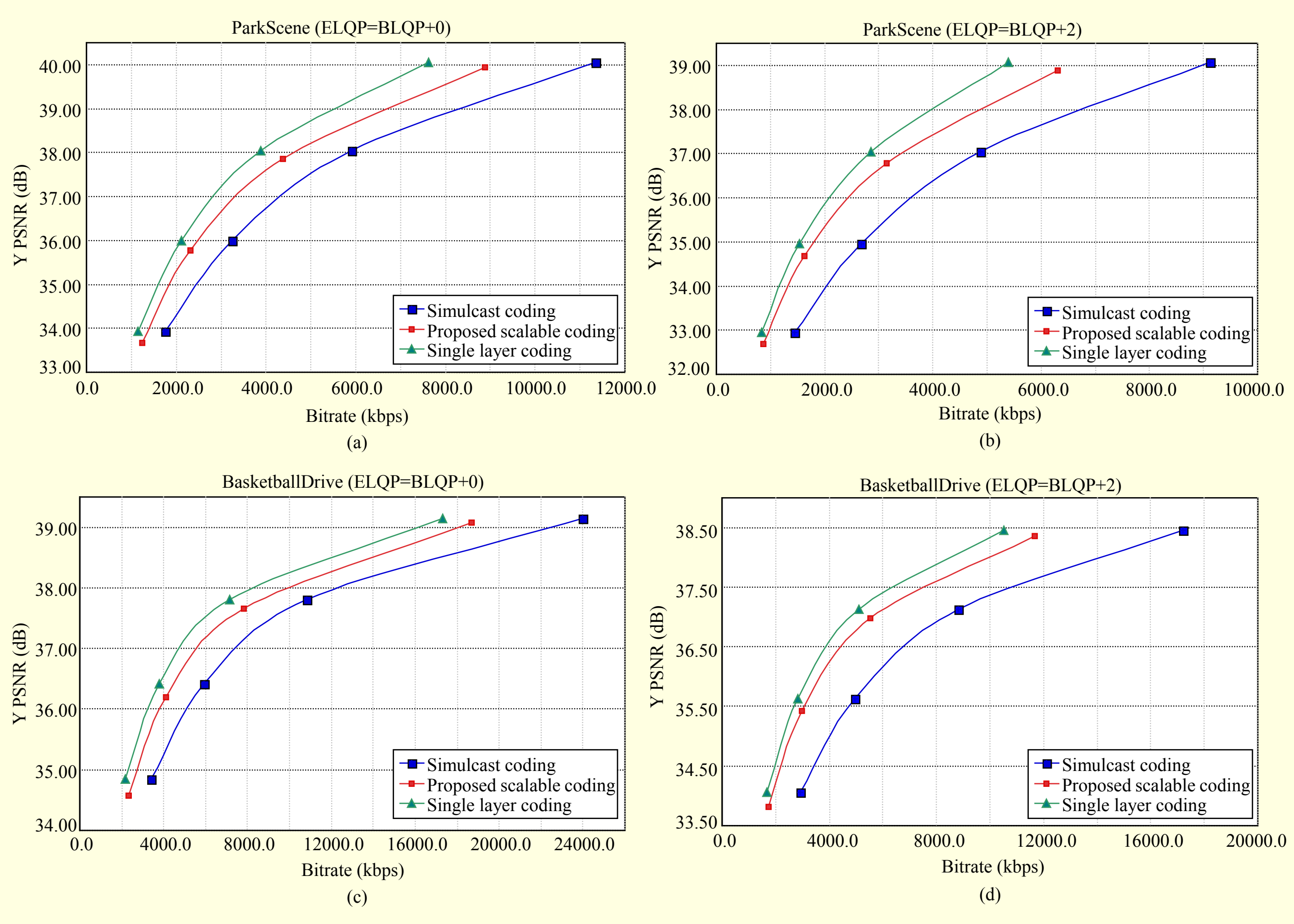


Fig. 7. RD-curves for RA spatial scalability with resolution ratio of 1.5.

scalability, owing to the multiple-loop decoding architecture.

References

[1] J. W. Kang et al., "Description of Scalable Video Coding Technology Proposal by ETRI and Kwangwoon Univ.," *JCTVC-K0037*, Oct. 2012.

[2] C.A. Segall and G.J. Sullivan, "Spatial Scalability with the H.264/AVC Scalable Video Coding Extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, Sept. 2007, pp. 1121-1135.

[3] ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for Generic Audio-Visual Services," *ITU T Rec. H.264 and ISO/IEC 14496-10 (AVC)*, May 2003.

[4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, Sept. 2007, pp. 1103-1120.

[5] G.J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, vol. 12, Dec. 2012, pp. 1649-1668.

[6] I.-K. Kim et al., "Block Partitioning Structure in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1697-1706.

[7] J. Lainema et al., "Intra Coding of the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1792-1801.

[8] P. Helle et al., "Block Merging for Quadtree-Based Partitioning in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1720-1731.

[9] V. Sze and M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1778-1791.

[10] A. Norkin et al., "HEVC Deblocking Filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1746-1754.

[11] C.-M. Fu et al., "Sample Adaptive Offset in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, Dec. 2012, pp. 1755-1764.

[12] J. Kim et al., "An SAD-Based Selective Bi-prediction Method for Fast Motion Estimation in High Efficiency Video Coding," *ETRI J.*, vol. 34, no. 5, Oct. 2012, pp. 753-758.

[13] ISO/IEC JTC1/SC29/WG11 N13333, *Text of ISO/IEC FDIS*

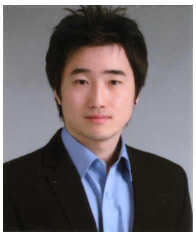
23008-2 High Efficiency Video Coding, Jan. 2013.

[14] ISO/IEC JTC-1/SC29/WG11 w12957, *Joint Call for Proposals on Scalable Video Coding Extensions of High Efficiency Video Coding (HEVC)*, July 2012.

[15] J. Dong et al, "Down Sampling Filters for Anchor Generation for Scalable Extensions of HEVC," *Tech. Rep. M24499*, *ISO/IEC JTC1/SC29/WG11 MPEG*, Geneva, Switzerland, May 2012.

Hahyun Lee received his B.S. degree in electronics engineering from Korea Aerospace University, Rep. of Korea, in 2002 and his M.S. degree in mobile communication and digital broadcast engineering from the University of Science and Technology (UST), Rep. of Korea, in 2007, respectively. Since 2008, he has been a senior member of the research staff in the Broadcasting & Telecommunications Media Research Laboratory of ETRI, Rep. of Korea. His research interests include video coding, image processing, and video communication.

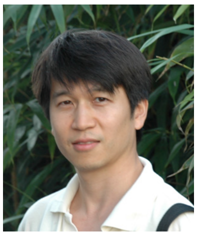
Jung Won Kang received her BS and MS degrees in electrical engineering in 1993 and 1995, respectively, from Korea Aerospace University, Seoul, Rep. of Korea. She received her PhD degree in electrical and computer engineering in 2003 from the Georgia Institute of Technology, Atlanta, GA, USA. Since 2003, she has been a senior member of the research staff in the Broadcasting & Telecommunications Media Research Laboratory, ETRI, Rep. of Korea. Her research interests are in the areas of video signal processing, video coding, and video adaptation.



Jinho Lee received his B.S. degree in electrical and electronics engineering from Korea University, Rep. of Korea, in 2007 and his M.S. degree in telecommunications and digital broadcast engineering from the University of Science and Technology (UST), Rep. of Korea, in 2009. Since 2009, he has been a member of the engineering staff at ETRI, Daejeon, Rep. of Korea. His research interests include video coding, HEVC, and SHVC.

Jin Soo Choi received his BE, ME, and PhD in electronics engineering from Kyungpook National University, Rep. of Korea, in 1990, 1992, and 1996, respectively. Since 1996, he has been a principal member of the research staff at ETRI, Rep. of Korea. He has been involved in developing the MPEG-4/HEVC codec system, data broadcasting system, and 3D/UHDTV broadcasting system. His research interests include visual signal processing and interactive services in the field of digital broadcasting technology.

Jinwoong Kim received his B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Rep. of Korea, in 1981 and 1983, respectively. He received his Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 1993. He has been working at ETRI since 1983, leading many projects in the telecommunications and digital broadcasting areas, such as the development of an MPEG-2 video encoding chipset and real-time HDTV encoder system and innovative technologies for data and viewer-customized broadcasting. He also carried out projects on multimedia search, retrieval, and adaptation technologies related to the MPEG-7 and MPEG-21 standards. Currently, the major focus of his research is on 3D and realistic media technologies and systems. He was the leader of the 3D DMB and multi-view 3DTV system development project and is now actively working on building practical digital holographic 3D systems. He was a keynote speaker at 3DTV-CON 2010 and has been an invited speaker at a number of international workshops, including 3D Fair 2008, DHIP 2012, and WIO 2013.



Donggyu Sim received his B.S. and M.S. degrees in electronics engineering from Sogang University, Seoul, Rep. of Korea, in 1993 and 1995, respectively. He received his Ph.D. degree from the same university in 1999. He was with the Hyundai Electronics Co., Ltd., from 1999 to 2000, where he was involved in MPEG-7 standardization. He was a senior research engineer at Varo Vision Co., Ltd., working on MPEG-4 wireless applications from 2000 to 2002. He worked for the Image Computing Systems Lab. (ICSL) at the University of Washington as a senior research engineer from 2002 to 2005. He conducted research on ultrasound image analysis and parametric video coding. In 2005, he joined the Department of Computer Engineering at Kwangwoon University, Seoul, Rep. of Korea, as an associate professor. He was elevated to IEEE Senior Member in 2004. His current research interests include image processing, computer vision, video communication, and video coding.

http://dx.doi.org/10.4218/etrij.13.2013.0040