IEIE Transactions on Smart Processing and Computing, vol. 4, no. 6, December 2015
http://dx.doi.org/10.5573/IEIESPC.2015.4.6.434

IEIE Transactions on Smart Processing and Computing

Scalable Multi-view Video Coding based on HEVC

Woong Lim, Junghak Nam, and Donggyu Sim

Department of Computer Engineering, Kwangwoon University, 447-1, Wolgye-dong, Nowon-gu, Seoul - Korea
{limwoong, qejixfyza, dgsim}@kw.ac.kr

***** Corresponding Author:

Received December 16, 2015; Accepted December 24, 2015; Published December 31, 2015

** Regular Paper*

Abstract**:** In this paper, we propose an integrated spatial and view scalable video codec based on high efficiency video coding (HEVC). The proposed video codec is developed based on similarity and uniqueness between the scalable extension and 3D multi-view extension of HEVC. To improve compression efficiency using the proposed scalable multi-view video codec, inter-layer and inter-view predictions are jointly employed by using high-level syntaxes that are defined to identify view and layer information. For the inter-view and inter-layer predictions, a decoded picture buffer (DPB) management algorithm is also proposed. The inter-view and inter-layer motion predictions are integrated into a consolidated prediction by harmonizing with the temporal motion prediction of HEVC. We found that the proposed scalable multi-view codec achieves bitrate reduction of 36.1%, 31.6% and 15.8% on the top of ×2, ×1.5 parallel scalable codec and parallel multi-view codec, respectively.

Keywords**:** High efficiency video coding, Scalable video coding, Multi-view video coding

1. Introduction

With recent remarkable development of multimedia technology, various high performance multimedia devices such as smart phone, tablet PC, smart TV, and so on have been widely deployed all over the world. Driven by the market environment, users demand multimedia contents in higher resolution and high quality videos for better experiences. In addition, the market demands have independently driven to the new display devices with diverse resolutions and multiple view-points for 3D. In the future, we can expect multimedia services supported for heterogeneous scalability of diverse resolution and multi-view ones. From the video coding point of view, scalable and multi-view video coding technique is required to support various display devices with diverse resolution and multiple view points over a limited bandwidth. According to these circumstances, the conventional H.264/AVC has its own scalable and multi-view extensions independently [1-11]. Beyond the separate scalable coding standards, a consolidated video coder will be expected for the joint scalability in the near future.

After finalization of H.264/AVC and its scalable and multi-view extensions, ISO/IEC MPEG and ITU-T VCEG founded the Joint Collaborative Team on Video Coding (JCT-VC) and started the standardization of new state-of-art video coding, named HEVC to achieve twice higher coding performance compared to the conventional H.264/AVC [12-23]. On January 2013, HEVC was released as Final Draft International Standard (FDIS) for the upcoming Ultra High Definition (UHD) videos beyond the Full-HD videos. After finalization of the standardization, it is expected that HEVC will be widely used for diverse video services such as wide-band streaming, broadcasting, and conferencing with very high resolution videos (e.g. over 4K×2K), etc. The extensions for scalable video and multi-view video are also being standardized for multiple layers (resolutions/quality) and multiple views by JCT-VC and JCT-3V groups, respectively, i.e. HEVC and its extensions consider not only the high resolution video but also the multiple layer/view scalability for various devices with diverse resolutions and view-points [24-30]. However, scalable and multi-view extensions of HEVC are being independently standardized in the similar manner of the standardization of H.264/AVC extensions. According to this fact, the international standard does not support the view/resolution scalable services, e.g. 3D video service for displays with different resolutions. Therefore, research on the unified scalable multi-view video coding could be

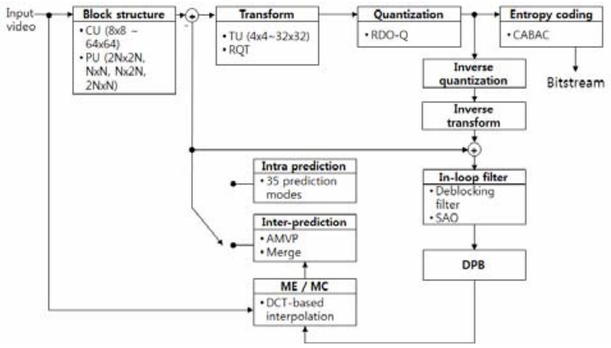


Fig. 1. Block-diagram of HEVC encoder.

introduced for the future media applications. In this paper, we propose an integrated scalable multi-view video coder that can support view/layer scalabilities in a consolidated framework based on HEVC. To maximize coding performance of the proposed joint codec, we designed high level syntaxes to indicate view and layer IDs, decoded picture buffer management algorithm, and adaptive inter-layer/inter-view prediction methods. Therefore, the proposed scalable multi-view codec supports not only the inter-layer prediction but also the inter-view prediction to improve coding performance for the picture of the enhanced layer in the extended view. Using inter-view and inter-layer prediction adaptively, we found that a significant coding gain is achieved for the extended view in the enhanced layer over test video sequences.

In Section 2, the extensions of HEVC will be introduced to summarize the key technologies of multi-view and scalable coders. Then, in Section 3, the proposed joint codec will be described in detail. To evaluate the performance of the proposed codec, experimental results will be shown in Section 4 and conclusion will be addressed in Section 5.

2. The Overview of HEVC Standard and Extensions

Overall coding structure of HEVC standardized by JCT-VC is similar to that of the conventional H.264/AVC e.g. block-wise coding, intra/inter prediction, transform, etc., as shown in Fig. 1. However, HEVC employs various improved coding tools to improve compression performance. For intra coding, more prediction directions were designed. For inter coding, 1/4 pixel resolution motion estimation based on DCT-based interpolation and new motion prediction were adopted. Context adaptive binary arithmetic coding (CABAC) for entropy coding is employed which is the same with one in H.264/AVC; however, syntax models and initialization were modified according to new syntax elements in HEVC. For in-loop filtering, not only deblocking filter but also sample adaptive offset (SAO) is designed to improve coding efficiency and subjective quality. Moreover, the hierarchical predictions and transformations over larger block partitions contribute to significant coding gain of 40% in bit-saving against H.264/AVC.

By JCT-3V group, the multi-view extension of HEVC is being standardized since July 2012. The major tool of HEVC-based multi-view extension is inter-view prediction for coding efficiency. Because multi-view videos are captured by slightly shifted multiple cameras at the same time, we can expect high correlation among views. Furthermore, even if temporal correlation decreases by scene change and very fast motion activity, the high inter-view correlation can be preserved. Therefore, using the inter-view prediction, additional redundancy can be removed and higher coding efficiency can be achieved. Note that the base view shall be compatible to HEVC for backward compatibility.

The scalable extension of HEVC is also being standardized with consideration of temporal, spatial and quality scalability under the JCT-VC. The scalable extension enables flexible and adaptive video services depending on various channel bandwidths and resolutions. To achieve these functionalities, the major tool of scalable coding is inter-layer prediction. Similarly to the inter-frame prediction, the inter-layer prediction removes redundancy between two layers and enables to achieve high coding efficiency. Especially, inter-layer prediction can be conducted by adding the up-sampled base layer frame into the DPB for spatial scalability, so called the reference index method. In the same manner, the view scalability of multi-view HEVC (MHVC) was developed by adding the reconstructed inter-view frames into the DPB. The SHVC and MHVC are independently developed but they have the same philosophy to achieve different scalabilities. In this paper, we develop a joint scalable multi-view video coding based on modification of DPB management on HEVC.

3. The Proposed Scalable Multi-view Video Coding Based on HEVC (SMHVC)

The proposed SMHVC provides not only the spatial scalability but also the view-scalability with compatibility for the base layer of the base view to HEVC single-view/layer coding. The proposed SMHVC can achieve higher coding performance for enhanced layers and extended views from advantages of inter-view and inter-layer predictions, respectively. In addition, for the enhanced layer of the extended view, the proposed codec can achieve higher coding performance by employing adaptive inter-layer/inter-view prediction method.

Fig. 2 shows a coding structure with the proposed SMHVC for two views and two spatial layers. The video at the base view of the base layer is coded by HEVC and the coded video can be decoded for the conventional 2D video services. With better network connections and devices, 2D spatial scalable and/or multi-view video services can be performed with the proposed SMHVC. As shown in Fig. 2, the higher spatial resolution multi-view video services are also possible with the proposed SMHVC.

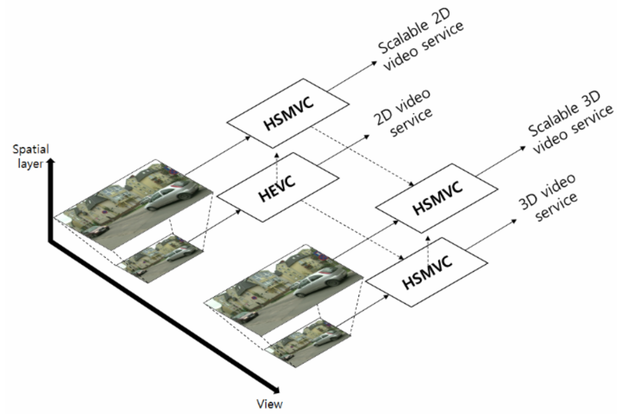
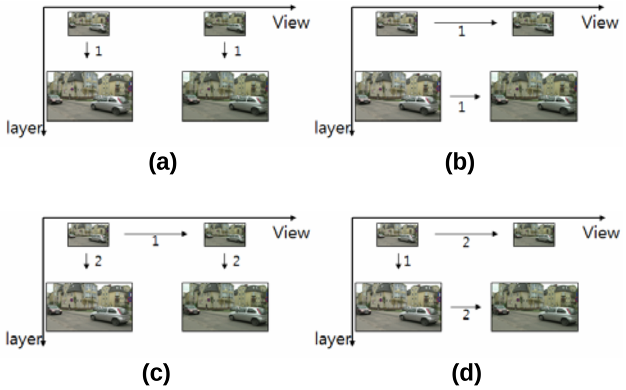


Fig. 2. Applications of the proposed SMHVC.



(a) (b)

(c) (d)

Fig. 3. Four coding structures for scalable and multi-view videos; (a) Parallel scalable codec, (b) Parallel multi-view codec, (c) Parallel scalable codec on multi-view codec, (d) Parallel multi-view codec on scalable codec.

3.1 Overall Structure of the Proposed Integrated Scalable Multi-view Video Codec

We have several possible scenarios to provide the scalable multi-view video services. Fig. 3 shows four coding structures for scalable multi-view videos. Figs. 3(a) and (b) show that parallel scalable and parallel multi-view codecs which are capable of providing the spatial/view scalable functionalities with two independent scalable codecs or two multi-view codecs in parallel, respectively, with minimal computational complexity. However, each codec employs only one dimensional prediction i.e. inter-layer or inter-view prediction. Figs. 3(c) and (d) show that the parallel scalable codec on multi-view codec and parallel multi-view codec on scalable codec, respectively, which compress the base layer or the base view first using scalable codec or multi-view codec. By using the parallel multi-view codec or parallel scalable structures, they can achieve additional coding gain. Both algorithm can use one of inter-layer or inter-view redundancy for the enhancement layer of the auxiliary view. However, the enhancement layer of the auxiliary view could be coded by using both predictions.

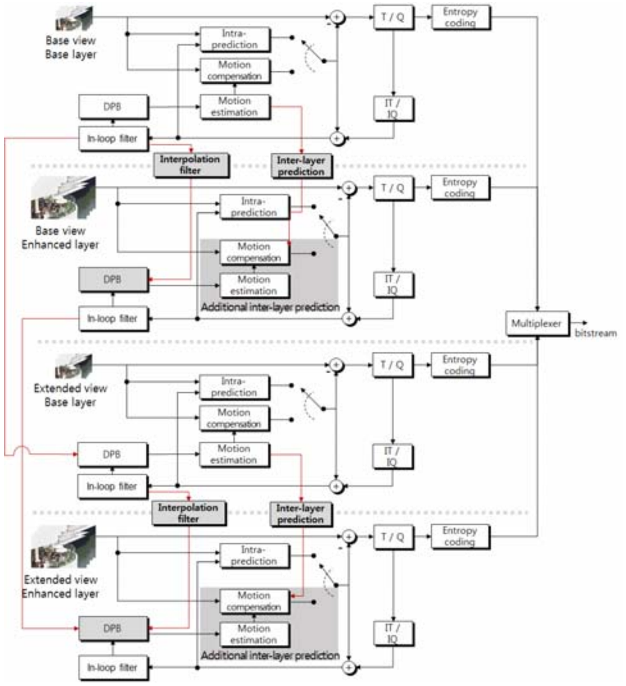


Fig. 4. Block diagram of the proposed SMHVC coding for the extended view in the enhanced layer.

Fig. 4 shows a block diagram of the proposed SMHVC for the extended view in the enhanced layer. In Fig. 4, the interpolated reconstructed picture of the base layer is fed into the decoded picture buffer of the enhanced layer for inter-layer prediction and the up-scaled motion vector of the base layer can be exploited in motion prediction of the enhanced layer. Likewise, in case of multi-view coding, the reconstructed picture of the base view is transferred to decoded picture buffer of the extended view without up-scaling and the motion vector is also exploited for the coding of the extended view. The proposed consolidated SMHVC conducts compression using both inter-layer and inter-view prediction with the same manners of the conventional scalable and multi-view coding structures. For the proposed SMHVC, both inter-layer and inter-view predictions are adaptively employed for coding of the enhancement layer in the extended view which is not supported in the independent scalable and multi-view codecs.

Fig. 5 shows the random access coding structure of SMHVC for a group of picture (GOP) in case that the size of GOP is equal to four using temporal (dotted line) and the proposed adaptive inter-layer / inter-view prediction (solid line). Differently from the conventional scalable coding and multi-view coding algorithms, the proposed SMHVC improves coding efficiency by maximizing removal of temporal, inter-layer and inter-view redundancies while coding the picture of the enhanced layer in the extended view. The best prediction mode is adaptively decided based on rate-distortion optimization (RDO) and this will affect the coding performance of the following encoding pictures.

For prediction with the reconstructed pictures from the

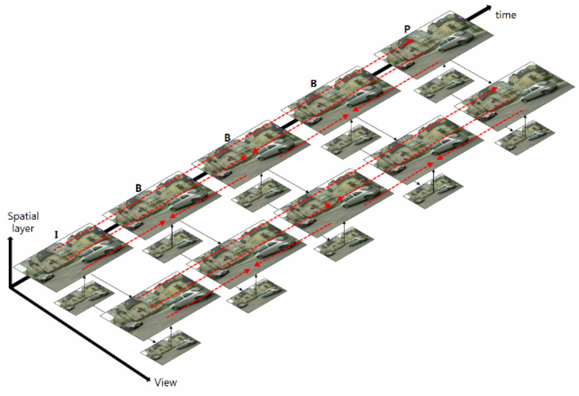
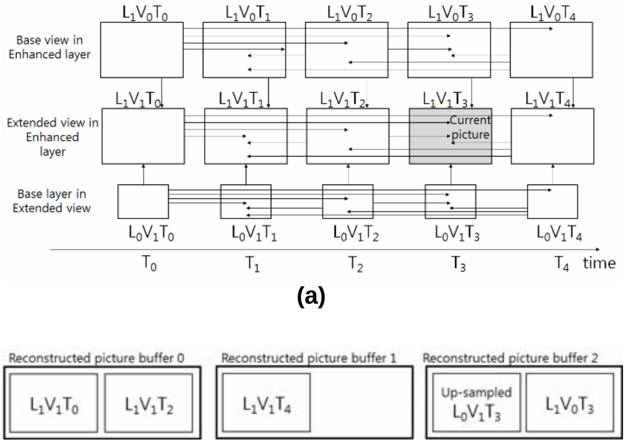


Fig. 5. Random access coding structure of SMHVC using temporal and the proposed adaptive inter-layer and inter-view prediction.

different layer or view, identification numbers of the view and layer should be signaled. If the reference picture comes from a lower spatial layer, up-sampling process should be conducted additionally with the proposed interpolation method. In SMHVC, we also propose a new motion prediction through modification of the AMVP and merge mode in HEVC with consideration of the adaptive inter-layer/inter-view prediction. The proposed motion vector prediction is conducted by adding the inter-layer and inter-view motion vector as additional candidates to AMVP and merge mode and it can improve the accuracy of motion vector prediction. It employs the high motion correlation among layers and views. To evaluate coding performance of the proposed SMHVC, we implemented the consolidated scalable multi-view codec on HEVC reference software and a coded bitstream for the base view in the base layer is completely compatible to HEVC.

3.2 Decoded Picture Buffer (DPB) Management

For the current scalable and multi-view coding standards, the reconstructed pictures of the base layer and the base view are used as reference pictures to improve coding performance of the enhanced layers and extended views, respectively. The SVC does not manage the reconstructed pictures of the base layer in DPB but employs a special decoding process so called intra-BL. In SHVC, the reconstructed picture of the base layer is used to conduct inter-layer prediction instead of intra-BL. This has an advantage preserving coding gain, compared to intra-BL without any change of decoding process by using the same inter-prediction syntaxes i.e. reference index and motion vector. However, due to the resolution difference between the base layer and the enhanced layer in case of spatial scalability, the reconstructed picture and the temporal motion of the base layer are up-scaled before putting the reconstructed picture into DPB. Then, inter-layer prediction conducted with the same fashion of motion estimation and motion vector prediction, respectively. If the inter-layer prediction is conducted, the inter-layer phase vector would not be large because the reference picture of the base layer and the encoding picture of the enhanced layer are at the same time, therefore, zero displacements frequently occur in the most prediction units and we can skip the phase estimation as an encoding issue. However, inter-layer phase shift occurs in down-sampling of the input picture and up-sampling of the reconstructed picture for scalable coding. Accordingly, the proposed SMHVC includes an independent decoded picture buffer management to employ aforementioned inter-layer prediction to improve coding performance. Furthermore, in case of inter-view prediction, the same decoding process can be also employed based on the proposed DPB management algorithm.



(a)

(b)

Fig. 6. An example of the proposed DPB management for the proposed SMHVC (a) Coding order and structure, (b) Decoded picture buffer.

Fig. 6 shows an example of coding structure and decoded picture buffer management for the proposed SMHVC. *L*, *V*, and *T* denote layer, view and time, respectively. As shown in Fig. 6(a), it is assumed that the current coding picture is $L_1V_1T_3$ of the extended view in the enhanced layer at T3 and the coding order is T_0, T_4, T_2, T_1, and T_3. For the current coding picture, it makes use of the pictures that are in the enhanced layer and have smaller Picture Order Count (POC) for forward prediction, i.e. $L_1V_1T_0$ and $L_1V_1T_2$ in reference list 0, as shown in Fig. 5(b). For backward prediction, $L_1V_1T_4$ in reference list 1 which has a larger POC than the current picture can be referenced. Finally, $L_0V_1T_3$ and $L_1V_0T_3$ are stored in reference list 2 with the same POC can be used for adaptive inter-layer and inter-view prediction. However, the reference pictures in the base layer have a lower resolution than the enhanced layer for the spatial scalable coding. Therefore, the reference picture in the base layer should be up-sampled for a reference picture of inter-layer prediction.

In the proposed adaptive inter-layer and inter-view prediction, temporal P, temporal B, inter-layer, inter-view, combined temporal-inter-layer, and combined temporal-inter-view prediction modes are adaptively selected. Note that temporal correlation is generally higher than inter-layer or inter-view correlation, therefore, we restricted prediction modes in SMHVC as above mentioned with consideration of the amount of the mode signal bits. To

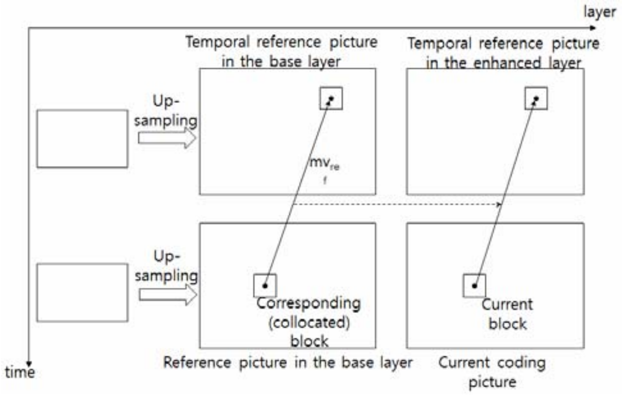


Fig. 7. Up-scaled inter-layer temporal motion vector derivation in spatial scalable coding.

achieve higher coding gain for the picture of the extended view in the enhanced layer, the listed prediction methods are supported using the proposed DPB management in SMHVC.

3.3 Motion Vector Prediction of the Proposed SMHVC

SMHVC supports merge mode and Advanced Motion Vector Prediction (AMVP) as the same with HEVC which derive motions from neighboring blocks based on motion correlation. For inter-layer and inter-view prediction, motions from the reference layer and the reference view can be additionally employed because those reference layer and view are captured at the same time and also highly correlated. In SMHVC, the available motion information from the reference layer or view are derived and employed for motion prediction. To find the corresponding motion information, up-scaling for inter-layer motion derivation and global disparity-based inter-view motion derivation are described in this chapter.

In case of coding a picture in the enhanced layer, the predictor of the temporal motion vector for the current block can be derived from the temporal motion vector of the corresponding block in the base layer. Fig. 7 shows the temporal motion derivation using up-scaling in case of the spatial scalable coding. As shown in Fig. 8, the enhanced layer coding picture employs the up-scaled motion vector of the corresponding block in the base layer. In the base view, the motion vectors are stored in each 4×4 block regardless of the size of CU or PU. Therefore, to employ the temporal motion of the base layer, we need to find the motion vector of the corresponding 4×4 block described as follows.

- Find out the position of the first pixel in 4×4 block (*x*, *y*) in the enhanced layer
- Find the corresponding pixel position (*x'*, *y'*) in the base layer
- Calculate the index of CTU and sub-block index of the corresponding block which include the pixel (*x'*, *y'*)
- Up-scale the motion vector of the corresponding block in the base view

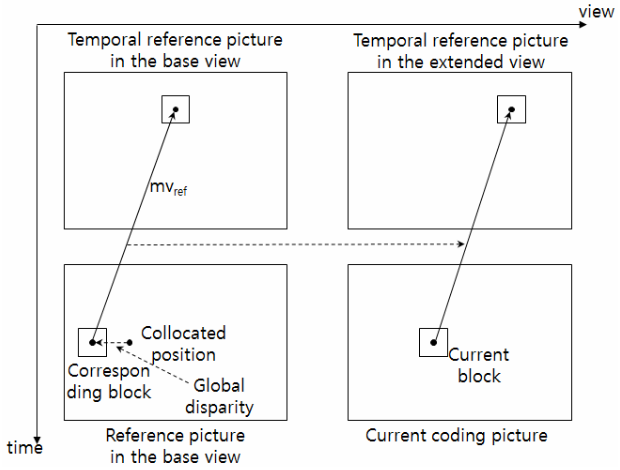


Fig. 8. Inter-view temporal motion vector derivation.

In the proposed SMHVC, the motion vector prediction for the inter-view prediction exploits the predicted motion with consideration of the global disparity as shown in Fig. 8. To predict the motion vector from the base view for the current block in the extended view, the corresponding position should be found in advance. Among pictures captured at the same time by cameras with different positions in multi-view video, the disparity exists. Therefore, by using the disparity, the corresponding position of a block in the reference picture can be found. This predicted inter-view motion vector can be replaced to each other to the predicted inter-layer motion vector for the scalable coding and then, merge mode and AMVP are performed in the same way. In SMHVC, the global disparity vector is explicitly signaled in bitstream. And decoder can calculate the predicted inter-view motion vector candidate as same in encoder. Because the global disparity vector calculation is an encoding issue, it doesn't affect the decoding process. Thus SMHVC calculates the global disparity vector by global disparity estimation on the 1/8 down-sampled picture with consideration of the computational complexity.

3.4 Inter-layer Interpolation Filter

The proposed SMHVC compresses scalable multi-view videos by estimating motion, disparity and phase using the reconstructed reference pictures in the DPB. Especially, all the reference pictures from the base layer for inter-layer prediction in spatial scalable coding should be interpolated to the same resolution of the enhanced layer. In the proposed SMHVC, Discrete Cosine Transform-based Interpolation Filter (DCT-IF) is employed. The filter used for half-pel positions of ×2 up-sampling is the same as that of HEVC. In HEVC, an 8-tap FIR interpolation filter, {-1, 4, -11, 40, 40, -11, 4, -1}, is used for luma 1/2 pixel position. In the proposed SMHVC, we derived DCT-based interpolation filter for 1/3 and 2/3 pixel positions in case of ×1.5 up-sampling, {-1, 4, -9, 26, 53, -12, 5, -1} and {-1, 5, -12, 53, 26, -9, 4, -1} are used, respectively. Chroma

Table 1. Test conditions.

Parameter	Configured value
Coding Mode	Random Access
Maximum CTU size	64
Maximum CU depth	4
Maximum TU depth	3
The number of reference pictures	4
Entropy coding	CABAC
RDO-Q	On

Table 2. The values of quantization parameter.

View/ Layer	QP for Exp. 1	QP for Exp. 2	QP for Exp. 3	QP for Exp. 4
V0L0	22	26	32	36
V0L1	24	28	34	38
V1L0	25	29	35	39
V1L1	27	31	37	41

samples also need to be interpolated and 4-tap filter, {-4, 36, 36, -4}, is used for 1/2 pixel position. For 1/3 and 2/3 pixel position, {-5, 25, 51, -7} and {-7, 51, 25, -5} are applied, respectively.

4. Performance Evaluations of the Proposed Scalable Multi-view Video Codec

In this paper, we proposed scalable multi-view video codec to improve the coding performance. To evaluate the performance, we implemented SMHVC on HM6.1, the reference software of HEVC, and conducted the experiments using full-HD multi-view video sequences recommended by JCT-3V group [31][32].

4.1 Test Conditions for the Performance Evaluation

For performance evaluation, we used JCT-3V multi-view test sequences with full-HD resolution with 25fps (frames per second). For the experiments, we generated the spatial down-sampled video using the down-sampling filter with 0.9πcut-off frequency recommended by JCT-VC. Using the filter, we generated 1/2 and 2/3 down-sampled video with 940×540 and 1280×720 resolution respectively. To evaluate the performance, the experiments performed under Common Test Condition (CTC) [33] and 2-view coding configuration of JCT-3V. Table 1 shows the detailed test condition and Table 2 shows the QPs for each experiment.

In the experiment, the left view is set as the base view and the picture of the base view in the base layer (V0L0) coded first then the picture of the base view in the enhanced layer (V0L1) and the picture of the extended view in the base layer are coded. Lastly, the picture of the extended view in the enhanced layer is coded. During coding of these pictures acquired at the same time, from the QP used for the picture of the base view in the base layer, the QP for multi-view coding is set to +3 and the QP for scalable coding is set to +2.

Table 3. Coding efficiency of SMHVC compared to parallel scalable coding and parallel multi-view coding.

	Parallel scalable coding ×2 spatial scalable coding	Parallel scalable coding ×1.5 spatial scalable coding	Parallel multi-view coding for ×1.5 spatial scalable video
	BD-Bitrate		
GT_Fly	-44.60%	-43.00%	-6.80%
Poznan_Hall2	-23.10%	-14.30%	-17.50%
Poznan_Street	-33.30%	-26.90%	-26.50%
Undo_Dancer	-43.50%	-42.20%	-12.30%
Average	-36.10%	-31.60%	-15.80%

4.2 Experimental Results of the Proposed SMHVC

In this section, the experimental results of SMHVC will be shown. To evaluate the coding performance of SMHVC, the reference software, SHVC, for standardization by JCT-VC was used. In SHVC, interlayer prediction is performed by referencing the corresponding block without further motion estimation. To improve the coding efficiency, SMHVC conduct additional motion estimation i.e. combined prediction, enhanced P and enhanced B, for prediction of more accurate motion. SMHVC is integrated and can provide scalable and multi-view coding. However, by using scalable and multi-view codecs each, the scalable multi-view video services can be provided, e.g. for coding of stereo video, parallel scalable coding can be possible without multi-view coding.

Table 3 shows the coding performance of SMHVC compared to the results of parallel scalable coding and parallel multi-view coding. As shown in Table 3, SMHVC outperforms the parallel scalable coding with about 36% and 32% bitrate reduction in average for ×2 and ×1.5 spatial scalable videos respectively based on Bjøntegaard-Delta (BD)-Bitrate [34]. The parallel scalable coding has an advantage for coding of scalable multi-view video with only scalable codec, however, it cannot exploit the inter-view correlation using inter-view prediction. Therefore, the high compression performance cannot be expected. Another alternative to provide the scalable multi-view video services using only 1 codec is parallel multi-view coding. Parallel multi-view codec compresses multi-view video in each layer without inter-layer prediction. By using the additional inter-layer prediction, SMHVC achieves 16% higher coding gain against parallel multi-view coding as shown in Table 3. However, the coding efficiency of SMHVC compared to parallel multi-view coding is less

Table 4. Coding efficiency of SMHVC compared to scalable-based parallel multi-view coding.

Sequence	BD-Bitrate		
	V0L1	V1L0	V1L1
GT_Fly	7.2%	3.8%	5.0%
Poznan_Hall2	12.7%	8.1%	9.1%
Poznan_Street	7.3%	7.6%	-31.8%
Undo_Dancer	1.3%	6.8%	6.1%
Average	7.1%	6.6%	-2.9%

than which it is compared to parallel scalable coding because it is with absence of removal of inter-view redundancy.

To provide the scalable multi-view video services using both scalable and multi-view codec, scalable-based parallel multi-view coding and multi-view-based parallel scalable coding can be used. The former compresses the base layer and the enhanced layer in the base view first then the layers in the extended view while the latter compresses the base view and the extended view of the base layer first then the view of the enhanced layer. As shown before, parallel multi-view codec outperforms parallel scalable codec. Therefore, in this paper, scalable-based parallel multi-view coding, the former, is used and performed experiments to be compared to SMHVC. Table 4 shows the coding performance of SMHVC compared to scalable-based parallel multi-view codec. As shown in Table 4, SMHVC shows about 2.9% better coding efficiency for the enhanced layer in the extended views while 7.1% and 6.6% worse for the V0L1 and V1L0 respectively. This is because SHVC and 3D-HTM are used to conduct the experiments and were improved during standardization independently. However, scalable-based parallel multi-view codec uses 2 separated codecs and they are not synchronized in coding loop. For coding of scalable multi-view video, the scalable coding has to be performed first prior to the parallel multi-view coding to reference the reconstructed picture and must be synchronized among platforms while SMHVC is synchronized in coding loop and integrated in a platform. Another advantage of SMHVC is that it can additionally exploit the syntax and compress the scalable multi-view video in a single-loop. If SMHVC includes additional coding tools in 3D-HTM, the better coding performance can be easily expected.

Table 5 shows the encoding computational complexity of SMHVC compared to parallel scalable codec and parallel multi-view codec respectively. SMHVC shows higher encoding complexity than both parallel scalable codec and parallel multi-view codec. SMHVC uses more reconstructed picture i.e. inter-layer and inter-view reference pictures and includes additional inter-layer motion estimation for scalable coding. Therefore, SMHVC needs more computational complexity. As shown in Table 5, encoding time is increased by about 66% and 59% when encoding using SMHVC in ×2 and ×1.5 spatial scalable coding with about 34% bitrate reduction. SMHVC shows about 8% complexity increment compared to parallel multi-view codec in case of ×1.5 spatial scalable coding with 16% bitrate gain.

Table 5. Encoding complexity of SMHVC compared to parallel scalable codec.

	Parallel scalable coding ×2 spatial scalable coding	Parallel scalable coding ×1.5 spatial scalable coding	Parallel multi-view coding for ×1.5 spatial scalable video
GT_Fly	153.5%	145.5%	104.7%
Poznan_Hall2	168.3%	159.9%	115.2%
Poznan_Street	185.8%	178.9%	115.0%
Undo_Dancer	158.1%	155.2%	98.7%
Average	166.0%	159.2%	108.2%

SMHVC needs additional memory to store the reference pictures. In the multi-view coding, SMHVC performs additional inter-layer prediction for combined prediction compared to parallel multi-view codec. If the resolution of the enhanced layer is full-HD, about 16.7% and 6.3% of additional memory is required in level 4.x and 5.x of HEVC respectively. In case of scalable coding, if the resolution of the base layer is 720p, about 11.2% and 4.2% of additional memory is required in level 4.x and 5.x. SMHVC integrated on a platform and most of the conventional tools can be shared and reusable for both scalable and multi-view coding while the scalable-based parallel multi-view codec needs both scalable codec and multi-view codec on multiple platform. This kind of codec on multiple platforms needs to transfer the data from one to another as a big overhead and is restricted to exploit syntax derivation.

5. Conclusion

In this paper, we proposed SMHVC overcome the high cost of existence of multiple codecs in a device. To integrate the scalable codec and multi-view codec on a platform, we defined high-level syntax to identify the view / layer and proposed the combined inter-prediction, motion prediction and DPB management to improve the coding performance. To reference the reconstructed picture of the base view in the base layer, the reference picture is managed in the separated picture buffer to combine the inter-view and inter-layer prediction. For motion prediction of the enhanced layer, up-scaled inter-layer motion prediction with consideration of spatial resolution ratio and global disparity-based inter-view motion prediction were proposed. To interpolate the reconstructed picture of the base view, DCT-based ×1.5 interpolation filter for 1/3 and 2/3 fractional positions.

To evaluate the coding performance, compared to scalable multi-view codec, the proposed SMHVC achieved about 34% coding gain with about 62% complexity increment compared to parallel scalable video codec and about 16% coding gain with about 16% complexity increment compared to parallel multi-view video codec. Lastly, we expect that if the proposed integrated scalable multi-view video codec is updated based on the new version of the reference software, it will show further improvement of coding efficiency.

Acknowledgement

This research was partly supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2014R1A2A1A11052210) and the Commercializations Promotion Agency for R&D Outcomes(COMPA) funded by the Ministry of Science, ICT and Future Planning(MISP).

References

[1] T. Wiegand, G.-J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE trans. on Circuits and Systems for Video Technology*, vol. 13, no.7, pp. 560-576, July 2003. Article (CrossRef Link)

[2] A. Tamhankar and K.-R. Rao, "An Overview of H.264/MPEG-4 part 10," *4th EURASIP Conference on Video/Image Processing and Multimedia Communications*, vol. 1, pp. 1-51, June 2003. Article (CrossRef Link)

[3] T. Wiegand, G. Sullivan, and A. Luthra, "Draft ITU-T recommendation and final draft international standard of joint video specification," *JVT Document*, JVT-G050r1, Geneva, CH, May 2003. Article (CrossRef Link)

[4] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand, "Constrained inter-layer prediction for single-loop decoding in spatial scalability," *IEEE International Conference on Image Processing*, pp. 870-873, Genoa, IT, Sep. 2005. Article (CrossRef Link)

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007. Article (CrossRef Link)

[6] M. Wien, H. Schwarz, and T. Oelbaum, "Performance analysis of SVC," *IEEE trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1194-1203, Sep. 2007. Article (CrossRef Link)

[7] K. De Wolf, D. De Schrijver, S. De Zutter, and R. Van de Walle, "Analysis and coding performance of inter-layer prediction," *International Symposium on Signal Processing and Its Applications*, pp. 1-4, Sharjah, AE, Feb. 2007. Article (CrossRef Link)

[8] C.-S. Park, S.-Jin. Baek, M.-S. Yoon, H.-K. Kim, and S.-J. Ko, "Selective inter-layer residual prediction for SVC-based video streaming," *IEEE Trans. Consumer Electronics.*, vol. 55, no. 1, pp. 235-239, Feb. 2009. Article (CrossRef Link)

[9] Y. Ho and K. Oh, "Overview of Multi-view Video Coding," *6th EURASIP Conference focused on Speech and Image Processing*, vol. 1, pp. 1-51, June 2007. Article (CrossRef Link)

[10] ISO/IEC JTC1/SC29/WG11, "Requirements on Multi-view Video Coding v.7," *The 77th MPEG meeting*, N8218, July 2006. Article (CrossRef Link)

[11] ISO/IEC MPEG & ITU-T VCEG, "Joint Draft 1.0 on Multi-view Video Coding," *The JVT meeting*, JVT-U209, Nov. 2006. Article (CrossRef Link)

[12] K. McCann, W.-J Han, I.-K Kim, J-H Min, E. Alshina, A. Alshin, T. Lee, J. Chen, V. Seregin, S. Lee, Y.-M Hong, M.-S. Cheon, and N. Shlyakhov, "Samsung's Response to the Call for Proposals on Video Compression Technology," *The 1st JCT-VC meeting*, JCTVC-A124, Dresden, DE, Apr. 2010. Article (CrossRef Link)

[13] M. Winken, S. Bobe, B. Bross, P. Helle, T. Hinz, H. Kirchhoffer, H. Lakshman, D. Marpe, S. Oudin, M. Preib, H. Schwarz, M. Siekmann, K. Suhring, and T. Wiegand, "Description of video technology proposal by Fraunhofer HHI," *The 1st JCT-VC meeting*, JCTVC-A116, Dresden, DE, Apr. 2010. Article (CrossRef Link)

[14] T. Wiegand, J. Ohm, G. Sullivan, W. Han, R. Joshi, T. Tan, and K. Ugur, "Special section on the joint call for proposals on high efficiency video coding (HEVC) standardization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1661-1666, Dec. 2010. Article (CrossRef Link)

[15] J. Kim and et al., "TE9: Report on large block structure testing," *The 3rd JCT-VC meeting*, JCT-VC C067 (m18090), Guangzhou, CN, Oct. 2010. Article (CrossRef Link)

[16] B. Bross, W.-J. Han, J.-R. Ohm, G.J. Sullivan, and T. Wiegand, "WD6: Working draft 6 of high-efficiency video coding," *The 8th JCT-VC meeting*, JCT-VC H1003, San Jose, CA, Jan. 2012. Article (CrossRef Link)

[17] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[18] I. Kim, J. Min, W. Han, and J. Park, "Block Partitioning Structure in HEVC Standard," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[19] P. Helle, S. Oudin, B. Bross, D. Marpe, M. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand, "Block Merging for Quadtree-based Partitioning in HEVC," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[20] J. Lainema, F. Bossen, W. Han, J. Min, and K. Ugur, "Intra Coding of the HEVC standard," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[21] R. Sjoberg, Y. Chen, A. Fujibayashi, M. Hanuksela, J. Samuelsson, T. Tan, Y. Wang, and S. Wenger, "Overview of the HEVC High-level syntax and reference picture management," *IEEE trans. on*

Circuits and Systems for Video Technology, to be published, 2012. Article (CrossRef Link)

[22] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, "HEVC deblocking filter," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[23] C. Fu, E. Alshina, Y. Huang, C. Chen, C. Tsai, C. Hsu, S. Lei, J. Park, and W. Han, "Sample adaptive offset in HEVC Standard," *IEEE trans. on Circuits and Systems for Video Technology*, to be published, 2012. Article (CrossRef Link)

[24] A. Luthra, "Scalable enhancement requirements for HEVC," *Document of Joint Collaborative Team on Video Coding*, JCTVC-E502, Geneva, CH, Mar. 2011. Article (CrossRef Link)

[25] A. Luthra, J. Ohm, and J. Ostermann, "Draft call for proposals on the scalable video coding extensions of HEVC," *The 98th MPEG meeting*, N12398, Geneva, CH, Nov. 2011. Article (CrossRef Link)

[26] H. Choi, J. Nam, D. Sim, and I. V, Bajić, "Scalable video coding based on high efficiency video coding (HEVC)," *IEEE Pacific Rim Conference on Communications*, Computers and Signal Processing, pp.346-351, Victoria, CA, Aug. 2011. Article (CrossRef Link)

[27] H. Choi, J. Nam, and D. Sim, "Scalable structures and inter-layer predictions for HEVC scalable extension," *The 6th JCT-VC meeting*, JCT-VC F096 (m20509), Torino, IT, Jul. 2011. Article (CrossRef Link)

[28] ISO/IEC JTC1/SC29/WG11, "Applications and requirements on 3D video coding," *The 96th MPEG meeting*, N12035, Geneva, CH, Mar. 2011. Article (CrossRef Link)

[29] ISO/IEC JTC1/SC29/WG11, "Call for proposals on 3D video coding technology," *The 96th MPEG meeting*, N12036, Geneva, CH, Mar. 2011. Article (CrossRef Link)

[30] J. Nam, W. Lim, D. Sim, and I. Bajić, "Multi-view video coding based on high efficiency video coding (HEVC)," *The 26th International Technical Conference on Circuits and Syst., Computers, and Communications*, pp. 169-172, Gyeongju, KR, Jun. 2011.

[31] http://hevc.kw.bbc.co.uk/trac/browser/tags/HM-6.1, HM6.1 S/W. Article (CrossRef Link)

[32] K. McCann, B. Bross, S. Sekiguchi, and W. Han, "High efficiency video coding (HEVC) test model 6 (HM 6) encoder description," and T. Wiegand, "WD6: Working draft 6 of high-efficiency video coding," *The 8th JCT-VC meeting*, JCT-VC H1002, San Jose, CA, Jan. 2012. Article (CrossRef Link)

[33] F. Bossen, "Common test conditions and software reference configurations," *The 8th JCT-VC meeting*, JCT-VC H1100, San Jose, CA, Jan. 2012. Article (CrossRef Link)

[34] G. Bjøntegaard, "Calculation of average PSNR differences between RD-Curves," *ITU-T SG16/Q.6*, VCEG-M33, Austin, TX, Apr. 2001. Article (CrossRef Link)

Woong Lim was born in Busan Metropolitan city, Republic of Korea, in 1981. He received the B.S. and M.S. degree in Computer Engineering from Kwangwoon University, Seoul, Korea, in 2008 and 2010, respectively. Now, he is working toward Ph.D. degree at the same University. His current research interests are high-efficiency video compression, Scalable video coding, multiview video coding and computer vision.

Junghak Nam was born in SangJu City, KyungBuk Province, Republic of Korea, in 1979. He entered the Kwangwoon University in 1998, which received the B.S. and M.S. degrees in Computer Engineering, in 2006 and 2008, respectively. And he received ph.D is currently a Ph.D. in 2013. He is currently working for LG Electronics since 2013. His interesting fields are H.264/AVC, HEVC, and parallel processing

Donggyu Sim was born in Chungchung Province, Korea, in 1970. He received the B.S. and M.S. degrees in Electronic Engineering from Sogang University, Seoul, Korea, in 1993 and 1995, respectively. He also received Ph.D. degree at the same University in 1999. He was with the Hyundai Electronics Co., Ltd. from 1999 to 2000, where 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 researched on the ultrasound image analysis and parametric video coding. He joined the Department of Computer Engineering at Kwangwoon University, Seoul, Korea, in 2005 as an Associate Professor. He was elevated to an IEEE Senior Member in 2004. His current research interests are image processing, computer vision, and video communication.