IEEK Transactions on Smart Processing and Computing

Low-complexity generalized residual prediction for SHVC

Kyeonghye Kim, Jiwoo Ryu, and Donggyu Sim

Department of Computer Engineering, Kwangwoon University / Seoul, Republic of Korea
{kimkhye428, jwRyu, dgsim}@kw.ac.kr

***** Corresponding Author: Donggyu Sim

Received July 14, 2013; Revised July 29, 2013; Accepted September 12, 2013; Published December 31, 2013

***Abstract*:** This paper proposes a simplified generalized residual prediction (GRP) that reduces the computational complexity of spatial scalability in scalable high efficiency video coding (SHVC). GRP is a coding tool to improve the inter prediction by adding a residual signal to the inter predictor. The residual signal was created by carrying out motion compensation (MC) of both the enhancement layer (EL) and up-sampled reference layer (RL) with the motion vector (MV) of the EL. In the MC process, interpolation of the EL and the up-sampled RL are required when the MV of the EL has sub-pel accuracy. Because the up-sampled RL has few high frequency components, interpolation of the up-sampled RL does not give significantly new information. Therefore, the proposed method reduces the computational complexity of the GRP by skipping the interpolation of the up-sampled RL. The experiment on SHVC software (SHM-2.0) showed that the proposed method reduces the decoding time by 10 % compared to conventional GRP. The BD-rate loss of the proposed method was as low as 1.0% on the top of SHM-2.0.

***Keywords*:** Scalable video coding, HEVC, SHVC, Generalized residual prediction

1. Introduction

With the recent advances of the network and smart devices, the production and consumption of high resolution videos are increasing rapidly. As the preceding video codecs are unsuitable for compressing such high resolution videos efficiently, the Video Coding Experts Group (VCEG) and Moving Picture Experts Group (MPEG) established a Joint Collaborative Team on Video Coding (JCT-VC) to initiate a standardization for High Efficiency Video Coding (HEVC), which is the next-generation video codec. In January 2013, the Final Draft International Standard (FDIS) of HEVC was released and the standardization of Scalable HEVC (SHVC) was started [1]. SHVC is an extension of HEVC standard to add spatial and quality scalability to HEVC on the top of the temporal scalability of HEVC [2].

SHVC is a multi-layered video codec standard with inter-layer prediction tools. In addition to encoding each layers separately, the inter-layer prediction tools exploit the inter-layer redundancy among multiple layers to increase the coding efficiency. As a result, the high computational complexity of SHVC has become an obstruction for the real-time service of SHVC for commercialization [3]. To reduce the computational complexity of SHVC, this paper proposes a simplified generalized residual prediction (GRP) for spatial scalability.

This paper is organized as follows. Section 2 introduces SHVC and GRP, and Section 3 presents the proposed simplified GRP. Section 4 analyzes the computational complexity and Bjontegaard Distortion-rate (BD-rate), and Section 5 concludes the paper.

2. Scalable HEVC (SHVC)

The coding structure of SHVC was composed of a base layer (BL) and enhancement layers (ELs), as shown in Fig. 1. The input video of SHVC is a multi-layered video that is generated by down-sampling the highest EL in terms of the spatial, temporal or quality scale. In the multi-layered video, the highest EL and BL have the finest and coarsest spatial, temporal or quality resolution, respectively. For example, in spatial scalability, if the highest EL has a spatial resolution of 1920x1080, then down-sampled videos of the highest EL with resolutions of, for example,

This work was supported by the IT R&D program of MKE/KEIT. [10039199, A Study on Core Technologies of Perceptual Quality based Scalable 3D Video Codecs]

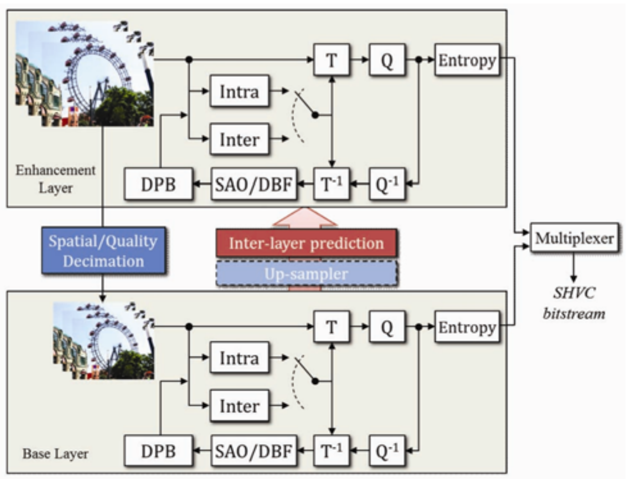

Fig. 1. Block diagram of the SHVC encoder.

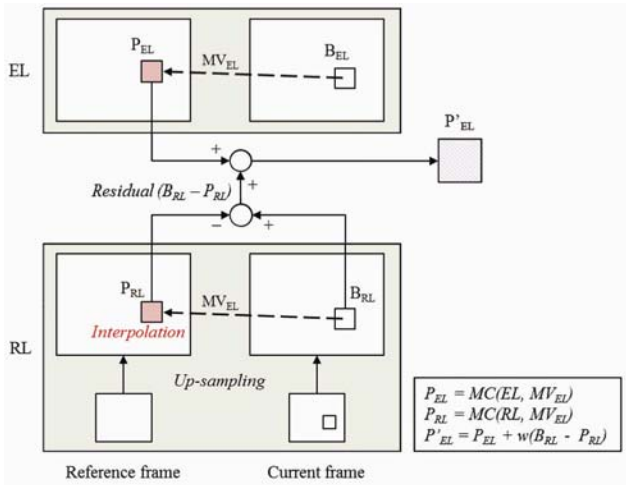

Fig. 2. Generalized residual prediction for spatial scalability.

960x540 and 480x240, can be used as the lower layers of scalable video. In this case, BL would be a layer with a resolution of 480x240.

BL of SHVC was coded in the same way as HEVC, and the enhancement layers are coded using inter-layer prediction tools. In the inter-layer prediction, an enhancement layer refers to the reference layer (RL), which could be either the BL or another enhancement layer on a lower scale. The enhancement layer may refer to the texture or syntax information of the RL (i.e. inter-layer texture prediction and inter-layer syntax prediction). The current SHVC reference software includes the prediction tools, which refer to the texture and motion information of collocated block in the RL. In spatial scalability, the inter-layer prediction is performed after texture and motion information of the RL are up-scaled according to the resolution of the enhancement layer. In texture up-sampling, DCT-IF based 8-tap filter is used.

2.1 Generalized residual prediction (GRP)

GRP is a technique in inter-layer prediction that predicts the residual signals of the EL using texture information of the RL. Although the conventional inter-layer prediction uses the collocated block in the up-sampled RL as a predictor, the inter-layer prediction with GRP improves the predictor by adding a predicted residual signal on the EL predictor.

Fig. 2 presents a typical process of GRP for spatial scalability [4]. The top rightmost frame represents the current frame, where the prediction is performed. The prediction signal (P'_{EL}) is a weighted sum of a predictor in EL (P_{EL}) and the residual predictor (B_{RL}-P_{RL}). w is the weight value used to construct the P'_{EL} (w = 0, 0.5, or 1). B_{RL} is a collocated block of B_{EL} in the RL. P_{EL} and P_{RL} are the motion-compensated signals from the EL and up-sampled RL with a motion vector MV_{EL}. The process of GRP includes a great deal of interpolations. In addition to the up-sampling of the RL, interpolation is needed in MC for P_{EL} and P_{BL} if the MV_{EL} has sub-pel accuracy. Such increased interpolation results in the additional computational burden on SHVC. For practical applications of GRP in SHVC, simplification of the GRP is required.

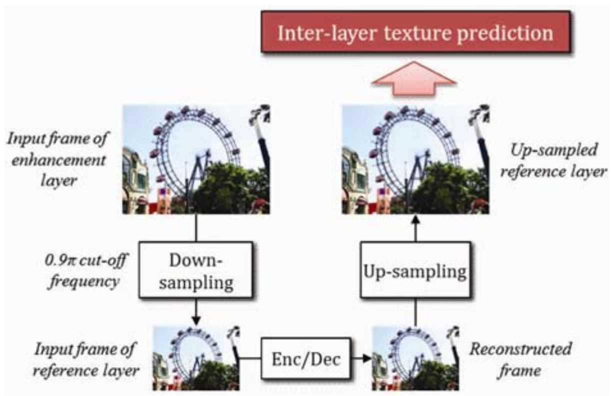

Fig. 3. Up-sampled RL generation.

3. The proposed simplified GRP

This paper proposes the simplified GRP for spatial scalable SHVC. The simplification was made by exploiting the properties of SHVC, where all RLs are created by down-sampling the higher ELs.

Fig. 3 shows the framework of the up-sampled RL generation. First, the input frame of the enhancement layer is down-sampled with a downsampling filter with a cut-off frequency of 0.9π to produce a frame with a lower layer [5]. The reconstructed frame of the lower layer is produced after coding the down-sampled frame. The up-sampled RL is created by up-sampling the reconstructed lower frame to be used as a reference in inter-layer prediction.

The most high-frequency textures in the up-sampled RL are removed by down-sampling because the up-sampled RL is obtained with the up-sampled layer of the RL. Because the up-sampled RL has a low cut-off frequency, the intensities among the neighboring pixels do not change rapidly. Therefore, this paper presumed that the sub-pel motion compensation in the up-sampled RL would have a minor influence on the prediction performance

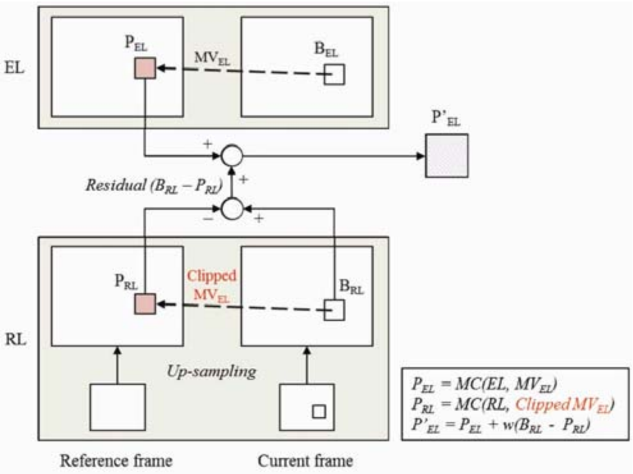

Fig. 4. Proposed generalized residual prediction.

because interpolation of the up-sampled RL does not produce significantly new information.

To reduce the computational complexity of the GRP while maintaining its prediction performance, the proposed simplified GRP forgoes interpolation of the up-sampled RL. If MV_{EL} has sub-pel accuracy, the proposed algorithm clips MV_{EL} to an integer unit to skip the interpolation, as shown in Fig. 4 [6]. The interpolation in HEVC uses 8-tap filter coefficients. Accordingly, 7 addition and 8 multiplication operations are needed to construct a one half-pixel. By removing the interpolation, only the pixel copy process is needed to construct a predictor and it can reduce the computational complexity of GRP.

4. Performance Evaluation

SHVC common test condition [7] sequences are used to evaluate the proposed algorithm. Table 1 provides details of the test sequences and QP settings. All the tests were conducted under a random access configuration. For a comparative study, three tests with different sets of inter-layer prediction tools were conducted, as shown in Table 2. Test 1 shows an anchor test with SHVC software SHM2.0. Tests 2 and 3 perform a conventional GRP and the proposed simplified GRP, respectively. Tables 3 and 4 compare the BD-rate and coding time for the three test cases.

As shown in Table 3, the conventional GRP has 1.4% and 2.3% of the BD-rate reduction for 2x and 1.5x spatial scalability (respectively) with an approximately 28% and 20% increment in the encoding and decoding time. For some cases, the conventional GRP increased the decoding time by up to 28%.

On the other hand, as shown in Table 4, the proposed simplified GRP has a significantly lower encoding time complexity, as low as the anchor, whereas the BD-rate gain is lower than the conventional GRP. The BD-rate gain of the simplified GRP for both 2x and 1.5x spatial scalability cases is 1.0%. Note that the BD-rate loss of the simplified GRP compared to the conventional GRP was 0.4% and 1.3% for the 2x and 1.5x spatial scalability cases,

Table 1. Test conditions for spatial scalability.

Seq.	Scalability	Resolution		QP	
		BL	EL	QP_{BL}	QP_{EL}
Class B	2x spatial scalability	960×540	1920×1080	22 26 30 34	QP_{BL}+ 0 QP_{BL}+ 2
	1.5x spatial scalability	1280×720			

Table 2. Summary of the test tools.

No.	Summary
Test 1	SHM2.0 software (Anchor)
Test 2	SHM2.0-based GRP with interpolation
Test 3	SHM2.0-based GRP without interpolation (Proposed method)

Table 3. RD and complexity performance of the conventional GRP.

Seq.	2x spatial scalability			1.5x spatial scalability		
	BD-rate	ENC Time	DEC Time	BD-rate	ENC Time	DEC Time
Kimono	-2.3%	114%	118%	-1.9%	111%	119%
ParkScene	-1.0%	117%	125%	-2.4%	114%	117%
Cactus	-2.2%	135%	124%	-3.9%	138%	117%
Basketball Drive	-1.3%	139%	121%	-1.9%	133%	115%
BQTerrace	-0.4%	136%	128%	-1.6%	152%	123%
Average	-1.4%	128%	123%	-2.3%	129%	118%

Table 4. RD and complexity performance of the proposed simplified GRP.

Seq.	2x spatial scalability			1.5x spatial scalability		
	BD-rate	ENC Time	DEC Time	BD-rate	ENC Time	DEC Time
Kimono	-1.9%	108%	108%	-1.2%	107%	107%
ParkScene	-0.5%	104%	110%	-0.7%	105%	107%
Cactus	-1.7%	115%	109%	-2.2%	112%	110%
Basketball Drive	-0.7%	105%	108%	-0.6%	112%	108%
BQTerrace	0.0%	107%	111%	-0.3%	115%	115%
Average	-1.0%	108%	109%	-1.0%	110%	109%

respectively. Although the BD-rate loss of the simplified GRP is relatively low for 2x spatial scalability, some BD-rate loss is observed on 1.5x spatial scalability.

For a comparative study, coding gain and complexity of the relevant study was also investigated. Table 5 lists the RD and complexity performance of the simplified GRP proposed in the JCT-VC contribution document JCTVC-M0260 [8]. This simplified GRP method uses a bilinear interpolation on MC and uses a 4-tap filter for upsampling. Compared to the proposed method, this method shows a higher BD-rate gain of 1.9% on both the x1.5 and x2 spatial scalability, but its complexity reduction appears to be rather insignificant, particularly for the decoder. Note

Table 5. RD and complexity performance of JCTVC-M0260.

Seq.	2x spatial scalability			1.5x spatial scalability		
	BD-rate	ENC Time	DEC Time	BD-rate	ENC Time	DEC Time
Kimono	**-2.8%**	113%	**121%**	**-2.0%**	**105%**	**125%**
ParkScene	-1.3%	119%	117%	-1.5%	113%	121%
Cactus	-3.2%	117%	122%	-3.2%	112%	126%
Basketball Drive	-1.8%	114%	121%	-1.9%	109%	123%
BQTerrace	-0.6%	121%	119%	-0.8%	118%	121%
Average	-1.9%	117%	120%	-1.9%	117%	120%

Table 6. The numbers of additions and multiplications in interpolation for residual prediction.

	Addition	Multiplication	Total
The number of operation	2,618,726,432	**2,992,830,208**	**5,611,556,640**

that the decoder time complexity of JCTVC-M0260 increased by 20% for both scalabilities, which is comparable to the conventional GRP.

Table 6 lists the mean number of operations, addition and multiplication, of interpolation in conventional GRP. The evaluation was conducted at the decoder side for the "BasketballDrive" sequence in 2x spatial scalability. As shown in Table 6, the conventional GRP requires approximately 5 million additional operations. The proposed GRP can reduce the computational complexity by removing the addition and multiplication operations, i.e. interpolation.

5. Conclusion

The residual prediction is efficient in improving the RD performance in SHVC. On the other hand, simplification of the residual prediction for SHVC is imperative, because the computational load of the residual prediction is relatively high. To reduce the computational complexity of GRP in SHVC, this paper proposed removing the interpolation based on the amount of information in the sampled data. The performance evaluation was conducted on SHM2.0 under common test conditions. With the simplified GRP, the encoding and decoding time were significantly lower than the conventional GRP, by 1.0% in BD-rate gain, compared to SHM2.0 without GRP. In addition, the proposed GRP decreased by approximately 5 million of the arithmetic operations in the GRP process.

References

[1] B. Bross et al., "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Last Call)," ISO/IEC JTC1/SC29 WG11 m28292 (JCTVC-L1003), Geneva, CH, Jan. 2013. Article (CrossRef Link)

[2] J. Chen et al., "SHVC Working Draft 3," ISO/IEC JTC1/SC29 WG11 m30670 (JCTVC-N1008), Vienna, AT, Jul. 2013. Article (CrossRef Link)

[3] E. Alshina et al., "AhG17: complexity analysis of SHM2.0," ISO/IEC JTC1/SC29 WG11 m29777 (JCTVC-N0150), Vienna, AT, Jul. 2013. Article (CrossRef Link)

[4] X. Li et al., "TE3: Results of Test 4.6.2.1 on generalized residual prediction," ISO/IEC JTC1/SC29 WG11 m27406 (JCTVC-L0078), Geneva, CH, Jan. 2013. Article (CrossRef Link)

[5] InterDigital Communications, "Downsampling filters for anchor generation for scalable extensions of HEVC," ISO/IEC JTC1/SC29 WG11 m24499, Geneva, CH, May 2012. Article (CrossRef Link)

[6] K. Kim et al., "Generalized residual prediction with motion vector clipping for SHVC," ISO/IEC JTC1/SC29 WG11 m29726 (JCTVC-N0106), Vienna, AT, Jul. 2013. Article (CrossRef Link)

[7] X. Li et al., "Common SHM test conditions and software reference configurations," ISO/IEC JTC1/SC29 WG11 m29547 (JCTVC-M1009), Incheon, KR, Apr. 2013. Article (CrossRef Link)

[8] X. Li et al., "SCE3: Results of Test 3.3 on generalized residual prediction," ISO/IEC JTC1/SC29 WG11 m28717 (JCTVC-M0260), Incheon, KR, Apr. 2013. Article (CrossRef Link)

Kyeonghye Kim received her B.S. degree in Computer Engineering from Kwangwoon University (KWU), Seoul, Republic of Korea, in 2012. She is working toward a M.S. degree in Computer Engineering at the same University. Since 2011, she has been a Research Assistant with the Department of Computer Engineering, KWU. Her research interests include video signal processing and scalable video coding.

Jiwoo Ryu received his B.S. Degree in Computer Engineering from Kwangwoon University (KWU) in Seoul, Republic of Korea, in 2011. His field of research includes next-generation video coding and perceptual-quality based video coding.

Donggyu Sim received his B.S. and M.S. degrees in Electronic Engineering from Sogang University, Seoul, Republic of Korea, in 1993 and 1995, respectively. He also received his Ph.D. degree at the same University in 1999. He worked at 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 ultrasound image analysis and parametric video coding. He joined the Department of Computer Engineering at Kwangwoon University, Seoul, South Korea, in 2005 as an Associate Professor. He was elevated to an IEEE Senior Member in 2004. His current research interests include image processing, computer vision, video communication and video coding.