An SAD-Based Selective Bi-prediction Method for Fast Motion Estimation in High Efficiency Video Coding

Jongho Kim, DongSan Jun, Seyoon Jeong, Sukhee Cho, Jin Soo Choi, Jinwoong Kim, and Chieteuk Ahn

As the next-generation video coding standard, High Efficiency Video Coding (HEVC) has adopted advanced coding tools despite the increase in computational complexity. In this paper, we propose a selective bi-prediction method to reduce the encoding complexity of HEVC. The proposed method evaluates the statistical property of the sum of absolute differences in the motion estimation process and determines whether bi-prediction is performed. A performance comparison of the complexity reduction is provided to show the effectiveness of the proposed method compared to the HEVC test model version 4.0. On average, 50% of the bi-prediction time can be reduced by the proposed method, while maintaining a negligible bit increment and a minimal loss of image quality.

Keywords: Bi-prediction, HEVC, motion estimation (ME), sum of absolute differences (SAD).

Manuscript received Mar. 22, 2012; revised May 14, 2012; accepted June 1, 2012.

The work was supported by the IT R&D program of KCC/KCA [KCA-2011-10-912-02-001, Development and Standardization of Terrestrial Stereoscopic 3DTV Broadcasting System technology].

Jongho Kim (phone: +82 42 860 5144, pooney@etri.re.kr), DongSan Jun (dschun@etri.re.kr), Seyoon Jeong (jsy@etri.re.kr), Sukhee Cho (shee@etri.re.kr), Jin Soo Choi (jschoi@etri.re.kr), and Jinwoong Kim (jwkim@etri.re.kr) are with the Broadcasting & Telecommunications Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Chieteuk Ahn (ahnc@yonsei.ac.kr) is with the School of Integrated Technology, Yonsei University, Incheon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.12.0112.0186

I. Introduction

As the demand for high-quality video services such as ultra high definition (UHD) is increasing and the available bandwidth for transferring high-resolution video remains limited, a new video compression standard with high coding performance is desired. The ISO-IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) recently formed the Joint Collaborative Team on Video Coding (JCT-VC), aiming to develop the next-generation video coding standard, called High Efficiency Video Coding (HEVC). Since JCT-VC requested a call for proposals on HEVC in January, 2010 [1], [2], it has deployed various coding tools that can provide higher coding efficiency than the state-of-the-art video coding standards. This improvement has mainly been realized through a highly flexible hierarchy of unit representation, which includes three block concepts: a coding unit (CU), prediction unit (PU), and transform unit (TU). This separation of the block structure into three different concepts allows each to be optimized according to its role; a CU is a macroblock-like unit that supports region splitting in a manner similar to a conventional quad-tree, a PU supports non-square motion partition shapes for motion compensation, and a TU allows various transform sizes to be defined independently from the PU size. In addition, there are several advanced tools, such as more sophisticated intra prediction, advanced motion vector prediction (AMVP), block merge mode, and adaptive loop filtering (ALF). While HEVC focuses on achieving high coding efficiency through these tools, its computational complexity has been dramatically increased.

In particular, inter prediction demands a heavy complexity burden of up to 80% in the entire encoding process, as shown

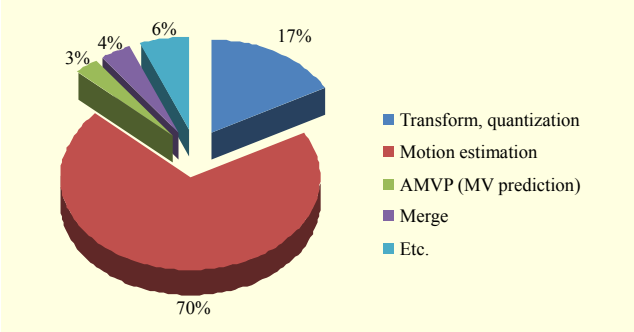

Fig. 1. Percentage of computation time for inter prediction in HEVC hierarchical-B structure.

in Fig. 1. Because this is mainly caused by motion estimation (ME), it is essential to reduce the heavy complexity of the ME.

Although several methods have been proposed to reduce the encoding complexity in conventional video coding standards [3]-[5], they have only been focused on uni-prediction and have been deployed on bi-prediction without adaptive modification. While bi-prediction occupies over 30% of the ME time, there are few fast algorithms considering the characteristics of the bi-prediction structure. Because video encoders for broadcasting services support a bi-prediction structure to obtain better coding performance of inter prediction, the efficient complexity reduction of bi-prediction is quite important while maintaining almost the same rate-distortion (RD) performance.

The remainder of this paper is organized as follows. In section II, we analyze the sum of absolute differences (SAD) properties according to the best prediction mode and various CU sizes. The proposed method is described in section III. Finally, experiment results and some conclusions are given in sections IV and V, respectively.

II. Analysis of SAD Property in Inter Prediction

When enabling bi-prediction, ME is performed according to four prediction modes: forward prediction mode (*list0*), backward prediction mode (*list1*), bi-prediction mode, and merge mode. In HEVC, merge mode can be considered a kind of special inter prediction in which no motion vectors are transmitted; however, it is different from skip mode in terms of residual signal transmission. Bi-prediction enhances the inter-frame coding performance and allows the encoder to search for a better reference block from forward and backward pictures [6], [7]. In particular, it is efficient for video sequences having scene or illumination changes, camera panning, zoom-in/out, and abrupt objects. For CU blocks coded using the bi-prediction mode as the best prediction mode, we observe that the SAD distribution of the uni-prediction mode, such as

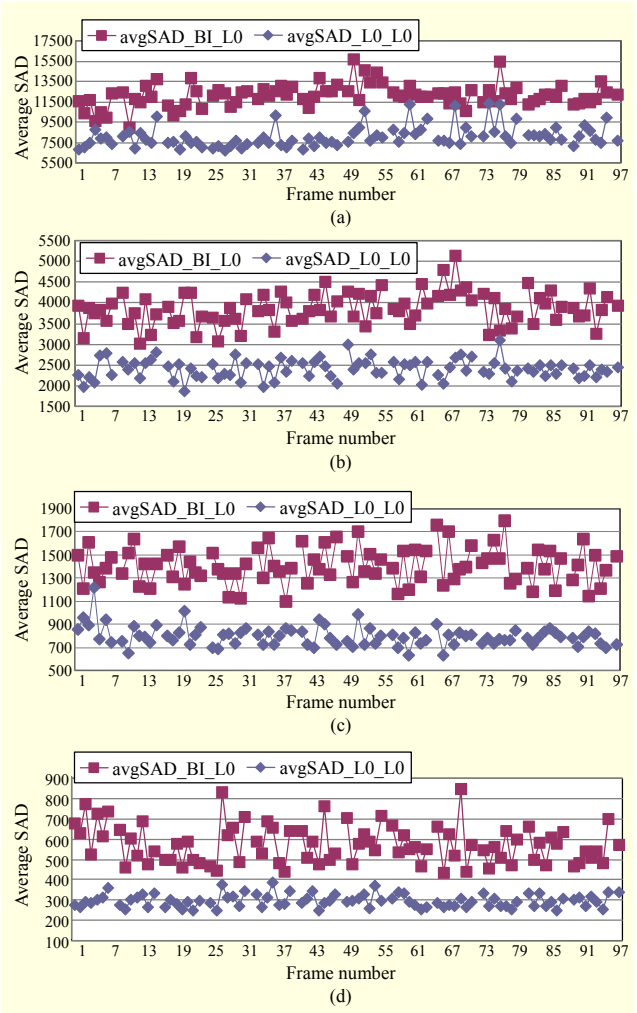

Fig. 2. Average SAD of forward prediction according to best prediction mode and various CU sizes: (a) 64×64, (b) 32×32, (c) 16×16, (d) 8×8.

forward or backward prediction, is different from that of the bi-prediction mode.

We confirm the results of our observation using the following experiments under the common test conditions described in [8]. First, after encoding one frame, we compute the average SAD per frame obtained from the specific uni-prediction mode according to the best prediction mode. This is denoted as *avgSAD_A_B*, where *A* and *B* indicate the best prediction mode and the specific uni-prediction mode, respectively. For example, *avgSAD_BI_L0* indicates the average SAD of the forward prediction calculated from the CU blocks determined as the bi-prediction mode in a frame. Figure 2 shows the average SAD of forward prediction according to the best prediction mode and various CU sizes. The average SAD of forward prediction, such as *avgSAD_BI_L0*, has a larger SAD compared to *avgSAD_L0_L0*. Similarly, Fig. 3 shows the average SAD of backward prediction according to

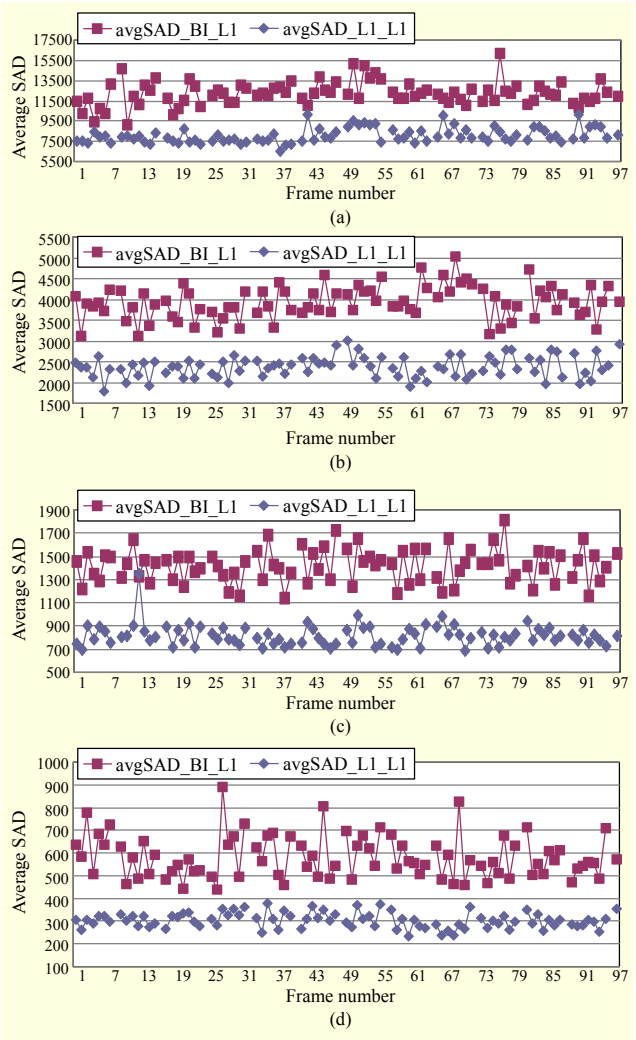

Fig. 3. Average SAD of backward prediction according to best prediction mode and various CU sizes: (a) 64×64, (b) 32×32, (c) 16×16, (d) 8×8.

the best prediction mode and various CU sizes. The average SAD of backward prediction, such as *avgSAD_BI_L1*, has a larger SAD compared to *avgSAD_L1_L1*.

As shown in Figs. 2 and 3, the bi-prediction process is conditionally performed for a current CU block only if the SAD of uni-prediction is larger than the predefined threshold, which is the average SAD of the CU blocks previously coded by the uni-prediction mode in the past frames. However, we cannot know the average SAD of the CU blocks until the entire encoding of a frame is completed. We measure the probability that the SAD of uni-prediction obtained from the current CU block coded by the bi-prediction mode is larger than the average SAD of previously coded CUs, the best mode of which is the uni-prediction mode. As Table 1 demonstrates, with a high accuracy of 83.5% on average, we can adaptively compute the threshold values before performing bi-prediction

Table 1. Probabilities that SAD of uni-prediction for CU blocks coded by the bi-prediction mode is larger than average SAD of previously coded CUs, best mode of which is the uni-prediction mode.

Class (width×height)	Sequences	Probabilities (%)
A (2,560×1,600)	Traffic	86.5
	PeopleOnStreet	90.8
	Nebuta	79.0
	SteamLocomotive	77.5
B (1,920×1,080)	Kimono	75.4
	ParkScene	86.1
	Cactus	86.6
	BasketballDrive	84.2
	BQTerrace	80.0
C (832×480)	BasketballDrill	81.9
	BQMall	88.5
	PartyScene	81.6
	RaceHorses	80.0
D (416×240)	BasketballPass	90.9
	BQSquare	92.6
	BlowingBubbles	79.1
	RaceHorses	78.8
Average		83.5

in the current CU.

III. Proposed Selective Bi-prediction Method

The proposed selective bi-prediction method is developed under the SAD analysis shown in section II. For CU blocks coded by the bi-prediction mode as the best prediction mode, we confirm that uni-prediction has a higher SAD than that of the previously coded CU blocks, the best prediction mode of which is the uni-prediction mode. Therefore, we propose the SAD-based selective bi-prediction method. The proposed method is conditional, determining whether or not bi-prediction for a current CU block is performed only if the SAD of uni-prediction is larger than the average SAD of the previous CU blocks coded by the uni-prediction mode.

Herein, we use the cost function, J_{Motion}, to consider not only SAD but also the required bit rate for motion information:

$$J_{\text{Motion}} = SAD(s, c(MV, ref\_idx)) + \lambda_{\text{Motion}} \cdot R(MVD, ref\_idx). \quad (1)$$

In (1), $SAD(s, c(MV, ref\_idx))$ is the sum of the absolute differences between the current block s and its reference block

Table 2. RD performance comparison of proposed method.

Class	Sequences	Proposed method	
		BDBR (%)	BD-PSNR (dB)
A	Traffic	0.09	–0.003
	PeopleOnStreet	0.08	–0.003
	Nebuta	0.05	–0.001
	SteamLocomotive	0.05	0
B	Kimono	0.09	–0.003
	ParkScene	0.02	–0.001
	Cactus	0.05	–0.001
	BasketballDrive	0.13	–0.003
	BQTerrace	0.10	–0.001
C	BasketballDrill	0.06	–0.002
	BQMall	0.16	–0.006
	PartyScene	0.11	–0.005
	RaceHorses	0.22	–0.008
D	BasketballPass	0.06	–0.003
	BQSquare	0.03	–0.001
	BlowingBubbles	0.05	–0.002
	RaceHorses	0.24	–0.011
Average		0.09	–0.003

Table 3. Complexity reduction for proposed method.

Class	Sequences	ΔT			
		QP=22	QP=27	QP=32	QP=37
A	Traffic	46.36	49.61	51.58	54.57
	PeopleOnStreet	42.02	46.57	48.67	50.72
	Nebuta	17.76	28.42	36.83	52.62
	SteamLocomotive	52.37	57.1	57.01	56.93
B	Kimono	41.71	42.95	45.72	45.14
	ParkScene	46.1	50.3	52.71	56.37
	Cactus	46.16	50.29	52.18	51.81
	BasketballDrive	43.08	47.55	49.1	50.02
	BQTerrace	43.62	53.51	54.41	54.61
C	BasketballDrill	44.82	48.82	49.85	52.39
	BQMall	45.29	50.65	53.03	54.26
	PartyScene	31.38	36.24	40.14	44.28
	RaceHorses	37.38	42.5	46.45	49.42
D	BasketballPass	43.02	42.23	46.97	49.87
	BQSquare	23.67	33.44	41.36	44.62
	BlowingBubbles	35.98	37.5	44.67	50.76
	RaceHorses	39.31	40.65	43.38	48.46
Average		40.00	44.61	47.89	50.99

c, whose motion vector is MV and reference frame index is ref_idx. MVD is the difference between the current MV and its predicted MV (PMV), λ_{Motion} is a Lagrangian multiplier [9], and $R(MVD, ref\_idx)$ is the encoded bit rate required for MVD and ref_idx.

The proposed method computes two cost functions for the current CU, $avgJ^{L0}_{\text{Motion}}$ and $avgJ^{L1}_{\text{Motion}}$, which are the average J_{Motion} for previous CU blocks coded by the uni-prediction mode:

$$avgJ^{L0}_{\text{Motion}} = \frac{\sum_{0}^{m-1} J^{L0}_{\text{Motion}}}{m}, \quad avgJ^{L1}_{\text{Motion}} = \frac{\sum_{0}^{n-1} J^{L1}_{\text{Motion}}}{n}. \tag{2}$$

In (2), m and n are the numbers of previous CUs coded by the forward prediction mode and the backward prediction mode, respectively.

After computing (2), the following conditions are checked to determine whether the bi-prediction is performed for the current CU. If one of the SADs computed by the forward prediction mode and the backward prediction mode of a current CU, J^{L0}_{Motion} and J^{L1}_{Motion}, is smaller than the average SAD of the previous CU blocks coded by the uni-prediction mode, bi-prediction is skipped. The conditions can be expressed as

$$\text{Bi-prediction} = \begin{cases} \text{On}, & \text{if } \{(J^{L0}_{\text{Motion}} > avgJ^{L0}_{\text{Motion}}) \\ & \quad \&\& (J^{L1}_{\text{Motion}} > avgJ^{L1}_{\text{Motion}})\}, \\ \text{Off}, & \text{otherwise.} \end{cases} \tag{3}$$

IV. Experiment Results

The proposed method is implemented in the HEVC test model (HM) version 4.0, which is set as an anchor. An experiment is conducted on the encoder configuration using random access and high efficiency, which are described in [8].

To evaluate the coding performance, we use the Bjøntegaard delta bit rate (BDBR) and the Bjøntegaard delta peak signal-to-noise rate (BD-PSNR) [10], as recommended by the JCT-VC. In general, a BDBR decrease of 1% corresponds to a BD-PSNR increase of 0.05 dB [10]. Table 2 shows that the average BDBR increment and BD-PSNR drop are 0.09% and –0.003 dB, respectively. Such an insignificant degradation does not manifest in any noticeable visual difference.

To verify the effectiveness of the proposed method, we

Table 4. Speed-up ratio of number of ME calculations for proposed method.

Class	Sequences	ΔNumber			
		QP=22	QP=27	QP=32	QP=37
A	Traffic	62.29	64.58	62.88	60.36
	PeopleOnStreet	57.66	60.61	60.94	59.84
	Nebuta	39.04	46.97	59.49	63.53
	SteamLocomotive	61.91	62.26	56.81	53.04
B	Kimono	61.4	59.52	54.97	51.66
	ParkScene	60.55	63.04	62.24	60.56
	Cactus	63.36	66.36	63.41	59.54
	BasketballDrive	61.49	62.76	59.53	56.06
	BQTerrace	56.09	64.93	64.39	60.95
C	BasketballDrill	61.6	62.83	62.68	61.24
	BQMall	60.26	62.34	62.03	60.84
	PartyScene	49.1	55.88	59.89	61.46
	RaceHorses	54.05	59.28	61.57	61.94
D	BasketballPass	59.19	62.09	63.96	64.56
	BQSquare	44.01	52.38	59.45	63
	BlowingBubbles	52.17	58.77	62.09	63.93
	RaceHorses	54.55	57.93	60.09	61.11
Average		56.40	60.15	60.97	60.21

measure the time reduction (percentage) for a comparison of the computational complexity:

$$\Delta T = \frac{Time_{\text{Anchor}} - Time_{\text{Proposed}}}{Time_{\text{Anchor}}} \times 100 . \quad (4)$$

In (4), $Time_{\text{Proposed}}$ and $Time_{\text{Anchor}}$ are the ME time for bi-prediction with and without the proposed method, respectively.

The proposed method shows a good performance for the complexity reduction. As shown in Table 3, the time reduction of the bi-prediction is from 40% to 50% at various bit rates when the proposed method is implemented in an HM 4.0.

In addition, since the execution time may not truly reflect computational complexity due to different programming skills or use of a different hardware platform, we examine how many times bi-prediction ME is performed in the encoder. Using $Number_{\text{Proposed}}$ and $Number_{\text{Anchor}}$, which represent the number of bi-prediction MEs with and without the proposed method, respectively, we measure the speed-up ratio as follows:

$$\Delta Number = \frac{Number_{\text{Anchor}} - Number_{\text{Proposed}}}{Number_{\text{Anchor}}} \times 100 . \quad (5)$$

As shown in Table 4, the proposed method can reduce the number of ME calculations for bi-prediction up to 60% with negligible loss of quality.

In particular, the time reduction of *Nebuta* and *BQ-Square* is relatively small, as those sequences having fast motion are frequently coded by the bi-predictive mode, as it's the best prediction mode.

V. Conclusion

The proposed method exploited the statistical property of SAD derived from the ME process and was adaptively applied to bi-prediction after satisfying the predefined conditions. The experiment results showed that the proposed method significantly reduces the computational complexity of bi-prediction while maintaining almost the same RD performance as an anchor for various bit rates and motion sequences. On average, 40% to 50% of the bi-prediction time and 56% to 60% of the number of bi-prediction ME executions are reduced by the proposed method.

References

[1] ISO/IEC JTC 1 SC29 WG11, "Joint Call for Proposals on Video Compression Technology," Doc. N11113, Jan. 2010.

[2] ISO/IEC JTC 1 SC29 WG11, "Vision, Applications and Requirements of High-Performance Video Coding," Doc. N11096, Jan. 2010.

[3] S. Jeong et al., "Highly Efficient Video Codec for Entertainment-Quality," *ETRI J.*, vol. 33, no. 2, Apr. 2011, pp. 145-154.

[4] D.S. Jun and H.W. Park, "An Efficient Priority-Based Reference Frame Selection Method for Fast Motion Estimation in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, Aug. 2010, pp. 1156-1161.

[5] B.G. Kim, J.H. Kim, and C.S. Cho, "Fast Inter Mode Decision Algorithm Based on Macroblock Tracking in H.264/AVC Video," *ETRI J.*, vol. 29, no. 6, Dec. 2007, pp. 736-744.

[6] M. Flierl and B. Girod, "Multihypothesis Motion Estimation for Video Coding," *Proc. Data Compression Conf.*, Snowbird, UT, USA, Mar. 2001, pp. 341-350.

[7] K. Lillevold, "B Pictures in H.26L," ITU-T Video Coding Experts Group, Document Q15-I-08, Oct. 1999.

[8] F. Bossen, "Common Test Conditions and Software Reference Configurations," JCT-VC document, JCTVC-F900, July 2011.

[9] G.J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, Nov. 1998, pp. 74-90.

[10] G. Bjøntgaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T SG16 Q.6 VCEG, Doc. VCEG-M33, 2001.

Jongho Kim received his BS from the Control and Computer Engineering Department, Korea Maritime University, Busan, Rep. of Korea, in 2005 and his MS from the University of Science and Technology (UST), Daejeon, Rep. of Korea, in 2007. In September 2008, he joined the Realist-Media Research Team at ETRI, Daejeon, Rep. of Korea, where he is currently a researcher. His research interests include video processing and video coding.

DongSan Jun received his BS in electrical engineering and computer science from Pusan National University, Busan, Rep. of Korea, in 2002 and his MS and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2004 and 2011, respectively. He has been a senior researcher at ETRI, Daejeon, Rep. of Korea, since 2004 and an adjunct professor in the Mobile Communication and Digital Broadcasting Engineering Department at the University of Science and Technology (UST) in Daejeon, Rep. of Korea, since 2011. His research interests include image computing systems, pattern recognition, video compression, and realistic broadcasting systems.

Seyoon Jeong received his BS and MS in electronics engineering from Inha University, Incheon, Rep. of Korea, in 1995 and 1997, respectively. Since 1996, he has been a senior member of the research staff at ETRI, Daejeon, Rep. of Korea, and he is presently working toward his PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea. His current research interests include video coding, video transmission, and UHDTV.

Sukhee Cho received her BS and MS in computer science from Pukyong National University, Busan, Rep. of Korea, in 1993 and 1995, and her PhD in electrical and computer engineering from Yokohama National University, Yokohama, Kanagawa, Japan, in 1999. Since 1999, she has been with the Broadcasting and Telecommunications Convergence Media Research Department at ETRI, Daejeon, Rep. of Korea. Her current research interests include the realistic media processing technologies, such as video coding and the systems of 3DTV and UHDTV.

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

Jinwoong Kim received his BS and MS in electronics engineering from Seoul National University, Seoul, Rep. of Korea, in 1981 and 1983. He received his PhD in electrical engineering from Texas A&M University, College Station, Texas, USA, in 1993. He has been working at ETRI, Daejeon, Rep. of Korea, since 1983 and is now a principal member of the research staff and the director of the Broadcasting and Telecommunications Convergence Media Research Department. He has been leading many government-funded R&D projects on digital broadcasting technologies, including data broadcasting, viewer-customized broadcasting, and MPEG-7 /MPEG-21 standard technology development. His recent research interest is in 3DTV technologies, such as stereoscopic 3DTV, multiview 3DTV, and digital holography, as well as UHDTV. He was the chair of the 3DTV project group of TTA. He was a Far-East Liaison of the 3DTV Conference in 2007 and 2008 and has been an invited speaker to a number of domestic and international workshops, including the EU-Korea Cooperation Forum workshop on ICT, the 3D Fair 2008 in Japan, and the 3DTV Conference 2010.

Chieteuk Ahn is a professor at the Yonsei Institute of Convergence Technology (YICT). He received his BE and ME from Seoul National University, Seoul, Rep. of Korea, in 1980 and 1982, respectively, and his PhD from the University of Florida, Gainesville, FL, USA, in 1991. He began working with ETRI, Daejeon, Rep. of Korea, in 1982 and served as a senior vice president from 2003 to 2010. At ETRI, he was involved in developing technologies related to digital switching systems, MPEG standardization, and broadcasting and telecommunications convergence. He joined the YICT in March 2011. His recent work has been focused on MPEG technology, as well as on developing multi-dimensional interactive multimedia technologies in broadcasting and telecommunications. His research interests include multimedia signal processing, multimedia systems in broadcasting, and telecommunications convergence.