On improving the Fast Mode Decision of the Enhancement Layer in Scalable Video Coding extension of H.264/AVC

Sangwoo Lee and Sang Ju Park

School of Electronic & Electrical Engineering, Hongik University,
72-1, Sangsu-Dong, Mapo-Gu, Seoul, Korea 121-791,
+82-2-320-1475, +82-2-320-1119 (FAX)
klassesw@mail.hongik.ac.kr, sjpark@hongik.ac.kr

Abstract

Scalable video coding (SVC) is an extension of H.264/MPEG 4 AVC approved by JVT on November, 2007. The characteristic of the SVC is the encoding of a high-quality video bitstream that contains one or more enhancement layer bitstreams in addition to the base layer bitstream. We propose a fast enhancement layer macro block (MB) mode decision algorithm for spatial scalable SVC utilizing statistical characteristics of lower layer. The MB mode of spatial enhancement layer is statistically highly correlated with MB mode of lower layer. The proposed algorithm intelligently limits the possible candidate MB modes of enhancement layer to the modes predicted from the base layer for spatial scalable coding. We implemented our algorithm on JSVM codec to verify the performance of our algorithm. Using our algorithm, we can reduce the encoding time while almost maintaining PSNR and bitrate.

Keywords: *H.264/AVC, scalable video coding, SVC, inter-layer prediction, fast mode decision*

1. Introduction

The emergence of various types of video devices triggered by the rapid development of the technology and its demands requires better compression schemes. Developers came out with scalable video coding (SVC) [1]. SVC became the standard for H.264/AVC as MPEG-4 AVC/H.264 Amd.3 Scalable Video Coding by JVT [2-4].

SVC supports the temporal, spatial, and quality scalabilities. Each scalability consists of one base layer and one or more enhancement layer(s). It can be used either by itself or combined together. Base layer is encoded by normal H.264/AVC. Enhancement layer encoder utilizes the coded information of lower layers.

The spatial scalability coding process finds the macro block (MB) mode that has the minimum rate distortion (RD) cost using the information from lower layers. It usually requires a very long encoding time since JSVM [4] requires an exhaustive searching best mode within all available MB modes.

Our previous study [5] achieved relatively good performance. However, it has some room for improvements in both encoding time (when quantization parameter (QP), value is small) and PSNR characteristics (when QP value is large). In this paper, we propose an additional fast mode decision algorithms for spatial scalability encoding that improves the previous results [5].

2. Summary of Previous Work

According to [5], $Mode_{BL\_Pred}$ which refers to predicted mode from base layer is mostly 16×16 regardless of the characteristic of video sequences. In addition, this tendency grows as QP becomes larger.

Table 1. Percentage of $Mode_{EL}$ along with $Mode_{left}$ and $Mode_{above}$ for SMVB blocks (Foreman, QP: 30) (%)

$Mode_{left}$	$Mode_{above}$	16 × 16	16 × 8	8 × 16	8 × 16
16 × 16	16 × 16	96.15	1.12	1.02	1.50
16 × 16	16 × 8	64.97	<u>30.52</u>	0.85	2.14
16 × 16	8 × 16	89.15	2.55	4.31	2.60
16 × 8	16 × 16	88.52	4.12	3.02	2.73
8 × 16	16 × 16	68.21	2.11	<u>27.12</u>	1.84

Table 2. $Mode_{EL,BL}$ for NSMVB's categorized as $Mode_{BL}$

$Mode_{BL}$	$Mode_{EL}$ belongs to $Mode_{EL,BL}$
16 × 16	16 × 16, 8 × 8
16 × 8	16 × 16, 16 × 8, 8 × 8
8 × 16	16 × 16, 8 × 16, 8 × 8
8 × 8	16 × 16, 8 × 8

Table 3. Probability of $Mode_{EL}$ belongs to $Mode_{EL,BL}$

QP \ Sequence	Mother & Daughter	Foreman	Harbor
20	.7943	.8434	.8399
25	.8233	.8747	.8632
30	.8912	.9551	.9318
35	.9654	.9891	.9754
40	.9832	.9951	.9992

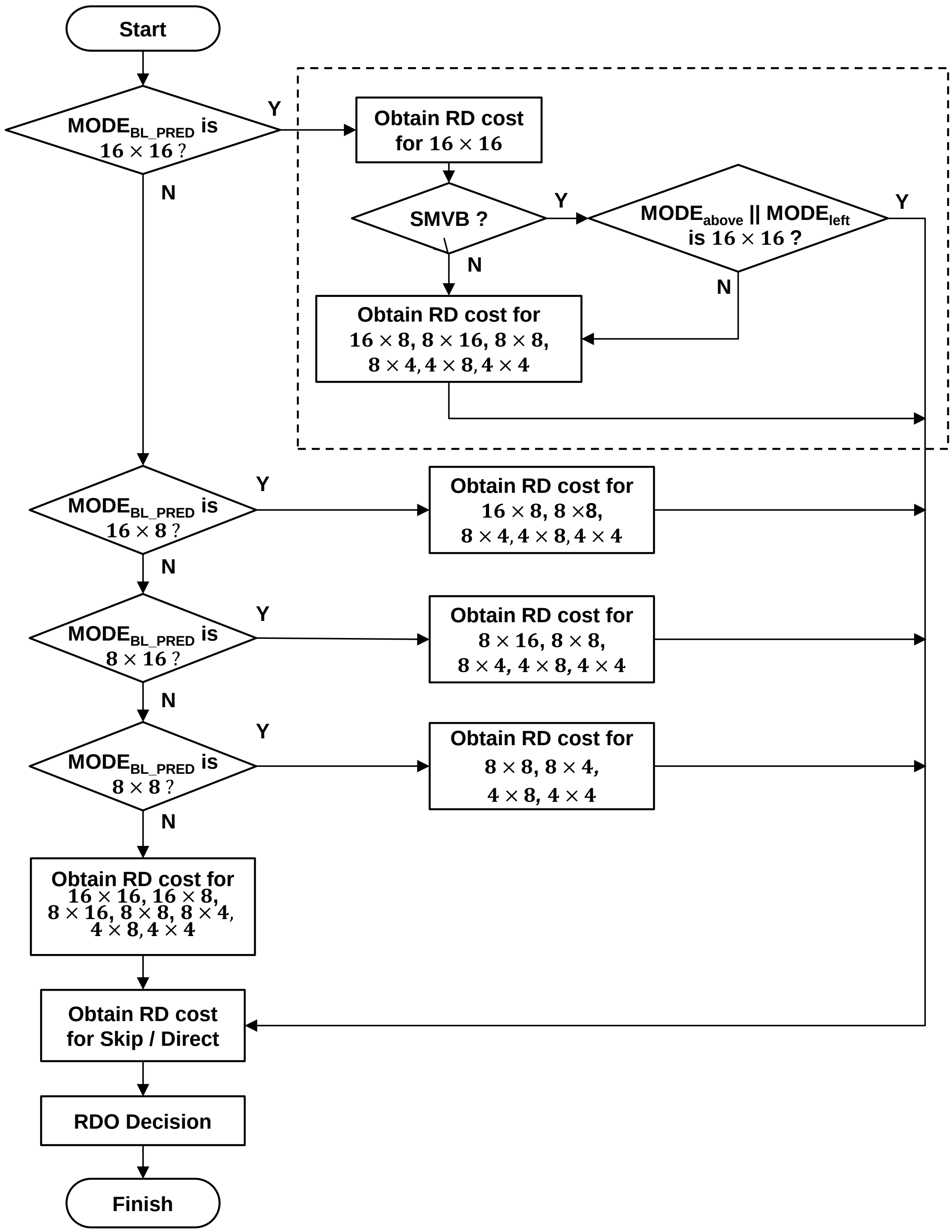


Figure 1. Flowchart of Fast Enhancement Layer MB Mode Decision Algorithm of [5]

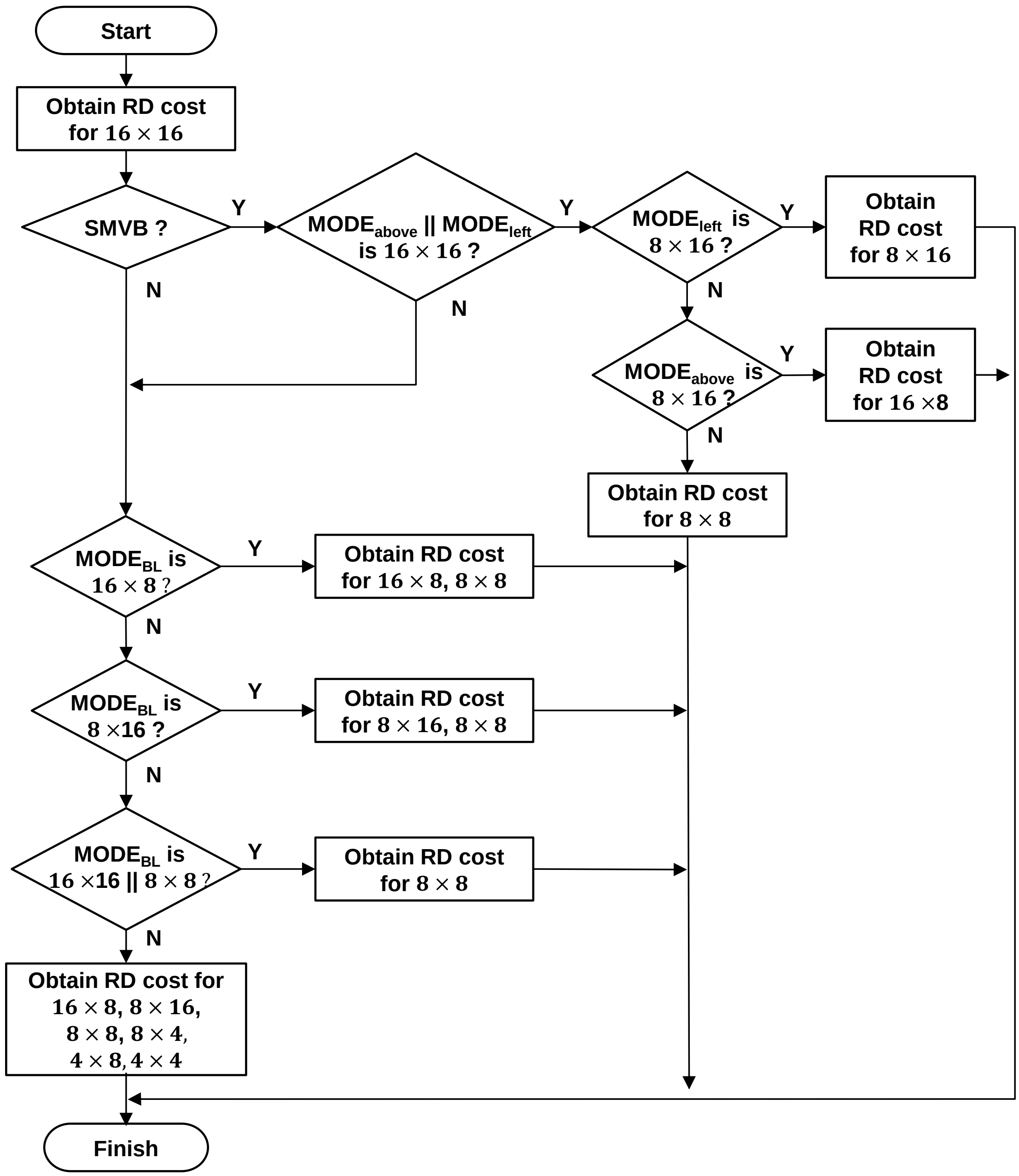


Figure 2. Flowchart of Proposed Fast Enhancement Layer MB Mode Decision Algorithm, this Fow Chart Replacing Dashed Box of Figure 1

Within this investigation [5], enhancement layer blocks having identical lower layer and enhancement layer motion vectors and whose $Mode_{BL\_Pred}$ is 16×16 are divided into two classes. One is called same motion vector block (SMVB) for which either left block mode $Mode_{left}$ or above block mode $Mode_{above}$ is 16×16 or SKIP. For SMVB, the mode is simply set to 16×16 with no further RD cost computation. The other is called not same motion vector block (NSMVB). For NSMVB's, the RD costs of all available modes (16×8, 8×16, and 8×8) are computed and final mode decision is made.

3. Statistical Analysis of MB Mode Decision of Base Layer and Enhancement Layer

Table 1 shows a sample $Mode_{EL}$ (denoting mode chosen at the enhancement layer without referencing the lower layer predicted mode) distribution of SMVB's. Each row represents the percentage of $Mode_{EL}$ categorized as $Mode_{left}$ and $Mode_{above}$, which are already the decided neighboring MB modes. It can be seen that $Mode_{EL}$ has a relatively high percentage of being 8×16 when $Mode_{left}$ is 8×16, and being 16×8 when $Mode_{above}$ is 16 × 8, as indicated by underlines.

Also, for NSMVB's, we noticed it is highly likely that $Mode_{EL}$ belongs to some subset of modes denoted as $Mode_{EL,BL}$ depending on the lower layer mode $Mode_{BL}$. Table 2. shows the subsets found experimentally for each $Mode_{BL}$. Table 3. summarizes the probabilities of $Mode_{EL}$ that actually belong to $Mode_{EL,BL}$ for various QP and sequences.

4. Proposed Algorithm

We propose an improved fast mode decision algorithm based on [5] utilizing the results in section 3. When the predicted mode $Mode_{BL\_Pred}$ is 16×16, the algorithm first calculates the RD cost of the 16×16 mode. Then it decides whether it is SMVB or not [5]. When it is classified as SMVB, it looks into $Mode_{above}$ and $Mode_{left}$. If $Mode_{above}$ is 16×8 the algorithm computes the RD cost of the 16×8 mode. Similarly, if $Mode_{left}$ is 16×8, it computes the RD cost of the 16×8 mode. The mode decision is made between the 16×16 mode and the additionally investigated mode. When it is classified as NSMVB, it checks that $Mode_{BL}$ and RD cost computations are done for modes only in the subsets $Mode_{EL,BL}$. By incorporating these two additional techniques, the proposed algorithm increases PSNR for large value of QP (by 1.) and improves encoding speed for the small value of QP (by 2.).

Table 4. Simulation Conditions

Reference Codec	JSVM 9.14
GOP size	8
Frames	100
Motion search range	32 pixel
Motion search accuracy	1/4 pixel
Motion search function	Full pixel : SAD Sub pixel: SATD
FGS scalability	Do not use
Input sequence	Base layer: QCIF 15fps Enhancement layer: CIF 15fps
`InterLayerPred` flag value	2

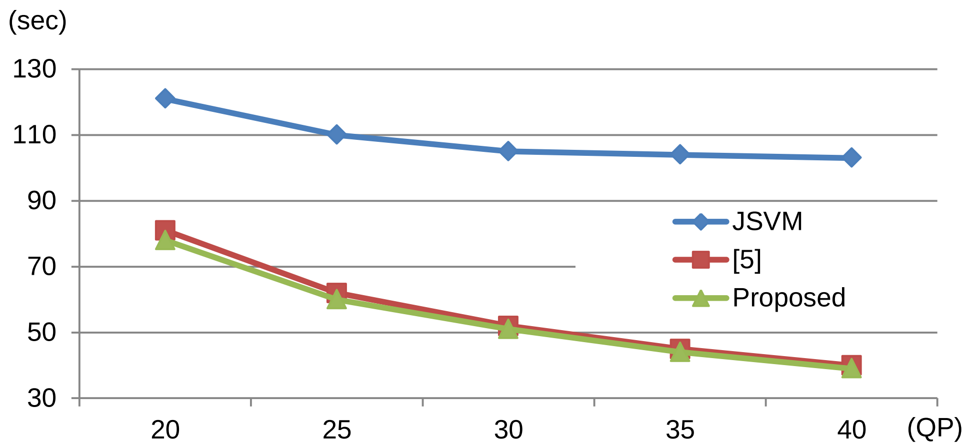
Figure 1 shows the flowchart for the overall enhancement layer MB mode decision from our previous report [5]. Figure 2 is the flowchart for the proposed algorithm replacing the dashed box of Figure 1.

5. Simulation Results

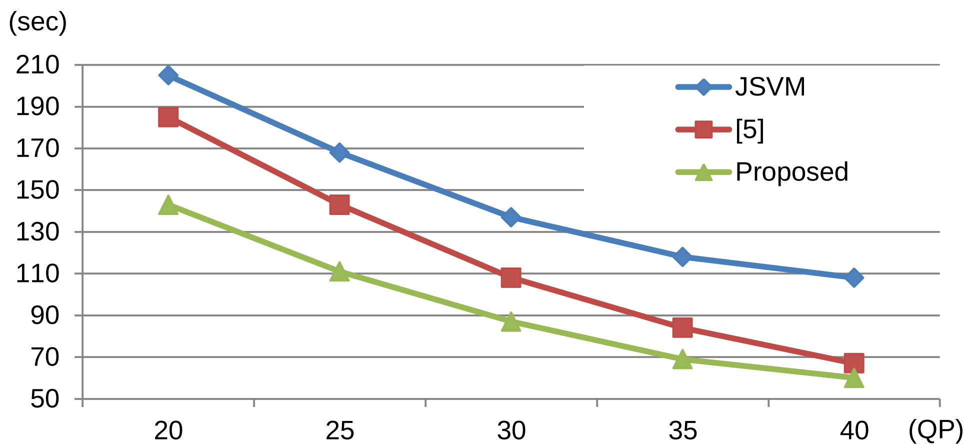
To justify the performance of the proposed algorithm, we tested our algorithm on three test sequences (Mother and Daughter, Foreman, and Harbor) with various quality characteristics. We compared the results with JSVM4, and [5]. Experiments are done on Intel Core2Quad 2.83GHz PC with 4GB of main memory running Windows 7. Other conditions are listed in Table 4.

Table 5. Encoding Time Comparison (sec)

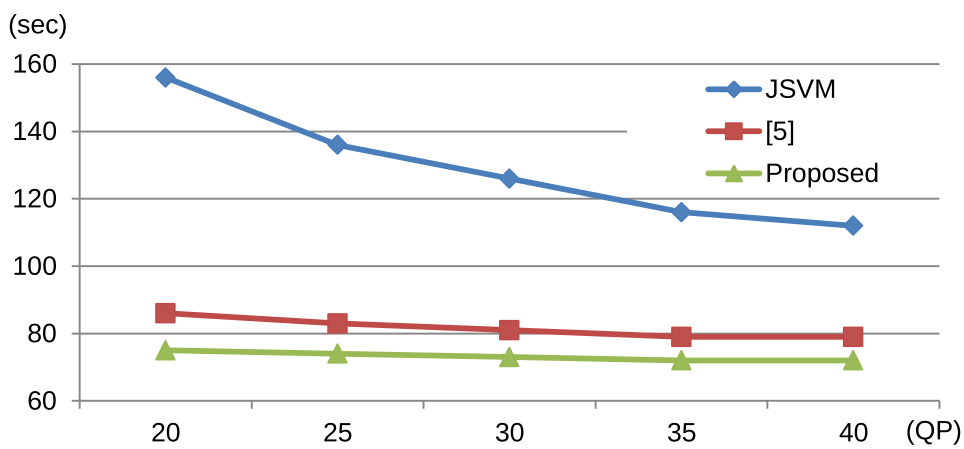
Sequence	Encoder	QP				
		20	25	30	35	40
Mother & Daughter	JSVM	121	110	105	104	103
	[5]	81	62	52	45	40
	Proposed	78	60	51	44	39
	Δ_{JSVM} (%)	33.54	45.45	51.43	57.70	62.14
	$\Delta_{[5]}$ (%)	3.70	3.23	1.92	2.22	2.50
Foreman	JSVM	156	136	126	116	112
	[5]	86	83	81	79	79
	Proposed	75	74	73	72	72
	Δ_{JSVM} (%)	51.28	45.59	42.06	37.93	35.71
	$\Delta_{[5]}$ (%)	12.79	10.84	9.88	8.86	8.86
Harbor	JSVM	205	168	137	118	108
	[5]	185	143	108	84	67
	Proposed	143	111	87	69	60
	Δ_{JSVM} (%)	30.24	33.93	36.50	41.53	44.44
	$\Delta_{[5]}$ (%)	23.12	22.38	19.44	17.86	10.45



(a) Mother & Daughter



(b) Foreman



(c) Harbor

Figure 3. Encoding Time Comparison

Figure 3 and Table 5 show the encoding time of three test sequences with various QPs. The table also shows encoding time percentage decreases of the proposed method with respect to JSVM and [5] denoted as Δ_{JSVM} and $\Delta_{[5]}$ respectively. The proposed algorithm reduced the encoding time of [5] up to by 23.12% (Harbor sequence, QP: 20). Improvement in terms of encoding time for small value of QP is relatively larger than that for large QP value.

On the average, the proposed algorithm improves PSNR value of [5] by 0.04dB. Maximum PSNR improvement over [5] is 0.12 dB for Foreman sequence with QP=40.

Finally, it should be noticed that there are no significant differences in bit rates. The proposed algorithm increased the bit rates of JSVM only by 0.88Kbps or by 0.0044% on the average. Considering the improvement in terms of encoding speed demonstrated on Table 5, this bit rate increase is negligible.

6. Conclusion

We proposed a new fast mode decision algorithm for spatial enhancement layer in SVC which improves our previous work. It reduces encoding time for the small value of QP and increases PSNR characteristic for the large value of QP. Compared to JSVM, the proposed algorithm reduces the complexity greatly with only a very small sacrifice in RD performances.

Acknowledgment

This work was supported by the Seoul R&BD Program no. 10555.

References

[1] J. -R. Ohm, “Advances in scalable video coding”, Proceedings of the IEEE, vol. 93, no.1, **(2005)** January, pp. 42-56.
[2] Joint Video Team (JVT) of ISO/IEC MPEG & ITUT VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), “Joint Draft 10 of SVC Amendment,” San Jose, **(2007)** April.
[3] Joint Video Team (JVT) of ISO/IEC MPEG & ITUT VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), “ITU-T Rec. H.264 ISO/IEC 14496-10 version 8,” **(2007)** November.
[4] Joint Video Team (JVT) of ISO/IEC MPEG & ITUT VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), “JSVM 9 Software,” Hannover, **(2008)** July.
[5] T. Kim, J. Boo and S. J. Park, “Fast Mode Decision on the Enhancement Layer in H.264 Scalable Extension”, IEICE Transactions on Information and Systems, Vol. E92-D, No.12, **(2009)** December, pp. 2545-2547.

Authors

Sangwoo Lee

He received the B.S., M.S. and Ph.D. degrees in School of Electronic & Electrical Engineering from Hongik University, Korea, in 2003, 2006 and 2012 respectively. His research interests are multimedia data processing and compression.

Sang Ju Park

He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Korea, in 1984 and 1986 respectively and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1992. He was an MTS at PRPA(Philips Research Palo Alto) and Philips Semiconductors, Sunnyvale during 1992-1995. Since March 1996, he has been at Hongik University, where he is currently a professor. His research interests are multimedia data processing and compression.