

일반논문-08-13-1-14

Switching Picture Added Scalable Video Coding and its Application for Video Streaming Adaptive to Dynamic Network Bandwidth

Jie Jia^{a)}, Hae Kwang Kim^{a)†}, and Hae Chul Choi^{b)}

Abstract

Transmission of video over Internet or wireless network requires coded stream capable of adapting to dynamic network conditions instantly. To meet this requirement, various scalable video coding schemes have been developed, among which the Scalable Video Coding (SVC) extension of the H.264/AVC is the most recent one. In comparison with the scalable profiles of previous video coding standards, the SVC achieves significant improvement on coding efficiency performance. For adapting to dynamic network bandwidth, the SVC employs inter-layer switching between different temporal, spatial or/and fidelity layers, which is currently supported with instantaneous decoding refresh (IDR) access unit. However, for real-time adaptability, the SVC has to frequently employ the IDR picture, which dramatically decreases the coding efficiency. Therefore, an extension of SP picture from the AVC to the SVC for an efficient inter-layer switching is investigated and presented in this paper. Simulations regarding the adaptability to dynamic network bandwidth are implemented. Results of experiment show that the SP picture added SVC provides an average 1.2 dB PSNR enhancement over the current SVC while providing similar adaptive functionality.

Keywords: Scalable Video Coding, SP picture, bitstream switching, bandwidth adaptation

1. Introduction

Video communication over heterogeneous network for diverse clients under various transmission environments is the predominant real-time multimedia application today, among which video conference, mobile video, and high-definition (HD) TV broadcast can be good examples.

Those applications require both coding efficiency and adaptivity for reliable video service. Driven by those applications, the Scalable Video Coding (SVC) extension of the H.264/AVC has been developed and recently standardized by the Joint Video Team (JVT) of the ITU-T Video Coding Expert Group (VCEG) and the ISO/IEC Moving Picture Expert Group (MPEG) [1]. The SVC provides coded stream scalability in temporal, spatial and fidelity. In comparison with scalable profiles of previous video coding standards, such as MPEG-2, H.263+ and MPEG-4 part II, the SVC achieves significant improvement on coding efficiency performance via extensively developed inter-layer prediction tools.

The SVC provides coded stream universal accessibility via bitstream extraction which extracts scalable coded

a) 세종대학교 컴퓨터과학부

Department of Computer Science, Sejong University

b) ETRI

Radio & Broadcasting Division, ETRI,

† 교신저자 : 김해광(hkkim@sejong.ac.kr)

* This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No.R01-2007-000-11078-0) and ETRI (Electronics and Telecommunications Research Institute) the IT R&D program of MIC/IITA [2005-S-103-03, Development of Ubiquitous Content Access Technology for Convergence of Broadcasting and Communications].

stream into another stream corresponding to a lower spatial, temporal and/or fidelity level. For adapting to dynamic network bandwidth, the scalably coded stream can be switched between different layers so that different bit rates could be provided for the varying network throughput. Current SVC supports bitstream inter-layer switching at instantaneous decoding refresh (IDR) access unit. However, a frequent usage of IDR picture for providing real-time adaptability to bandwidth variation dramatically decreases the coding efficiency performance of SVC. Referring to the H.264/AVC, a new picture type, the SP picture, was developed and adopted for the efficient bitstream switching functionality [2] - [4].

Similar to the P picture, the SP picture employs an inter picture coding scheme where motion compensated prediction is applied. While different from the P picture, the SP picture consists of a primary SP picture (SP picture for non-switching) and a secondary SP picture (SP picture for switching). The two SP pictures are predicted from different reference pictures while provide an identical reconstruction, which guarantees a drift-free switching. Note, usually, only the primary SP picture is coded and transmitted in the bitstream. The secondary SP picture is only transmitted and decoded when switching happens. At that time, the corresponding primary SP picture will neither be transmitted nor decoded. There are two quantization processes employed for the primary SP picture coding. One is for residual signal coding. The other is for construction picture re-quantization. The second finer quantizer requantizes the constructed signal so that a reasonable number of bits can be used to represent the secondary SP picture, where a lossless coding is employed with the requantized primary construction picture as input. For further description about the SP picture, readers are referred to [5] - [8]. As illustrated, the SP picture provides significant performance improvement over the IDR picture. Therefore, in this paper, an SP picture added scalable video coding is investigated

for efficient bitstream switching as well as inter-layer switching.

This paper is organized as follows. Sec. 2 presents an overview of the SVC scalable coding structure. Sec. 3 investigates the SP picture added SVC as well as its application for video streaming adaptive to dynamic network bandwidth with inter-layer switching. Sec. 4 analyzes performance of the SP picture added SVC, followed by conclusion drawn in Sec. 5.

II. Overview of the Scalable Video Coding

The SVC has been developed as a scalable extension of the H.264/AVC. It provides coded stream scalability in spatial, temporal and fidelity domain. Basically, the SVC is a layer-based coding scheme, where each spatial, temporal and fidelity level is indicated by a unique layer identification (ID). The SVC inherits most of the efficient coding tools from the AVC. Moreover, it exploits extensive inter-layer prediction techniques for enhanced coding efficiency. Generally, the SVC provides coded stream temporal scalability via a hierarchical prediction structure, and spatial and fidelity scalability by the layered coding structure with enhanced inter-layer prediction. In this section, a brief overview of the SVC is given.

1. Temporal Scalability

A hierarchical prediction structure provides SVC flexible temporal scalability, which can be either dyadic or non-dyadic, as illustrated in Fig. 1, where (a) illustrates the prediction structure with hierarchical B picture for dyadic temporal scalability, and (b) illustrates a non-dyadic case. For temporal scalability, each scalable layer is identified by a temporal level ID, T_i ($i = 0, 1, 2 \dots$), which starts from zero, and is increased by one for every enhancement on

the temporal layer. Fig. 1 (a) illustrates the case where two temporal scalable layers are available with 1/2 and 1/4 of the full temporal resolution. And Fig. 1 (b) illustrates the case where two sub-sequences with 1/3 and 1/9 of the full temporal resolution are scalably available.

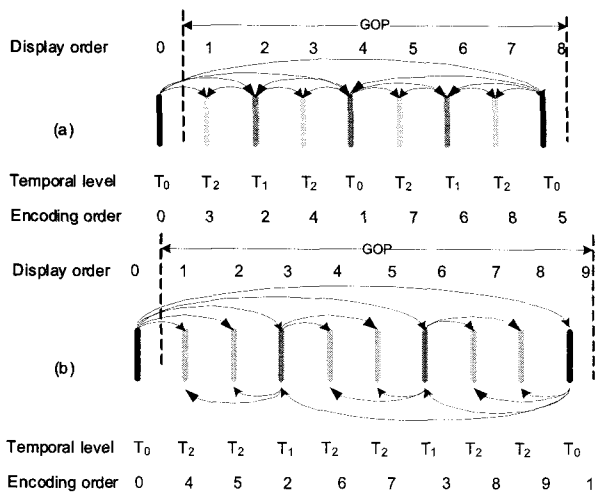


Fig. 1 Illustration of hierarchical prediction structure (a) dyadic temporal scalability (b) non-dyadic temporal scalability

A basic principle for the hierarchical prediction structure is that pictures with temporal level T_i can only be predicted from previously decoded pictures with the same or lower temporal level T_j , where $i \geq j$. Note, the multiple reference prediction can also be combined into the hierarchical prediction structure. Moreover, the hierarchical prediction structure is also possible to provide a zero delay prediction with hierarchical P picture. For detailed description, readers are referred to [9][10].

2. Spatial Scalability

The SVC provides spatial scalability via layered coding scheme, which has also been utilized in the scalable profiles of MPEG-2, H.263+, and MPEG-4 visual. In comparison with the previous standards, the SVC employs effi-

cient inter-layer prediction for enhancement layer coding that greatly enhances the SVC coding efficiency performance. Regarding the inter-layer prediction, it consists of inter-layer motion prediction, inter-layer intra prediction and inter-layer residual prediction [9][11].

Basically, for spatial scalable coding, the spatial base layer is firstly encoded, followed by the enhancement layer coding. In order to employ information from base layer as much as possible, a new macroblock (MB) type, BISkip was developed for the enhancement layer coding. When this MB type is employed and the corresponding MB in base layer is inter picture coded, the MB in enhancement layer derives both MB partition and motion information including reference picture and motion vector from the collocated MB in base layer. This is the inter-layer motion prediction.

It is found that when motion compensated prediction is employed for both the base and enhancement layer, residual signal from the enhancement layer can be further predicted from that of the base layer. This is called the inter-layer residual prediction, where the block-wise up-sampled base layer residual signal is used as the reference for the enhancement layer residual signal coding.

Regarding the inter-layer intra prediction, it is employed when the BISkip mode is utilized and the corresponding block in base layer is fully located within an intra picture coded MB. Note, in order to provide a single loop decoding, constrained intra picture prediction is always required for layers which are used for inter-layer prediction. When the inter-layer intra prediction is used, the corresponding reconstructed intra block in base layer is up-sampled and used as the reference for the enhancement layer MB coding.

3. Fidelity Scalability

The recently standardized SVC supports fidelity scalability in the form of coarse gain scalability (CGS) and me-

dium gain scalability (MGS) [12]. The CGS is obtained by coding video with the same scalable coding structure as that for the spatial scalability, but without the up-sampling of base layer signal when the inter-layer prediction is employed. Therefore, the number of quality levels supported by the CGS is identical to the number of scalable layers. The MGS is provided by an FGS-like scalable coding structure, where the motion compensated prediction is employed but without bit-plane coding [13]. For drift control purpose, a concept of key picture was utilized where the motion compensation can only be applied based on previously decoded picture at quality base layer. Switching between different quality levels for MGS is virtually possible in any access unit. Hence, the MGS provides more flexible fidelity scalability than the CGS.

III. SP Picture Added Scalable Video Coding

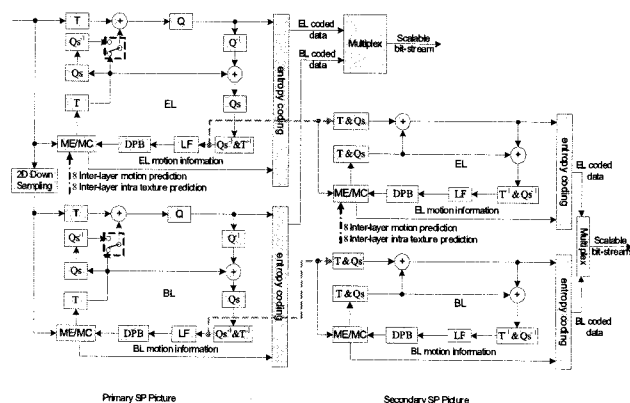


Fig.2 Encoding process illustration of SP picture for SVC

For efficient bitstream and inter-layer switching, SP picture for SVC was proposed and presented in [14]. In this section, the coding process of SP picture for SVC is firstly described, followed by an illustration of the SVC inter-layer switching scheme based on SP pictures.

1. Encoding process of SP picture for SVC

In comparison with the SP picture for AVC, the SP picture for SVC employs scalable coding scheme where the inter-layer prediction is utilized for enhancement layer coding. Fig. 2 illustrates a schematic block diagram of the SP picture coding process for SVC, where the figure on the left side presents the coding process for the primary SP picture, and the figure on the right side illustrates the coding process for the secondary SP picture. In Fig. 2, two spatial scalable layers are assumed to be coded. When the primary SP picture coding process is concerned, the original picture is firstly down-sampled into picture with lower resolution, which is encoded in the base layer. Following that, picture with the original resolution is encoded in the enhancement layer. Then the coded streams from the base and enhancement layer are multiplexed into one scalable bitstream.

In Fig. 2, T and T-1 represent the transform and inverse transform process, respectively. Q, Q-1, Qs and Qs-1 are quantizer and dequantizer used for the residual signal coding and construction signal requantization, respectively. Decoded picture buffer (DPB) holds reconstructed pictures which are used for the following inter picture coding. Regarding the base layer SP picture coding process, a similar coding scheme as that used for the AVC is employed. This provides the SP picture added SVC an AVC compatible base layer encoding.

Regarding the enhancement layer SP picture coding, besides the intra-layer motion compensated prediction, the inter-layer prediction is also employed. Different from the inter-layer prediction for P picture coding, where motion, residual and texture information from base layer are used as the reference for enhancement layer coding, the inter-layer prediction for SP picture coding employs only inter-layer motion and intra prediction. As Fig. 2 shows, the residual signal for SP picture is obtained in the transform domain.

Therefore, in the proposed SP picture for SVC coding scheme, the inter-layer residual prediction is not employed.

For obtaining an identical reconstruction between the primary SP picture and the secondary SP picture while , the constructed primary SP picture is requantized with a finer quantizer Q_s and used as the input for the secondary SP picture coding. Note, the same quantizer Q_s is employed for both the primary SP picture requantization and the secondary SP picture coding. Therefore, the secondary SP picture coding process is actually a lossless coding process. This guarantees the drift-free switching between different streams using SP picture.

2. SP Picture Added SVC Inter-layer Switching

Current SVC allows inter-layer switching at IDR access unit. However, frequent usage of IDR picture decreases the coding efficiency performance. SP picture for SVC was proposed for enhanced coding efficiency while providing similar inter-layer switching functionality. The enhanced coding efficiency performance is obtained from employing intra-layer inter picture prediction and the inter-layer prediction for primary SP picture coding. Fig. 3 illustrates an SP picture added SVC inter-layer switching and the corresponding prediction structure.

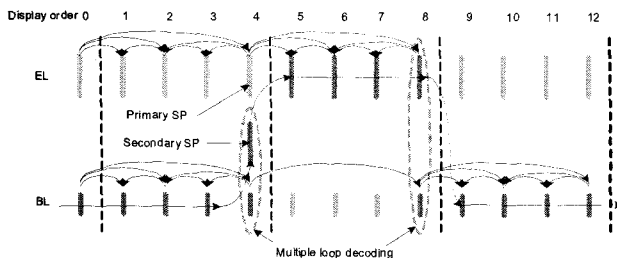


Fig. 3 Illustration of SP picture added SVC inter-layer switching

In Fig. 3, two spatial scalable layers are coded. During the transmission, inter-layer switching happens at the 4th and 9th frame, corresponding to switching up and switch-

ing down, respectively. At the 4th frame, the bitstream is switched from base layer to enhancement layer. In this case, at the switching point, a secondary SP picture for the target bitstream is transmitted and decoded, which provides an identical reconstruction to that of the primary SP picture that is originally coded in the target bitstream. Regarding the switching down, it can be easily obtained via multiple-loop decoding. For instance, in the above example, in order to do the bitstream switching from the enhancement layer to the base layer, all that needed is to do the multiple-loop decoding at each temporal level 0 point. With multiple-loop decoding, picture 8th in the base layer is available. Then, it can be further used as the reference for the following picture decoding in the order of 12, 10, 9, and 11 and so on. In this way, the SP picture added SVC is capable of doing efficient bitstream switching as well as inter-layer switching.

IV. Simulation and Performance Analysis

This section analyzes the coding efficiency performance of the SP picture for SVC and its application for video streaming adaptive to dynamic network bandwidth.

1. Adaptive video streaming

Video transmission over various network conditions requires coded stream capable of adapting to dynamic network bandwidth variation. Fig. 4 presents an illustrative dynamic network variation on bandwidth, where the bandwidth fluctuation is quantized and presented in terms of bit rate which can be obtained from different scalable layers. Basically, it can be seen that the more the scalable layers, the finer the adaptability, which can also be observed from the performance comparison between the CGS and the MGS [12]. With those scalable layers, the SP picture added

SVC provides network real-time bandwidth adaptability via dynamic inter-layer switching, as shown in Fig. 5.

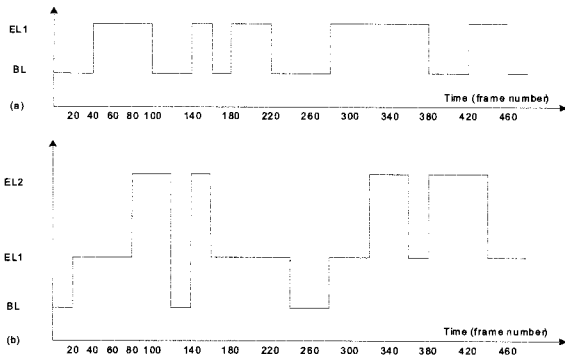


Fig. 4 Example of dynamic network bandwidth (a) rate variation when 1 enhancement layer is utilized; (b) rate variation when 2 enhancement layers are utilized

Fig. 5 presents the adaptability comparison between scalable coded streams with different GOP size. The solid line indicates the dynamic network bandwidth variation, while the dashed line represents the adapted video streams, where the SP picture added SVC inter-layer switching is employed. The comparison shows that the smaller the GOP size, the better the adaptability.

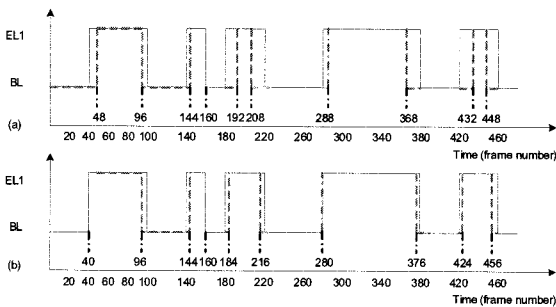


Fig. 5 Illustration of scalable coded stream adaptability to dynamic network bandwidth (a) adaptation illustration when GOP size is 16 (b) adaptation illustration when GOP size is 8

2. Performance Analysis

To illustrate the coding efficiency performance of the SP picture added SVC. Simulations based on the JSVM 8 [15]

were performed for different standard test sequences under the SVC coding efficiency test condition [16]. In the simulations, CABAC is used as the entropy coding method, and rate-distortion optimization is turned on.

Performance comparison between the SP picture added SVC and the conventional SVC with period P and IDR picture are presented in Fig. 6 to Fig. 8, Tab. 1 and Tab. 2.

From the simulation results, it can be observed that the SP picture added SVC generally outperforms the conventional SVC while providing a similar switching functionality.

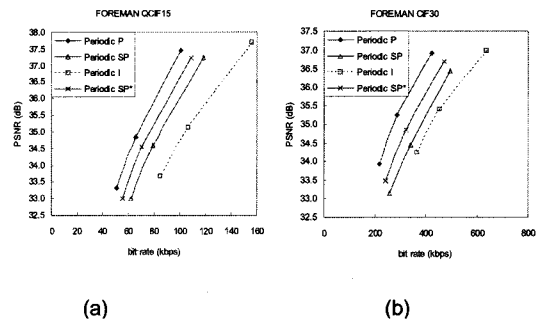


Fig. 6 Performance comparison for Foreman sequence with GOP size = 8, the SP and SP* curve presents the SP picture performance which are encoded with different QPsp and Qs values, respectively. (a) base layer (b) enhancement layer

Fig. 6 presents the performance comparison between the periodic SP, periodic P and periodic I pictures. Moreover, two periodic SP coding with different quantization parameters are compared. Basically, it can be seen that the periodic SP picture achieves significant coding efficiency enhancement in comparison with the periodic I picture. Regarding the two periodic SP coding, the periodic SP is coded with $QP_{sp} = QP_p - 2$ and $Q_s = QP_p - 5$. The periodic SP* is coded with $QP_{sp} = QP_p - 1$ and $Q_s = QP_p - 10$. From the simulation results, it can be observed that the lower the Q_s , the better the coding efficiency of the primary SP picture. However, as it can be expected, lower Q_s increases the number of bits used for representing

the secondary SP picture. Therefore, in the following simulation, quantization parameters are set as $QP_{sp} = QP_p - 2$ and $Q_s = QP_p - 5$.

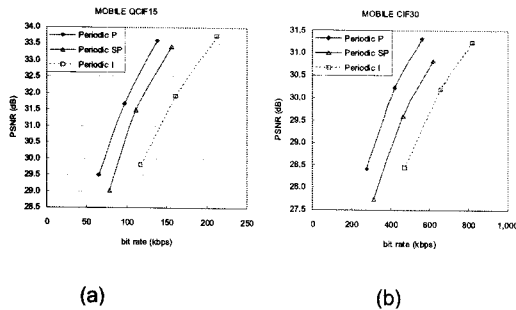


Fig. 7 Performance comparison for Mobile sequence with GOP size = 16, key picture is encoded as Inter P, SP and Intra picture, respectively. $QP_{sp} = QP_p - 2$, $Q_s = QP_p - 5$ (a) base layer, (b) enhancement layer

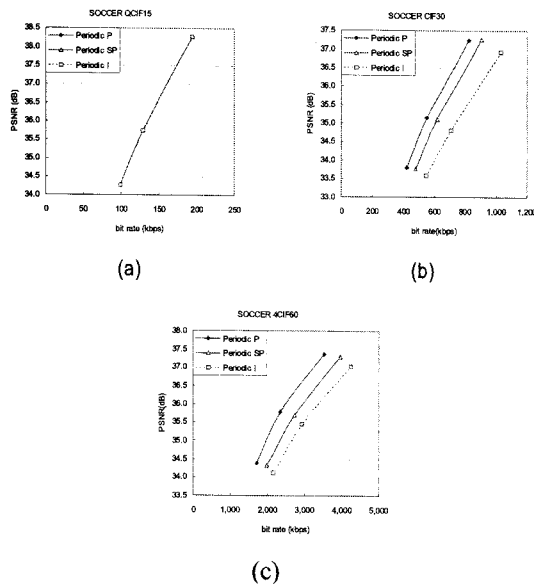


Fig. 8 Performance comparison between SP picture and EIDR picture for Soccer sequence, GOP size = 32, $QP_{sp} = QP_p - 2$, $Q_s = QP_p - 5$ (a) base layer: QCIF format at 15Hz (b) enhancement layer 1: CIF format at 30Hz (c) enhancement layer 2: 4CIF format at 60Hz

Fig. 7 presents the performance comparison for Mobile sequence with GOP size 16. Similar performance as that has been observed from Foreman sequence is given. Moreover,

a detailed observation on the performance of the SP picture for the base layer and enhancement layer shows that the SP picture achieves less performance improvement for the enhancement layer than that for the base layer. This is due to the reason that the coding efficiency performance of the base layer influences the performance of the enhancement layer. The higher the coding efficiency of the base layer, the better the performance of the enhancement layer.

Fig. 8 compares the performance between the SP picture and the EIDR picture. As illustrated in the figure, pictures in the base layer are similarly encoded, therefore, the same performance can be observed from the base layer performance figure. Regarding the performance of the enhancement layer, it can be observed that the SP picture provides enhanced coding efficiency about 1 dB for the first enhancement layer, and 0.6 dB for the second enhancement layer.

Table 1. Comparison of average coded I and SP picture size for Foreman sequence

Foreman	29	32	35
IDR	20370	15235	11314
	26	21249	29
SP	32	18486	35
	35	20092	38
		14760	41
			11913

Table 2. Comparison of average coded I and SP picture size for Soccer sequence

Soccer	27	30	33
IDR	41438	55002	69538
	24	32051	27
SP	30	26274	33
	33	27338	36
		38690	39
			51253
			52082
			53696

Tab. 1 and Tab. 2 compares the average coded picture size between IDR picture and secondary SP picture. In this simulation, "foreman" and "Soccer" sequences are coded at QCIF format. Various switching between different quality layers are supposed to be employed. The QP parameters listed on the first row are quantization parameters which

are employed for the target bit stream coding. The quantization parameters listed on each column are quantization parameters employed by the stream from which the target stream switches. From the table, it can be seen that, generally, the secondary SP picture costs less number of bits than that used for representing the IDR picture. Besides, it can be observed that, regarding the switching up and the switching down cases, generally, switching up requires more bits than that used for the switching down cases. Moreover, it can be further observed that switching between closer quality levels requires less number of bits to represent the secondary SP picture.

V. Conclusions

This paper presents a switching picture added scalable video coding and its application for video streaming adaptive to dynamic network bandwidth. The SVC is briefly reviewed, followed by a detailed description of the encoder design of the SP picture for SVC, where the inter-layer prediction is investigated for the enhancement layer SP picture coding. After that, the SP picture added SVC inter-layer switching and its application for the dynamic adaptation to the network bandwidth variation are illustrated. Simulation results show that an average 1.2 dB PSNR enhancement can be obtained by the SP picture added SVC in comparison with the conventional SVC with IDR picture.

Reference

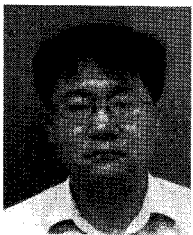
- [1] Wiegand, T., Sullivan, G.J., Reichel, J., Schwarz, H., and Wien, M. (eds): Joint Draft 10 of SVC Amendment. Joint Video Team, Doc. JVT-W201, San Jose, CA, USA, April 2007
- [2] ITU-T and ISO/IEC JTC 1: Advanced Video Coding for Generic Audio-Visual Services. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG4-AVC), Version 4: July 2005
- [3] Wiegand, T., Sullivan, G.J., Bjøntegaard, G., and Luthra, A.: 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
- [4] Karczewics, M., and Kurceren, R.: A Proposal for SP-frames. Video Coding Expert Group Meeting, Doc. VCEG-L27, Eibsee, Germany, Jan. 2001
- [5] Karczewics, M., and Kurceren, R.: The SP- and SI- Frames Design for H.264/AVC. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp.637-644, July 2003
- [6] Setton, E., and Girod, B.: Rate-distortion Analysis and Streaming of SP and SI Frames. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 16, No. 6, pp.733-743, June 2006
- [7] Sun, X., Wu, F., Li, S., Shen, G., and Gao, W.: Drift-free Switching of Compressed Video Bitstreams at Predictive Frames. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 16, No. 5, pp.565-576, May, 2006
- [8] Stockhammer, T., Liebl, G., and Walter, M.: Optimized H.264/AVC-Based Bit Stream Switching for Mobile Video Streaming. *EURASIP Journal on Applied Signal Processing*, Vol. 2006, pp. 1-19, Sept. 2006
- [9] Schwarz, H., Marpe, D., and Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. Joint Video Team, Doc. JVT-W132, San Jose, CA, USA, April 2007
- [10] Schwarz, H., Marpe, D., and Wiegand, T.: Hierarchical B Pictures. Joint Video Team, Doc. JVT-P014, Poznan, Poland, July 2005
- [11] Schwarz, H., Marpe, D., and Wiegand, T.: SVC Core Experiment 2.1: Inter-layer Prediction of Motion and Residual Data. ISO/IEC JTC 1/WG11, Doc.M11043, Redmond, WA, USA, July 2004
- [12] Kirchhoffer, H., Schwarz, H., and Wiegand, T.: SVC Core Experiment 1: Simplified FGS. Joint Video Team, Doc. JVT-W090, San Jose, CA, USA, April 2007
- [13] Winken, M., Schwarz, H., Marpe, D., and Wiegand, T.: Adaptive Refinement of Motion Information for Fine-granular SNR Scalable Video Coding. *Proceeding of EuMob'96*, Alghero, Italy, Sept. 2006
- [14] Jia, J., Kim, H.-K., Choi, H.-C., and Kim J.-G.: SP Picture for SVC Switching. Joint Video Team, Doc. JVT-V045, Marrakech, Morocco, Jan. 2007
- [15] ITU-T and ISO/IEC JTC1: JSVM8 Software. Hangzhou, China, Oct. 2006
- [16] Wien, M., and Schwarz, H.: Test Conditions for SVC Coding Efficiency and JSVM Performance Evaluation. Joint Video Team, Doc. JVT-Q205, Nice, France, Oct. 2005

저 자 소 개



Jie Jia

- 1996년 8월 ~ 2000년 7월 : Xidian University 통신공학과 (학사)
- 2000년 8월 ~ 2003년 4월 : Xidian University 통신공학과 (석사)
- 2003년 4월 ~ 2006년 2월 : Xidian University 통신공학과 (박사과정 수료)
- 2005년 1월 ~ 2007년 2월 : 세종대학교 전임연구원
- 2007년 3월 ~ 현재 : 세종대학교 컴퓨터공학과 (박사과정)
- 주관심분야 : 비디오 코딩, 컴퓨터 비전



Hae Kwang Kim

- 1985년 2월 : 한양대학교 전자공학과 (석사)
- 1985년 12월 ~ 1991년 12월 : 삼성전자 종합연구소 대리
- 1992년 9월 ~ 1994년 8월 : 프랑스 Paul-Sabatier 대학교 전산학 (D.E.A)
- 1994년 9월 ~ 1997년 3월 : 프랑스 Paul-Sabatier 대학교 전산학 (박사)
- 1997년 4월 ~ 2000년 2월 : 현대전자 정보통신연구소 책임연구원
- 2000년 3월 ~ 현재 : 세종대학교 컴퓨터공학과 부교수
- 주관심분야 : 비디오 코딩, 멀티미디어 시스템



Hae Chul Choi

- 1997년 2월 : 경북대학교 전자공학과 (학사)
- 1999년 2월 : KAIST 전기 및 전자공학과 (석사)
- 2004년 8월 : KAIST 전기 및 전자공학과 (박사)
- 2004년 9월 ~ 현재 : ETRI 전파방송연구단 방송미디어연구그룹 선임연구원
- 2007년 3월 ~ 현재 : 과학기술연합대학원 이동통신 및 디지털방송공학 겸임조교수
- 주관심분야 : 영상통신, 영상분할, 비디오 부호화, Scalable Video Coding, MPEG Video, ITU-T VCEG