DOI: 10.3969/j. issn. 1673-5188. 2016. 01. 003
http://www.cnki.net/kcms/detail/34.1294.TN.20160119.0957.002.html, published online January 19, 2016

Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard

Ming Li and Ping Wu
(ZTE Corporation, Shenzhen 518057, China)

Abstract

Multi-layer extension is based on single-layer design of High Efficiency Video Coding (HEVC) standard and employed as the common structure for scalability and multi-view video coding extensions of HEVC. In this paper, an overview of multi-layer extension is presented. The concepts and advantages of multi-layer extension are briefly described. High level syntax (HLS) for multi-layer extension and several new designs are also detailed.

Keywords

HEVC; multi-layer extension

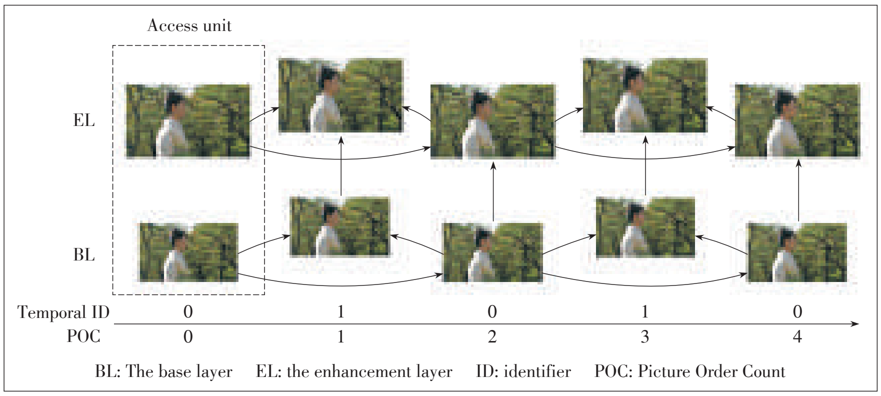
1 Introduction

High Efficiency Video Coding (HEVC) standard is the newest video coding standard of the ITU-T Q6/16 Video Coding Experts Group (VCEG) and the ISO/IEC JTC 1 SC 29/WG 11 Moving Picture Experts Group (MPEG). The first version of HEVC standard was released in 2013 [1] and referred to as "HEVC Version 1" standard. It is the next generation video coding standard after H.264/AVC, and achieves a dramatic improvement of coding efficiency relative to existing H.264/AVC. Testing results demonstrate that HEVC brings the same subjective quality by consuming an average of 50% fewer coding bits than that of H.264/AVC [2], [3]. HEVC standard is believed to be adopted in most of the potential applications employing video coding including broadcast, storage, streaming, surveillance, video telephony and etc.

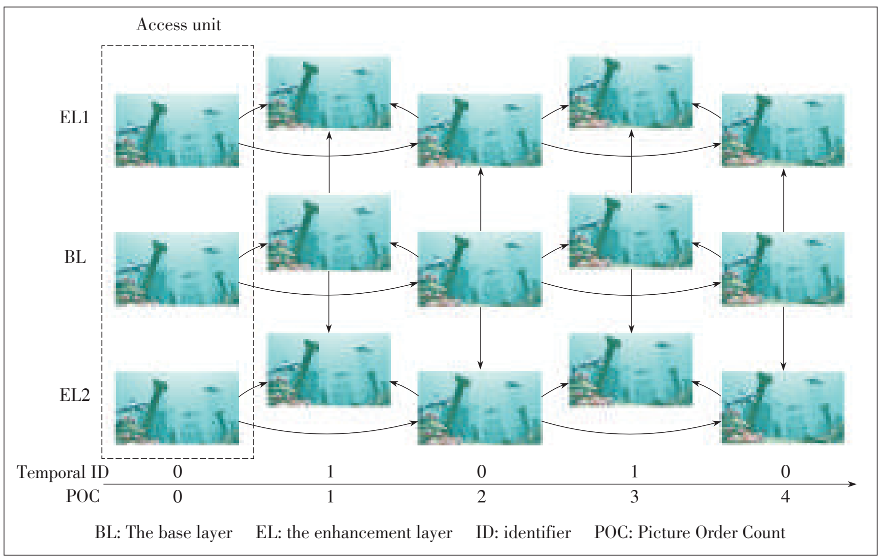
To address the requirements of a wider range of applications, key extensions of the HEVC Version 1 standard have been introduced by the Joint Collaboration Team on Video Coding (JCT-VC) and Joint Collaboration Team on 3D Video Coding Extension Development (JCT-VC) of VCEG and MPEG [4]. Range extensions (RExt), multiview extension (MV-HEVC) and scalable extension (SHVC) were introduced and included in the second version of HEVC standard [5]. 3D high-efficiency video coding extension (3D-HEVC) was finalized as the latest extension in the third version of HEVC standard [6] to enable high-coding of the representative 3D video signal of "multiview video + multiview depth". Currently, new extensions for Screen Content Coding (SCC) [7] and high dynamic range and wide color gamut (HDR & WCG) [8] are being developed in VCEG and MPEG and are scheduled for release in the coming one or two years.

In the second version of the HEVC standard, the concept of layer refers to a scalable layer (e.g. a spatial scalable layer) in SHVC or a view in MV-HEVC. **Fig. 1** shows an example SHVC bitstream of spatial scalability with two layers. The base layer (BL) is of lower resolution, and the enhancement layer (EL) higher resolution. In both BL and EL, temporal scalability, which is already supported in HEVC Version 1, is ensured by using hierarchical B-pictures, and the pictures with temporal identifier (ID) equal to 0 and 1 form sub-layers of BL and EL, respectively. A similar layer concept is also applied to the MV-HEVC bitstream in **Fig. 2**, where a layer corresponds to a view and one base view and two dependent views are referred to as BL (central view), EL1(right view) and EL2 (left view), respectively. In both SHVC and MV-HEVC, BL provides backward compatibility to single layer HEVC codec.

For both SHVC and MV-HEVC, the inter-layer prediction is the key to superior coding efficiency compared with simulcast. MV-HEVC is based on HEVC Version 1 standard and follows the same principle of multiview video coding (MVC) extension of H.264/AVC [9], which does not introduce changes to block-level algorithms. In MV-HEVC, inter-layer prediction is carried out by high-level operations to put the reconstructed pictures from reference views to the reference lists of the current picture. Block level tools are adopted in 3D-HEVC for further improving coding efficiency. In the development of SHVC in JCT-VC, in-depth study and testing has been conducted to evaluate the overall performance of two inter-layer prediction



▲**Figure 1. An example of SHVC bitstream.**



▲**Figure 2. An example of MV-HEVC bitstream.**

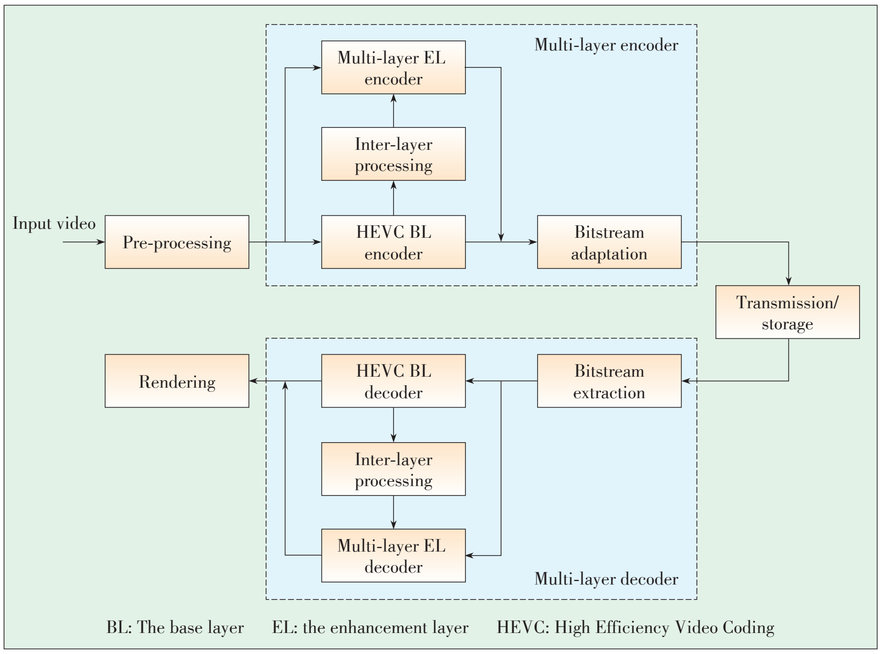
schemes of high-level extension using reference index to signal inter-layer reference and block level extension with dedicated tools for coding EL [10]. Considering the trade-offs among design advantage, coding efficiency and complexity, JCT-VC selects high-level extension approach for SHVC. Therefore, a multi-layer extension of HEVC is established and employed as the common structure for both MV-HEVC and SHVC, as well as other future extensions using the layered structure.

2 Multi-Layer Extension

Multi-layer extension represents scalable and multiview structures by layers and offers flexibility to combinations of different types of layered structures. Compared to Scalable Video Coding (SVC) and MVC extensions of H.264/AVC standard, multi-layer extension provides an identical framework for SHVC and MV-HEVC extensions, as well as future extensions employing multi-layer structure, of HEVC standard. From the perspective of decoding operations, the main feature is that the decoding operations at block level are kept the same as those specified for single layer profiles of HEVC standard. For example, an SHVC decoder conforming to Scalable Main profile is implemented using the same block level decoding algorithms as the ones specified in HEVC Main Profile. In multi-layer extension, inter-layer prediction is enabled by putting the reconstructed pictures of lower layers into the reference lists of the pictures in higher layers within the same access unit.

An example of an end-to-end system employing multi-layer bitstream with a BL and one EL is shown in **Fig. 3**. At the source side, the pre-processing module is used to get the BL video and EL video for BL encoder and EL encoder, respectively. For example, when the multi-layer encoder is an SHVC encoder of spatial scalability with two layers, the pre-processing module generates a lower resolution video for BL encoder by down-sampling the input video. When the multi-layer encoder is an MV-HEVC encoder with its input of stereo video of two views, the pre-processing module will choose one view as BL and the other as EL according to the configurations. The bitstream adaptation module forms the multi-layer bitstream by combining the coding bits of BL and EL following the specifications of multi-layer extension. At the destination side, the bitstream extraction module in multi-layer decoder is to separate BL and EL stream from the received multi-layer bitstream, e.g. by running the bitstream extraction process. The rendering module at destination is to show the decoded video according to the requests from users. In the above mentioned examples, the rendering module displays the desired video from an SHVC decoder of spatial scalability, or constructs a stereo pair from the decoded videos from an MV-HEVC decoder for 3D viewing.

At both sides of source and destination, when inter-layer prediction is used, the inter-layer processing module accesses the decoded picture buffer (DPB) of the BL encoder (or BL decoder) to get the corresponding reconstructed BL picture to generate the inter-layer reference picture for encoding (or decoding) the EL picture in the same access unit. When one or more parameters of resolution, bit depth and colour gamut of the BL reconstructed picture are different from the parameters of EL picture, the inter-layer processing module performs necessary operations on the BL reconstructed picture that may include conversions of texture, color and motion field. The output picture of the inter-layer processing module is then put into the inter-layer reference picture set and marked as "used for long-term reference" in encoding (or decoding) the EL picture. In the pro-



▲**Figure 3. End-to-end structure of system employing multi-layer bitstream.**

cess of constructing the reference picture lists for the EL picture, the inter-layer reference picture is added to the reference picture list, and assigned with a reference index along with the temporal reference pictures of the EL picture. In HEVC multi-layer extension, the parameters for inter-layer processing and inter-layer prediction are signalled in parameter set and slice segment header. Inter-layer prediction is signalled by setting the values of the syntax elements of reference index in prediction unit (PU) equal to the corresponding reference index of the inter-layer reference picture, and carried out without changing any operations below slice level specified in HEVC Version 1 standard. This is referred to as "high level syntax (HLS) extension scheme".

The multi-layer codec is of a multi-loop coding structure. The major advantage, especially compared to SVC employing a single-loop design, is the HLS extension scheme that reuses the block level algorithms already designed for HEVC Version 1 codec. The additional operations newly introduced to the EL codec is to interpret the dependency among layers for inter-layer prediction and the inter-layer processing to generate the inter-layer reference picture to be involved in reference lists for EL pictures. Accordingly, the EL codec needs to access the DPB of BL codec for BL reconstructed picture and maybe also the associated motion information of BL picture to derive motion predictor for EL PUs. As the interface for BL motion information already exists for motion prediction at BL, EL can reuse such interface to get BL motion information when an EL PU referencing to the inter-layer reference picture for motion prediction. In this way, a multi-layer codec can be conveniently designed and implemented, for example, by taking the already existing HEVC Version 1 codec as BL codec and integrating an inter-layer processing module as well as high-level interpretation for multi-layer structure signalled in parameter sets and slice segment header in HEVC Version 1 codec to form EL codec. In comparison with SVC codec, the HLS extension scheme avoids a completely new design of EL by reusing most parts of the HEVC Version 1 design, and also saves a large amount of extra interfaces to be implemented on already available single layer design to meet EL's accessing. Therefore, the HLS extension scheme greatly brings down the workload for SHVC and MV-HEVC codec design and implementation, which is believed to push wide adoption of layered coding extensions of HEVC to applications.

3 HLS for Multi-Layer Extension

To describe the common layered structure of SHVC and MV-HEVC, HLS specified in HEVC Version 1 standard is further extended, including network abstraction layer (NAL) unit header, parameter sets, slice segment header and supplement enhancement information (SEI). New designs are being introduced to make the multi-layer extension more flexible for applications and future extensions using layered structure.

3.1 NAL Unit Header and Parameter Sets

Multi-layer extension shares the same NAL unit header as that specified in HEVC Version 1 standard. In NAL unit header, a syntax element namely nuh_layer_id is coded in 6 bits to signal the layer to which a video coding layer (VCL) NAL unit or non-VCL NAL unit belongs to. In HEVC Version 1 standard, the value of nuh_layer_id in the conformed bitstream shall be 0 and the conformed decoder ignores all NAL units with nuh_layer_id not equal to 0. In the multi-layer extension, the value of nuh_layer_id is always 0 in the BL NAL units, which are backward-compatible with HEVC Version 1 codec. With nuh_layer_id distinguishing the NAL units of different layers, the NAL unit types defined in HEVC Version 1 standard are re-used to indicate the type of raw byte sequence payload (RBSP) data structure contained in the EL NAL units and signalled by the existing syntax element nuh_unit_type in NAL header. Therefore, no new NAL types are introduced by the multi-layer extension.

Video parameter set (VPS) is adopted in the development of HEVC Version 1 standard. In multi-layer extension, VPS is further extended to signal the common information for the layers. VPS could be used in session negotiation to provide the characteristics of the multi-layer bitstream and decoding capability. Layers indicated by VPS can be a spatial/quality scalable layer, a view, and an auxiliary picture layer. VPS describes the number of layers and dependency relationship among the layers. The dependency relationship indicates the reference layers for inter-layer prediction when decoding the current layer. In multi-layer extension, a layer can only reference lower layers. VPS signals the representation format for each layer. VPS provides the information for bitstream conformance and operation points, including profile, tier, level, layer sets, hypothetical reference decoder (HRD) parameters, and etc. Video usability information (VUI) for multi-layer bitstream is also signalled in VPS.

Extension of sequence parameter set (SPS) for multi-layer extension introduces only one syntax element inter_view_mv_vert_constraint_flag, which is to signal whether the vertical components of motion vectors used for inter-layer prediction are constrained. Extension of picture parameter set (PPS) for multi-layer extension includes parameters for picture processing to derive inter-layer reference pictures, including reference picture scaling offsets, reference region, reference phase offsets and colour mapping.

3.2 Layer-Wise Decoding and Picture Order Count (POC) Resetting

In SVC, the decoding process can only correctly start from an access unit with all pictures coded as instantaneous decoding refresh (IDR) pictures. At the encoder side, coding an access unit with all IDR pictures always leads to an instantaneous bit-rate increment. By comparison, multi-layer extension releases the constraint that the intra random access point (IRAP) pictures are aligned within one access unit. A device can start decoding a multi-layer bitstream from an access unit with the BL picture being a random access picture, and make an access to EL layers later, for example, when a random access picture is in the EL layer.

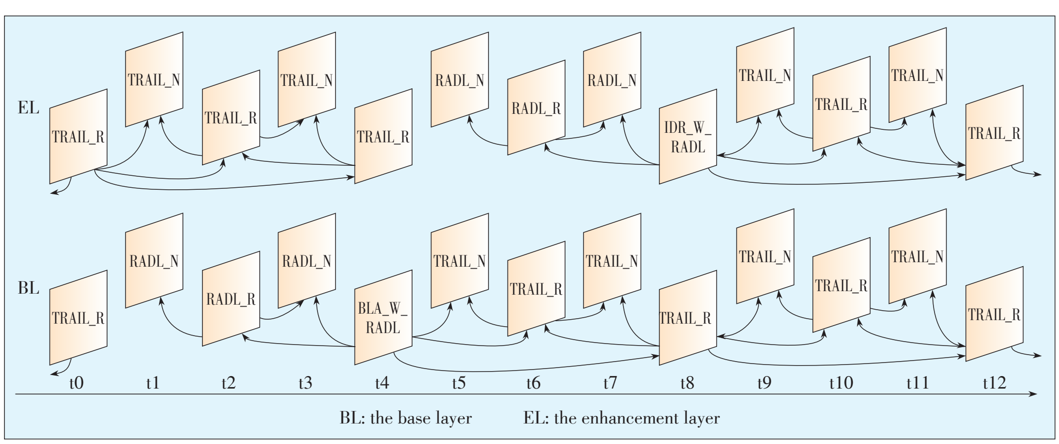
Fig. 4 shows a multi-layer bitstream structure supporting layer-wise decoding. The access unit AU(t4) is an access unit with BL picture coded as a broken link access (BLA) picture. The multi-layer decoder can make an access to the BL of this multi-layer bitstream from AU(t4) first, and then access the EL from AU(t8) in which the EL picture is an IDR picture. Note that in the example shown in Fig. 4, the EL picture in AU(t4) is not decodable, because its temporal reference picture in EL (i.e. the EL picture in AU(t0)) is not available when accessing from AU(t4).

As with the HEVC Version 1 standard, POC is used in multi-layer extension to represent the relative output order of pictures within each layer, and generally the derivation of POC value of a picture does not depend on the POC values of pictures in other layers. As IDR picture and BLA picture will force the complete POC value or the most significant bits (MSB) of the POC to be 0, the POC values of pictures within an access unit may be different (e.g. BL and EL pictures in AU(t4) in Fig. 4), which violates the constraint that the pictures in the same access unit have the same POC value. To solve this, a POC resetting process is designed for multi-layer extension, which resets the POC values for the pictures in an access unit when such pictures in different layers would get different POC values following the normal POC derivation process as specified in HEVC Version 1 standard [11]–[13]. In addition, to keep the consistency of the POC differences in reference picture set (RPS) operations, POC shifting operations are performed after resetting on previous pictures in decoding order as a decrement in POC values. The parameters for POC resetting are signalled in slice segment header extension.

3.3 Hybrid Coding

Unlike SVC, which has the BL coded in H.264/AVC and the BL bitstream embedded in the SVC bitsteam, multi-layer extension enables the BL bitstream to be provided by external means not specified in the second version of HEVC standard. Furthermore, the BL bitstream provided by external means can be generated by any single layer encoder besides HEVC, such as H.264/AVC, MPEG-2, and etc. This feature of multi-layer extension can be referred to as hybrid codec scalability or hybrid coding. In this case, decoding of the external BL is out the scope of multi-layer extension, and hybrid coding is carried out following the decoding operations of EL specified in multi-layer extension by forwarding the reconstructed pictures after decoding external BL to EL and inserted into the EL reference lists for inter-layer prediction.

One use case for hybrid coding is long-term gradual upgrading a system by appending an HEVC EL to the existing stream coded by other standards (e.g., MPEG-2, H.264/AVC). The HEVC EL may provide higher resolution, higher dynamic



▲**Figure 4. Example multi-layer bitstream structure with layer-wise decoding.**

range and/or wider color gamut to enhance the viewing quality or provide a view other than the view represented by BL to form a stereo pair for 3D viewing. Accordingly, the devices with hybrid coding will provide more vivid viewing experience, while legacy devices can still provide basic perceptual quality by discarding the HEVC EL bits. The main advantage is that a lot of bandwidth can be saved compared to simulcast solution of two separate bitstreams while maintaining backward compatibility during upgrade. However, the cost is that the devices with hybrid coding need to support a number of standards and conduct exact synchronization of HEVC EL pictures and BL pictures in both inter-layer prediction and picture output process (especially for stereo pair in 3D viewing).

3.4 Independent Non-Base Layer (INBL)

Multi-layer extension supports INBL. The INBL is an EL in multi-layer bitstream, which is coded without using inter-layer prediction and conforms to a single layer profile. That is, the only difference of an INBL and ordinary single layer bitstream is that the nuh_layer_id in NAL units in INBL stream is not equal to 0. In VPS, a flag is signalled along with profile, tier and level of a layer to indicate whether this layer is an INBL. INBL provides a simulcast layer in the multi-layer extension. This flag is also used to signal the capability of a decoder whether INBL can be processed, which is used in, for example, session negotiation. An INBL rewriting process is also designed for multi-layer extension to convert the INBL bitstream extracted from multi-layer bitstream into a bitstream conforming to a single layer profile.

4 Conclusions

This paper gives an overview of the concepts and HLS in multi-layer extension of HEVC Version 1 standard. Multi-layer extension is developed based on HEVC Version 1 standard and serves as a common architecture for HEVC extensions using layered structure, including SHVC and MV-HEVC in the second version of HEVC. High-level extension approach is used to multi-layer extension without changing the block level decoding operations already specified in single layer HEVC profiles. This design principle enables the implementation of SHVC and MV-HEVC to be built on existing single layer HEVC codec with additional inter-layer reference picture processing operations, which dramatically alleviates the workload of codec design. Additionally, several new designs are also developed for multi-layer extension to achieve more flexibility for applications and future extensions using layered structure. The benefits of multi-layer extension will facilitate widespread adoption of layered coding extensions of HEVC to applications.

Acknowledgment

The authors thank the experts of ITU-T VCEG, ISO/IEC, MPEG, the ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) and the ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development (JCT-VC) for their contributions.

References

[1] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v1, Apr. 2013.
[2] J. -R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.22, no.12, pp. 1669–1684, Dec. 2012. doi: 10.1109/TCSVT.2012.2221192.
[3] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, "Report on HEVC compression performance verification testing," ITU, JCT-VC Document JCTVC-Q1011, Mar. 2014.
[4] G. J. Sullivan, J. M. Boyce, Y. Chen, *et al.*, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013. doi: 10.1109/JSTSP.2013.2283657.
[5] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v2, Oct. 2014.
[6] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v3, Apr. 2015.
[7] R. Joshi, S. Liu, G. J. Sullivan, *et al.*, "HEVC screen content coding draft text 4," ITU, JCT-VC Document JCTVC-V1005, Oct. 2015.
[8] K. Minoo, P. Yin, T. Lu, *et al.*, "Exploratory test model for HDR extension of HEVC," MPEG Document N15792, Oct. 2015.
[9] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Recommendation H.264, Jun. 2011.
[10] L. Guo, Y. He, D.-K. Kwon, *et al.*, "TE 2: summary report on inter-layer texture prediction signaling in SHVC", ITU, JCT-VC Document JCTVC-L0022, Jan. 2013.
[11] Y. Chen, Y.-K. Wang, and A. K. Ramasubramonian, "MV-HEVC/SHVC HLS: cross-layer POC alignment," ITU, JCT-VC Document JCTVC-N0244 & JCT-3V Document JCT3V-E0075, Jul. 2013.
[12] G. J. Sullivan, "JCT-VC AHG report: multi-layer picture order count derivation (AHG10)," ITU, JCT-VC Document JCTVC-P0010, Jan. 2014.
[13] Hendry Fnu, A. K. Ramasubramonian, Y.-K. Wang, *et al.*, "MV-HEVC/SHVC HLS: on picture order count," ITU, JCT-VC Document JCTVC-P0041 & JCT-3V Document JCT3V-G0031, Jan. 2014.

Manuscript received: 2015-11-10

Biographies

Ming Li (li.ming42@zte.com.cn) received the BEng degree in telecommunication engineering and PhD degree in communication and information systems from Xidian University, China, in 2005 and 2010, respectively. He has been a standardization engineer in video coding in ZTE since 2010. His current research interests include video coding and multimedia communication.

Ping Wu (ping.wu@zte.com.cn) received the BEng degree in Electrical Engineering from Tsinghua University, Beijing, China in 1985 and received the PhD degree in signal processing from Reading University, United Kingdom in 1993. From 1993 to 1997, he was a research fellow in the area of medical data processing in Plymouth University, United Kingdom. From 1997 to 2008, he was a consultant engineer in News Digital Systems Ltd, Tandberg Television, and Ericsson. He participated in the development of ISO/IEC MPEG and ITU-T video coding standards. He also supervised the engineering team to build the High Definition H.264 encoder products for broadcasters. From 2008 to 2011, he joined Mitsubishi Electric Research Centre Europe and continued to participate in High Efficiency Video Coding (HEVC) standard development with contributions in Call for Evidence and Call for Proposal. From 2011, he has been a senior specialist in video coding in ZTE. He has many technical proposals and contributions to the international standards on video coding over past 18 years.