EXTENSIONS OF HIGH EFFICIENCY VIDEO CODING STANDARD: AN OVERVIEW

Zoran Miličević[1], Zoran Bojković[2]
[1]Department of Telecommunications and IT Genaralstaff of the SAF in Belgrade
[2]University of Belgade

Abstract: *For extension of High Efficient Video Coding (HEVC), a high quality experience is offered by using three-dimensional (3-D) video. Also, the primary usage scenario for multiview video is to support 3-D video applications, where 3-D depth perception of a visual scene is provided by a 3-D display system. The efficient representation and compression of stereo and multiview video is a central component of any 3-D or multiview system, since it defines the format to be produced, transmitted and displayed. On the other hand, scalable extension of the HEVC standard allows for the scaling of video to different spatial resolution, bit depth, and/or color gamut while maintaining high coding efficiency. In the first part of this article, we deal with 3-D and multiview extensions of HEVC. Coding architecture together with corresponding for the efficient compression is included, too. In the second part, scalable extension of HEVC standard with the migration path from high-definition (HD) to ultra HD is presented.*

Key words: *High efficient video coding, Coding architecture, Coding tools, Efficient compression, Scalability.*

1. Introduction

In the family of video coding standards, High Efficient Video Coding (HEVC) has the potential to replace/supplement all the existing standards i.e., MPEG and H.26x series including H.264/Advance Video Coding AVC. The main goal was to achieve a compression gain higher when compared to the H.264/AVC at the same video quality. Also, the focus was on monoscopic video. Today, a primary usage of HEVC is to support the delivery of ultra-high definition (UHD) video. For extension of HEVC, a Join Collaborative Team on 3-D Video Coding Extensions Development (JCT-3V) has been formed between ISO/IEC and ITU-T. The first version of the standard was approved in 2013.

A high-quality and immersive multimedia experience (which has been feasible on consumer electronics platforms through advances in display technology, signal processing, transmission technology and circuit design) are offered by 3-D video. In addition to advances on the display side, notable increase has been in the production of 3-D content. As for the delivery of 3-D video, it is essential to determine an appropriate data format taking into consideration the constraints imposed by each delivery channel including bit rate and compatibility replacement [1]. The 3-D representation compression formats together with signaling protocols define the interoperability of the system.

Since developing the H.264/MPEG-4 AVC standard, the Joint Video Team of the ITU-T Video Coding Experts Group (VCEG) and ISO/IEC MPEG have standardized an extension of the technology that is referred to as multiview video coding (MVC). Multiview video formats [2] are able to provide depth perception of a visual scene through the appropriate 3-D display system. Also, video formats are enhancing not only the viewing experience through depth, but enable free view-point video, which may be useful in surveillance or teleconference applications. The H.264/MPEG-4 AVC standard is the basis for multiview video coding formats. The compressed multiview stream includes a base view bit stream that is coded independently from all other views in a manner compatible with decoders for single-view profile of the standard.

In July 2012, the ISO/IEC MPEG and ITU-T VCEG issued a joint call for proposal for scalable video coding extensions of HEVC standard. After three month, 20 responses to this call were received from companies, research institutes, and universities worldwide. In that way scalable HEVC (SHVC) project was established [3]. In scalable video coding, interlayer prediction is a powerful tool. Scalable extensions the HEVC video coding standard support up to 8 layers. There are two different approaches. The first approach relies on new block-level signaling to indicate whether the enhancement-layer block is predicted from the base layer or the current layer. On the other side, the second approach treats the reconstructed picture of the base layer as an interlayer reference picture. It uses the existing reference index signaling that is already part of the block-single layer HEVC codec to identity whether the block-level prediction comes from the base layer in the current layer.

The first part of this article deals with 3-D and multiview extensions of HEVC. Coding architectures together with coding tools for the efficient compression are included. The second part is devoted to scalable extension with the migration path from HD to UHD video.

2. 3-D an multiview extensions of HEVC

Recent improvements in 3-D technology led to a growing interest in 3-D video. The availability of 3-D capable TV sets and Blue-ray players, the introduction of 3-D broadcast channels, and the release of 3-D Blue-ray discs bring 3-D video into consumers' homes. Also, the number of cinema screens capable to showing 3-D movies as well as the number of movies produced in 3-D has been constantly increased. Autostereoscopic displays, which provide a 3-D viewing experience without glasses, are improved and are considered as a promising technology, for future 3-D home entertainment. In contrast to common stereo displays, autostereoscopic displays require, not only two but a multitude of different views for providing the 3-D viewing experience [1]. Multiview extensions of HEVC utilized the same design principles of multiview coding in H.264/AVC framework. This scheme provides backward compatibility for monoscopic decoding and utilizes inter-view prediction between the texture views. The bit rate required for multiview video coding (MVC) extensions of H.264/AVC increases approximately linearly with the number of coding views. MVC is not appropriate for delivering 3-D content for autostereoscopic displays. A promising alternative is the transmission of 3-D video in the multiview video plus depth (M.V.D) format.

Going back to the history, MVC has been active research area from 1986 with early work on disparity constructed prediction [4], followed by other coding schemes in the late 1980s and early 1990s [5], [6]. In 1996, the International Video Coding Standard H.262/MPEG-2 Video [7], [8], was amended to support the coding of multiview video by means of design features originally intended for temporal scalability [9]. However, the multiview extension of H.262/MPEG-2 video was never displayed in actual products. It did not offer a very compelling compression improvement due to limitation in the coding tools enabled for inter-view prediction in that design [10], [11]. Backward compatibility for monoscopic decoding and utilization of inter-view prediction between the texture views are provided. The basic block level decoding process remains unchanged. Single layer codec designs that have been designed for 2-D applications are extended without major implementation changes to support stereo and multiview applications.

2.1. Coding architecture alternatives

To achieve higher compression efficiency an alternative coding architecture would leverage the benefits of block level coding tools [2]. In this architecture, the base view is fully compatible with HEVC in order to extract monoscopic video. Only the dependent views would utilize additional tools. It has been recognized that there is significant correlation between motion and mode parameters between the base and dependent views. When this correlation is exploited, this leads to notable bit rate savings.

In terms of compression formats, it is anticipated that extension of the HEVC standard would be developed that support the efficient inclusion of depth information [2]. One of the desirable characteristics of this format is for stereo video to be easily extracted to support existing stereoscopic displays. In such case, the dependency between the video data and depth data may be limited. Allowing for a greater degree of dependency between the different components, may provide more significant benefits in term of compression capability and rendering performance.

In the context of depth-based 3-D formats, there are many variations. For example in hybrid, AVC and depth-base 3-D formats, the base view can be coded with AVC, while additional texture views and supplemented depth video can be encoded with HEVC.

2.2. Coding tools for the efficient compression

For the efficient compression of 3-D video data with multiple video and depth components, a number of coding tools are used to exploit the dependencies among the components. A 3-D video encoder can select the best coding method for each block from a set of conventional 2-D coding tools and additional new coding tools such as inter-new prediction, motion/mode parameter prediction, and depth coding [2].

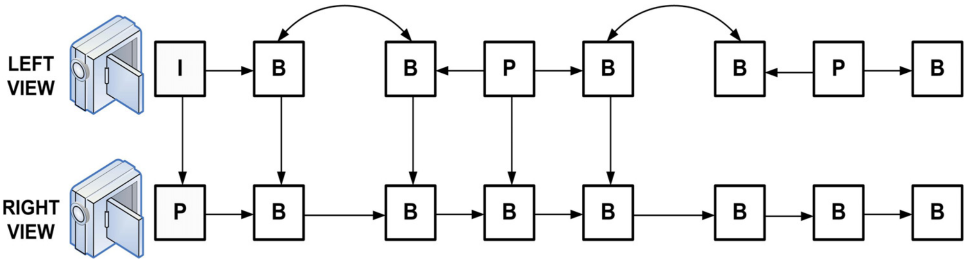

Figure 1. *Inter-view prediction presentation.*

Intra-view prediction is employed in all standardized designs for efficient multiview video coding. The concept is to exploit both inter-new and temporal redundancy for compression. Since the cameras of a multiview scenario capture the same scene from nearly view points, substantial inter-view redundancy is present. This holds for both texture views and the corresponding depth map images associated with each view. Thus, inter-prediction can be applied to both types of data independently. Illustration of inter-view prediction is shown in Figure1. Inter-view prediction in enabled in video coding standards such as AVC and HEVC, through the flexible reference picture management capabilities of these standards. The decoded pictures from other views are made available in the reference picture lists for use by the inter picture prediction processing. As a result, the reference picture lists includes the temporal reference pictures that may be used to predict the current picture along with the inter-view reference pictures from neighboring views. Block-level decoding modules remain unchanged. Only small changes to the high-level syntax are required. As for the prediction, it is adaptive. It means that the best predictor among temporal and inter-view references can be selected on a block basis in term of rate-distortion cost.

In the context of multiview video coding, it is possible to inter side information used in the decoding process e.g. motion vector for a particular block, based on other available data e.g. motion vectors from other blocks. Illustration of motion prediction between views is shown in Figure 2. The motion vector of view 1 is inferred from the motion vector of view 0 from corresponding, blocks at time 1 based on the display between blocks. Interference of coded block data between views could be considered an extension of the basic principle of direct mode prediction in AVC for 2-D video coding. In 2012, it has been proposed to estimate the disparity from sample locations instead of using a global disparity vector [12]. In this way, motion parameters of a block in the reference view can be used as a motion candidate for the current view. Advanced schemes based on the concept of motion vector competition have also been shown to provide competitive performance and further gains [13].

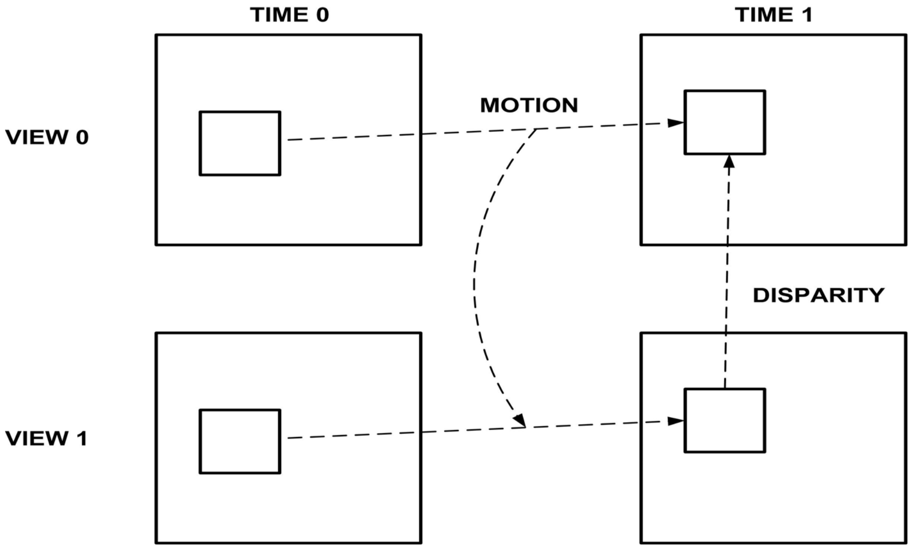

Figure 2. *Motion prediction between views.*

A coding mode that infers the block partitioning of sub-blocks and associated motion parameters from the co-located block in the associated video picture has been described [14]. Then technique decides whether partitioning and motion information are inherited from the co-located regions of the video pictures for each depth block, or whether new motion data should be transmitted. Its such information is inherited, no additional bits for partitioning and motion information are required.

Depth information can be used at the encoder to realize more efficient compression with view synthesis prediction schemes. A shift value in texture samples from the original views represents a depth sample. Coding errors in depth maps result in wrong pixel shifts in synthetized views. This leads to artifacts especially along visible object boundaries. Thus, a depth compression algorithm needs to preserve depth edges much better than traditional coding method. A depth signal consists of larger homogeneous areas inside scene objects and sharp transitions along boundaries between objects at different depth values. Therefore, in the frequency spectrum of a depth map, low and very high frequencies are dominant [2]. Video compression algorithm are designed to preserve low frequencies, while image blurring occurs in the constructed video at high compression rates. The need for compression techniques that are adopted to these special characteristics of the depth signal as well as the requirement on maintain the fidelity of edge information in the depth maps has motivated research in this area.

3. Scalable extensions of HEVC standard: the migration path from HD to UHD video

Scalable extensions of the HEVC video coding standard, referred to as SHVC supports up to 8-layers [3]. Two-layer scalable system consists of the base layer (BL) and an enhancement layer (EL). The scalable system is shown in Figure 3.

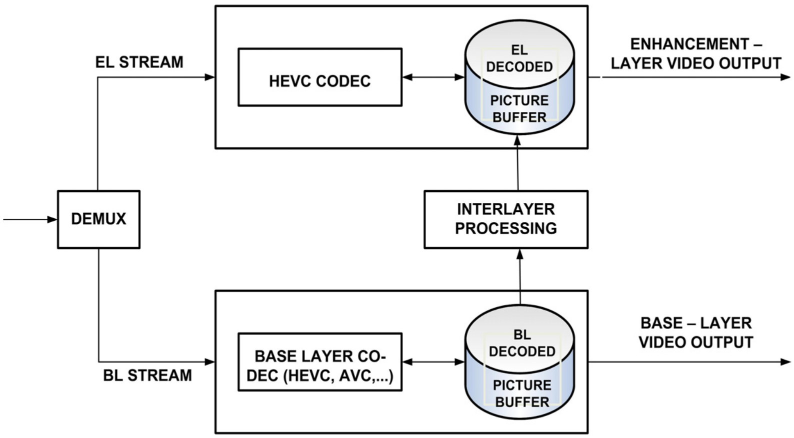

Figure 3. *Scalable system architecture.*

The BL reconstruction is retrieved from the base layer decoded picture buffer (BL DPB). Appropriate interlayer (ILR) processing is applied to the reconstructed BL picture to obtain the ILR picture. The ILR picture is then put into the EL decoded picture buffer (EL DPB) as a long term picture and use along with the EL temporal reference pictures for EL coding. The system has a few designs advantages. First at all, block level logic of the EL codec remains the same as a single layer HEVC codec. Any necessary changes to the EL-decoders are allowed only at the slice header and above, that is high-level syntax (HLS) changes only. Keeping the detailed block-level operations compatible with a single layer-code is beneficial because existing design of HEVC codec can be maximally reused, thereby reducing the implementation costs of SHVC codec. As for the base layer codec, it can operate as a "black box" because the scalable coding of the EL needs only the reconstructed BL pictures. This makes it easier to support a multilayer video coding system where different codecs are used to code different layers. In that way, we arrive the hybrid codec scalability. This feature allows previous generation codec to be used in the BL for backward capability, while more efficient HEVC codec is used in the EL to help improve coding performance.

The scalable system represents an architecture design that is compatible with multiview extensions to HEVC (MV-HEVC). A unified architecture between the two extensions SHVC and MV-HEVC is desirable. Unified design carries a coherent message and can increase the chance of commercial adoption of both extensions. It becomes easier to combine these two extensions in the future allowing, for example, combined view and spatial scalability [15].

The interlayer processing block is necessary only for SHVC, but not for MV-SHVC. In SHVC, the video signal represented in two layers can have different parameters such as spatial resolution, sample bit depth, color gamut, etc. As for the MV-HEV, it does not allow differences in way of these video parameters. Depending on which parameters differ between the BL and EL, appropriate forms of interlayer processing are applied to the BL reconstructed pictures to derive the interlayer pictures for efficient EL coding. Table 1 summarized the scalability features supported by SHVC and compares them with the prior generation scalable coding standard. SHVC supports a richer set of scalability features.

Table 1. Compression of scalability features in SVC and SHVC.

Scalability features	**SCALABLE CODINE STANDARD**	
	SVC	SHVC
Temporal	X	X (HEVC version 1)
Spatial	X	X
SNR	X	X
Hybrid codec		X
Bit depth		X
Color gamut		X

In addition to temporal, spatial and signal-to-noise (SNR) scalability, SHVC also supports hybrid codec scalability and bit depth scalability. Temporal scalability is already supported by HEVC, revision 1 published in 2013. It should be noted that the combination of hybrid, spatial, bit depth will enable the migration path from HD to UHD. The UHD video format defines high spatial resolution and high frame rates, together with a wide color gamut.

To achieve higher coding efficiency, appropriate forms of interlayer processing are needed to desire the interlayer (IRL) pictures for coding the EL. Interlayer processing in SHVC includes the resampling process for spatial and bit depth scalability and the color mapping process for color gamut scalability. The standard supports resampling of both BL and reconstructed texture and BL motion.

4. Conclusion

Extensions of high efficient video coding provide an attractive operating point in terms of coding efficiency and complexity. Within the HEVC framework, multiview and 3D extensions are enabled. The inclusion of depth information is another key target of the current 3D video coding extensions development activity. When inter-view prediction is enabled, there exists dependency between the different views. Additionally, decoded information for texture components may be used in the decoding of the depth.

Coexistence of formats, coupled with an increased level of network heterogeneity, makes scalable video coding suitable choice for delivery because it offers storage and bandwidth efficiency together with channel error robustness.

The scalable extensions of HEVC, SHVC, adopts a scalable coding architecture with only high-level syntax changes relative to its base codec, which allows SHVC to be deployed with significantly reduced information cost. By supporting a rich set of scalability features with high scalable coding efficiency, SHVC addresses the increasing market demand for higher quality and higher value video content delivery to the home.

References

[1] A. Vetro, T. Wiegand, G. J. Sullivan, "Overview of the Stereo and Multuview Video Coding Extensions of the H.264/MPEG-4 AVC Standard", *Proceeding of the IEEE*, vol. 23, no. 4, pp. 626-641, April 2011.

[2] A. Vetro, D. Tien, "Analysis of 3D and Multuview Extensions of Emerging HEVC Standard", *Mitsubishi Electric Research Laboratories,* TR 2012-068, August 2012.
[3] Y. Pe, P. Andrivon, "The Scalable Extensions of HEVC for Ultra-High-Definition Video Delivery", *IEEE Multimedia*, vol. 21, no. 3, pp. 58-64, July-September 2014.
[4] M. P. Lukacs, "Predictive coding of multi-view point image sets", *Proceeding of IEEE International Conference Acoustic Speech Signal Processing*, Tokyo, Japan, vol. 1, pp. 521-524, 1986.
[5] I. Dinstain et al., "On the compression of stereo images: Preliminary results", *Signal Processing, Image Communications*, vol. 17, vol. 4, pp. 373-382, August 1989.
[6] M. G. Perkins, "Data Compression of stereo pairs", *IEEE Transactions Communications*, vol. 40, vol. 4, pp. 684-696, April 1992.
[7] ITU-T and ISO/IEC JTC, "Generic coding of moving pictures and associated audio information – Part 2: Video", ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video), 1994.
[8] K. R. Rao, Z. S. Bojkovic, D. A. Milivanovic, "*Multimedia communication systems: Techniques*, *Standards and Networking* ", Prentice Hall, 2002.
[9] A. Puri, R. V. Kollaritis, B. G. Haskell, "Stereoscopic video compression using temporal scalability", *Proc. SPIE Conf. Vis. Commun. Image Processing*, vol. 3501, pp.745-756, 1995.
[10] J.-R. Ohm, "Stereo/multiview video encoding using the MPEG family of standards", *Proc. SPIE Conf. Stereoscopic Display Virtual Reality Syst., VI*, San Jose, CA, Jan. 1999.
[11] G. J. Sullivan, "Standards-based approaches to 3D and multiview video coding", *Proc. SPIE Conf. Appl. Digital Image Process., XXXII*, San Diego, CA, August 1999.
[12] H. Schwartz, T. Wiegand, "Inter-view prediction of motion data in interview video coding", *Proceedings Picture Coding Symposium*, Krakov, Poland, 2012.
[13] J. L. Lin, Y. W. Chen, Y. L. Charg, Y. P. Tsai, Y. W. Huang, S. Lei, "3D-CES a resolution on the motion vector competition – based skip/Direct mode with explicit signaling", Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) Doc. JCT.3V-A0045, Stockholm, SE, 2012.
[14] M. Winkon, H. Schwartz, T. Wiegand, "Motion Vector Interference for High Efficient 3D Video plus Depth Coding", *Proceedings Picture Coding Symposium*, Krakov, Poland, 2012.
[15] G. J. Sullivan et al., "Standardized Extensions of High Efficiency Video Coding (HEVC)", *IEEE J. Selected Topic in Signal Processing*, vol. 7, no. 6, pp. 1001-1016, 2013.

Sadržaj: *Za proširenje standarda vrlo efikasnog video kodovanja (HEVC) koristi se bogato iskustvo stečeno primenom trodimenzionalnog (3-D) videa. Primarni scenario ogleda se u podršci 3-D aplikacji, gde je kompresija višestrukog (tzv. "multiview") videa centralno pitanje. S druge strane, skalabilno proširenje HEVC standarda, obezbeđuje skaliranje videa u pogledu zauzete prostorne rezolucije, dubine i/ili boje, uz istovremeno efikasno kodovanje. U prvom delu rada analizirano je proširenje HEVC standarda uključujući kodne arhitekture i kodne alate. Drugi deo rada predstavlja skalirano proširenje HEVC standarda sa prelaskom od videa visoke rezolucije na video ultra visoke rezolucije.*

Ključne reči: *Vrlo efikasno video kodovanje, kodne arhitekture, kodni alati, efikasna kompresija, skalabilnost.*

PROŠIRENJE STANDARDA ZA VRLO EFIKASNO VIDEO KODOVANJE: PREGLED

Zoran Miličević, Zoran Bojković