HEVC 기반 삼차원 영상의 스케일러블 전송을 위한 확장 시스템

최병두, 조용진, 박민우, 이진영, 위호천, 김찬열
삼성전자
b.d.choi@samsung.com

High-level framework for scalable 3D video coding based on HEVC

Byeongdoo Choi, Yongjin Cho, Min Woo Park, Jin Young Lee, Hocheon Wey, Chanyul Kim
Samsung Electronics

요 약

A HEVC-based scalable 3D video coding system is proposed. The proposed system supports scalable transmission of multiview video data with depth maps. Key technologies in this system are reference picture management, reference picture list construction, and cross-layer dependency signaling. All the proposed technologies are used for the development of video coding system for UHD stereo display and glassless 3D display.

1. Introduction

Growing interests in realistic 3D video services for movies and broadcastings have accelerated the development of 3D display devices, e.g. UHD stereo display, glassless 3D display and holographic real 3D display, and the related video processing technologies, e.g. multiview video data acquisition and pre-/post- processing. Even if those recent technologies enable 3D video services, the quality of service to end-users is mainly determined by how to deliver more qualified video data via a limited channel bandwidth.

The ITU-T Visual Coding Experts Group (VCEG) and the ISO/IEC Moving Pictures Experts Group (MPEG) have developed an improved video coding standard with the name High-Efficiency Video Coding (HEVC), which provides about 50% bit rate savings in comparison to the H.264/AVC at the same fidelity [1]. The remaining issue is how to extend the HEVC to deliver 3D video contents to various devices having different resolution capacities.

In the paper, we propose a novel multi-dimensional video coding system based on HEVC. It enables scalable transmission of multiview video coding with depth map to cope with various display types, e.g. UHD stereo display and auto-stereoscopic display. Key technologies in this system are reference picture management system, reference picture list construction, and cross-layer dependency signaling. All the technologies were proposed and adopted in JCT-VC and JCT-3V, which are joint groups of VCEG and MPEG for developing HEVC extensions.

2. Proposed video coding system

The proposed multi-dimensional video coding system is based on HEVC coder with modifications to support depth map coding and view/spatial scalability. Fig. 1 demonstrates an operation of the proposed system, where multiple YUV views and corresponding depth maps are fed into an HEVC-based multi-dimensional video encoder, and multi-layered coded video streams are output. To cope with auto-stereoscopic display as well as stereo display with varying display sizes, and different viewing preferences, supplementary data, i.e, depth map, are coded together with multi-resolution YUV texture views. A bitstream extractor can re-generate a proper bitstream from the original bitstream by extracting target views and depth maps.

On top of the functionality of legacy multiview video coders, e.g. MVC, the proposed coder provides a spatial and/or a quality scalable transmission of 3D contents. Hence, multiple views with depth maps are coded in multiple layers with various spatial resolution and/or quality levels. Fig. 2 illustrates an access unit containing multiple layers with multi-dimensional scalabilities, where sub-bitstreams of a single HD view, stereo HD views, or stereo UD views with depth maps can be extracted by a bitstream extractor.

The basic coding mechanism of the proposed coder is not quite different from that of the HEVC-based coder, but mainly the composition of access units, reference picture management in decoded picture buffer and reference picture list construction are improved with changes of

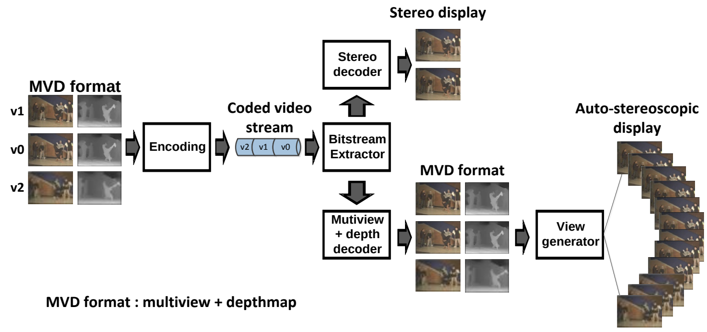

Fig. 1. General structure of multiview-plus-depth coding

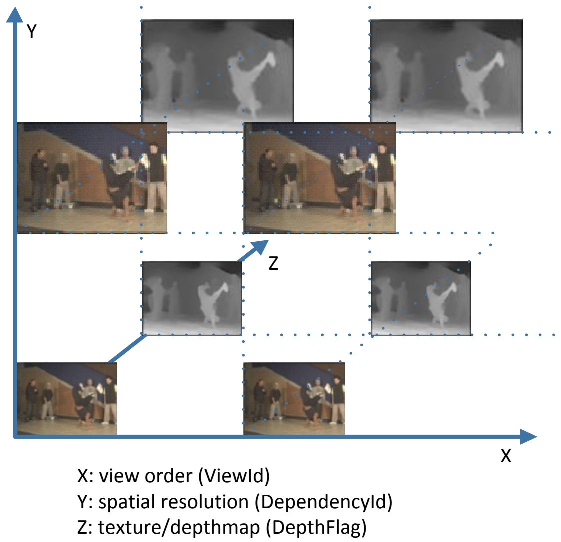

Fig. 2. Multi-dimensional layered structure in an access unit

high−level syntax to allow inter−layer prediction between views.

As illustrated in Fig. 2, all video pictures and depth maps that represent a video scene at the same time instance build an access unit. Inside an access unit, a base view picture is coded without any dependency from other views and the associated depth map is coded as an enhancement layer. Thereafter, video pictures and depth maps of other views are coded as view scalable layers. To enable spatial or quality scalability on top of view scalability, higher resolution and/or quality pictures of the same views are coded as enhancement layers.

Basically, the inter−layer prediction is allowed between pictures inside an access unit. The dependency information across layers is signaled in a video parameter set that leads all coded video sequences of all layers and contains useful information for session negotiation and bitstream extraction [2].

Within a coded video sequence, the dependency between layers can be changed by a layer dependency change SEI message, inter−layer prediction indication flag in slice header or RPS signaling for inter−layer prediction [3].

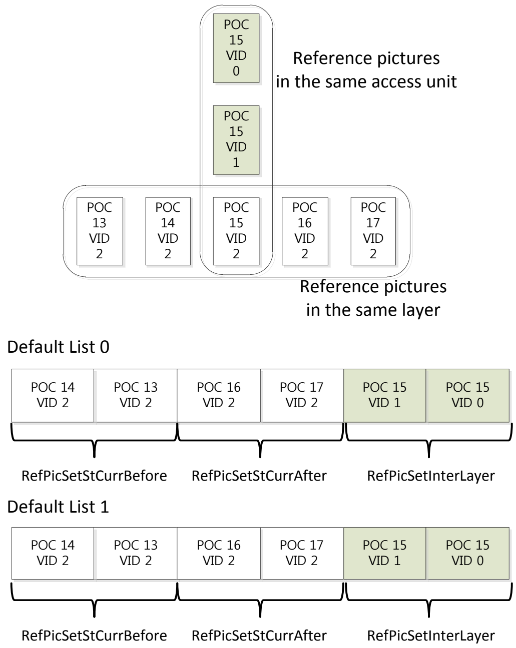

Fig. 3. Reference picture list initialization

RefIdx	list_entry_l0
0	0
1	4
2	5
3	1
4	2
5	3

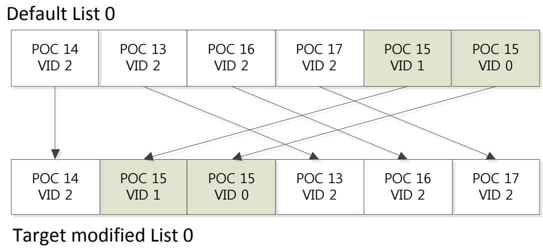

Fig. 4. Reference picture list modification

3. Reference picture management

In a single layer coding, decoded pictures are stored in decoded picture buffer (DPB) to be used for inter

prediction, and reference pictures that are used for inter prediction are organized in one or more reference picture lists according to the picture type.

In the proposed multi−dimensional video coding system, decoded pictures are stored for both inter prediction and inter−layer prediction. If a decoded picture is not used for inter prediction or inter−layer prediction, the picture is removed from a DPB immediately after its decoding. It enables to minimize the DPB size [4].

Table 1. Coding efficiency of the proposed coder (vs. simulcast)

sequence	2 views (BD-rate)			3 views (BD-rate)		
	Y	U	V	Y	U	V
Balloons	28.1%	31.5%	30.7%	37.4%	39.0%	40.1%
Kendo	26.1%	28.7%	28.4%	34.3%	34.7%	36.5%
Newspaper	27.3%	24.9%	26.0%	36.8%	36.1%	35.9%
GhostTownFly	40.6%	43.4%	43.8%	54.3%	57.2%	57.7%
PoznanHall2	27.1%	28.2%	27.3%	36.2%	37.1%	32.1%
PoznanStreet	32.8%	38.9%	27.6%	44.3%	51.8%	47.4%
UndoDancer	37.4%	41.6%	41.6%	50.9%	54.7%	54.6%
Average	**31.4%**	**33.9%**	**32.2%**	**42.0%**	**44.4%**	**43.4%**

By default, the inter-layer references are put after inter prediction references, as shown in Fig. 3. To improve the coding efficiency, the pictures in List 0 and List 1 can be reordered. Fig. 4 shows the reordering mechanism by explicit signaling [5].

4. Simulation results and conclusion

Table 1 shows the coding performance in comparison to HEVC simulcast. The proposed system outperforms up to 33% BD-rate for 2-view and 44% BD-rate for 3-view.

In this work, a multi-dimensional video coding system based on HEVC was proposed. It supports multiview-plus-depth data coding, enabling spatial/view/quality scalable transmission. Key technologies in this system are reference picture management, reference picture list construction, and cross-layer dependency signaling.

All the technologies for the development of multi-functional video coding system were proposed and adopted in the international video coding standardization bodies JCT-VC and JCT-3V, which work for ISO/IEC MPEG-H video (HEVC) and ITU-T Rec. H.265. It is expected that the proposed technologies are utilized in various 3D video services in near future.

5. References

[1] B. Bross, et al., " High efficiency video coding (HEVC) text specification draft 10 (for FDIS & Consent)," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-L1003, 12th Meeting:Geneva, CH, Jan. 2013.

[2] B. Choi, et al., " On video parameter set extension," ISO/IEC JTC1/SC29 WG11, Apr. 2013.

[3] B. Choi, et al., " View dependency change SEI message for MV-HEVC," ISO/IEC JTC1/SC29 WG11, Oct. 2012.

[4] B. Choi, et al., " Unused reference picture management for MV-HEVC and other HEVC extension," ISO/IEC JTC1/SC29 WG11, Apr. 2013.

[5] B. Choi, et al., " 3D-HEVC HLS: On Reference list," ISO/IEC JTC1/SC29 WG11, Apr. 2012.