LETTER

A Synchronization and T-STD Model for 3D Video Distribution and Consumption over Hybrid Network

Kugjin YUN[†], ***Student Member***, **Won-sik CHEONG**[††], ***Nonmember***, ***and*** **Kyuheon KIM**[†a)], ***Member***

SUMMARY Recently, standard organizations of ATSC, DVB and TTA have been working to design various immersive media broadcasting services such as the hybrid network-based 3D video, UHD video and multiple views. This letter focuses on providing a new synchronization and transport system target decoder (T-STD) model of 3D video distribution based on heterogeneous transmission protocol in a hybrid network environment, where a broadcasting network and broadband (IP) network are combined. On the basis of the experimental results, the proposed technology has been proved to be successfully used as a core element for synchronization and T-STD model in a hybrid network-based 3D broadcasting. It has been also found out that it could be used as a base technique for various IP associated hybrid broadcasting services.

key words: *hybrid 3DTV, heterogeneous transmission, T-STD model, DASH, MPEG-2 TS*

1. Introduction

The effort preoccupying the core technology around the 3D and UHD is proceeding in recent years. Especially, many countries are accelerating their efforts to provide households with these immersive media by using the hybrid networks combined with IP networks around standard organizations such as ATSC, DVB, and TTA [1]. In this respect, the ATSC has been standardizing the 3DTV broadcasting service on the hybrid networks in order to provide high-quality 3D video without deterioration after completion of the service-compatible 3DTV broadcasting service. With the development of 3DTV broadcasting service, one of the main requests of ATSC 3.0 is focused on support for hybrid broadcasting service such as multiple view service. TTA also has recently carried on discussion about immersive media services including hybrid network-based 3D video, 3D UHD video and multi-view video, etc [2], [3].

However, due to technical characteristic differences of the delivery systems and the networks, problems reside in quality, synchronization and T-STD model of hybrid network-based 3D broadcasting.

This letter proposes a hybrid 3D broadcasting system using MPEG-2 TS and MPEG-DASH that provides full HD quality 3D presentation. In specific, the proposed system delivers each view of the 3D video over two different physical networks to maintain full HD quality, and provides core technologies of synchronization and T-STD model.

The remainder of this letter is structured as follows. Section 2 gives an overview of hybrid network-based 3D broadcasting service and proposed core technologies such as synchronization and T-STD model. Section 3 shows the developed hybrid network-based 3D broadcasting system and experimental results, which are followed by some concluding remarks in Sect. 4.

2. Hybrid Network-Based 3D Broadcasting Service

In order to keep backward compatibility with the existing DTV broadcasting, the base video is encoded by MPEG-2 or H.264/AVC, and it is transmitted through the existing broadcasting infrastructure. On the other hand, the additional video is encoded and multiplexed using multi-bitrate supported H.264/AVC and MPEG-2 TS, respectively. And then it is transmitted through the ISO/IEC 23009-1 Dynamic Adaptive Streaming over HTTP (DASH) [4] to provide stable media stream according to the IP status. The DASH facilitates an approach to segment files with various qualities through Media Presentation Description (MPD). Figure 1 shows a conceptual diagram of hybrid network-based 3D broadcasting service.

2.1 Synchronization of Hybrid Network-Based 3D Video

The proposed 3D broadcasting service is realized by an arbitrary pairing of both real-time broadcasting streams and IP-based streams, which do not have the same Presentation Time Stamp (PTS). It is therefore necessary to provide a robust synchronization method. In hybrid network-based 3D broadcasting service, a SMPTE timecode scheme [5] has introduced to synchronize the base video and corresponding additional video. This scheme provides backward compatibility with MPEG-2 and H.264/AVC standards by inserting the identical timecode value in MPEG-2 Group of Picture

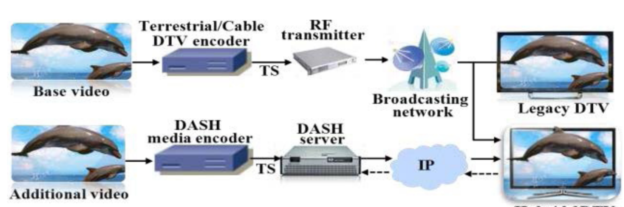

Fig. 1 A conceptual diagram of hybrid network-based 3D broadcasting service.

Manuscript received December 5, 2014.
Manuscript publicized July 13, 2015.
[†]The authors are with Kyung Hee University, Yongin, 446701, Korea.
[††]The author is with ETRI, Daejon, Korea.
a) E-mail: kyuheonkim@khu.ac.kr
DOI: 10.1587/transinf.2014EDL8242

(GoP) header and AVC picture timing Supplemental Enhancement Information (SEI). However, it is required to add an additional memory for the synchronization as recognizing timecode values after those two video decoding process. That is, it is not able to guarantee the backward compatibility with the existing T-STD model [6].

This letter proposes the metadata-based synchronization that fully resolves the previously mentioned problems.

Table 1 represents the proposed metadata structure that provides the linkage between a base video stream and additional video one, and stable synchronization information in a hybrid network environment.

In this metadata, **referenced_media_filename_length** indicates the length of MPD which includes additional video information associated with a base video stream. In addition, **referenced_media_filename _byte** is a MPD file name, which includes additional video information. The term **frame_number** provides the frame number of each Access Unit (AU) of a base video stream and additional video one. The frame number starts with '0' at the beginning of video stream and increase monotonically.

Figure 2 shows the transmission structure for a hybrid network-based 3D broadcasting service, and synchronization process using the metadata proposed in Table 1. The metadata is transmitted as a Packetized Elementary Stream (PES) like an AV data stream with the stream type 0x06 (PES packets containing private data) [6]. It is also required to assign 45kbits per second for the metadata in one Transport Stream (TS) packet of 188bytes. This metadata has not increased TS data size such as that it is about 0.0023% data size to general TS transmitted through a broadcasting network. It therefore has the merit of backward compatibility with the existing T-STD model by applying the metadata specified in ISO/IEC 13818-1 and minimum system's complexity without any additional memory after a decoding process in a 3D receiver.

As shown in Fig. 2, PTS(i) and PTS(k) represent the PTS value of i-th and k-th AUs in the base video stream and additional video one, respectively. The same PTS value is assigned with a frame number in the proposed metadata stream. Thus, both the based and additional video streams can be synchronized through a same frame number in the metadata stream even though the base and additional video TS streams have different PTS values due to different clock references. In the 3D receiver, the PTS values to AUs in additional video TS stream with the same frame number is compensated into the PTS values to AUs in the base video TS stream by comparing the frame number assigned to AUs in the metadata information.

Table 1 Metadata structure.

Syntax	Num. of bits
media_pairing_information() {	
referenced_media_filename_length	8
for(i=0;i<referenced_media_filename_length;i++) {	
referenced_media_filename_byte	8
}	
reserved	7
frame_number	25
}	

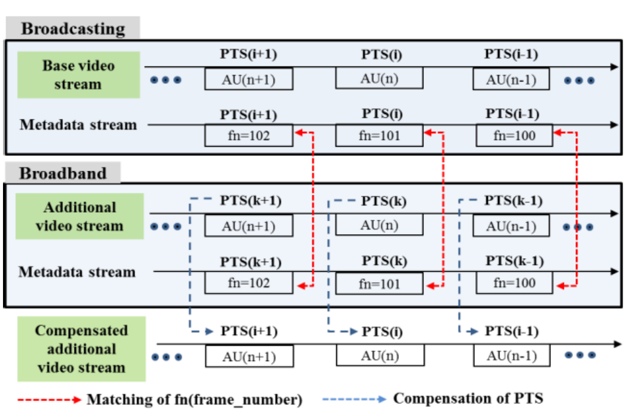

Fig. 2 Transmission structure of both video and metadata-based synchronization process.

2.2 T-STD Model for Hybrid Network-Based 3D Broadcasting Service

In general DTV broadcasting, a virtual reference model, which is called a T-STD, is specified in order to regulate the buffering and timing of decoders for stable media playback. However, the hybrid network-based 3D broadcasting service is to use not only the broadcasting network but also IP network, and thus, it is required to set up a new T-STD model with consideration of variable bandwidth, jitter and delay of IP network.

Figure 3 shows the proposed T-STD model for a hybrid network-based 3D broadcasting service. The proposed T-STD model is extended on the basis of ISO/IEC 13818-1 T-STD specification [6] with newly defined two buffers.

Firstly, as being explained above, the hybrid network-based 3D broadcasting service is designed to use the DASH to transmit the additional video stream. The outputted TS from H.264/AVC multi-encoder is put into the DASH server where the TS is converted into temporally segmented TS files satisfying the DASH specification [4]. Since the DASH server cannot start a transmission before the completion of

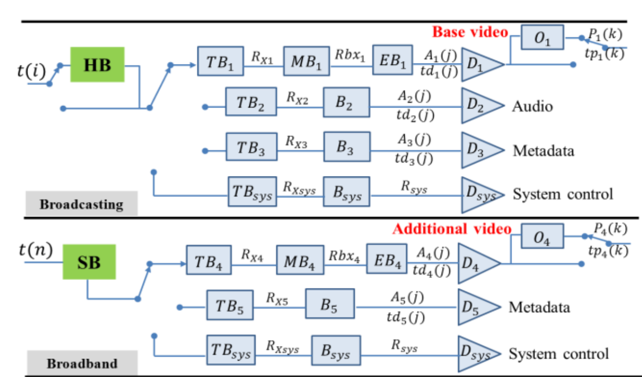

Fig. 3 T-STD model for hybrid network-based 3D broadcasting service.

a segmented TS file conversion, an initial transmission delay is inevitable. In order to overcome such an initial delay and guarantee stable 3D video playback, it is required to set up a Hybrid Buffer (HB) for buffering of base video stream transmitted through a broadcasting network. HB is located in front of the Transport Buffer (TB) specified in ISO/IEC 13818-1 T-STD. The buffering time of HB, H_b, can be defined as follows:

$$H_b = S_d + \alpha + minbufferTime \tag{1}$$

where S_d is the duration of a segmented TS file and α $(0 < \alpha < 1)$ is the DASH server processing delay. The *minbufferTime* is defined in MPD of the DASH specification, and calculated as follows:

$$minbufferTime = S_d + \beta \tag{2}$$

where β $(0 < \beta < 1)$ is MPD processing delay of the 3D receiver. Since a segmented TS file shall be completely received before it is passed onto the decoder, S_d is required in (2).

Secondly, it is required to set up a Streaming Buffer (SB) to buffer the additional video stream due to delay and jitter over IP network. The size of SB is determined according to the *minbufferTime* for an initial buffering and *bandwidth* described in the MPD. The size of SB is calculated as follows:

$$\mathrm{SB} = minbufferTime \times bandwidth \tag{3}$$

where *bandwidth* represents the encoding rate of additional video.

As shown in Fig. 4, the proposed T-STD model has a merit that maintains the T-STD model applied to a legacy DTV after buffering a TS of base video transmitted through a broadcasting network.

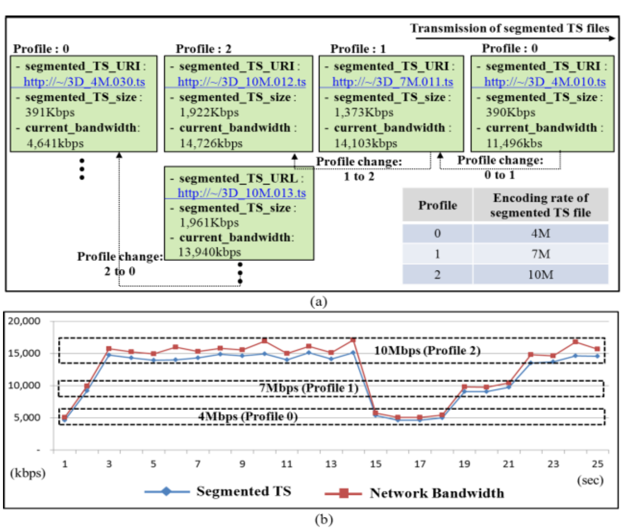

Fig. 4 Experimental results: (a) switching results of segmented TS files for additional video and (b) seamless switching according to IP network status.

3. Experimental Results

The test conditions used to verify the proposed technologies are described in Table 2. The 3D video sequences consist of various types of contents including a documentary, music, and sports. The base video is encoded with MPEG-2 video at 17Mbps, which was applied to the existing terrestrial DTV broadcasting [7]. The additional video is encoded with H.264/AVC at 4Mbps, 7Mbps and 10Mbps to provide an appropriate picture quality in accordance with the IP network status.

The DASH server supports not only the HTTP/1.1 [8] but also uploading of segmented TS files and MPD file for an additional video. In this experiment, we applied the DASH TS main profile and *minbufferTime* = 1.4 sec for an initial buffering in the 3D receiver.

Figure 4 (a) shows the reactive switching result of segmented TS files, which is transmitted according to the IP status. The **segmented_TS_URI** indicates a path of segmented TS file. The **segmented_TS_size** and **current_bandwidth** show the size of each segmented TS file and the current network bandwidth, respectively. The profile represents an encoding rate of each segmented TS file.

As shown in Fig. 4 (a), the DASH server transmits a GoP-based segmented TS file encoded with 4Mbps after a transmission of an initial segment. And then, it transmits a segmented TS file encoded with 7Mbps and 10Mbps as a current network status is getting better. On the other hand, when the IP status reduces to 4Mbps, the DASH server transmits the segmented TS file encoded with 4Mbps again. The experiment result shown in Fig. 4 (b) indicates that segmented TS files with each encoding rate are switched and transmitted seamlessly in proposed the T-STD model according to the IP status.

Table 3 shows the result of synchronization process by the proposed metadata in Sect. 2.1, which evaluates the performance by extracting actual values within encoded and decoded stream. A degree of synchronization on the timeline in this paper is described by the PTS Gap, which means PTS difference between both video streams. As shown in this table, it is confirmed that the proposed metadata successfully performs its functionality by doing the timestamp compensation process.

Figure 5 shows a diagram of newly developed hybrid

Table 2 Test conditions.

Test sequence	Encoding method	Encoding rates	Quality (dB)	Resolution
Base video	MPEG-2	17M	38.6	1,920x 1,080
Additional video	H.264/AVC	4M	35.7	
		7M	38.2	
		10M	39.7	

Table 3 Result of synchronization by the proposed metadata.

3D video	Encoded stream	Decoded stream
PTS Gap	317 ms	0 ms

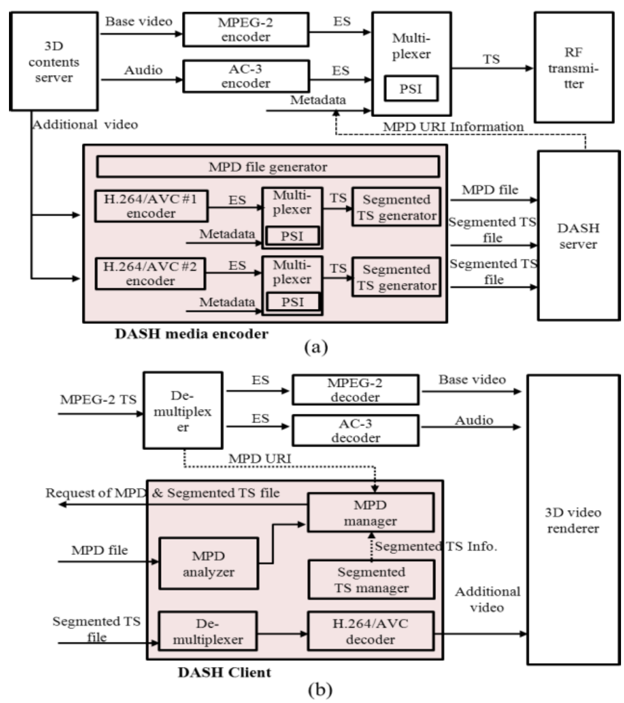

Fig. 5 Block diagram of hybrid network-based 3D broadcasting system: (a) 3D transmitter and (b) 3D receiver.

network-based 3D broadcasting system. The 3D transmitter is composed of an MPEG-2 video and AC-3 encoder for base video and audio encoding, a DASH media encoder for generating segmented TS files and MPD file of additional video stream using a H.264/AVC multi-encoder, a multiplexer for MPEG-2 TS generation, and a DASH server for registration of segmented TS files and MPD file, and transmission of them according to the IP status. As mentioned in Sect. 2.1, the proposed metadata contains URI information of the MPD generated from the DASH media encoder.

The 3D receiver is broadly composed of a demultiplxer to parse MPEG-2 TS, an MPEG-2 and AC-3 decoder, a DASH client, and a 3D video renderer as being shown in Fig. 5 (b). In the DASH client, The MPD analyzer checks the validity of transmitted MPD file using a MPD schema specified in ISO/IEC 23009-1, and obtains attributes of each segmented TS file. The segmented TS manager plays a role in transmitting segmented TS file information according to network conditions to the MPD manager. The MPD manager requests a corresponding segmented TS file on current network status, and manages a MPD file constantly.

Figure 6 shows the snapshot of a stable 3D video rendering by the developed hybrid network-based 3D broadcasting system to verify the proposed technologies, which has a backward compatibility to a legacy broadcasting system. Based on the experimental results,

Fig. 6 Snapshot of the developed system and a stable 3D video rendering in hybrid networks.

it is confirmed that the proposed technologies provide a simple and efficient 3D video synchronization and stable transmission in hybrid networks while maintaining the existing T-STD model.

4. Conclusion

In this letter, we propose a novel synchronization method and T-STD model that can be efficiently applied in hybrid network-based 3D broadcasting service. It provides not only a stable delivery scheme of 3D video but also extensibility of being easily combined with a legacy T-STD model. Furthermore, it is confirmed that these proposed methods can be applied as a core technology for various hybrid network-based media synchronization and transmission.

Acknowledgments

This work was partly supported by the ICT R&D program of MSIP/IITP [B0126-15-1013, Development of generation and consumption of Jigsaw-liked Ultra-Wide Viewing Spacial Media] and the ICT Standardization program of MSIP.

References

[1] S. Dogan, et al., "Real-time immersive multimedia experience for collaborating users over hybrid broadcast networks," IEEE Conf. Future Network Summit, pp.1–10, 2013,

[2] T. Schierl and S. Narasimhan, "Transport and storage systems for 3-D video using MPEG-2 systems, RTP, and ISO file format," Proc. IEEE, vol.99, no.4, pp.671–683, April 2011.

[3] http://www.atsc.org/cms/index.php/subcommittees/42-subcommittees/283-technology-and-standards-group-tg3

[4] ISO/IEC 23009-1:2012, Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats.

[5] B. Kim, S. Hahm, J. Kang, K. Lee, J. Lee, H. Park, Y. Jang, and H.J. Kim, "World's first hybrid 3DTV broadcasting experiment," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp.1–5, June 2012.

[6] ISO/IEC 13818-1:2013, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: System, June 2013.

[7] A/53, ATSC Digital Television Standard, Part 1 – Digital Television System, Aug. 7, 2013.

[8] RFC 2616: Hypertext Transfer Protocol – HTTP/1.1, June 1999.