IEICE TRANS. INF. & SYST., VOL.E96–D, NO.9 SEPTEMBER 2013

LETTER

A Delivery Format for Unified Stereoscopic Video Content Transmissions over Dynamic Adaptive Streaming Scheme

Jangwon LEE[†], ***Student Member***, **Kugjin YUN**[†], **Doug Young SUH**[†], ***Nonmembers***, ***and*** **Kyuheon KIM**[†a)], ***Member***

SUMMARY This letter proposes a new delivery format in order to realize unified transmissions of stereoscopic video contents over a dynamic adaptive streaming scheme. With the proposed delivery format, various forms of stereoscopic video contents regardless of their encoding and composition types can be delivered over the current dynamic adaptive streaming scheme. In addition, the proposed delivery format supports dynamic and efficient switching between 2D and 3D sequences in an interoperable manner for both 2D and 3D digital devices, regardless of their capabilities. This letter describes the designed delivery format and shows dynamic interoperable applications for 2D and 3D mixed contents with the implemented system in order to verify its features and efficiency.

key words: *delivery format, unified transmissions, stereoscopic video content, dynamic adaptive streaming service*

1. Introduction

In today's technical arena, 3D content services are considered to be one of the most promising applications in the home and mobile video entertainment fields. There are already various types of digital devices on the market that support 3D content, such as laptops, mobile phones, tablet PCs, and digital TVs These devices enable users to enjoy a 3D visual experience with ease; however, stereoscopic video content services still have difficulties with respect to their management. This is because the various digital devices have different performance capabilities and stereoscopic video content is very sensitive to variable network conditions.

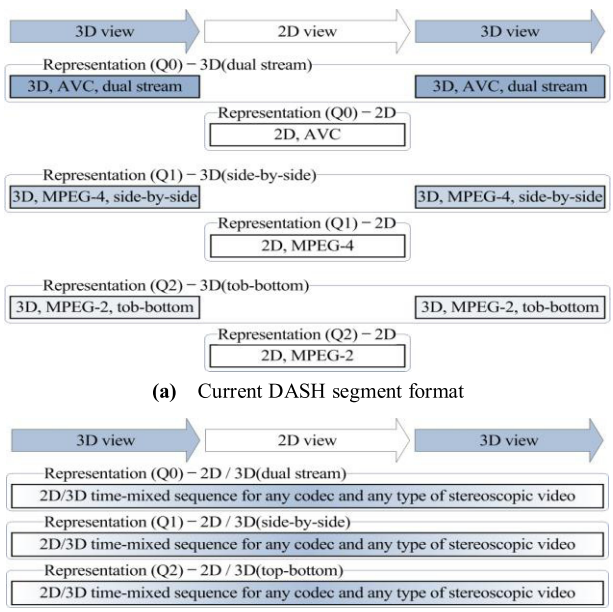
In order to overcome these difficulties, a HTTP adaptive streaming scheme is proposed as a delivery method for stereoscopic video content, which provides a seamless multimedia service over a network with a throughput bandwidth variable [1] Dynamic Adaptive Streaming over HTTP (DASH) developed by Moving Picture Experts Group (MPEG) is a representative standard technology for these schemes. In DASH, a media source is prepared with small pieces of variable quality, and the delivery format of these pieces, which are called segments, is defined. These segments are selected according to devices and network conditions and then delivered to a client repeatedly [2]. Thus, multimedia content services can be provided, which are not only suited to the client's own device performance but also dynamically adapted to the network conditions at any given moment.

Manuscript received December 13, 2012.
Manuscript revised March 13, 2013.

[†]The authors are with Kyung Hee University, Yongin, 446701, Korea.
a) E-mail: kyuheonkim@khu.ac.kr

DOI: 10.1587/transinf.E96.D.2162

This letter focuses on providing a stereoscopic video content service based on the DASH framework. For this service, the additional specific requirements for stereoscopic video content should be considered in addition to the current DASH service. Firstly, the service should be able to support various types of stereoscopic video content which is used in current technical areas such as side-by-side, top-bottom, vertical/horizontal line interleaved and frame sequential, and dual stream types [3]–[5] The current DASH services have limitations in terms of stereoscopic video composition and usage of codecs, since it supports stereoscopic video signaling method in codec dependent ways For example, MPEG-2 codec only provides the signaling method for side-by-side and top-bottom types, thus the content encoded with MPEG-2 can be delivered in these two types only. In addition, as shown in Fig. 1 (a), every representation for 3D view must be prepared by individual methods according the used codec. Secondly, 2D and 3D sequences should be dynamically switchable in a time-mixed service [6]. As shown in Fig. 1 (a), the current DASH technology does not offer a delivery format for 2D/3D time-mixed sequence, and thus 2D and 3D sequences are handled as individual representations



(a) Current DASH segment format

(b) Proposed delivery format

Fig. 1 Configuration of representations for 2D/3D time-mixed service using various types of stereoscopic video content.

Thus, overhead bytes and redundant operations are caused in media preparing, delivering, and decoding processes when switching 2D and 3D representations

Therefore, as shown in Fig. 1 (b), this letter proposes a delivery format for unified transmissions of various types of stereoscopic video composition and of codecs which also support dynamic switching in 2D/3D time-mixed service in one representation. This letter is organized as follows: firstly, Sect. 2 describes the proposed delivery format. In Sect. 3, the implemented system based on the designed delivery format is shown, together with evaluation results. Finally, Sect. 4 provides an analysis of the proposed technology with a view to future work on this topic.

2. Delivery Format Design for Stereoscopic Video Content

The proposed delivery format is designed based on Stereoscopic Video Application Format (SVAF, ISO/IEC 23000-11) [7], which is a standardized MPEG-A technology. SVAF involves formatting various stereoscopic video content into a unified storage format such as ISO base media file format (ISO BMFF, ISO/IEC 14496-12) [8]. This design concept of SVAF enables the provision of stereoscopic video content service. SVAF is also compatible with the current DASH service, since DASH adopts the base structure of ISO BMFF as a delivery format.

As shown in Fig. 2, the delivery format supports two types of stereoscopic video contents: frame packing and dual stream. In the frame packing type such as side-by-side, vertical line interleaved, frame sequential, and topbottom, left and right view sequences are composed in one frame packing sequence, and then the sequence is encoded into one elementary stream. In the dual stream type, left and right view sequences are independently encoded into two elementary streams as shown in Fig. 2 (b). The delivery formats of these types are composed of both an initialization and a media segment. The initialization segment is delivered at the beginning of a presentation and carries the initialization parameters of the presentation. Thereafter, several media segments are followed, and each of them carries partial elementary streams with their respective timing information.

The initialization segment includes the stereoscopic video media information box (***svmi***), which describes the unified initialization parameters of various types of stereoscopic video contents. As shown in Table 1, the ***svmi*** box includes the ***stereoscopic_composition_type***, which indicates the individual types of stereoscopic video contents with an integer value. For example, the side-by-side, vertical line interleaved, frame sequential, topbottom, and dual stream types are assigned as 0, 1, 2, 3, and 4 values, respectively. The ***is_left_first*** indicates whether the left view image appears first or not. As shown In Table 1, this box does not contain any information associated with specific codec, thus it is possible to provide unified signaling method for various types of stereoscopic video content regardless of their codec types.

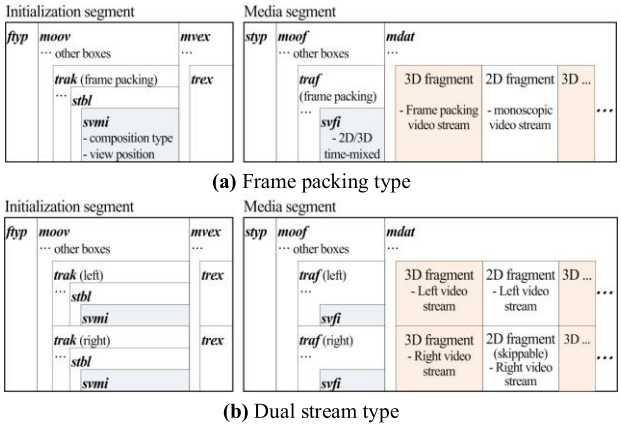


Fig. 2 Delivery format for stereoscopic video content.

Table 1 Boxes for stereoscopic video description.

Syntax	No. of Bits
aligned(8) class StereoscopicVideoMediaInformationBox extends FullBox('svmi', version = 0, 0){	
stereoscopic_composition_type	8
reserved = 0	7
is_left_first	1
}	
aligned(8) class StereoscopicVideoFragmentInformationBox extends FullBox('svfi', version = 0, 0){	
stereo_mono_change_count	32
for(i=0; i<=stereo_mono_change_count; i++){	
sample_count	32
reserved=0	7
stereo_flag	1
}	
}	

To support the 2D/3D time-mixed services explained in Sect. 1, the media segment should be able to contain both 2D and 3D fragments as shown in Fig. 2. The media segment includes the stereoscopic video fragment (***svfi***) box, which is a newly designed box for describing timing information regarding the 2D or 3D fragments contained in the media segment. As shown in Table 1, the ***svfi*** box includes the ***stereo_mono_change_count***, which is the number of 2D and 3D fragments contained in the media segment. Each fragment is signaled in the following loop, where the ***sample_count*** indicates the number of frames in the fragment, and the ***stero_flag*** shows whether the fragment is 2D or 3D.

As shown in Fig. 2, the ***svmi*** and ***svfi*** boxes are located in the ***trak*** box in the initialization segment and the ***traf*** box in the media segment for each elementary stream of stereoscopic video content, respectively. The elementary streams themselves are contained in the ***mdat*** box in the media segment. In the dual stream type, one of the left and right view sequences can be displayed as a 2D sequence. Thus, as shown in Fig. 1, only one elementary stream can be contained in the 2D fragment. In addition, it is possible to provide interoperable services for both 2D and 3D devices by using the dual stream type since 2D devices can display only

one elementary stream regardless of the other stream.

3. Experimental Results

On the basis of the proposed delivery format in Sect. 2, this section shows the implementation system for dynamic adaptive streaming of various stereoscopic video contents. As shown in Fig. 3, this system is broadly composed of the content generator, server and client. The content generator creates initialization and media segment sets with multiple qualities. It also generates the Media Presentation Description (MPD) [2], which describes the directory of the segments. Both the generated segments and MPD are stored in the server. When the client requests a service, the MPD and its segments are delivered to the client via HTTP messages. While the client communicates with the server, the quality selector in the client consistently checks the bandwidth and selects the media segment with the proper quality. The received segments are parsed, and then the elementary streams of the stereoscopic content are decoded into a 3D or 2D/3D time-mixed sequence. The sequences are displayed by using both the initialization parameters and timing information that are provided by the ***svmi*** and ***svfi*** boxes explained in Sect. 2.

Figure 4 shows the configuration and the presentation of the test segment sets. For a comparative experiment, two segment sets for the same service were prepared as shown in Fig. 4. Set A is constructed according to the delivery format proposed in Sect. 2, while Set B is applied by the existing DASH technology [2] without the ***svmi*** and ***svfi*** boxes proposed in this letter. Both Set A and Set B commonly contain three stereoscopic video streams with different qualities, which comprise two frame packing streams with 640×480 and 1920×1080 resolutions, and one dual stream content with $1920 \times 1080 \times 2$. All of the streams are encoded with MPEG-4 AVC and have a 18 seconds duration, and each media segment has a two-second duration.

To verify the 2D/3D time-mixed service, two 2D fragments of 5–8 seconds and 11–14 seconds are inserted into both Set A and Set B. However, Set A and Set B have different configuration methods in terms of switching 2D and 3D fragments. In Set A, the initialization parameters of the stereoscopic video content are stored in the ***svmi*** box in each initialization segment. In addition, two 2D and three 3D fragments are contained in one sequence for each quality. The ***svfi*** box for describing the 2D/3D time-mixed information is stored in all of the media segments. On the other hand, Set B comprises five sequences for each quality, where each sequence contains one fragment of 2D or 3D, since DASH does not support methods for constructing 2D and 3D video sequences together within one sequence. Consequently, five initialization segments are needed for one quality.

As shown in Fig. 4, the two test segment sets show the same presentation. Throughout the 18 seconds of the presentation, three different bandwidth limitations are imposed on the client every six seconds, such as 200 Kbps,

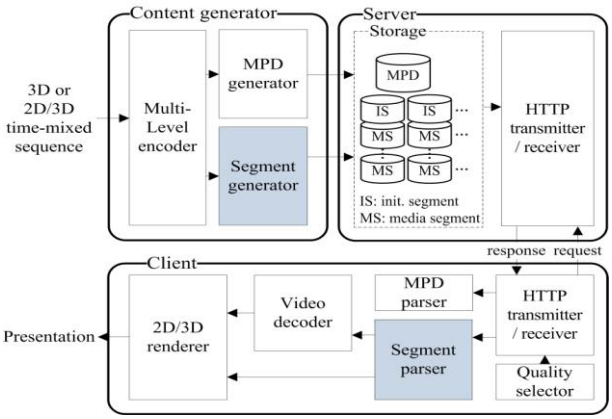


Fig. 3 Functional diagram of the implemented system.

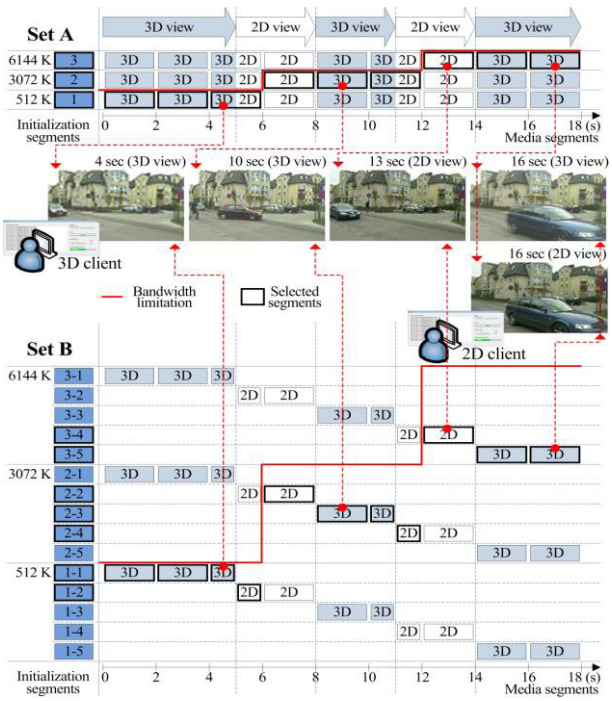


Fig. 4 Configuration and presentation of test segment sets.

4000 Kbps, and 9000 Kbps, respectively. The five pictures in Fig. 4 are captured at 4, 10, and 13 seconds in the 3D client, and at 16 seconds in both the 3D and 2D clients, respectively. In the presentation, the segments with the best qualities under the bandwidth limitations are selected in the 3D client. For example, the sequences with 512 Kbps, 3072 Kbps, and 6144 Kbps are sequentially selected during the first, second, and final six seconds This result shows the successful bandwidth dynamic streaming of the stereoscopic video content by the proposed delivery format. In addition, a 2D/3D interoperable service is also shown as the presentation of the 2D and 3D clients at 16 seconds.

Figure 5 shows the comparison results of the test segment sets in each process during the presentation. Firstly, Set A can be composed of a smaller number of segments in

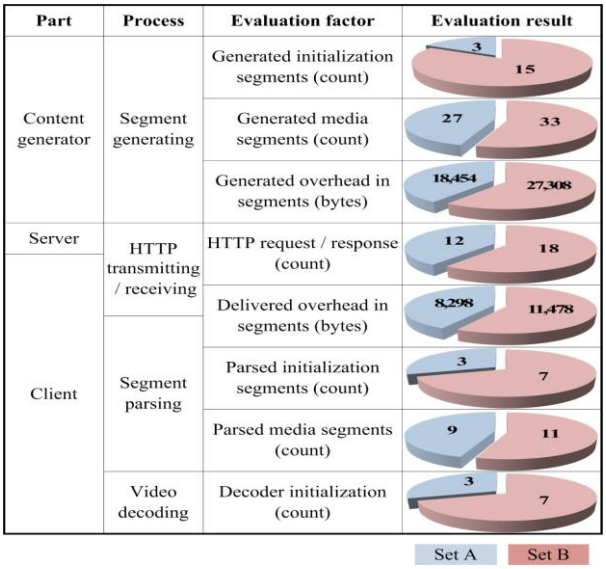


Fig. 5 Comparison results of test segment set in each process.

the content generating process. As shown in Fig. 4, Set A contains only three initialization segments while Set B has 15 segments. In addition, in Set B, there are three more media segments for each quality than in Set A. Set B also includes approximately 9000 more overhead bytes to be generated than Set A, where the overhead is measured as the total size of segments excluding the elementary streams. In the transmission and receiving processes, the number of HTTP messages available to carry the segments is six fewer in Set A than Set B. Set A also has approximately 3000 smaller overhead bytes to be delivered and six fewer segments to be parsed than Set B. Finally, in the video decoding process, Set A can be handled by four fewer decoder initialization operations than Set B. These comparison results of Set A and Set B in each process show that the proposed delivery format can support dynamic streaming of 2D and 3D video content with less overhead bytes and redundant operation than the current technology.

4. Conclusion

With the rapid development of 3D digital technologies in recent years, various types of 3D digital devices are appearing on the market. DASH is considered to be a suitable technology to provide interoperable services that cover various device capabilities and variable network bandwidths. However, DASH did not support interoperable services for various forms of stereoscopic video content, since individual delivery formats are used according to codec and composition types of stereoscopic video content. This letter proposed a new delivery format for unified transmissions of any stereoscopic video content types regardless of their composition types and codecs In addition, the proposed delivery format offers an effective 2D/3D time-mixed streaming service over DASH framework while maintaining compatibility with the legacy services. The interoperability of the proposed delivery format with 2D and 3D devices was proven with the implemented system. Also, in dynamic switching between the 2D and 3D sequences, the proposed delivery format also showed greater efficiency than the current technology. The suggested technology has been accepted as a working item of MPEG-A standards, and then it is in standardization progress in MPEG [9]. Therefore, it is expected to enable integrated dynamic services to 3D as well as 2D digital devices in the market through the use of the proposed technology.

Acknowledgments

This research was supported by the MKE (the Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2012-H0301-12-1006) and supervised by the NIPA (National IT Industry Promotion Agency).

References

[1] R. Rejaie, Yu Haobo, M. Handley, and D. Estrin, "Multimedia proxy caching mechanism for quality adaptive streaming applications in the Internet," Proc. IEEE Infocom 2000, vol.2, pp.980–989, March 2000.
[2] Text of ISO/IEC 2nd DIS 23009-1 Dynamic Adaptive Streaming over HTTP", N12166, Torino, Italy, July 2011.
[3] Text of ISO/IEC 13818-1:2007/AMD 7 Signaling of stereoscopic video in MPEG-2 systems," N12462, San Jose, USA, Feb. 2012.
[4] Stereoscopic 3D Full Resolution Contribution Link Based on MPEG-2 TS, SMPTE ST 2063-2012, 2012.
[5] Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV), DVB Document A154, Feb. 2011.
[6] K. Yun, K. Kim, N. Hur, S. Lee, and G. Park, "Efficient multiplexing scheme of stereoscopic video sequences for digital broadcasting services," ETRI J., vol.32, no.6, Dec. 2010, pp.961–964.
[7] Information technology – Multimedia application format (MPEG-A) – Part 11: Stereoscopic video application format, ISO/IEC 23000-11:2009 First ed., Nov. 2009.
[8] Information technology – Coding of audio-visual objects – Part 12: ISO base media file format, ISO/IEC 14496-12:2008 Third ed., Oct. 2008.
[9] ISO/IEC 23000-11 PDAM3 Support fragment structure for SVAF, N13049, Shanghai, China, Oct. 2012.