CARRIAGE OF 3D AUDIO-VISUAL SERVICES BY T-DMB

Sukhee Cho, Namho Hur, Jinwoong Kim, Kugjin Yun, and Soo-In Lee

Electronics and Telecommunications Research Institute
161 Gajeong-dong, Yuseong-gu, Daejeon, 305-350 Republic of Korea

ABSTRACT

In this paper, we introduce our experience on the development of a three-dimensional audio-visual(3D AV) service system based on the terrestrial digital multimedia broadcasting (T-DMB) system. 3D AV service is now much more feasible than before with the fast advancement of hardware technologies, especially 3D flat panel display, processors and memory. 3D AV service over DMB system is very attractive due to the facts that (1) glassless 3D viewing with small display is relatively easy to implement and more suitable to the single user environment like DMB, (2) DMB is a new media and thus has more flexibility in adding new services on the existing ones, (3) 3D AV handling capability of 3D DMB terminal has lots of potential to generate new types of services if it is added with other components like built-in stereo camera. In order to provide successful 3D DMB services over existing DMB system, we need to solve several issues like (1) guaranteeing *backward compatibility* with the T-DMB system, (2) minimizing the overhead on the transmitted bit-rate and the required processing power of the terminal, (3) providing good 3D depth perception without a noticeable eye strain. We propose a very efficient and backward compatible system architecture for the 3D DMB, and show how we can get better depth perception with the limited bit budget of the DMB system.

1. INTRODUCTION

Mobile reception of broadcasting services has recently got much attention worldwide. DMB, Digital Video Broadcasting-Handheld (DVB-H), MediaFLO are such examples. Among them, Korea commenced commercial T-DMB broadcasting for the first time to provide mobile multimedia services in 2005. Telecommunication Technology Association (TTA) of Korea and ETSI in Europe established a series of specification for T-DMB video and data services based on the EUREKA-147 Digital Audio Broadcasting(DAB) system [1, 6].

Increasing the reality is another direction of future multimedia services. There have been a lot of research activities on 3DTV and UD(Ultra-high Definition) TV concepts and systems, though most of them stayed at experimental level. Providing a viable 3DTV service to the level of current 2D television is not yet feasible due to several technological limitations. A wide flat panel 3D display for multiple viewers without using glasses is not mature yet, among other things. However, a small size auto-stereoscopic display has recently been mature to be commercialized for single user environment.

Providing mobility and increased reality is thus a promising direction for new multimedia services. Specifically, 3DAV service over T-DMB system is attractive in that (i) it is aimed for a single user; (ii) it adopts a small-sized auto-stereoscopic 3D display which can be made with a pretty mature technology at a very reasonable price; (iii) visual fatigue problem is reduced due to a small range of binocular disparity. We believe that a portable, personal 3DTV will be a valuable stepping stone towards realizing the ideal 3DTV for multi-users at home.

In this paper, we propose a cost-effective method of carrying 3DAV services by the T-DMB system. One of the most important requirements of the 3D DMB system is *backward compatibility* with the existing T-DMB system. The *backward compatibility* means that existing T-DMB receivers can identify a 3D DMB signal and recover the 2D DMB audio-visual signals without any problem. To satisfy such a requirement, we define a new object descriptor(OD) of MPEG-4 Systems. Specifically we regard a pair of stereoscopic video signals and multi-channel 3D audio signals as two objects. Then we consider that the two elementary streams(ESs) of (V_l, V_a) constitute the video object and similarly, the two ESs of (A_s, A_a) constitute the audio object. To verify the proposed scheme of carrying 3DAV services over T-DMB, we implemented a prototype 3D DMB system. Through the experimental results, we show that the proposed system satisfies the required *backward compatibility* and provides acceptable depth impression even under the limited bandwidth of 3D video (allocated less than or equal to 0.8 Mbps) and a small-sized auto-stereoscopic 3D LCD display. In Section 2, we explain the proposed system architecture in detail. In Section 3, experimental results of the subjective quality(depth perception) test with different bit-rate combinations are presented, and we conclude the paper in Section 4.

This work was supported in part by the Ministry of Information and Communication of Korea under the title of "The development of SmarTV technology."

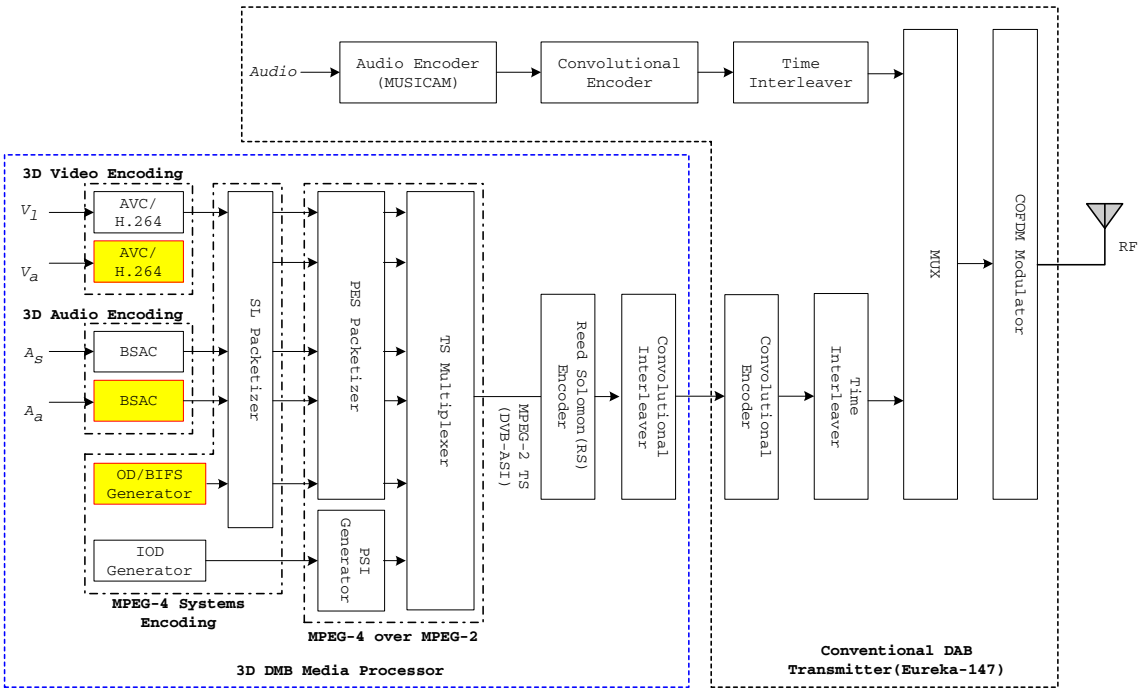

Fig. 1. Block diagram of the T-DMB system.

2. PROPOSED METHOD OF CARRYING 3DAV SERVICES OVER T-DMB

Figure 1 shows a conceptual block diagram of the T-DMB system consisting of DAB system, MPEG-4 Systems, and MPEG-2 Systems. The DAB system, originally designed for the mobile reception of audio and data, offers the effective data rate of 1.5Mbps in 1.536MHz [1]. To deliver multimedia data under the limited bandwidth of the DAB, the advanced video codec (AVC) and the bit sliced arithmetic coding (BSAC) [2] were adopted as the standards of video and audio coders for the T-DMB, respectively.

Note that both the Reed Solomon(RS) encoder and the convolutional interleaver are widely used as the channel coding to enhance error correction at the receiver. The role of the MUX depicted in Figure 1 is to effectively insert the channel coded MPEG-2 transport stream (TS) into the output of the ensemble multiplexed audio of Eureka-147 DAB system in the ensemble transport interface (ETI) frames at every 24ms.

Figure 1 also shows the added (shaded) blocks in the media processor to deliver 3DAV services over T-DMB system. The proposed 3D DMB media processor consists of 3D video encoding, 3D audio encoding, MPEG-4 Systems encoding, MPEG-4 over MPEG-2 encapsulator, and channel coding. As input signals, we are considering a pair of stereoscopic video signals and multichannel 3D audio signals. A monoscopic video and associated depth information are another way to represent stereoscopic video, though the acquisition of accurate depth information in various shooting conditions is still a challenge. In this paper, the left and right (additional) video signals are denoted by V_l and V_a respectively. Similarly, the multichannel 3D audio signals are denoted by A_s and A_a where A_s denotes the normal stereo audio signals for the sake of simplicity.

The 3D video encoding is performed with two separate AVCs. In order to guarantee backward compatibility, we fix the size of V_l to 320×240(QVGA). On the other hand, two other sizes of V_a, 160×240 and 160×120, are tested to find the best compromise between bit-rate overhead and quality of the depth perception. Actually we need to reduce the horizontal resolution to a half, since the interleaved 3D image will be used for 3D LCD. It should be noted that an optimized stereoscopic codec based on AVC is required for the further reduction of the bit-rates by exploiting the spatial correlation between the left and right view signals.

Similarly, the 3D audio encoding is performed by applying two separate BSACs for A_s and A_a, respectively. Here. the multichannel audio signals are acquired with a special 3D microphone and then mixed based on the specification described in ITU-R Rec. BS. 775-1 [4].

The MPEG-4 Systems encoding is to generate MPEG-4 initial object descriptor (IOD) as well as OD/BIFS. Next, it performs Sync Layer (SL) packetization for 3D audio-visual elementary streams and OD/BIFS based on MPEG-4 Systems. The MPEG-4 over MPEG-2 encapsulator converts SL packets to MPEG-2 TS packets. Note that the program specific information(PSI) is also utilized in making MPEG-2 TS packets in the same block.

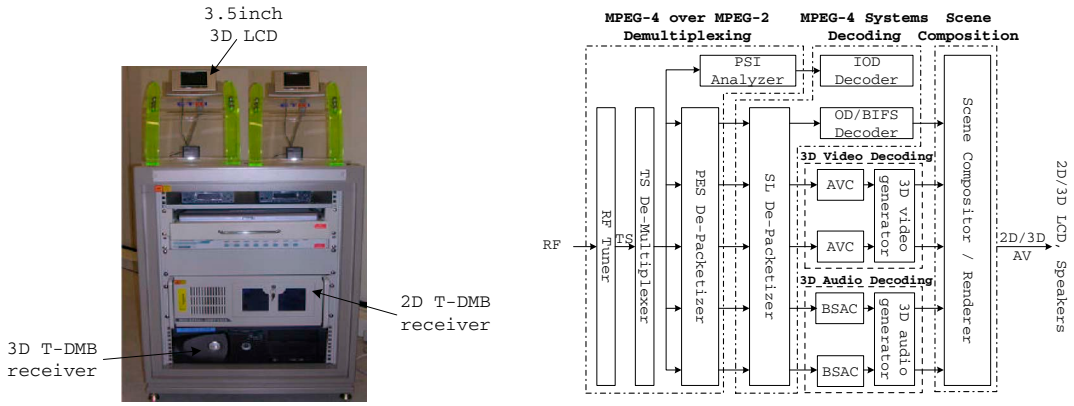

Fig. 2. Prototype of a 3D DMB receiver(left) and its structure(right).

Now we explain the crucial idea of describing AV objects in MPEG-4 Systems in more detail. Let us assume that in the case of 3D DMB system, there are four objects in total, i.e., V_l, V_a, A_s, and A_a. To meet the backward compatibility, we propose a scheme of using two ODs, each OD consisting of two ESs. Additionally, two ODs are assumed to be independent, but two ESs in each OD are assumed to be dependent. The dependence of ESs can be indicated simply by assigning 'TRUE' boolean value to `StreamDependenceFlag` and by assigning the same `ES_IDs` to `dependsOn_ES_ID` in ES_ Descriptor. Next, according to the definition of the MPEG-4 Systems, we assign 0x21 and 0x40 to `objectTypeIndication` of V_l and A_s in `DecoderConfigDescriptor`, respectively. On the other hand in the case of V_a and A_a, we assign 0xC0 and 0xC1 to the `objectTypeIndication` indicating *user private* in MPEG-4 Systems, respectively. Note that the dependent ESs of V_a and A_a are ignored with the current 2D T-DMB receivers, but are identified with the proposed 3D DMB receiver.

With the use of Pentium-IV PCs and auto-stereoscopic displays, we implemented a prototype 3D DMB terminal. Figure 2 shows a photograph of the prototype. The 3D DMB terminal consists of MPEG-4 over MPEG-2 de-capsulator, MPEG-4 Systems decoder, 3D video decoder, 3D audio decoder, and scene compositor.

The MPEG-4 over MPEG-2 de-capsulator is to reconstruct SL packets of V_l, V_a, A_s, A_a, and OD/BIFS. The PSI analyzer parses IOD information and then hands over to IOD decoder. The MPEG-4 Systems decoding is to recover the ESs of 3D audio-visual signals as well as OD/BIFS from the SL packets. According to the OD information, both the 3D video decoder and the 3D audio decoder determine how to decode the signals. Next, the decoded video and audio signals enter the 3D video generator and the 3D audio generator, respectively. Finally, the scene compositor produces the required video and audio signals in 2D or 3D format.

If the display mode is set to '3D', the scene compositor interleaves the left and right video signals to synthesize 3D video signals and 3D audio signals as the required 3D format. In the case of '2D' mode, the left video (V_l) and the stereo audio (A_s) are normally presented at the terminal.

We produced 3D DMB bitstreams according to the new syntax and semantics explained above, and tested them on our prototype system. We have verified that the proposed system satisfies the required backward compatibility with the T-DMB system. As we mentioned previously, the conventional T-DMB receiver ignored the elementary streams for the additional video (V_a) and the additional audio (A_a) because 'OD/BIFS Decoder' in 'MPEG-4 Systems Decoding' block does not identify the `objectTypeIndication` of ODs.

Table 1. Experimental conditions used in the experiments.

Sequences	Input Formats of V_a	Bit Rates	Display
'Soccer'	320×240 ('R')	512kbps	3.5 inch
'FlowerPot'	160×240 ('RH')	768kbps	200 cd/m^2
	160×120 ('RQ')		640×480

3. VIDEO CODING: EXPERIMENTAL RESULTS

While keeping the backward compatibility with the T-DMB system, we also need to achieve the best image quality of the stereoscopic video under the limited bit budget. With the limited bit budget, we may get better image quality if we reduce the resolution of V_a before coding and interpolate the decoded image at the receiver for display. In this section, we present the experimental results of the subjective quality tests on the various sizes of V_a for different available total bit-rates. Experimental conditions used in the experiments are summarized in Table 1. The input format of V_l is 320×240 (QVGA), and we used 3.5 inch 3D LCD display of 200 cd/m^2 in the experiment. For two video sequences 'Soccer' and 'FlowerPot', we evaluate the depth impression with regards to various total bit-rates and different input formats of V_a. Note that we

consider the three cases for the input format of V_a: 'R', 'RH', and 'RQ'. Simply, we would say that this is to find a way of reducing the bit-rates required for the compression of the additional video. As for the bit-rates, we consider two cases: 512kbps and 768kbps in total. Under the fixed total bit-rates, we change the bit-rates allocated for (V_l, V_a) in the following ways:(342, 170), (384, 128), (416, 96), and (448, 64) in case of 512kbps and (448, 320), (512, 256), (576, 192), and (640, 128) in case of 768kbps.

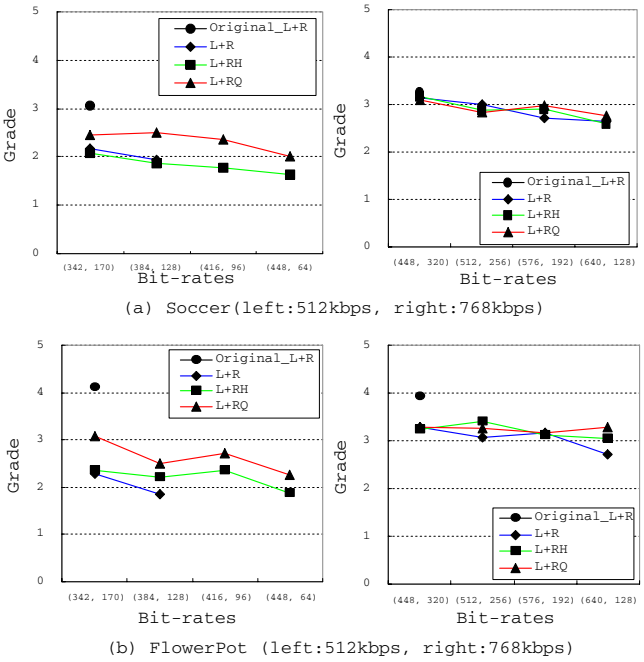

Fig. 3. Ratings of depth perception at the receiver for various bit-rates allocated to encode V_l and V_a: (a) 'Soccer' sequence; (b) 'FlowerPot' sequence.

For the subjective evaluation tests, we adopted the double-stimulus continuous quality-scale (DSCQS) method that is a standard procedure described in ITU-R Recommendation 500[5]. The viewers are asked to rate the depth impression of the 3D video sequences presented on the 3.5 inch auto-stereoscopic display.

Figure 3 shows the ratings of depth perception at the receiver for several cases. 10 subjects participated in the test. In this figure, 'Original_L+R' stands for the rating result of depth impression obtained without compressing V_l and V_a. For both video sequences, we did not consider the cases of (416,96) and (448,64) for 'L+R', just because the bit-rates below 96kbps is insufficient to compress the sequences. It is normally thought that the resolution of the additional video should be equal to that of V_l to achieve the highest depth impression. However, the present study suggests that this does not hold when we have a very limited bit budget: in case of 512kbps total bit-rates, it is clearly shown that 'L+RQ' gives the best depth perception for all different bit allocations; in case of 768kbps, the three resolutions of V_a show the same ratings for all bit-rates allocations. It should also be noted from the Figure 3 that the ratings of depth perception are almost inversely proportional to the difference between the bit-rates of (V_l, V_a). From the above results, we may conclude that reducing the resolution of V_a to its quarter size before compression and allocating more bit-rates than proportional to its size gives the best stereoscopic image quality in terms of depth perception.

4. CONCLUDING REMARKS

We introduced a 3D DMB prototype system which can provide 3DAV services over T-DMB while maintaining the backward compatibility with the T-DMB system. With transmission and reception experiments, we have verified that the proposed concept of 3DAV services over T-DMB works well. And under the limited bandwidth and the small-sized display, the subjective tests have shown that the developed system can provide acceptable depth and video quality. Similar approach could be applied to various applications such as terrestrial digital television, digital cable television, IPTV, and so on. T-DMB is a very attractive platform for successful commercial 3DTV trials due to its service characteristics: small display, single viewer and a new media. If the 3D DMB service is widely accepted, 'big' 3DTV at home will naturally be the next step.

5. REFERENCES

[1] ETSI EN 300 401 (2000): Radio broadcasting systems; Digital Audio Broadcasting(DAB) to mobile, portable and fixed receivers.

[2] ISO/IEC 14496-1:2001, Information Technology–Generic Coding of Audio-Visual Objects-Part 1: Systems.

[3] ISO/IEC 13818-1:2000/Fdam7, Information Technology–Generic Coding of Moving Pictures and associated Audio Information: Systems, Amendment 7: Transport of ISO/IEC 14496 data over ISO/IEC 13818-1.

[4] ITU-R Rec. BS. 775-1 (1994): Multichannel Stereophonic Sound System with and without accompaning Picture.

[5] ITU-R Recommendation BT.500-10 (2000): Methodology for the Subjective Assessment of the Quality of Television Picture.

[6] TTAS.KO-07.0026 (2004): Radio Broadcasting Systems; Specification of the video services for VHF Digital Multimedia Broadcasting (DMB) to mobile, portable and fixed receivers.