

Novel Hybrid Content Synchronization Scheme for Augmented Broadcasting Services

Soonchoul Kim, Bumsuk Choi, Youngho Jeong, Jinwoo Hong, and Kyuheon Kim

As a new hybrid broadcasting service, augmented broadcasting shows enhanced broadcasting content on a large TV screen, while augmented reality (AR) on a mobile device augments additional graphical content onto an input image from the device's own camera to provide useful and convenient information for users. A one-sided broadcasting service using AR has already been attempted in virtual advertisements during sport broadcasts. However, because its augmentation is preprocessed before the video image is transmitted, the viewer at home may have no influence on this formation; and no interaction for the user is possible unless the viewer has a direct connection to the content provider. Augmented broadcasting technology enables viewers to watch mixed broadcasting content only when they want such service and to watch original broadcasting content when they do not. To realize an augmented broadcasting service, the most important issue is to resolve the hybrid content synchronization over heterogeneous broadcast and broadband networks. This paper proposes a novel hybrid content synchronization scheme for an augmented broadcasting service and presents its implementation and results in a terrestrial DTV environment.

Keywords: Augmented reality, augmented broadcasting, hybrid contents synchronization, next-generation smart TV.

Manuscript received Mar. 20, 2014; revised July 16, 2014; accepted July 31, 2014.

This work was supported by the ICT R&D program of MSIP/IITP, Rep. of Korea (11921-03001, Development of Beyond Smart TV Technology).

Soonchoul Kim (corresponding author, choulsim@etri.re.kr), Bumsuk Choi (bschoi@etri.re.kr), Youngho Jeong (yhcheong@etri.re.kr), and Jinwoo Hong (jwhong@etri.re.kr) are with the Broadcasting & Telecommunications Media Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Kyuheon Kim (kyuheonkim@khu.ac.kr) is with the Department of Communications, Kyung Hee University, Yongin, Rep. of Korea.

I. Introduction

We are currently living in a hybrid era in various areas, including cars, energy, clothes, sports, and electronics. Today's television broadcasting is presented on hybrid TVs or smart TVs, which are basically equipped with a network interface. A television can connect to the Internet and cooperate with a smart device. This is an important change in that TVs can now deliver information other than typical broadcasting content through a broadcast network. Most modern TV viewers can watch a broadcast program while simultaneously finding additional information they may be curious about using a secondary device such as a tablet PC or a smartphone. However, such TV viewing experience may distract the viewer's attention. If TV can augment related content at an interesting position of the current program, and TV viewers can manipulate such content with a secondary device, the immersion effect will increase. As an example, when watching a player putt a golf ball, a 3D graphic map of the green can be augmented onto the real green. Viewers can clearly see the slope by manipulating the 3D map without taking their eyes off the program. In addition, viewers can actually read the green while in the comfort of their own home. The augmentation of a graphic object onto a viewing screen has recently been applied to mobile devices such as smartphones and tablet PCs. Mobile applications are based on progressive sensors, a notable example of which is augmented reality (AR). AR is a kind of mixed reality in which 2D/3D graphics are integrated into the real world to enhance the user experience and to enrich user knowledge, which is more attractive when interactivity is allowed [1]–[2].

Similar to an AR-based broadcast, a one-sided broadcasting service using AR has already been attempted in the virtual

advertisement of sport broadcasts. However, because its augmentation is preprocessed before the video images are transmitted, the viewer at home may have no influence on this formation; and no interaction for the user is possible unless the viewer has a direct connection to the content provider. To overcome this limited viewing experience without interaction for TV viewers, we studied a novel augmented television service technology (hereinafter referred to as *augmented broadcasting*) that conducts rendering to blend broadcasting content with augmented content (2D/3D graphic objects) in real time on receiving terminals such as TVs or set-top boxes. Augmented broadcasting is used to broadcast enhanced content on a large TV screen while AR on a mobile device augments additional graphical content onto an input image from the device's own camera to provide useful and convenient information for users, enabling viewers to watch mixed broadcasting content only when they want the service and to watch original broadcasting content when they do not. We previously presented the concept and model for augmented broadcasting and implemented a test-bed system, its relevant elements based on PC platforms, and an Intranet without a real broadcasting transmission scheme [3]–[4]. To realize an augmented broadcasting service in a real broadcasting environment, the most important issue is that of trying to resolve hybrid content synchronization over heterogeneous broadcast and broadband networks. In the recent research, content synchronization schemes for the provision of interactive services in hybrid broadcast/broadband environments and service-compatible 3D video services in DTV environments have been introduced in [5]–[6] and [7], respectively. Because these previous schemes were designed and structured according to particular service characteristics and policies, they cannot resolve the issue of content synchronization in a hybrid broadcasting service that includes new service properties such as augmented broadcasting. Therefore, this paper proposes a novel hybrid content synchronization scheme for an augmented broadcasting service and, presents its implementation and results for a terrestrial DTV environment.

The remainder of this paper is organized as follows. In Section II, the service model and structure of the augmented broadcasting system are presented in detail. In Section III, the design of the proposed hybrid content synchronization is provided. In Section IV, we describe the functional components implemented for verifying the backward compatibility and effectiveness of the proposed service technology. Finally, some concluding remarks regarding this proposal are provided in Section V.

II. Augmented Broadcasting System Structure

1. Service Model

Figure 1 shows the service model of the augmented broadcasting system. The augmented broadcasting provider defines the augmented region from the audio/video content and the information expressed in the augmentation region. The information is formatted as augmented broadcasting metadata and generated by an authoring tool. The metadata includes the unique name, object type, position, presentation/lifetime, resource location of the augmented objects, and rendering attributes for mixing the augmented objects according to the augmented broadcasting service scenario. A service scenario can be planned during the content production process. Content providers produce various augmented content databases harmonized with a planned scenario. A broadcasting program is transmitted together with metadata to set-top boxes or smart TVs. When viewers want an augmented broadcasting service while watching TV, they can launch it by clicking a service button located on a TV remote control or mobile application on a smart device. Viewers can, therefore, watch mixed broadcasting contents when they want such service and the original broadcasting content when they do not. In addition, viewers can select a preferred content provider and even view the same video content again while rewatching the same broadcast from a different content provider.

2. System Architecture

Consuming augmented broadcast content on a TV is not a passive viewing experience rather it is a relatively active viewing experience through viewer selection and manipulation

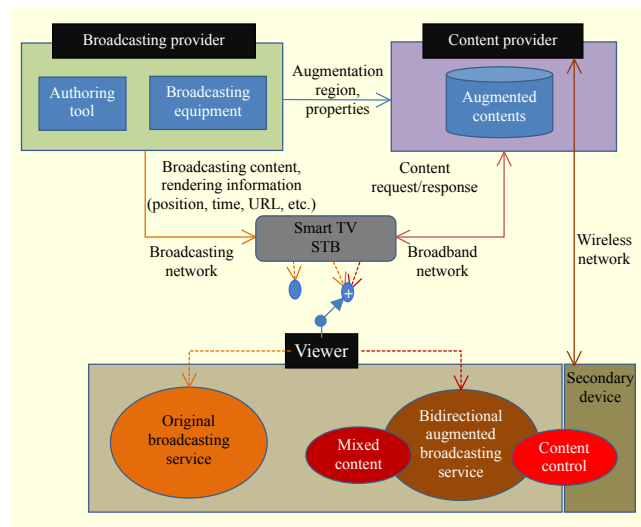


Fig. 1. Bidirectional augmented TV broadcasting model.

of preferred content. Accordingly, we should consider the major requirements (that is, content synchronization and user interface) in view of implementing and providing such service.

A. Real-Time Graphics Rendering Performance Requirements

Recent smart TVs are equipped with a high-performance GPU/CPU but still do not offer a better performance than a mobile device. The receiver for the proposed service should be able to decode audio/video streams in real time and simultaneously overlay 2D/3D graphic objects on all 30 video frames of high-definition quality video per second. The graphic objects need to be modeled under adequate polygons to avoid a delay in video playback.

B. Metadata Presentation and Network Delivery Requirements

Metadata include a descriptive expression (position, time schedule, coordinates, camera sensory data, texture properties, and so on) for augmented broadcasting. Thus, metadata should be presented as a formal format that the interpreter in the receiver can translate easily and quickly. Most of all, metadata should flow in a minimum amount of traffic because a broadcasting channel has a relatively limited bandwidth compared with broadband. Metadata occasionally need to be delivered in a broadband network.

C. Augmented Content Synchronization Requirements

Augmented content should be positioned properly in the frame sequence when a particular space of the scene is substituted with another object, and the position of the object might change with the scene for several seconds. Timing information for content synchronization should, therefore, be inserted into the signaling data. For example, digital television data broadcasting similarly uses a synchronized/asynchronous trigger. In the rendering process of graphic objects, a receiver can utilize this timing information.

We suggest an augmented broadcasting system for next-generation smart TVs, which requires a hybrid broadcasting and broadband network, and provide an optimized solution for realizing such service. This system is divided into two subsystems in the broadcasting domain, a content provisioning server and smart device in the broadband domain, as shown in Fig. 2.

The broadcasting transmission system consists of an AV encoder, signaling server, authoring tool, and multiplexer. The AV encoder outputs the compressed audio/video stream of the MPEG-2 TS and maintains the original broadcasting program from the audio/video source. The signaling server generates signaling information for identifying the application service at the receiver. In the MPEG-2 system, signaling information

should be delivered into the Program and System Information Protocol (PSIP)/program-specific information (PSI) sections to identify a new broadcasting service and describe a channel guide. The authoring tool is used for generating information for the real-time rendering of augmented content on a TV or set-top box. The information is generated as metadata, which includes the regions where the augmented content will be instructed and presented; the multiplexer generates timing information for content synchronization, which enables a receiver to overlay augmented content based on a time schedule. Lastly, all data are collected together in a multiplexer and transmitted as a single stream.

The service platform is loaded onto a set-top box or smart TV and has the following major functions: the de-multiplexer (DEMUX) separates a broadcasting signal from a tuner according to the program information, and the decoder makes an original audio/video stream. The service manager parses a metadata stream extracted from the DEMUX and constructs the instruction, which describes the augmentation content, such as a 2D/3D object, that should be placed on the related video frames at a synchronized time.

The content server registers objects of 2D/3D graphical content in a database. Upon receiving a download request from the service platform or smart device, the server delivers the content of the identifier within the request's URL. Additionally, the server might request a service subscription or device information for authentication prior to delivering the content.

The smart device works as a supplementary screen and loads the related content on the device's own screen. In addition, it synchronizes with the service platform in real time by exchanging the coordinates of scale, rotation, and transition format.

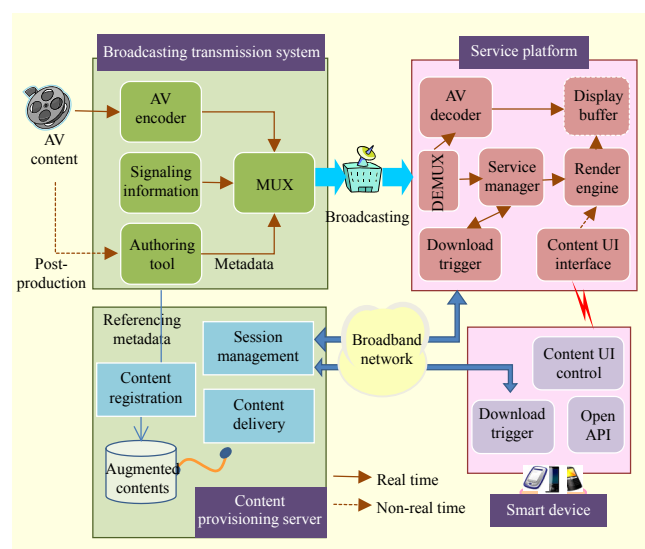


Fig. 2. Bidirectional augmented broadcasting system structure.

III. Design of Hybrid Content Synchronization Scheme for Augmented Broadcasting

1. Synchronization Elements of Augmented Broadcasting

Augmented broadcasting basically utilizes both broadcast and broadband content. The audio/video over a broadcast network is emitted continuously in real time, and the content through a broadband network is acquired when needed. Both types of content should be harmonized at near-frame precision in the TV, appearing as if the mixed content was originally emitted from one source. Figure 3 shows an example of hybrid content synchronization based on augmented broadcasting technology. As an example, consider a student watching a TV program on the history of Egypt. While an expert describes the structure of a pyramid, the relevant 3D content is augmented onto the TV screen at the targeted position (augmentation region) and at the appointed presentation time (augmentation time). In addition, the student may manipulate the 3D object by handling a remote control device with a motion sensor. The 3D

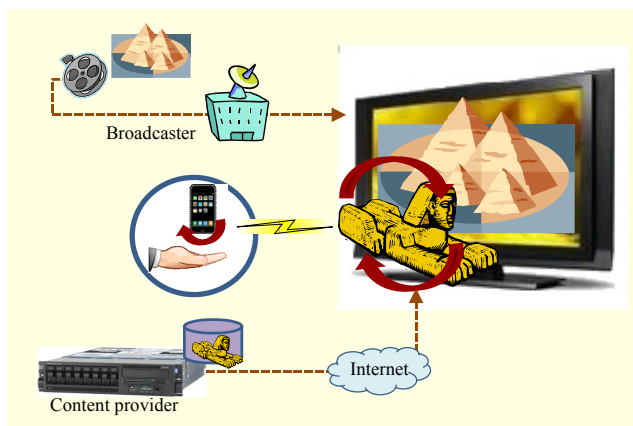


Fig. 3. Example of hybrid content synchronization based on augmented broadcasting technology.

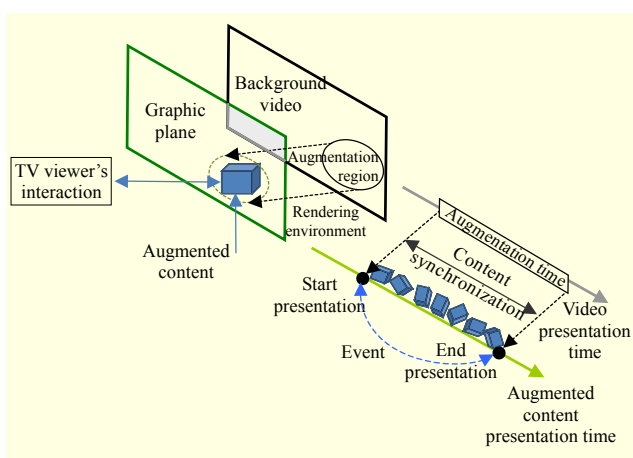


Fig. 4. Relationship among augmentation region, augmentation time, and augmented content.

object then disappears at the end of the appointed presentation time.

The relationship among augmented content, augmentation region, augmentation time, and background video on a TV is shown in Fig. 4. The augmentation region defines a targeted area for the augmented object to be overlaid in a video scene. The augmentation time is the duration of the augmented content — from the time it first appears to the time it fades. During the augmentation time, the augmentation region can move or rotate during every scene, and each event affects the presentation of the augmented content. An event consists of the augmentation region and augmentation time for the augmented content. The rendering environment defines the rendering attributes (light position and colors) for naturally matching the augmented content with the scene.

2. Synchronization Scheme for an Augmented Broadcasting Service

A chain of synchronized events occur with non-continuity within a broadcasting program. The timing synchronization for the events in the current TV broadcasting environment, where it is not easy to estimate the current broadcast position of a broadcast program, needs to be newly defined to enable the receiver to proceed effectively with the synchronization process. The real-time broadcast stream to be provided to the receiving terminal equipped with a service platform includes the metadata. The receiving terminal may determine whether to receive the augmented broadcasting service by an augmented broadcasting descriptor contained in the program initialization information of the broadcast stream. Table 1 illustrates the syntax of an augmented broadcasting descriptor including the descriptor_tag, descriptor_length, and augmented_broadcasting_service_type. More specifically, the descriptor_tag is information that identifies an associated descriptor as an augmented broadcast descriptor. The descriptor_length indicates the total length of the descriptor. The augmented_broadcasting_service_type is used to define

Table 1. Augmented broadcasting descriptor syntax.

Syntax	Semantics
AugmentedBroadcasting_descriptor() {	
descriptor_tag	descriptor identifier
descriptor_length	descriptor length
Reserved	
ABM_service_type	service type
}	

which are arranged based on the transmission time to be transmitted at a particular time, with the stored initial PCR, in units of Packetized Elementary Stream (PES), and it then transmits the metadata at the time when a null packet is detected during multiplexing TS streams. The PSIP generator creates an augmented broadcasting descriptor that describes the augmented broadcasting, and it then inputs it to the ReMUX. The input augmented broadcasting descriptor is contained in the program initialization information of the real-time broadcast stream.

When an MPEG-2 TS stream including internal video data for an augmented broadcasting program is emitted in real time, the ReMUX continuously reads the MPEG-2 TS stream by one TS packet unit. At this time, the value of the initial PCR of the input MPEG-2 packet as timing information for augmented broadcasting synchronization is stored in memory. The ReMUX multiplexes the metadata with the stored initial PCR in units of PES of metadata and detects whether there are metadata to be transmitted in real time during this process. If there are metadata to be transmitted at a particular time, then the ReMUX executes efficiently transmission without additional bandwidth load by carrying the metadata at the time when a null packet is detected. As the metadata has properties of randomly ordered streams, rather than consecutive data streams of the broadcast program, the timing information for the augmented broadcasting synchronization is transmitted only when the metadata is sent; thus, the receiving terminal does not need to refer to other broadcasting or data-packet information when receiving and parsing the metadata.

A receiving terminal interprets the augmented_broadcasting_descriptor contained in the TS to determine whether to receive an augmented broadcast. In addition, when the receiving terminal receives the augmented broadcasting metadata packet before the download of the augmented content is ensured, it extracts the initial PCR information for synchronization and XML data for augmented broadcasting from the metadata PES. The extracted XML data are buffered in the data buffer. As one example, a timing comparer calculates the presentation time stamp, which is an MPEG-2 system time clock, in consideration of the current status of the program, using the initial PCR and the activating augmentation time (hour:minute:second:frame) contained in the XML data. The presentation time stamp (PTS) may be calculated using the equation below.

$$\text{PTS for activating the augmentation time} = \text{Initial PCR base} + \{\text{number of frames converted from the activation time defined by XML data} \times \text{Interval of PTS (3,000)}\}$$

Assuming that the initial PCR is 30,000 and the program play time (for which content A defined by XML is activated) is 1 minute: 10 seconds: 10 frames, the augmented content is activated after 2,110 frames. Applying this to the above

equation, 6,330,000, as an interval value, is obtained. Then the PTS value is produced by adding the value of the initial PCR (that is, 30,000) and the obtained interval value ($6,330,000 + 30,000 = 6,360,000$).

The data buffer buffers the augmented content received from the Internet until the decode time stamp (DTS) reaches the activating augmentation time, and then it inputs it to an augmented data decoder. The augmented data decoder decodes the augmented data received. Once the timing comparer calculates the activating augmentation time and compares the calculated value with the current PCR value, it applies a start signal to the augmented content decoder when the DTS arrives. The frame buffer renderer renders the broadcast data of the video decoder together with the augmented content elementary stream (ES) decoded by the augmented content decoder.

IV. Implementation and Experimental Results

Finally, we implemented an authoring tool, a ReMUX, and a set-top box with the core technologies for augmented broadcasting. Then we constructed a test platform to verify the proposed hybrid content synchronization architecture for an augmented broadcasting service in a DTV broadcasting environment, as shown in Fig. 6. The broadcast network between a transmission system and a service platform equipped with a set-top box is connected by a terrestrial DTV RF signal. In addition, the legacy DTV terminal and an MPEG-2 TS analyzer are constructed together in the test bed to prove that the newly proposed scheme does not affect the quality of a current DTV broadcasting service. The augmentation metadata are generated by the authoring tool that defines the augmentation region and augmentation time and are then forwarded to a ReMUX. The ReMUX transmits an MPEG-2 TS stream after inserting the augmented broadcasting descriptor and packetizing metadata including the initial PCR value defined in Section II. We applied the proposed synchronization scheme to broadcasting programs, such as education and travel, for consideration as an effective augmented broadcasting service.

As the test results in Fig. 6 show, the proposed scheme overlays precisely augmented content whose augmentation region and augmentation time are reserved by an authoring tool in advance. In addition, it shows the normal operation of a legacy DTV terminal without any fault or error of the video stream during testing. The augmented contents are also loaded within a smart device at the same time and synchronized by exchanging 3D coordinates, such as those relating to scale, rotation, translation, and play-speed control as manipulated by the user.

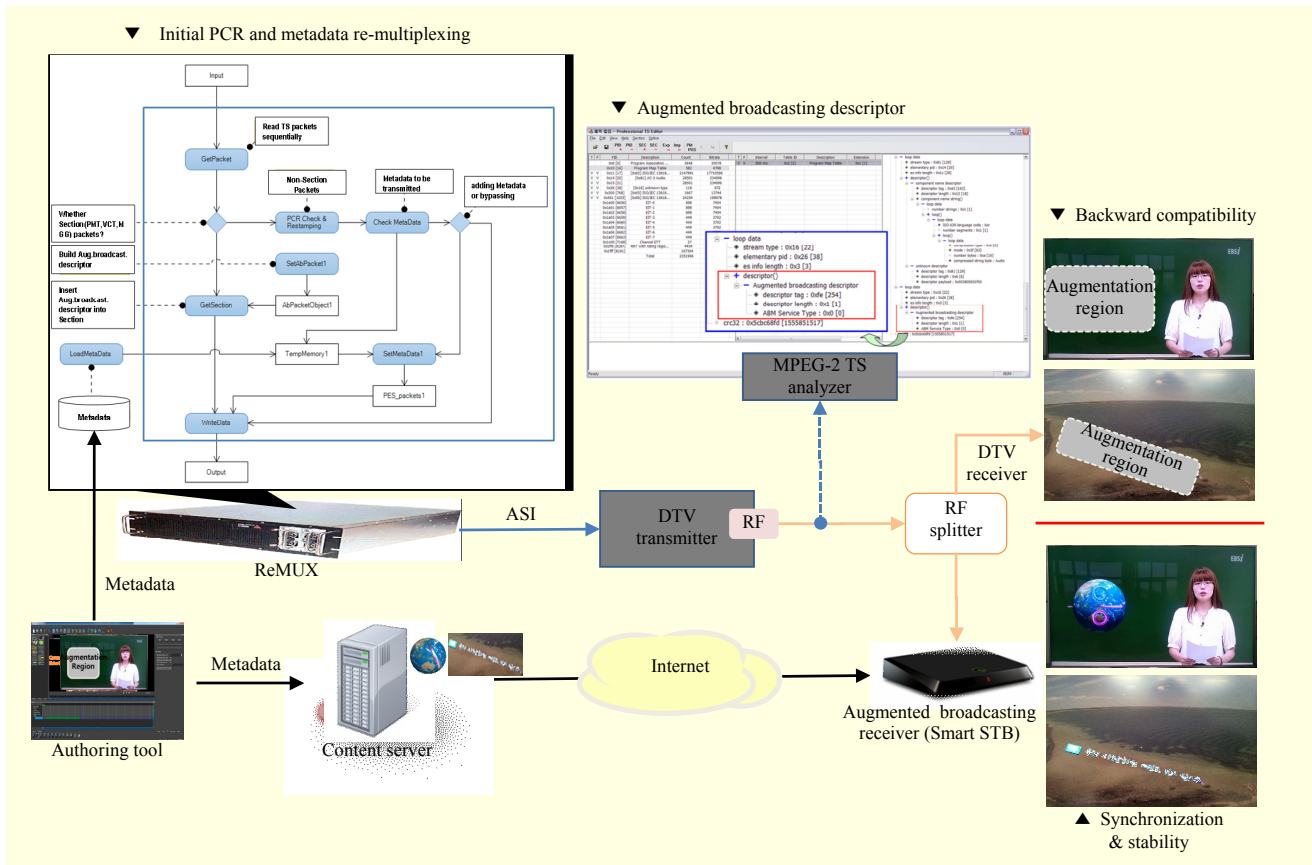


Fig. 6. Test platform for verifying hybrid content synchronization scheme based on augmented broadcasting technology.

V. Conclusion

Next-generation smart TV has gone beyond the classic viewing or constrained UI/UX of the current smart TV and aims at truly viewer-centric broadcasting, straightforward multimedia consumption, and social network participation. Augmented broadcasting as a new hybrid broadcasting service for next-generation smart TV is used to broadcast enhanced content on a large TV screen, while augmented reality (AR) on a mobile device shows the rendered results on a mobile screen after constructing a mixture of additional graphical content and an input image from the device's own camera.

This paper proposed a novel hybrid content synchronization scheme for an augmented broadcasting service and described its implementation and results in a terrestrial DTV environment. In addition, by verifying its backward compatibility and stability, the feasibility for a terrestrial DTV-based augmented broadcasting service has increased.

References

[1] R. Azuma, "A survey of Augmented Reality Presence," *Teleoperators Virtual Environments*, vol. 6, no. 4, Aug. 1997, pp.

355–385.

- [2] P. Milgram and A.F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Trans. Inf. Syst.*, vol. E77-D, no. 12, Dec. 1994, pp. 1321–1329.
- [3] S. Kim et al., "An Architecture of Augmented Broadcasting Service for Next Generation Smart TV," *Broadband Multimedia Syst. Broadcasting*, Seoul, Rep. of Korea, June 27–29, 2012, pp. 1–4.
- [4] B. Choi et al., "A Metadata Design for Augmented Broadcasting and Testbed System Implementation," *ETRI J.*, vol. 35, no. 2, Apr. 2013, pp. 292–300.
- [5] ETSI TS 102 809, *Digital Video Broadcasting (DVB): Signaling and Carriage of Interactive Applications and Services in Hybrid Broadcast/Broadband Environments*; V 1.1.1, Jan. 2010.
- [6] A. Baba et al., "Seamless, Synchronous, and Supportive: Welcome to Hybridcast: An Advanced Hybrid Broadcast and Broadband System," *IEEE Consum. Electron. Mag.*, vol. 1, no. 2, Apr. 2012, pp. 43–52.
- [7] K. Yun et al., "Repetitive Delivery Scheme for Left and Right Views in Service-Compatible 3D Video Service," *ETRI J.*, vol. 36, no. 2, Apr. 2014, pp. 264–270.



Soonchoul Kim received his BS and MS degrees in electrical, electronics and computer engineering from Sungkyunkwan University, Seoul, Rep. of Korea, in 1998 and 2000, respectively. Since 2000, he has been at the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea. Currently, he

has participated in developing smart TV technology as a principal researcher while pursuing a doctoral degree in computer engineering from Sungkyunkwan University. His current interests include network management and security; augmented broadcasting; and hybrid media transmission technology.



Bumsuk Choi received his BS and MS degrees in computer science from Chungnam National University, Daejeon, Rep. of Korea, in 1997 and 2001, respectively. He joined the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, in 2001. Recently, he

has participated in developing Beyond Smart TV Technology. His current interests include 3D audio, digital rights management, Smart TV technology, and 4D cinema technology.



Youngho Jeong received his BS and MS degrees in electronics engineering from Chonbuk National University, Jeonju, Rep. of Korea, in 1992 and 1994, respectively and his PhD degree in electronics engineering from Chungnam National University, Daejeon, Rep. of Korea, in 2006. Since 1994, he has been at

the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, as a director of the Smart TV Service Research Team. He has also been an adjunct professor with the Department of Mobile Communication & Digital Broadcasting Engineering at the University of Science and Technology, Daejeon, Rep. of Korea. His research interests include augmented broadcasting, realistic media, smart TV, smart advertisement, and emotion media.



Jinwoo Hong received his BS and MS degrees in electronic engineering from Kwangwoon University, Seoul, Rep. of Korea, in 1982 and 1984, respectively. He received his PhD degree in computer engineering from the same university in 1993. Since 1984, he has been with the Electronics and Telecommunications

Research Institute, Daejeon, Rep. of Korea, as a principal researcher, where he is currently a managing director of the Next-Generation Smart TV Research Department. From 1998 to 1999, he was at the Fraunhofer Institute, Erlangen, Germany, as a visiting researcher. His research interests include technologies for multimedia frameworks; smart TV; smart media and services; realistic media; and emotion media.



Kyuheon Kim received his BS degree in electronics engineering from Hanyang University, Seoul, Rep. of Korea, in 1989 and his MPhil and PhD degrees in electrical and electronics engineering from Newcastle University, Newcastle upon Tyne, UK, in 1996.

From 1996 to 1997, he was at Sheffield University, UK, as a research fellow. From 1997 to 2006, he worked as the head of the Interactive Media Research Team at the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, where he standardized and developed the T-DMB specification. He was the head of the Korean delegates for the MPEG standard body from 2001 to 2005. Since 2006, he has conducted research at Kyung Hee University, Seoul, Rep. of Korea. His research interests include interactive media processing, digital signal processing, and digital broadcasting technologies.