

Rich Media Contents Authoring System based on MPEG-4 LAsER

Injae Lee*, Jihun Cha*, Myungseok Ki*, Hankyu Lee*, Jinwoo Hong*

**Broadcasting&Telecommunications Media Research Dept., ETRI*

161 Gajeong-Dong, Yuseong-Gu, Daejeon, 305-350, Korea

{ninja, jihun, serdong, hklee}@etri.re.kr

Abstract— This paper presents an interactive rich media contents authoring system based on MPEG-4 LAsER. Compared to the old and plain AV centric multimedia services, rich media services provide attractive characteristics include dynamic updates of contents and interactivities among various forming multimedia data such as audio, video, graphics and text. This paper introduces an authoring system, which can easily and conveniently produce an interactive rich media contents by using MPEG-4 LAsER. Therefore, presented authoring system supports dynamic and active characteristics of rich media services. Detailed implementation of the interactive rich media authoring system and its capabilities are presented along with usage examples.

Keywords— Interactive, Rich media, MPEG-4 LAsER, Authoring

I. INTRODUCTION

Rich media service provides a fresh and active user experience based on enriched contents including audio, video, text, and graphics animation. It ranges from a movie enriched with vector graphics overlays and interactivity to complex multi-step services with fluid interaction and different media types at each step. The demand for such rich media service is increasing at a high pace spurred by the development of broadened next generation mobile infrastructures and the attempts to consume TV content in new environments.

Several technologies are competing to provide technical specifications for rich media services such as Flash, BiFS, LAsER, SVG, etc. Flash is very successful on the PC; however, it is not an open standard. Thus it suffers from serious drawbacks to attract other industry bodies and their requirements. Two important requirements for mobile rich media services are being an open standard and the supply of a mechanism to convert from various rich media formats (such as Flash and other proprietary formats) to the new standard. Therefore, two international standardization groups, MPEG and W3C, have worked on specifying standards which would satisfy these requirements.

MPEG developed composition and coding technologies for rich media contents. The first specification developed is MPEG-4 BIFS (Binary Format for Scene) [1]. BIFS was

developed to provide virtually all the functionalities rich media can provide to date on a fairly resource abundant environment: PC and the Internet. Such advanced functionalities include 3D presentation of the scene, an extensive support of animation, an arbitrary shaped video presentation, etc. Over the time the demand on richer multimedia services on mobile application domain gets increasing rapidly. However, the inherent characteristics of BIFS from VRML (Virtual Reality Meta Language) and its binary encoding structure make it inappropriate applying for services using resource constrained devices. Instead of compromising the performances, MPEG reached to the conclusion that an optimum between feature richness/compression efficiency and device constraints needed to be found, thus decided to create a new standard for rich media in a resource constrained environment. Based on these consideration and decision, MPEG developed a new composition coding technology called LAsER (Lightweight Application Scene Representation) [2] [3].

W3C also made attempts to define description languages for rich media contents: SMIL (Synchronized Multimedia Integration Language) [4] and SVG (Scalable Vector Graphics) [5] recommendations. However, neither SMIL nor SVG supports a streaming method. Therefore, these recommendations are inappropriate for fast, dynamic, interactive and interoperable applications. Inspired by the best concepts of the state-of-the-art open solutions such as SVG and BiFS, LAsER (Lightweight Application Scene Representation) tunes and optimizes each feature required by rich media services to effectively fulfill the need of an open standard. The main goal of LAsER is to provide a representation of scene data that is well suited for constrained environment, and it is defined to be the rich-media format targeted at constrained devices, such as mobile, embedded and consumer electronics devices. In order to satisfy these requirements, LAsER adopts subsets of SMIL and SVG recommendations and BiFS standard to support use scenarios suitable for rich media streaming and broadcasting.

The presented rich media contents authoring system is based on the newly developed open standard, MPEG-4 LAsER. This paper is organized as following. Section II presents the overview of MPEG-4 Part 20 (LAsER) specification. In section III, the architecture of the rich media contents authoring system

is presented. Section IV shows the implementation results with examples and section V concludes with the suggestion of further researches.

II. OVERVIEW OF MPEG-4 LAsER

Interactive rich media contents can provide rich information to users because it is equipped with powerful technologies to represent spatio-temporal composition of multimedia resources while assuming interactive consumption of them based on user demands. MPEG-4 Part 20 (ISO/IEC 14496-20) specifies the coded representation of multimedia presentations and the simple aggregation format for rich media services such as advanced mobile TV services, rich media portal services and graphics-centric services. It is divided into two specifications: LAsER and SAF(Simple Aggregation Format) [2].

A. LAsER

LAsER is a specification designed for representing rich media services to resource constrained devices. In other words, the main goal of LAsER is to provide a representation of scene data that is well suited for a resource constrained environment. The rich media presentation is a collection of a scene description and media. A scene description specifies the spatio-temporal layout of objects such as a text, graphics, animation, an individual audiovisual content of the form image, video, audio, font data and interactive relationships of them. Binary format is also defined to provide better transport efficiency.

LAsER is based on SVG Tiny which is the W3C Recommendation for mobile vector graphics services. Upon this basis, MPEG added LAsER Extensions; Dynamic Updates, Binary Encoding, Audio/Video support, and Font Information to employ various functionalities required for rich media services in a resource constraint environment. The specification defines a LAsER engine, which is shown in Figure 1, as the viewer for LAsER presentations. A LAsER engine has rich media composition capabilities and these capabilities rely on the usage of an SVG scene tree.

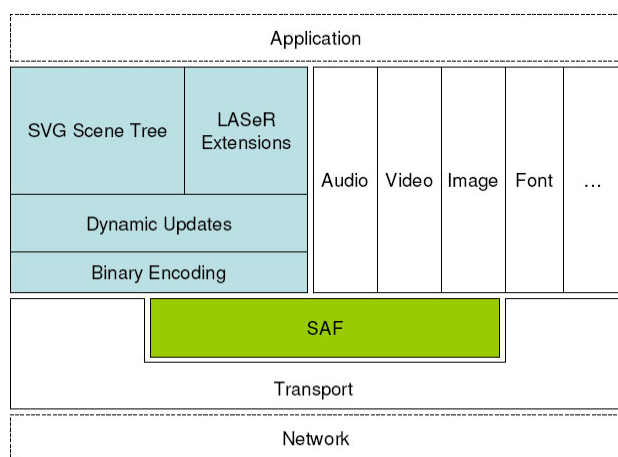


Figure 1. Architecture of a LAsER engine

B. SAF

SAF is a multiplexing scheme which provides a simple aggregation mechanism of comprising media for efficient transport of LAsER data together with audio/video content over various existing delivery channels.

SAF defines the binary representation of a compound data stream composed of different ES (Elementary Stream) such as LAsER scene description, video, audio, image, font, metadata streams. Data from these various data elementary streams results in one SAF stream by multiplexing them for simple, efficient and synchronous delivery. A SAF stream is made of SAF AU (Access Units) as shown in Figure 2. It is designed to be simple to implement to allow efficient demultiplexing on low-end devices.

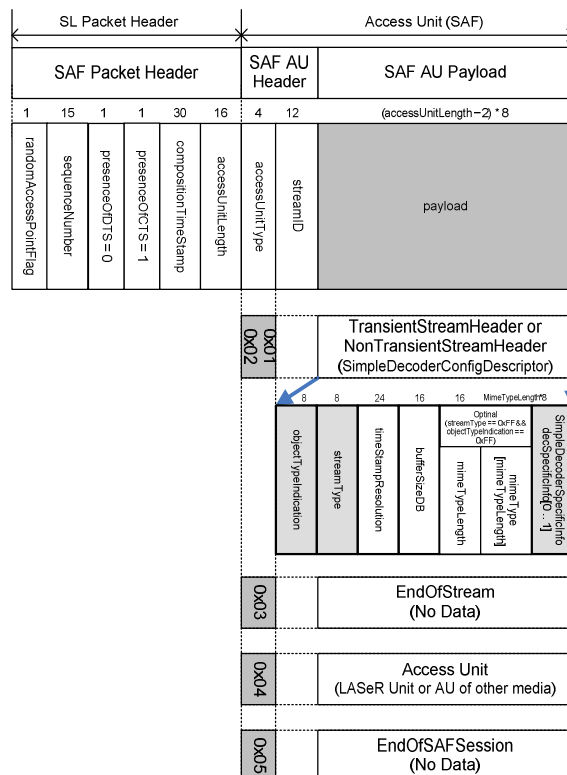


Figure 2. Architecture of the SAF packet

MPEG and IETF have jointly specified, in MPEG-4 Part 8 [6] and RFC 3640 [7], the transport over IP of any kind of MPEG ES. Any MPEG-4 ES which is packaged using the SL (Synchronization Layer) syntax can be transported using this specification. SAF streams can be wrapped into SL packets and transported into RTP packets as specified in RFC. There is currently no specification for the carriage of SAF packets over RTP.

III. ARCHITECTURE OF RICH MEDIA CONTENTS AUTHORING SYSTEM

The rich media contents authoring system enables an author to easily and conveniently produce rich media contents by

using MPEG-4 LAsER. It provides powerful graphic user interface and produces rich media contents in the form of SAF. The architecture of the presented authoring system is shown in Figure 3. It consists of the following parts; User Interface, Information Editor & Processor, Richmedia Encoder and Media Library.

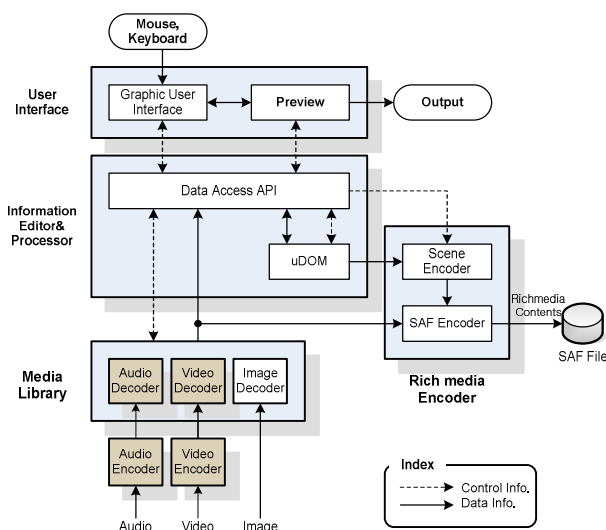


Figure 3. Architecture of the presented authoring system

A. User Interface

The presented authoring system supports a powerful graphic user interface to help an author conveniently create and edit rich media contents. And it provides the spatio-temporal authoring environment. Figure 4 shows the graphic user interface in the presented authoring system. It consists of menu and tool bars, scene editor, scene tree, event, attribute, time line windows.

The menu and tool bars provides fundamental functions for the authoring system such as open, save, publish, etc. In the scene editor window, a scene of contents can be created and edited by inserting, deleting and modifying audio-visual objects. It supports multimedia data such as AVC/BSAC/JPEG/PNG and graphics such as circle/rectangle/text/line.

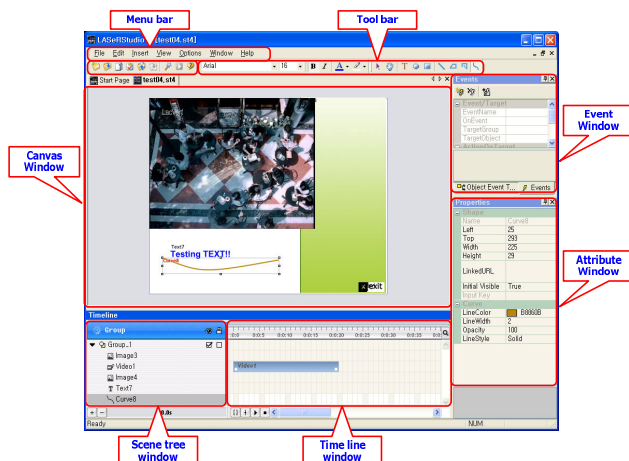


Figure 4. GUI of the presented authoring system

The scene tree window shows scene description composed of a number of different objects, and can edit spatial relationship manually. Many event properties of an object such as show, hide, animation can be realized in the event window. Corresponding attributes of events on individual objects are shown and can be edited in the attribute window. The temporal attributes of individual audio-visual objects in a scene are displayed as a track in the time line windows.

B. Information Editor & Processor

Information Editor & Processor provides the interfaces of easy access to all the functionalities of the presented authoring system using Data access APIs. Data access APIs contains the interfaces that construct and edit an internal data structure such as the uDOM(micro Document Object Model). The uDOM is based on XMT DOM, and thus has a tree structure with a variety of elements for scene description and reference to media object. It also allows hierarchical access to the scene information.

C. Rich media Encoder

Rich media Encoder consists of Scene Encoder and SAF Encoder. In Scene Encoder module, internal scene data based on uDOM is converted to a binary LAsER stream. As explained before, the LAsER stream describe a scene composition, however, it does represent only a spatio-temporal composition of objects, not contain object data itself. In order to display proper contents, it is required to add objects data corresponding to their composition information. Thus, the SAF Encoder encodes the LAsER stream and objects data such as audio and video data used in a scene in order to generate a SAF file.

D. Media Library

Media Library is the collection of the codecs for decoding audio-visual objects such as AVC for visual sequences, BSAC for audio sequences, JPEG, PNG for still images. Since the Media Library separately manages media codes from other parts of the presented authoring system, it is much easier to extend the scope of the Media Library by adding new media codecs.

IV. IMPLEMENTATION

In this paper, an interactive rich media comemts authoring system which is conformed to MPEG-4 LAsER specification is presented. The authoring results of the presented authoring system are shown in Figure 5, 6, 7. They are interactive rich media contents based on MPEG-4 LAsER including various media and events.

Figure 5 shows a cooking show including additional data in the form of images and graphics along with the main audio-visual information. This content can be offered additional information to viewers. Therefore, viewers can select interested menu options. When a viewer selects the "Today's recommended cooking" menu, an image containing the information of the food title and recipe is triggered and displayed on the screen (illustrated in Figure 5 (b) and (c)).

And the famous restaurant information also can be provided when a viewer selects a “recommended restaurant” menu as shown in figure 5(d).

And the additional information such as a singer /album information and a weekly music chart can be offered to viewers as shown in figure 6. The game ranking and entry team information can be displayed on the screen when a viewer clicks the menu (illustrated in Fig. 7 (a) and (b)).



Figure 5. An example of rich media contents for a cooking show



Figure 6. An example of rich media contents for a music program

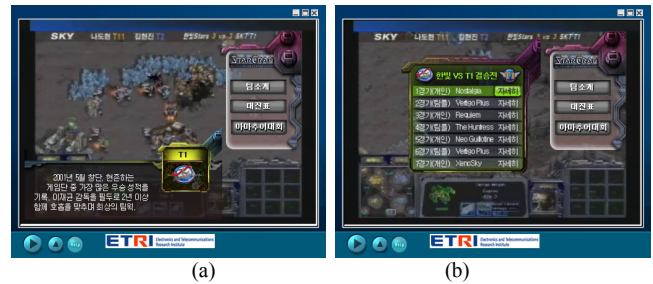


Figure 7. An example of rich media contents for a game program

As shown in figure 5, 6, 7, rich media contents are easily produced through GUI by temporal and spatial composition of objects, and by events managements.

V. CONCLUSION

The interactive rich media service is one of new paradigms for mobile broadcasting services. LASer is a newly introduced open standard for rich media service that provides interactive functionalities in a lightweight contents format. Providing interactive rich media contents make user behaviors from passive viewing to active playing on the contents. By interacting directly on the contents, users can acquire additional information on the terminal.

Presented rich media authoring system provided a noble approach to produce interactive rich media contents by using MPEG-4 LASer. Users can enjoy more versatile value added services compared to plain AV service with rich media contents. Interactive rich media contents is to provide detailed information using a telecommunication channel while broadcasting contents provides a selective menu along with the main AV such as advanced mobile TV service, advanced VoD service, etc. The way of effective transmission of the LASer-based rich media content needs to be further investigated.

ACKNOWLEDGMENT

This work was supported by the Ministry of Knowledge Economy under the title of “Development of UHD/3D Realistic Video and Interactive View Control Technologies for IPTV”.

REFERENCES

- [1] ISO/IEC 14496-1, Information Technology - Coding of Audio-Visual Objects - Part 1: Systems-First Edition, Dec. 1999.
- [2] ISO/IEC 14496-20, Information Technology - Coding of audio-visual objects, Part 20: lightweight Application Scene Representation (LASer) and Simple Aggregation Format (SAF), June 2006.
- [3] ISO/IEC JTC1/SC29/WG11 N9023, Part 20: lightweight Application Scene Representation (LASer) and Simple Aggregation Format (SAF) AMENDMENT 2: Extensions to support SVG1.2, Oct. 2007.
- [4] W3C, SMIL 2.0 Specification. <http://www.w3c.org/TR/smil20>.
- [5] W3C, SVG Tiny1.1 Specification. <http://www.w3.org/TR/SVG11/>.
- [6] ISO/IEC 14496-8:2003 - “Information technology - Coding of audio-visual objects - Part 8: Carriage of ISO/IEC 14496 contents over IP networks”
- [7] RTP Payload Format for Transport of MPEG-4 Elementary Streams, RFC 3640, <http://www.ietf.org/rfc/rfc3640.txt>