The Content Analyzer Supporting Interoperability of MPEG-4 Content in Heterogeneous Players

Hyunju Lee and Sangwook Kim

Department of Computer Science, Kyungpook National University,
Sankyuk-dong 1370, Bukgu, Daegu, Korea
{hyunju, swkim}@cs.knu.ac.kr

Abstract. MPEG-4 supports object-based interactive multimedia applications. To compose complex multimedia scene, MPEG-4 supports BIFS and XMT as scene description. For interoperability of the MPEG-4 content and object-based rendering information management, this paper proposes the content analyzer that supports multiple scene description. In order to improve content management's efficiency, the content analyzer provides module optimization from scene description's parsing to scene information management. This analyzer supports exact analysis of the scene description and object-based core information management. And it can be applied to XML-based heterogeneous players.

1 Introduction

MPEG-4(Moving Picture Experts Group, ISO/IEC 14496) [1] targets interactive multimedia applications and also offers content-based functionality, i.e. the possibility of accessing and manipulating individual objects in the scene.

MPEG-4 describes the composition of a complex multimedia scene which relies on the concept of both BIFS and XMT as scene description. The scene description describes the spatial-temporal composition of objects in a scene and provides media data to the presentation layer of the MPEG-4 terminal. The BIFS [2] is a binary format of VRML-based, and it is usually used when MPEG-4 content is played in the MPEG-4 player. In contrast, the XMT [3] is a XML-based textual format. The main advantage of XMT is exchangeability of content and interoperability with other media players of XML-based. Other existing MPEG-4 players focus on the BIFS as scene description. Therefore, if MPEG-4 content contains XMT as content's scene description, MPEG-4 content can not be reused by other purpose.

In this paper, we propose the content analyzer that supports two types of MPEG-4's scene description sufficiently and it also supports description conversion for interoperability of content. It consists of four core components: the description separator, the BIFS/XMT engine and the information manager. To improve content management's efficiency, the content analyzer provides module optimization from scene description's parsing to scene information management. To expand module easily and improve throughput of player, each module in content analyzer is independently managed. Specially, the XMT engine is adaptively designed for scene description change.

This paper is organized as follows. Section II, we propose the content analyzer in MPEG-4 player. In section III, presents experimental results. Finally, section IV presents conclusions.

2   The Content Analyzer

The content analyzer supports two types of MPEG-4's scene descriptions that are defined in MPEG-4 standard sufficiently.

In this chapter, we describe the content analyzer in detail.

2.1   The Content Analyzer Architecture in MPEG-4 Player

The MPEG-4 player's action processing is as follows. The MPEG-4 file formatter analyzes the header of an MPEG-4 file, which separates several media data that are included in a file and delivers it to the system decoder. The system decoder reads media data in decoding buffer, and decodes according to decoding time. The content analyzer composes a scene tree and arranges data of buffer on screen through the presenter. The player receives various user events in the user interface.

Figure 1 shows the architecture of the content analyzer in MPEG-4 player and the processing step (from scene description's parsing to scene information management) of the content analyzer.

The content analyzer consists of three major components. These are the description separator for analysis and classification of scene description, the BIFS/XMT engine for parsing process, the information manager for independent management of object information according tov object's properties.
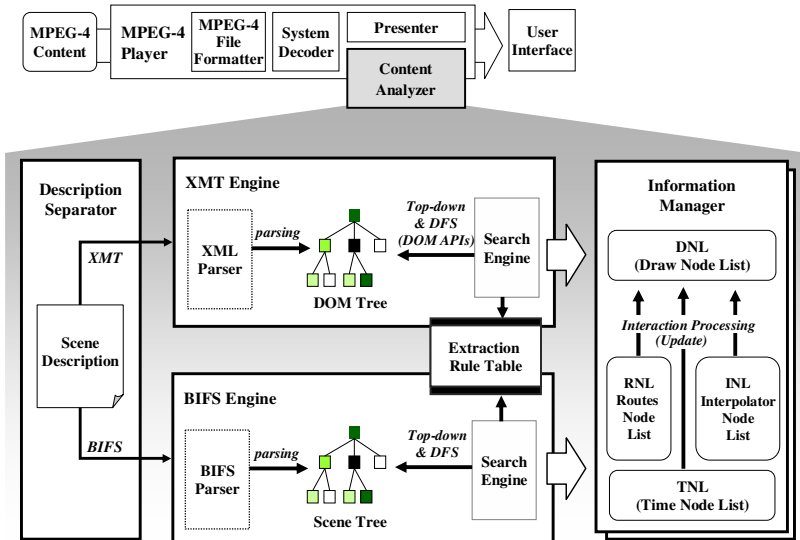

Fig. 1. The architecture of the content analyzer in MPEG-4 player and the processing step

2.2   The Description Separator

The description separator analyzes the scene description that is defined in MPEG-4 content, and calls each engine.

The BIFS's purpose is content playback in MPEG-4 player only. Therefore, it is usually included in MPEG-4 content. But, it is difficult to modify MPEG-4 content because BIFS is binary format. In contrast, the XMT is independently transmitted with MPEG-4 content. The XMT has two creation reasons. First, it supports representation of MPEG-4 scene description using a textual syntax. Therefore, a user can modify media content more easily. Second, it facilitates interoperability with the VRML, the X3D and the SMIL through content exchange. Therefore, if the MPEG-4 content contains an XMT file as scene description, the MPEG-4 content can be played in other XML-based players.

The description separator analyzes two types of scene descriptions exactly, and various analyzed information are used to playback of media data on screen according to use purpose. The description separator has three analysis types according to the content's scene description. First, the description separator calls the BIFS engine when the content contains BIFS only. Second, the description separator calls the XMT engine when the content contains XMT only. Third, the content can include both BIFS and XMT. In this case, the MPEG-4 content can be played to other media players as well as MPEG-4 player. At this time, the description separator calls the BIFS engine and the XMT engine together, and each information manager is created. The player can use parsing result according to contents' playback purpose and playback environment properly.

2.3   The BIFS Engine

If MPEG-4 content contains BIFS as scene description, the MPEG-4 player presents media according to BIFS (BInary Format for Scenes) and OD (Object Descriptor) information. The OD [4-5] identifies and describes elementary streams and associates these streams with corresponding audiovisual scene data. It is used to connect BIFS with media stream resources.

The BIFS engine analyzes the BIFS of MPEG-4 content exactly, searches and extracts the core information of objects that need to rendering. It consists of three components: the BIFS parser, the search engine and the extraction rule table.

The MPEG-4 describes objects and their behavior in hierarchical models. The MPEG-4 uses the concept of a scene tree with object nodes. The scene tree can be obtained through BIFS's parsing and it is used to hierarchically organize and manage the content of scene data. The multimedia information that is defined in scene tree can extract the necessary information through various search mechanisms and manage media data in user's point of view. To support sufficiently object's interactive information that is defined in tree, the MPEG-4 player needs to search the tree more exactly and fast.

In this paper, we use a top-down method and DFS (Depth First Search) mechanism for object-based search. The search engine receives a scene tree's header, searches an object node and judges whether searched object node can draw. When the search engine searches an object, we extract object information according to extraction rules.

The extracted object information is stored in the information manager. The fast search mechanism will be described in evaluation results section. The extraction rule table is based on the search engine and the information manager.

Figure 2 shows an example of the extraction rule table.

Extraction Rule Table

Geometry		Text		Video/Audio	
Field	**Field Type**	**Field**	**Field Type**	**Field**	**Field Type**
whichChoice	mpeg4_int32 (short)	whichChoice	mpeg4_int32 (short)	url	mpeg4_url (char)
center	mpeg4_vec2f (float x, y)	center	mpeg4_vec2f (float x, y)	whichChoice	mpeg4_int32 (short)
scale	mpeg4_vec2f (float x, y)	scale	mpeg4_vec2f (float x, y)	center	mpeg4_vec2f (float x, y)
drawingOrder	mpeg4_float (float)	drawingOrder	mpeg4_float (float)	scale	mpeg4_vec2f (float x, y)
translation	mpeg4_vec2f (float x, y)	translation	mpeg4_vec2f (float x, y)	drawingOrder	mpeg4_float (float)
diffuseColor	mpeg4_color (float r, g, b)	transparency	mpeg4_float (float)	translation	mpeg4_vec2f (float x, y)
filled	mpeg4_bool (char)	width	mpeg4_float (float)	**Image**	
transparency	mpeg4_float (float)	string	mpeg4_string (char)	**Field**	**Field Type**
lineColor	mpeg4_color (float r, g, b)	length	mpeg4_float (float)	url	mpeg4_url (char)
width	mpeg4_float (float)	style	mpeg4_string (char)	whichChoice	mpeg4_int32 (short)
		spacing	mpeg4_float (float)	center	mpeg4_vec2f (float x, y)
		size	mpeg4_float (float)	scale	mpeg4_vec2f (float x, y)
				drawingOrder	mpeg4_float (float)
				translation	mpeg4_vec2f (float x, y)

Fig. 2. An example of the extraction rule table

In extraction rule table, object and object's main property field are defined, which do playback in player, and is possible to extend according to playback environment. For example, if search engine searches a video object, it extracts property information such as url, scale, translation that is defined to extraction rule table. The player achieves search optimization through extraction rules.

2.4 The XMT Engine

The XMT (eXtensible MPEG-4 Textual format) is a framework for representing MPEG-4 scene description using a textual syntax. The XMT facilitates interoperability with XML-based other applications. The XMT format can be interchangeable among heterogeneous players such as SMIL players, VRML players and MPEG-4 players.

The XMT framework consists of two levels of textual syntax and semantics: the XMT-A format and the XMT-Ω format. The XMT-A is an XML-based version and provides one to one mapping between the textual format and binary format. The XMT-Ω is a high level abstraction of MPEG-4 features designed based on the SMIL. In addition, the XMT-C (Common) section contains the definition of elements and

attributes that are used within either XMT-A or XMT-Ω. The XMT can be analyzed by XML parser if XMT observes XML's basis grammar and grammar of DTD (Document Type Definition) or schema that is defined in document.

In this paper, the XMT engine manages XMT file when MPEG-4 content contains XMT as scene description and it supports both XMT-A and XMT-Ω.

The XMT engine analyzes the XMT of MPEG-4 content exactly, searches the DOM tree and extracts the core information of objects that need to rendering. It consists of three core components: the XML parser, the search engine and the extraction rule table. The DOM tree is created after XMT parsing through the XML parser. It is a data structure used to hierarchically organize and manage the content of scene data. The DOM tree is very similar to scene tree of BIFS in case of media representation.

The search mechanism for DOM Tree of XMT is the same as scene tree of BIFS. We search the DOM tree through top-down and DFS method. Extraction information and search method are the same as scene tree but extraction order has different approach because the DOM tree's content description form differs with the scene tree. The search engine can approach to the DOM tree using various DOM APIs. We extract basic information of access in extraction rule table like the BIFS engine.

2.5   The Information Manager

The main characteristic of MPEG-4 is the object-based coding and representation of an audiovisual scene. The BIFS/XMT engine creates interiorly the tree through scene description parsing. We applied the optimized search technique to the tree. However, a tree is only possible hierarchic management of all object information that is included in MPEG-4 content. For reusability and flexibility of object-based, we propose the information manager that manages object information efficiently. It consists of five data structure for basic object rendering and dynamic scene composition processing of objects. It is adaptively designed for rendering.

For basic management of rendering object, we create the DNL (Draw Node List). The DNL consists of heads and subs. An object consists of a single head and several subs. It is a structure that is added to the head and subs whenever an object is searched. The head contains ObjectID and the sub contains several property nodes.

The scene description's ROUTEs mechanism describes interactivity and behavior of objects. ROUTEs are comprised of connections that assign the value of one field to another field in the tree interiorly. It is information that processes dynamic user events that are defined in the MPEG-4 standard. We create the RNL (Routes Node List) for interaction processing. The RNL consists of SourceObjectID, EventType, TargetObjectID, ActionType and ModificationValues.

The MPEG-4 provides stream animation to the scene. We compose two INLs (Interpolator Node List) for interpolator processing. If an object includes interpolator node, the search engine divides whether interpolator is color or position, and store it in each List (CINL: Color Interpolator Node List, PINL: Position Interpolator Node List). The INL consists of ObjectID, key (time value) and keyValue (changing value according to key).

The MPEG-4's another mechanism for dynamic scene composition is time-related interaction. We create the TNL (Time Node List) for time-related interaction. An

object needs playback start time and end time. The TNL consists of ObjectID, two time values (StartTime and EndTime) and PropertyValue. The PropertyValue is activity or inactivity information of object. The MPEG-4 player preferentially processes time information than other user events. Each object of the first screen is played according to time information that is defined to the TNL, and receives other user events in addition.

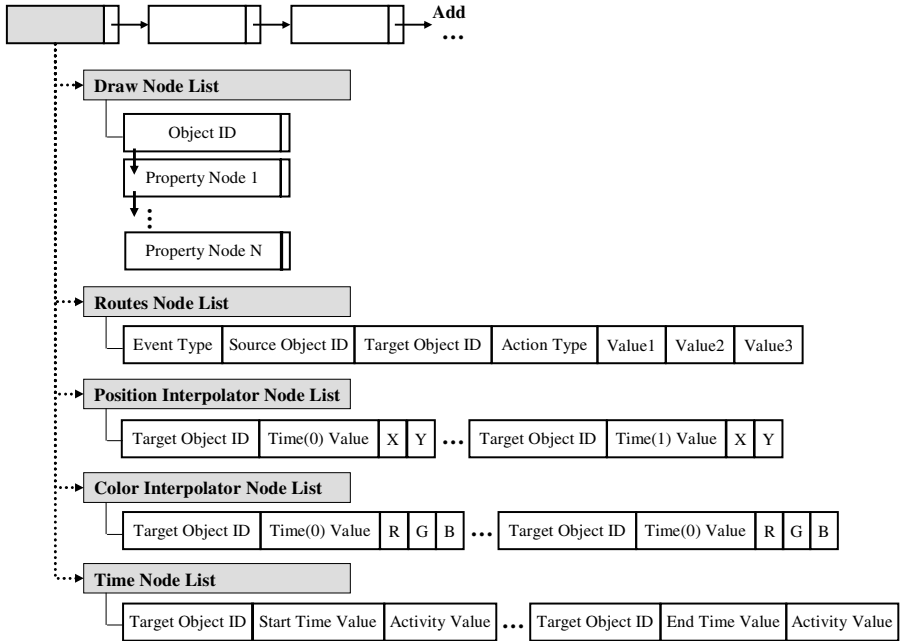Figure 3 shows five data structure of the information manager.

Fig. 3. The information manager

The information manager is dynamically created according to scene description's kind. If MPEG-4 content contains two types of scene description, two information managers are created. The information manager reduces tree re-composition and frequent search of object information because it manages interaction information as well as basic rendering object information by object-based. And, it reduces the load of system-side for additional scene update processing. Finally, whenever scene is updated, it can offer high-level quality of scene to user through reusability of object.

3   Experimental Results

We developed the content analyzer that supports two scene description formats and complete2D profile that is defined in MPEG-4 standard. The proposed content analyzer can manage the MPEG-4 content that consists of various media objects.

3.1   Implementation and Application

In the implementation, we used Xerces-C++ version 2.5.0 as XML parser and XMT API is written in C++. The content analyzer's other modules are written in C/C++ and Visual C++ based on Ms-Windows.

Figure 4 shows two XMT formats and an execution example of MPEG-4 content in the KNU MPEG-4 player. The KNU MPEG-4 Player [6] is an object-based interactive MPEG-4 player which is based on MS-Windows. Two XMT files are matched each schema.
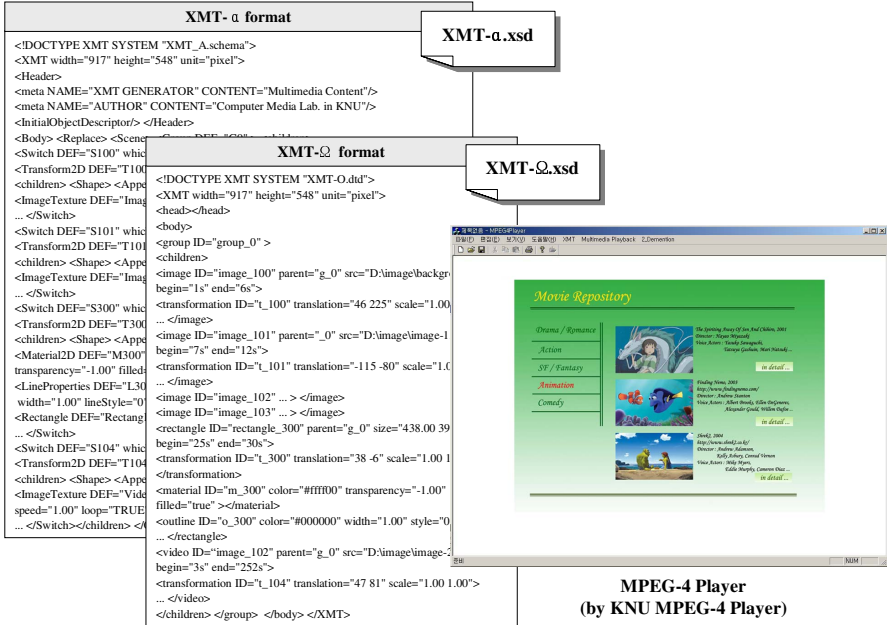

Fig. 4. Two XMT formats and an execution example of MPEG-4 content in the MPEG-4 player

In figure 4, the MPEG-4 content consists of various objects such as .mp4 video format, .gif image format and other geometry formats. The XMT also supports interoperability about XML-based other players such as VRML player or SMIL player. We contain a simple converter in the content analyzer for experiment. This converter [7] is a description conversion tool that can convert scene description from the XMT to other media description languages such as SMIL or VRML.

Figure 5 shows scene description format change processing through the converter and two application examples.

The XMT of MPEG-4 content is changed into VRML or SMIL description through the converter. And, each description is separately played to the VRML player and SMIL player. For SMIL description experiment, we used RealOne player. For VRML description experiment, we used Cortona player.
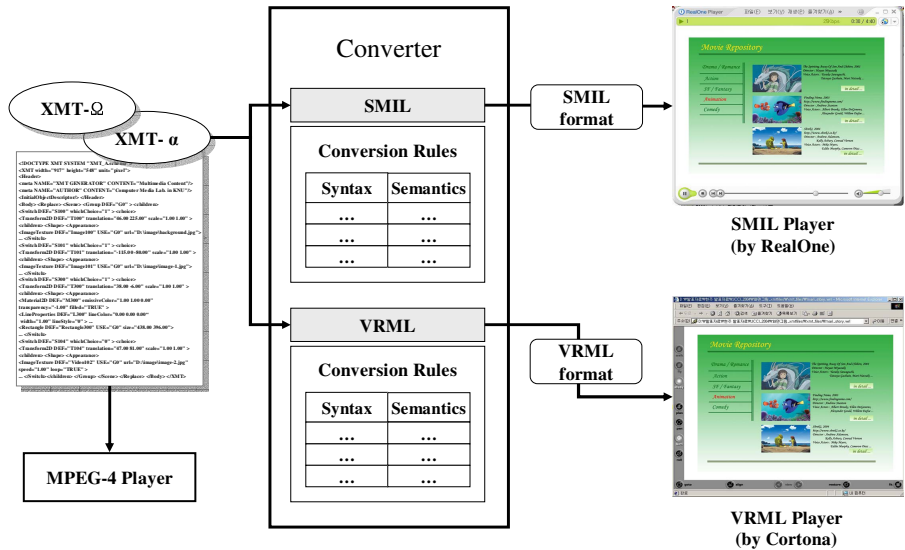
Fig. 5. The scene description format change and two application examples

The XMT format can be played in each player as well as the MPEG-4 player through language conversion. The converter changes XMT file to each description according to conversion rules of SMIL (or VRML). Conversion rule should be defined according to SMIL (or VRML) syntax and semantics. Two converted description formats can be played to each player.

The XMT can change more easily scene description's contents and attributes in user's point of view. If user knows basic syntax of XML, a user can easily change MPEG-4 content's attribute after calling various editors.

3.2 Evaluation Results

To evaluate the content analyzer's capability, the content analyzer is embedded in KNU player and we compared our content analyzer in KNU player with IM1-2D player (ver 5.0). Two players are based on MPEG-4 standard which can play the MPEG-4 content. Specially, the IM1-2D player is an open architecture player. The core comparison item is the number of searches for each tree for comparison of effective scene information management. The content analyzer contains the search engine for effective tree search and the IM1-2D player also contains search module.

The number of nodes of the MPEG-4 content as follows:

O_n: The number of objects which is defined in MPEG-4 content.
C_n: The number of conditional nodes
I_n: The number of interpolator nodes (including Color and Position2D)
R_n: The number of Routes nodes
AT_n: The number of time nodes
S_n: The sum of search of object information

$$S_{n1} = O_n + C_n + (R_n \times 2) + (AT_n \times 2) \tag{1}$$

$$S_{n2} = O_n + C_n \tag{2}$$

S_{n1} is a sum of the number of searches for all object information which contains in MPEG-4 content. It contains the number of objects, the number of conditional nodes, the number of routes nodes and the number of time nodes. Routes information is searched two times for information search because it consists of source object and target object. Time information is also searched two times for its processing because it consists of start time and end time individually.

In S_{n2}, additional tree search for routes and time information is unnecessary because all information which needs to rendering and event processing is stored in the information manager. If information manager is created, it can reduce frequent tree search that is performed whenever scene changes.

Table 1 shows several performance capability of the proposed content analyzer.

Table 1. The support capability of the content analyzer

Requirement	Support Item
Support Objects	Video(MPEG-1, MPEG-4 Video(.mp4), H.263) Audio(MPEG-4 AAC(mono, stereo), G.723) Image(BMP, JPEG, GIF format) 2D Geometry (Rectangle, Circle, Line) Text(English and Korean)
Spatial-temporal interaction	Fully supported
Independent management of each module	Fully Supported
Spatial-temporal composition	2D profiles (simple and complex scenes)
Independence of description	Fully supported
Integration with other players	Supported (with VRML player or SMIL player)
Interoperability of content	Supported (by the Converter)

4   Conclusions

In this paper, we proposed the content analyzer supporting interoperability of MPEG-4 content in heterogeneous players that is optimized for multiple scene description and management of the scene information. It consists of the description separator, the BIFS/XMT engine and the information manager. For multiple scene description support and extensibility of MPEG-4 content, we defined the description separator and the BIFS/XMT engine. For scene information management, we defined the informa-

tion manager. The information manager improves reusability and flexibility by object-based that is MPEG-4 scene's characteristic, and it also reduces the number of tree search. In order to expand module easily and improve player's throughput, each module in content analyzer is independently managed.

In the future, to support sufficiently spatial-temporal relationships of content, we will improve the content analyzer's composition capability, various events processing and playback quality in MPEG-4 player that requires high QoS.

References

1. R. Koenen, "MPEG-4 Overview," ISO/IEC JTC/SC29/WG11/N4668, March 2002, http://www.m4if.org/resources/Overview.pdf
2. A. Puri and T. Chen, Multimedia Systems, Standards, and Networks, Marcel Dekker, New York, USA, March 2000
3. M. Kim, M. Bourges-Sevenier, J. Dufourd, R. Rafey, S. Wood, L. Herrmann, Y. Fisher, Z. Lifshitz and C. Herpel, "Study Text of WD2.0 ISO/IEC 14496-1:2002 – Chapters 14-17," ISO/IEC JTC/SC29/WG11/N4848, May 2002
4. A.E. Walsh and M. Bourges-Sevenier, MPEG-4 Jump-Start, Prentice Hall, Upper Saddle River, NJ, USA, December 2001
5. K. David, K. Padhi, S. George, "Implementation of MPEG 4 systems and delivery layer," Information Technology: Research and Education, Proceedings of the ITRE2003, pp. 27-31, August 2003
6. H. Lee, S. Kim, ""An Adaptive Scene Compositor Model in MPEG-4 Player for Mobile Device" LNCS 3331: Proceedings of 5th pacific Rim Conference on Multimedia, pp. 461-469, Tokyo Japan, November 2004
7. H. Lee, Y. Lim and S. Kim, "An MPEG-4 System supporting XMT Scene Description," Proceedings of the 14th Joint Conference on Communications and Information, pp. II-D-2.1-4, Chungmu, Korea, April 2004