MPEG-4 BIFS Optimization for Interactive T-DMB Content

(지상파 DMB 컨텐츠의 MPEG-4 BIFS 최적화 기법)

차 경 애*
(Kyung-Ae Cha)

Abstract The Digital Multimedia Broadcasting(DMB) system is developed to offer high quality multimedia content to the mobile environment. The system adopts the MPEG-4 standard for the main video, audio and other media format. For providing interactive contents, it also adopts the MPEG-4 scene description that refers to the spatio-temporal specifications and behaviors of individual objects. With more interactive contents, the scene description also needs higher bitrate. However, the bandwidth for allocating meta data, such as scene description, is restrictive in the mobile environment. On one hand, the DMB terminal renders each media stream according to the scene description. Thus the binary format for scene(BIFS) stream corresponding to the scene description should be decoded and parsed in advance when presenting media data. With this reasoning, the transmission delay of the BIFS stream would cause the delay in transmitting whole audio-visual scene presentations,although the audio or video streams are encoded in very low bitrate. This paper presents the effective optimization technique in adapting the BIFS stream into the expected bitrate without any waste in bandwidth and avoiding transmission delays inthe initial scene description for interactive DMB content.

Key Words : Description Optimization, DMB, BIFS, Interactive Content, MPEG-4

1. INTRODUCTION

In this paper, effective techniques in adapting the encoded MPEG-4 scene description (called BInary Format for Scene) into suitable bitrate of the DMB transmission packet size with minimum bandwidth waste are presented. The scene descriptions for the same visible scene and scenario may differ in object ID assignments, node hierarchy organizations, etc., depending on the scene description authors' intentions. So a scene description can be optimized to adopt its encoded bit size into the given resource constraints. The research focuses on developing an efficient scene description optimization technique that makes it possible to match the expected bit size with the encoded bit stream of the modified scene description without any major loss in the original scene.

Section 2 presents the overview of the scene description for interactive DMB multimedia contents. Section 3 explains the optimization of the proposed method, while section 4 presents the experimental results. In Section 5, conclusions and further research direction are

* 대구대학교 정보통신공학부

discussed.

2. Scene Description for T-DMB contents

The DMB standard dopts MPEG-4 system's object-based coding scheme for audio-visual scene[1-4]. The scene consists of multiple objects, each of which can be of any type, such as video, audio, text, graphics, etc. as well as scene description refers to the spatio-temporal composition of these objects following the MPEG-4 Core 2D Profile[3,5-8].

Figure 1 shows an example of the portion of a scene description text, which depicts a rectangular object as geometry.

```
1) Group {
2) children [
3) DEF Switch3000 Switch {
    4) whichChoice 1
    5) choice [
    6) DEF Transform2D3000 Transform2D {
        7) translation -187.00 142.00
        8) scale 1.00 1.00
        9) rotationAngle 0.00
         10) children [
         11) Shape {
              12) appearance Appearance {
              13) material DEF Material2D3000 Material2D
                        {
              14) emissiveColor 0.75 0.75 0.75
              15) filled TRUE
              16) transparency -1.00
              17) LineProperties {
                   18) lineColor 0.00 0.00 0.00
                   19) width 1.00
                   20) lineStyle 0
                   }
         }
              21) geometry Rectangle {
                   22) size 100.00 50.00
```

<Fig. 1> An example of scene description text

The geometry node (line number 21) identifies a rectangular object with a width of 100 pixels and a height of 50 pixels. The attributes of the size of the rectangular node are mandatory to render the corresponding rectangular object when the scene is presented. The LineProperties node (line number 17) is used for describing the linear strokes of the geometry object in its parent Material2D node. When the scene description does not define the node, the presentation draws the real rectangular object with the default value as its line properties. Thus, there is no difference between the original scene and the optimized scene, which removed the LineProperties node when the node describes with default values.

The scene description text in Figure 1 would be optimized as shown in Figure 2, where the rectangular object can be presented in the same dimension drawing with the default values for its line stroke and filling attributes.

```
1) Group {
  2) children [
     6) DEF Transform2D3000 Transform2D {
         7) translation -187.00 142.00
           10) children [
           11) Shape {
               12) appearance Appearance {
               13) material DEF Material2D3000 Material2D
                          {
               15) filled TRUE
               21) geometry Rectangle {
                   22) size 100.00 50.00
```

<Fig. 2> An optimized of scene description text for Fig.1

With this approach of optimization, the removable attributes of the MPEG-4 scene description syntax are categorized. The categorized node type and their field types of the description, based on the characteristics of individual objects, are the criterion from which optimization is made.

3. Optimization of Scene Description for Interacrive T-DMB Contents

In this section, the optimization process of the scene description for adapting available bitrate is explained in detail.

3.1 BIFS(BInary Format For Scene description) Parsing

The BIFS parsing is the main part of the proposed system. The modules use the nested loop approach to analyze the scene description text by reading the whole input scene description text.

As mentioned above, object nodes in the scene description are composed of various fields and their assigned values. We categorize three kindsof tokens following with their characteristics.

Reserved Simple Tokens: Simple tokens are already reserved strings in the node parsing table(NODE table) that is constructed on the pre-process stage for scene description parsing. For example, the followings are included this category: "children," "choice," and so on. These tokens should be matched one of the node parsing table elements.

Related Tokens: Related tokens should have their actual attribute values so that they can determine the visible attributes of corresponding objects. For example, "emissiveColor," "lineStyle," "size," "radius," "URL," etc. belong to this category. These are also parsed with their assigned values by checking their following numeric, string or boolean type variables.

Combination Tokens: The combination tokens are composed of more than two different tokens that are associated by the dot (.) notation. In this case, each sub token is considered as a Related Token. These combination tokens are used in the condition node, rout nodes, and time related nodes.

The BIFS parsing module gets an original BIFS text and reads strings line by line through the scene description. According to the criterion for splitting tokens such as backspace, parentheses or square brackets, the parser recognizes individual tokens and then compares the node table entries and the tokens. The result from token matching process, tokens are classified one of the three types.

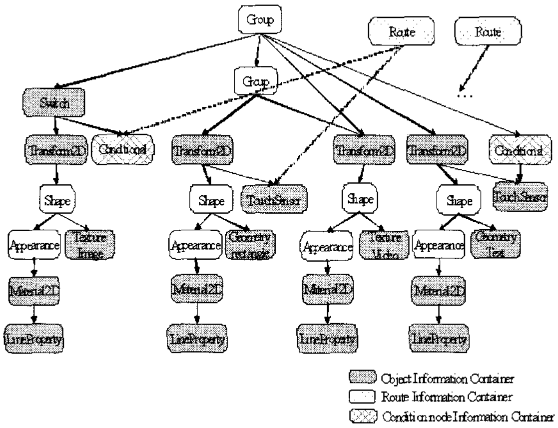
In order to detect the interactive scenario of the scene description as well as the spatio-temporal relationship of the objects, the following data structures are defined. The tokens and their related information such as field values are stored in the containers as following their features.

Multimedia object information container: Objects' attribute information and their related sensor information which are used to realize the animated scenario of objects are written in this container. The actual assigned values of Transform2D, Material2D, LineProperty, Touch Sensor, and other link nodes are in this container.

Rout information container: This has the assigned values of nodes that describe user interactive events and responsible actions. These are source object IDs for route node, action command, and their corresponding destination object ID.

Condition information container: The container is composed of action event types of condition nodes, such as their node ID, action command, destination node ID, and field values to represent the corresponding results of the action.

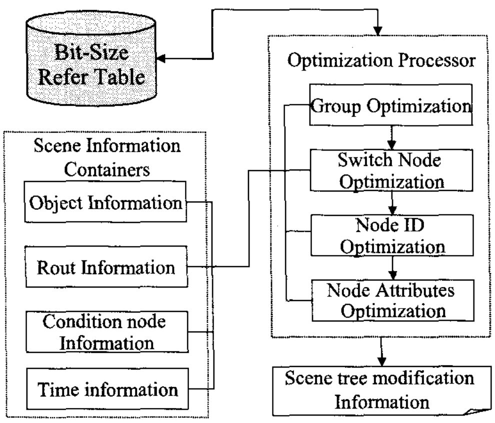
Figure 3 depicts the relationship between the containersand scene tree attributes. The object nodes and fields are constituted hierarchally so the scene description parsing result can be represented as tree structure. However, the optimization process doesn't need the whole information of the scene tree. Thus the selective values are stored in the related containers as described above. Moreover, information containers provide the prospective view of the scene organization to the optimization process.



<Fig. 3> Scene Tree and related Containers

3.2 Optimization of Scene Description

The optimal selection process of the objects in an MPEG-4 scene is performed in multiple steps. First, a set of attribute nodes, which affects the visual point of the original scene at a lesser degree, is removed in the sense that the modified scene is under the condition that the sum of the bit rates does not exceed the available bit size. If this condition cannot be met, then the next step removes a set of attributes that can be represented by using some default values without any object dropped under the same constraint. Figure 4 and the following expiations represent the optimization steps.



<Fig. 4> Optimization processing module

The Optimization Processor detects and removes the information from the original scene description, the group nodes and switchnodes in turn. And then, IDs of object nodes, rout nodes and conditional nodes are renamed in order to save bits for scene description. Finally, some attribute of object nodes are removed if they are not necessary. These processes are performed with information from containers and Bit-Size Refer Table.

Group Optimization: There are many BIFS visual nodes to manage scenes different ways in grouping and rendering. In scene description, the group concept is used to handle a set of objects more effectively and convenientlyon the aspect of authoring. In rendering aspects, the grouping objects do not affect the rendering, thus, grouping information can be removed from the scene description. In this step, it should be checked whether the description contains a sub-group or group ID, and, if contained, it would be deleted from the scene description text.

Switch node filtering: Different from other nodes, the switch node can be deleted from the scene description text even though it has its children nodes, and if the switch node is notused in other condition nodes or time related nodes. If there is a switch node on an object node, this step checks whether the switch node ID is used in the condition node or in the time node. When it is determined that the ID has not been used again, it removes the switch node.

Node ID Optimization: During authoring time, authors can declare their own intentions on the object node IDs, thus every scene description text may have different node ID assignments. Every node ID uses up 1 byte per character, thus it is efficient in reducing the bit size of BIFS to give all node IDs with shorter strings. Based on the rule that the first letter should start as a letter of thealphabet, we give two cipers ID per node, which is composed of one leading alphabet letter and one succeeding digit number. The letter identifies each object, and the digit number identifies the attributes node of the object.

Node Attribute Optimization: As in Figure 1, some node or fields with default values can be removed. This process detects this kind of attributes from the object information container and removes the related nodes or fields. At this stage, the removal ensures that low level attributes are deleted first. For example, material2D attribute node cannot be removed as long as its line property node exists.

4 . Experiment Results

The proposed system is implemented by using Microsoft Visual C++ under the Windows platform.

The sample contents used in the experiment are composed of various objects and its scene description describes several sub-groups and sensor nodes for interactive scenarios, as seen in Table 1.

Each content is composed of one video and one audio object as main audio-visual content.

<Table 1> Sample content specifications

Content No.	Content No.1	Content No.2	Content No.3
Original BIFS Size(bytes)	3058	4069	2724
Number of Group	8	10	8
Multimedia Object Information	Video: 1	Video: 1	Video: 1
	Audio: 1	Audio: 1	Audio: 1
	Image: 7	Image: 9	Image: 7
	Geometry : 12	Geometry : 16	Geometry : 10
	Total: 21	Total: 27	Total: 19
Number of Sensor	13	18	11

Several image and geometry object that are used to depict menu buttons and hot-spots are specified in the scene description. In addition, the interactive scenario also described using sensor nodes and rout nodes. For example, if user clicks a certain image object(button form) while the presentation, the corresponding text object which shows a products' information appears in the scene.

The scene descriptions of the sample contents are reconstructed by following the optimization process. At each step, the optimized scene description's bitrate are checked as seen in Table 2. Although the descriptions can be specified as different constitution types, samples followed the generally published scene description specification ways. As a result of the proposed optimization technique, the scene descriptions are reduced in terms of their encoded bit size down to 53.2 % at the most, compared to the original descriptions.

<Table 2> The optimization results of the sample descriptions

Content No.	Content No.1	Content No.2	Content No.3
Original Size	3058 bytes (100%)	4069 bytes (100%)	2724 bytes (100%)
Group Optimization	3001 bytes (98.1%)	4000 bytes (98.3%)	2667 bytes (97.9%)
Node ID Optimization	2196 bytes (71.8%)	2925 bytes (71.9%)	1961 bytes (72%)
Node Optimization	1664 bytes (54.4%)	2251 bytes (55.3%)	1450 bytes (53.2%)

Figure 5 shows a portion of the scene description text for content No.1. Figure 5(a) represents the original description while (b) shows the optimized description for an audio object and image object. As we can see, the fields that are assigned with default values and orderedGroup node are removed. Moreover the node IDs are re-assigned with strings in two ciphers.

```
Sound2D {
source AudioSource {
pitch 1.0000000
speed 1.0000000
startTime 0.0000000
stopTime -1.0000000
url 4  }  } ]  }
DEF L1 OrderedGroup {
children [
DEF I3_SWT Switch {whichChoice 0
choice [
 DEF I3_T2D Transform2D {
  scale 1.0000000 1.0000000
  translation -122.0000000 -92.0000000
  children [
  Shape {
  geometry Bitmap {
      scale 1.0000000 1.0000000  }
   appearance Appearance {
   material DEF I3_M2D Material2D {
    emissiveColor      0.0000000      0.0000000
0.0000000
    filled true
    transparency 0.0000000  }
texture ImageTexture { url 5
```

(a)

```
Sound2D {
        source AudioSource {
                        url 4
                        }              }
DEF B0 Switch {
        whichChoice 0
        choice [
        DEF B1 Transform2D {
              translation -122.00 -92.00
            scale 1.00 1.00
                children [
                Shape {
        appearance Appearance {
        material DEF B2 Material2D {
        emissiveColor 0.00 0.00 0.00
                filled TRUE
                transparency 0.00
        }
        texture ImageTexture {
                url 5
        }
}
```

(b)

<Fig. 5> A portion of Scene Description text for content No.1 :
(a) original scene description
(b) optimized scene description

5. Conclusion

In this paper, the effective algorithm to the TS bit size of the payload adaptive technique for the scene description of the interactive DMB content is proposed.

The features of the object nodes and their encoded bit size when the scene description text is transformed to the binary format were studied. The adaptive tool, based on optimization rules, was also designed and implemented. The research showed that the proposed optimization method reduced the scene description bitrate efficiently without losses in the original information. Basically, optimization should proceed in the scope of not changing the original scene. But if the size of the bitstream becomes bigger than expected after optimization, inevitable losses through optimization should be processed. Thus, when there is a difference between the optimized scene and the original, the scene should be updated in real time by using BIFS commands.

참 고 문 헌

[1] Lee, S. Cho, K. Yang, Y. Hahm and S. Lee, "Development of Terrestrial DMB Transmission System based on Eureka-147 DAB System." IEEE Transactions on Consumer Electronics, Volume 51, Issue 1, pp.63 - 68, Feb. 2005.

[2] V. Ha, S. Choi, J. Jeon, G. Lee, W. Jang and W. Shim, "Real-time Audio/Video Decoders for Digital Multimedia Broadcasting," Proceedings of the 4th International Workshop on System-on-Chip for Real-Time Applications, 2004.

[3] The basic interface between the transmitter and the receiver for video service for Terrestrial digital multimedia broadcasting in the VHF

band, Telecommunications Technology Association, 2005.
[4] J. Signes, Y. Fisher, and A. Eleftheriadis, "MPEG-4's binary format for scene description,"Signal Processing: Image Communication, vol.15, issues 4-5, pp. 321-345, January 2000.
[5] Olivier Avaro, Alexandros Eleftheriadis, "MPEG-4 Systems: Overview" Signal Processing: Image Communication 15 (2000), pp. 281 - 298.
[6] WG11 (MPEG) MPEG-4 Overview document, ISO/IEC JTC1/SC29/WG11 N4668, March 2002.
[7] A. Puri and A. Eleftheriadis, "MPEG-4: An Object-Based Multimedia Coding Standard Supporting Mobile Applications," Mobile Networks and Applications, vol. 3, pp.5-32, 1998.
[8] C. Herpel and A. Eleftheriadis, "MPEG-4 Systems: elementary stream management," Signal Processing: Image Communication, vol. 15 no. 4-5, pp. 299-320, January 2000.

차 경 애 (Kyung-Ae Cha)

- 정회원
- 1996년 2월 : 경북대학교 컴퓨터과학과 (이학사)
- 1999년 2월 : 경북대학교 컴퓨터과학과 (이학석사)
- 2003년 8월 : 경북대학교 컴퓨터과학과 (이학박사)
- 2001년 3월 ~ 2004년 2월 : 경북대학교 초빙교수
- 2004년 3월 ~ 2005년 2월 : 한국정보통신대학교 연구교수
- 2005년 3월 ~ 현재 : 대구대학교 정보통신공학부 교수
- 관심분야 : 멀티미디어시스템, 디지털방송, 컨텐트 저작 및 스트리밍