

특집논문 (Special Paper)

방송공학회논문지 제21권 제3호, 2016년 5월 (JBE Vol. 21, No. 3, May 2016)

<http://dx.doi.org/10.5909/JBE.2016.21.3.330>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 다중 감각 미디어 응용 포맷의 구성 방법 연구

오 정 엽<sup>a)</sup>, 김 상 균<sup>b)†</sup>

### A Study on Multiple Sensorial Media Application Format

Jung Yup Oh<sup>a)</sup> and Sang-Kyun Kim<sup>b)†</sup>

#### 요 약

본 논문은 MPEG-A(ISO/IEC 23000) 표준으로 새로이 제정된 다중 감각 미디어 응용 포맷의 구조와 구현에 대해 설명한다. 다중 감각 미디어 응용 포맷은 비디오/오디오 데이터와 더불어 감각 효과 메타데이터의 효과적인 저장, 재생, 관리를 용이하게 한다. 포맷을 구성하기 위해, MPEG-4 파트 12(ISO/IEC 14496-12)의 ISO 기본미디어 파일 포맷과 MPEG-V 파트 3(ISO/IEC 23005-3)의 표준 메타데이터를 사용하였다. 본 논문에서는 MPEG-V 파트 3의 감각효과 메타데이터(SEM)의 샘플화를 위해 SEM XML 문서의 분할과, 분할된 샘플을 이진화하는 방법에 대해 비교 분석한다. 이진화 방법에 대한 비교 분석 결과 MPEG-V 이진표현 방법과 Binary MPEG format for XML(BiM)이 gzip방법에 비해 압축율과 처리시간 등에서 우수한 성능을 보였다.

#### Abstract

This paper explains about the structure of multiple sensorial media application format (ISO/IEC 23000-17), which is newly standardized as a project of MPEG-A. This format facilitates effective storage, playing, and management of media with multiple sensorial effects. The ISO base media file format from MPEG-4 Part 12 and sensory effect metadata (SEM) from MPEG-V Part 3 are used to composed the multiple sensorial media application format. In this paper, a fragmentation method to break a SEM XML document into valid SEM samples is presented. Several binarization methods to compress the SEM samples are compared and evaluated as well. The compression ratio and processing time using the MPEG-V binary representation and the Binary MPEG format for XML (BiM) are superior to the gzip compression.

Keyword : Sensorial effects, MulSeMedia, MPEG-V, MPEG-A, ISO base media file format

a) 한국정보통신기술협회 표준기획부(Telecommunications Technology Association)

b) 명지대학교 컴퓨터공학과(Myongji Univ. Computer Eng. Dept.)

† Corresponding Author : 김상균(Kim, Sang-Kyun)

E-mail: goldmunt@gmail.com

Tel: +82-31-330-6443

ORCID:<http://orcid.org/0000-0002-2359-8709>

※ 이 논문의 연구결과 중 일부는 "IWAIT 2016"에서 발표한 바 있음.

※ 본 연구는 산업통상자원부 국가기술표준원에서 지원하는 국가표준기술력향상사업(과제번호: 10053655)의 일환으로 수행되었음.

· Manuscript received March 21, 2016. Revised May 24, 2016; Accepted May 24, 2016.

## 1. 서론

최근 IMAX나 4D 영화를 비롯하여 고해상도 HD 방송 및 3D 방송과 같이 영상 콘텐츠 관련 산업 분야에서 좀 더 현실감 있는 영상을 제공하는 방법에 대한 관심이 높아지고 있다. 사용자의 콘텐츠 몰입도 향상을 위해서는 시각, 청각적 일체감뿐만 아니라 후각, 촉각과 같은 여타 감각에 의 자극을 통한 일체감의 극대화가 필요하다. 이러한 사용

자 체험의 실감형 콘텐츠를 위해서 2008년부터 시작된 국제 표준으로 ISO/IEC 23005(이하 MPEG-V)가 있다. 이 중 파트 3(Sensory information)는 바람, 진동, 조명들과 같은 감각효과들을 서술하기 위한 기본 요소들과 실제 사용되는 감각효과들을 표현할 수 있는 XML 스키마인 Sensory Effect Metadata (이하 SEM)를 정의하고 있다.

하지만 SEM이 산업에서 쉽게 이용되기에는 많은 문제점이 있다. 기존의 SEM을 이용하는 방법은 영상콘텐츠에 맞게 저작한 감각효과 데이터를 영상콘텐츠와 함께 4D 미디어 플레이어로 재생하는 것이다. 이는 감각효과를 서술하는 XML file과 영상콘텐츠를 분리하여 관리하기 때문에 하나의 미디어를 저장, 교환, 관리, 재생하는 것이 어렵고 복잡하다. 이에 MPEG 표준화 그룹에서는 서로 다른 여러 표준을 조합하여 특정 응용 서비스를 목적으로 하는 표준을 새롭게 제정하고 있다. 이러한 움직임의 하나가 ISO/IEC 23000 (이하 MPEG-A) 멀티미디어 응용 포맷 표준화 활동이다<sup>[1]</sup>. MPEG-A는 Multimedia Application Format (이하 MAF)이라는 새로운 파일 포맷을 정의하는 표준이다. MPEG-A는 모두 13개의 파트로 구성되어 있다. 이 중 영상콘텐츠에 관련된 MAF는 파트 8의 Portable Video MAF (ISO/IEC 23000-8), 파트 9의 Digital Multimedia Broadcasting MAF ((ISO/IEC 23000-9)), 파트 11의 Stereoscopic Video MAF (ISO/IEC 23000-11)가 있다. 이들 MAF는 여러 개의 비디오와 관련 텍스트, 이미지 데이터를 동시에 저장하거나, DMB, 스테레오스코픽 비디오와 관련된 콘텐츠의 생성, 저장, 전송, 재생 등을 가능케 하거나, 3D 비디오와 대화형 서비스 데이터를 제공한다. 그러나 현재 제정된 MAF 중 어느 것도 감각효과 데이터를 비디오와 함께 저장할 수 있는 포맷을 지원하지 않는다. 본 논문은 다중 감각 미디어를 DVD와 같은 저장 플랫폼뿐만 아니라 방송, 인터넷 같은 스트리밍 플랫폼에도 쉽게 적용 할 수 있는 응용 포맷을 제안한다. 이를 위해 다중 감각 미디어 응용 포맷의 구조를 정의한다. 또한 응용 포맷의 인코딩 과정에서 필요한 XML의 여러 분할 방법 및 이진화 방법을 비교한다. 마지막으로, 제안 방법을 이용한 다중 감각 미디어 응용 포맷의 패커(packer)와 언패커(unpacker)의 구현에

대해 설명한다.

본 논문은 다음과 같은 구조를 가진다. II장은 SEM 데이터를 XML 문서로 표현하였을 때 이를 감각효과 단위에 따라 적법하게 분할하여 프로세스 유닛으로 만드는 방법을 설명한다. III장은 분할된 프로세스 유닛을 이진화하여 액세스 유닛으로 만드는 방법에 대해 설명한다. IV장은 분할된 SEM 데이터를 저장하기 위한 다중감각미디어 응용포맷의 구조에 대해 설명하고, V장은 III장에서 제시되었던 이진화 방법에 대해 비교 및 분석한 결과를 제시한다.

## II. SEM XML 분할

다중감각 미디어의 컨테이너 파일포맷으로는 ISO 기본 미디어 파일 포맷 (ISO base media file format)을 사용한다. ISO 기본미디어 파일 포맷의 미디어 데이터 박스는 비디오, 오디오, 메타데이터 샘플들로 구성된다. SEM도 미디어 데이터 박스에 넣기 위해서는 샘플단위로 만들어야 한다. 감각효과를 정의하고 있는 MPEG-V Part 3 (ISO/IEC 23005-3) 내에 SEM을 서술하는 주요 엘리먼트는 Effect 엘리먼트, GroupOfEffects 엘리먼트, ReferenceEffect 엘리먼트 등이 있는데, 간단하게는 이를 시간에 따라 잘라 내어 샘플을 만드는 방법을 생각 할 수 있다. 하지만 이렇게 시간에 따라 잘라낸 XML 문서는 유효(validate)하지 않아 디코더가 해석할 수 없는 문제가 발생한다. 본 장에서는 XML Streaming Instruction을 이용하여 유효성을 유지하면서 SEM XML 문서를 분할하는 방법에 대해 설명한다. 표 1은 응용포맷 내 저장되어야 하는 SEM XML 문서를 예시한다. 표 1의 XML 문서는 1개의 온도효과와 온도, 바람, 진동 효과를 동시에 포함하는 1개의 폭발 효과(그룹 효과)로 구성되어 있다. 이는 ISO/IEC 23005-3 (MPEG-V Part 3) 문서 내 감각효과 메타데이터 서술 표준 방식을 이용하여 저작되었다.

MPEG에서는 XML 문서를 분할하는 방법으로 MPEG-21 파트 7(ISO/IEC 21000-7) XML Streaming Instruction (이하 XSI)이 있다. MPEG-V는 XSI를 사용하여 SEM문서

표 1. 분할될 소스문서

Table 1. A source SEM XML document to be fragmented

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<SEM
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="urn:mpeg:mpeg-v:2010:01-SEDL-NS"
xmlns:sev="urn:mpeg:mpeg-v:2010:01-SEV-NS"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2004"
xmlns:ct="urn:mpeg:mpeg-v:2010:01-CT-NS"
xmlns:si="urn:mpeg:mpeg21:2003:01-DIA-XSI-NS"
xsi:schemaLocation="urn:mpeg:mpeg-v:2010:01-SEV-NS MPEG-V-SEV.xsd"
si:puMode="ancestorsDescendants" si:anchorElement="true" si:timeScale="1000">
<Effect xsi:type="sev:TemperatureType" si:anchorElement="true" intensity-range="0.0 100.0" intensity-value="40.0" duration="5000" activate="true" si:pts="6000" si:ptsDelta="0"/>
<GroupOfEffects si:puMode="ancestorsDescendants" si:anchorElement="true" id="explosion_effect" si:pts="20000">
  <Effect xsi:type="sev:TemperatureType" intensity-range="0.0 100.0" intensity-value="40.0" duration="5000" activate="true" si:ptsDelta="0"/>
  <Effect xsi:type="sev:WindType" intensity-range="0.0 100.0" intensity-value="50.0" duration="5000" activate="true" si:ptsDelta="0"/>
  <Effect xsi:type="sev:VibrationType" intensity-range="0.0 100.0" intensity-value="60.0" duration="5000" activate="true" si:ptsDelta="0"/>
</GroupOfEffects>
</SEM>

```

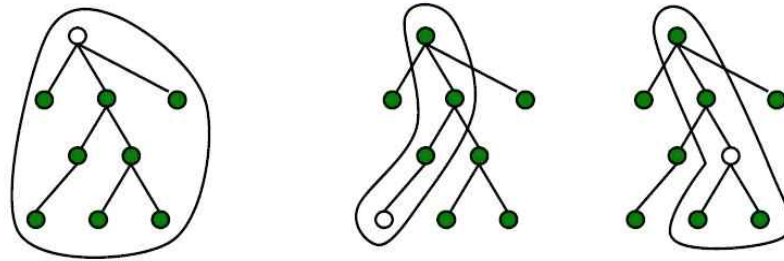


그림 1. process unit mode가 ancestorsDescendants일 경우의 분할 예  
 Fig. 1. A example case of a process unit mode - ancestorsDescendants

를 분할하고 시간 정보를 할당한다. MPEG-V 파트 3의 감각효과 메타데이터는 루트 엘리먼트, Group-OfEffects 엘리먼트, Effect 엘리먼트, ReferenceEffect 엘리먼트는 si-AttributeList 라는 속성을 갖고 siAttributeList 속성의 타입이 XSI다. XML 문서를 프로세스 유닛으로 나누는 방법은 XSI의 anchor element 속성과 process unit mode 속성으로 결정되며 프로세스 유닛의 임의 접근성 여부는 encodes-AsRAP속성으로 결정 된다.

process unit mode의 종류는 총 7개이며 이 중 ancestorsDescendants인 경우, anchor 엘리먼트와 그것의 부모

엘리먼트, 자식 엘리먼트를 모두 포함하여 분할된다. 그림 1은 process unit mode가 ancestorsDescendants인 경우의 분할 예이며 흰색 엘리먼트는 anchor 엘리먼트를 나타낸다.

Fragmenter는 XML 문서의 XSI 속성값을 이용하여 여러 개로 프로세스 유닛으로 분할하는 모듈을 말한다. 일반적인 Fragmenter는 깊이 우선으로 XML 문서를 파싱하며 anchorElement를 만나면 새로운 프로세스 유닛으로 만들 준비를 한다. 이후 puMode 속성값에 따라 분할할 범위를 정하여 새로운 프로세스 유닛으로 분할한다. 그림 2는 일반적인 Fragmenter의 역할을 나타낸다.

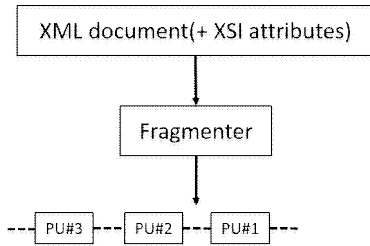


그림 2. 일반적인 fragmenter의 기능  
 Fig. 2. Normative functions of a fragmenter

표 2는 표 1을 XSI Fragmenter를 사용하여 분할한 결과이다. 표 1은 SEM 루트 엘리먼트 1개, Effect 엘리먼트 1개, GroupOfEffects 엘리먼트 1개의 앵커 엘리먼트가 있다. 이들은 Fragmenter를 통해 2개의 프로세스 유닛으로 분할된다. SEM 루트 엘리먼트의 XSI 속성인 puMode와 timeScale은 자식 앵커 엘리먼트들로 상속된다. SEM 루트 엘리먼트의 timeScale 속성과 각 앵커 엘리먼트의 pts 속성으로 프로세스 유닛의 시간 정보가 된다.

표 2. XSI fragmenter로 표1의 소스문서를 분할한 결과

Table 2. Process units of the XML document in Table 1 generated by the XSI fragmenter

PU #1
<pre> &lt;?xml version="1.0" encoding="UTF-8" standalone="yes"?&gt; &lt;SEM xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="urn:mpeg:mpeg-v:2010:01-SEDL-NS" xmlns:sev="urn:mpeg:mpeg-v:2010:01-SEV-NS" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2004" xmlns:ct="urn:mpeg:mpeg-v:2010:01-CT-NS" xmlns:si="urn:mpeg:mpeg21:2003:01-DIA-XSI-NS" xsi:schemaLocation="urn:mpeg:mpeg-v:2010:01-SEV-NS MPEG-V-SEV.xsd" si:puMode="ancestorsDescendants" si:anchorElement="true" si:timeScale="1000"&gt; &lt;Effect xsi:type="sev:TemperatureType" si:anchorElement="true" intensity-range="0.0 100.0" intensity-value="40.0" duration="5000" activate="true" si:pts="6000" si:ptsDelta="0"/&gt; &lt;/SEM&gt;                     </pre>
PU #2
<pre> &lt;?xml version="1.0" encoding="UTF-8" standalone="yes"?&gt; &lt;SEM xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="urn:mpeg:mpeg-v:2010:01-SEDL-NS" xmlns:sev="urn:mpeg:mpeg-v:2010:01-SEV-NS" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2004" xmlns:ct="urn:mpeg:mpeg-v:2010:01-CT-NS" xmlns:si="urn:mpeg:mpeg21:2003:01-DIA-XSI-NS" xsi:schemaLocation="urn:mpeg:mpeg-v:2010:01-SEV-NS MPEG-V-SEV.xsd" si:puMode="ancestorsDescendants" si:anchorElement="true" si:timeScale="1000"&gt; &lt;GroupOfEffects si:puMode="ancestorsDescendants" si:anchorElement="true" id="explosion_effect" si:pts="20000"&gt;   &lt;Effect xsi:type="sev:TemperatureType" intensity-range="0.0 100.0" intensity-value="40.0" duration="5000" activate="true" si:ptsDelta="0"/&gt;   &lt;Effect xsi:type="sev:WindType" intensity-range="0.0 100.0" intensity-value="50.0" duration="5000" activate="true" si:ptsDelta="0"/&gt;   &lt;Effect xsi:type="sev:VibrationType" intensity-range="0.0 100.0" intensity-value="60.0" duration="5000" activate="true" si:ptsDelta="0"/&gt; &lt;/GroupOfEffects&gt; &lt;/SEM&gt;                     </pre>

### III. SEM 샘플의 이진화

프로세스 유닛을 특정 법으로 인코딩하여 압축 혹은 패킷화(packetizing)한 형태를 액세스 유닛(access unit)이라고 한다. SEM 액세스 유닛 또한 이진화 할 필요가 있다. 모션 제어에 주로 쓰이는 Rigid body motion effect의 경우 1초에 수십 프레임의 움직임 정보를 갖고, 해당 XML문서의 크기는 기존 XML문서에 비해 상당히 커진다. 이러한

데이터를 전송하기 위해서는 효율적인 이진화 방법에 대한 고민이 필요하다. 본 장에서는 여러 이진화 방법 중 다중 감각효과 미디어에 가장 적합한 방법을 알아본다.

#### 1. 스키마 기반 이진화

Binary MPEG format for XML(이하 BiM)은 MPEG-7과 MPEG-B 국제 표준의 일부로 표준화한 XML 이진화 방식

표 3. 다중 감각 미디어 응용 포맷의 박스 구성

Table 3. Box structures of the multiple sensorial media application format

Mandatory(*)	Box Type					Box description
*	ftyp					file type and compatibility
	pdin					Progressive download Information
	moov					container for all the metadata
*		mvhd				movie header, overall declarations
*		trak				container for an individual track or stream
*			tkhd			track header, overall information about the track
			edts			edit list container
				elst		an edit list
*			mdia			container for the media information in a track
*				mdhd		media header, overall information about the media
				hdlr		handler, declares the media (handler) type "soun" for audio data "vide" for visual data "semh" for SEM
*				minf		media information container
					vmhd	video media header, overall information (video track only)
					smhd	sound media header, overall information(sound track only)
					hmhd	hint media header, overall information (hint track only)
					nmhd	null media header, overall information(some tracks only)
*					dinf	data information box, container
*					dref	data reference box, declares source(s) of media data in track
*					stbl	sample table box, container for the time/space map
*					stsd	sample descriptions (codec types, initialization etc.) "mpms" for SEM data
*					stts	(decoding) time-to-sample
					ctts	(composition) time to sample
*					stsc	sample-to-chunk, partial data-offset information
*					stsz	sample sizes (framing)
					stz2	compact sample sizes (framing)
*					stco	chunk offset, partial data-offset information
					co64	64-bit chunk offset
					stss	sync sample table (random access points)
		ipmc				IPMP Control Box
	mdat					media data container

이다. 원래는 MPEG-7 스키마를 사용한 XML 문서를 이진화하는 표준이었으나 제정 과정에서 XML 스키마를 기반으로 작성된 XML 문서를 범용적으로 이진화할 수 있는 기법으로 확장되어 MPEG-B(ISO/IEC 23001)에 편입되었다.

BiM은 스키마 문서의 정의에 따르는 오토마타를 사용하여 XML 문서를 바이너리 데이터로 이진화한다. BiM의 특징은 XML 스키마의 simple type의 경우에는 텍스트 기반 이진화 방식을 사용하고 complex type의 경우에는 FSA (Finite State Automata)를 사용하는 것에 있다. 예를 들어 일반 텍스트는 gzip을 사용하고 낱짜와 시간에 해당하는 텍스트 스트링은 비트 스트링으로 변환하여 크기를 줄인다. 그러나 엘리먼트가 complex type으로 정의된 경우는 FSA를 사용한다<sup>[2]</sup>.

## 2. 텍스트 기반 이진화

XML 데이터는 텍스트 파일로 저장되기 때문에 텍스트 기반 압축 방법은 XML 문서를 압축하기 위해 많이 쓰이는 방법이다. 그중에서 gzip은 가장 대표적인 텍스트 파일 압축 기법으로 디플레이트(deflate)알고리즘을 사용한다. 디플레이트 알고리즘은 LZ77 알고리즘과 허프만 코딩이 혼합된 형태의 무손실 압축 기법이다. LZ77 알고리즘을 이용하여 문자열의 중복을 제거하고, 중복된 데이터의 이진데이터의 길이를 줄이기 위해 허프만 코딩을 사용한다. 일반적으로 40-50%의 효율을 보이는 것으로 알려져 있다<sup>[3]</sup>.

## 3. 이진 규칙 기반 이진화

이진 규칙 기반 인코딩은 일종의 사전식 이진화의 일종으로 규칙에 맞는 비트 변환을 통해 데이터의 크기를 줄이는 방법이다. XML 문서의 태그들은 많은 경우에 반복되는 패턴들로 이루어진다. 이 경우, 자주 발생하는 태그에 대한 사전 테이블을 구성하여, 태그를 사전 테이블에 부여된 특정 이진코드로 변환함으로써 데이터의 크기를 줄일 수 있다. 즉, 사전 이진화 방법은 새로운 단어 마다 다른 정수를 부여하여 데이터를 변환한다. MPEG-V의 감각효과 메타데이터는 이 사전식 이진화 방법을 통한 압축 규칙을 제공한다. 이를 MPEG-V 이진화 표현(MPEG-V binary representation)이라 표현한다<sup>[4]</sup>.

## IV. 다중 감각 미디어 응용 포맷의 구조

다중 미디어 응용 파일 포맷은 ISO 기본 미디어 파일 포맷을 바탕으로 구성된다. ISO 기본미디어 파일 포맷은 박스 단위로 이루어지며 파일 포맷의 설계 시 필요한 박스를 사용하거나, 제거 혹은 생성할 수 있다. 표 3은 다중 감각 미디어 응용 포맷의 박스 순서를 나타내며 (\*)표시는 필수 박스를 의미한다.

다중 감각 미디어 응용 파일 포맷은 프로세스 유닛 단위로 만든 SEM을 샘플단위로 취급한다. 이는 마치 오디오와 비디오의 프레임처럼 SEM을 취급하는 것이다. 그림3은

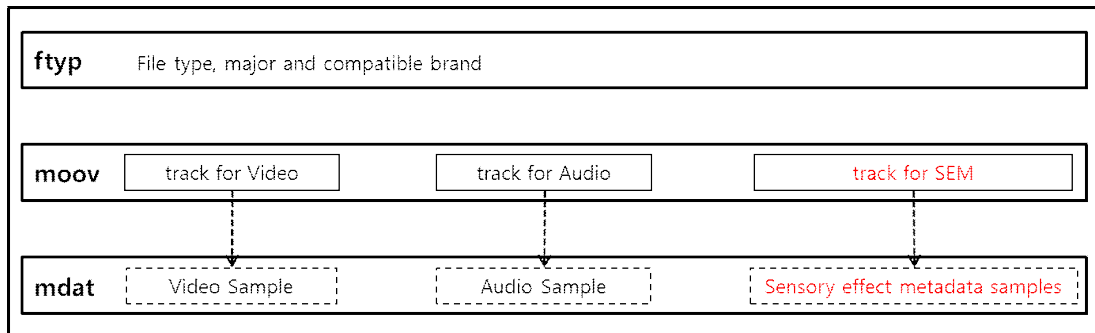


그림 3. SEM 트랙의 SEM 샘플 참조

Fig. 3. A SEM track and its references to the SEM samples

SEM 트랙이 미디어 데이터 박스('mdat')안의 SEM 샘플을 참조하는 모습을 나타낸다.

SEM 트랙은 샘플 디스크립션 박스(stsd), 샘플 디코딩 시간 박스(stts), 샘플 크기 박스(stsz), 샘플 청크 오프셋 박스(stco)를 통해 SEM 샘플의 정보를 트랙 내부에 갖고 있다. 샘플 디스크립션 박스(stsd)는 III장에서 설명한 SEM 샘플의 이진화 방법에 대해 기술한다. 샘플 디코딩 시간 박스(stts)는 SEM 샘플의 시간 정보인 pts를 담고 있다. 샘플 크기 박스(stsz)는 SEM 샘플의 크기를 담고 있으며 크기는 이진화 방법에 따라 달라진다. 샘플 청크 오프셋 박스(stco)는 파일의 시작점부터 SEM 샘플까지의 오프셋을 바이트 단위로 기록한다. 이러한 SEM 트랙의 정보를 사용하면 오디오, 비디오 데이터의 싱크에 맞춰 해당 시간에 디코딩 되어야 할 SEM 샘플을 미디어 데이터 박스(mdat)에서 찾을 수 있다.

## V. 다중 감각효과 샘플의 이진화 성능 실험

본 장에서는 이진화 실험을 통해 SEM 샘플에 적합한 이

진화 방법을 알아본다. 감각효과 메타데이터를 XML분할을 이용하여 SEM 샘플로 만든 후 III장에서 제시한 BiM, gzip, MPEG-V 기반 이진화 방법으로 이진화 하였다. 본 실험은 5,000개의 이펙트 엘리먼트로 구성된 예제와 Rigid body [4]효과로 구성된 예제로 진행하였다.

이진화 실험 예제1은 5천개의 이펙트 엘리먼트로 구성된 예제로, 1,053,068 바이트 크기의 XML 문서이다. 예제를 Effect 엘리먼트 단위로 분할하고 5천개의 SEM 샘플로 만든 후 이진화하여 압축률 비교 실험을 진행하였다. 이는 II장에서 설명된 다중 감각효과 응용 포맷의 샘플 생성 과정을 따른 것이다.

이진화 실험 예제 2는 Rigid Body 이펙트 엘리먼트로 구성된 23,156 바이트 크기의 XML 문서이다. 한 개의 이펙트 엘리먼트로 구성되었기 때문에 SEM 분할 없이 압축률 비교 실험을 진행하였다. 두 개의 예제 인스턴스 문서를 스키마 기반 이진화인 BiM, 텍스트 기반 이진화인 gzip, 이진 규칙 기반 이진화인 MPEG-V 이진 표현 방법으로 이진화하여 압축률을 비교하였다. 표 4는 실험에 이용된 실험 환경을 설명한다.

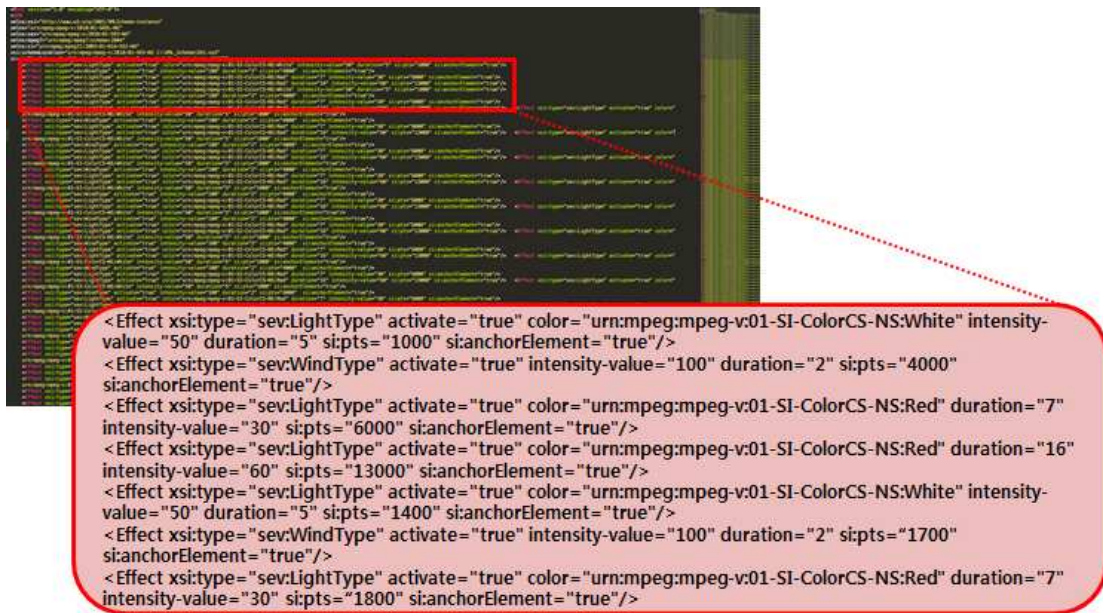


그림 4. (이진화 실험 예제1) 5,000개의 이펙트 엘리먼트로 구성된 SEM 인스턴스  
 Fig. 4. A SEM instance composed of 5,000 effect elements

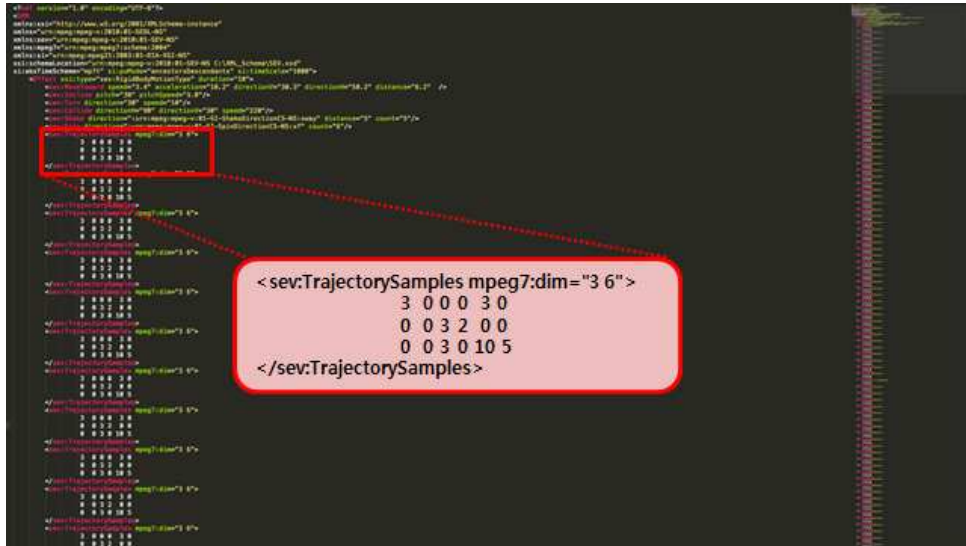


그림 5. (이진화 실험 예제 2) 한 개의 긴 이펙트 엘리먼트로 구성된 SEM 인스턴스  
 Fig. 5. A rigid body SEM instance composed of one huge effect element

표 4. 다중 감각효과 샘플 이진화를 실험 환경  
 Table 4. Test environment

Experimental setup	
CPU	2.4 GHz Intel Core i5
OS	Windows 7 Enterprise edition
IDE	Visual Studio 2010 Service Pack 1

BiM의 이진화 실험을 위해서 MPEG-7 reference software를 사용하였으며, gzip을 구현하기 위해 zlib을 사용하였다. MPEG-V 이진 표현의 구현은 국제 표준 문서를 참조하여 c++로 구현하였다. 또한 SEM 분할을 위해서는

MPEG-21 Part 7 reference software DIA-Fragment-A-1 라이브러리를 사용하였다. 인코딩, 디코딩에 사용한 라이브러리의 특징은 표 5와 같다.

MPEG-V 이진 규칙 기반 인코딩/디코딩의 경우 본 실험에서는 직접 구현한 소프트웨어를 사용하였지만 JAVA로 구현된 MPEG-V reference software를 사용할 수도 있다. zlib만이 유일하게 멀티코어를 지원하여 인코딩/디코딩 성능을 향상시킬 수 있는 여지가 있으며 C++로 구현된 라이브러리가 JAVA로 구현된 라이브러리보다 셋탑박스 같은 임베디드 시스템 환경에 적합하다. 또한 zlib은 오픈소스로 배포되기 때문에 라이브러리의 최신화 측면에서 유리한 점이 있다. LOC(Lines Of Code)는 MPEG-V reference soft

표 5. XML 이진화 라이브러리 별 특징  
 Table 5. Characteristics of the binarization libraries

Comp. Alg.	Library	LOC (Lines Of Code)	Multi-core support?	SW Lang.	Open source?
BiM	MPEG-7 reference software	57,290	X	JAVA	X
gzip	zlib	9,063	O	C++	O
MPEG-V	Implemented	14,846	X	C++	X
MPEG-V	MPEG-V reference software	100,304	X	JAVA	X



표 6. 다중 감각효과 샘플의 이진화 방법 별 압축률

Table 6. Experimental results of compression ratio per binarization method

Test set	SEM size (byte)	Fragmented SEM size (byte)	Compression method	Compressed size (byte)	Comp. ratio
1	1,053,068	3,071,250	BiM	112,280	99.34%
			gzip	2,435,000	20.71%
			MPEG-V	122,800	96%
2	23,156	-	BiM	1,389	94.0%
			gzip	19,283	16.7%
			MPEG-V	977	95.7%

ware가 100,304 라인으로 가장 많았으며 zlib이 9,063 라인으로 가장 적었다. 표 6은 이진화 방법 별 압축률 비교 결과를 보여준다.

예제 1의 경우 BiM과 MPEG-V 규칙 기반 이진화가 각각 99.34%, 96%로 높은 압축률을 보였다. 반면에 gzip의 경우 20.71%의 압축률을 보였는데 이는 SEM 샘플 내의

중복되는 부분이 많지 않기 때문이다. 따라서 텍스트 기반 압축법인 gzip은 BiM과 MPEG-V 규칙 기반 이진화에 비해 SEM 샘플의 이진화에 적합하지 않다는 결론을 얻을 수 있다.

예제 2의 경우 한 개의 긴 엘리먼트로 구성된 모션체어의 움직임 궤적 좌표를 표현하는 Rigid body 이펙트다. Rigid

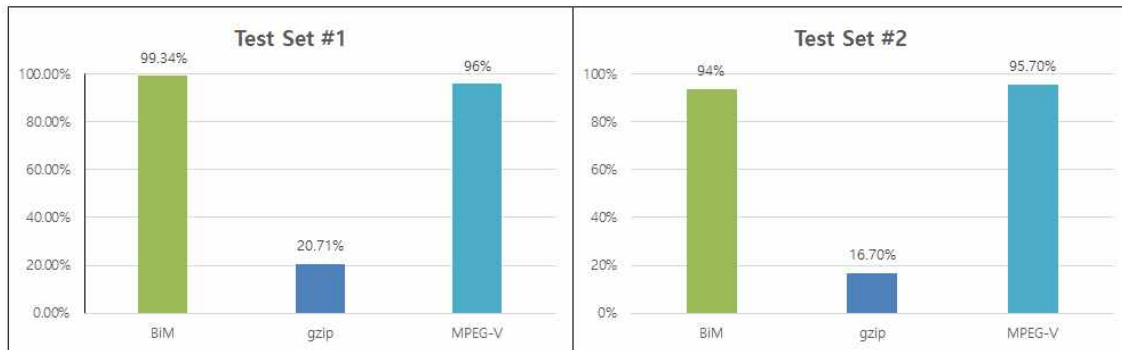


그림 6. 이진화 방법별 압축률 비교  
Fig. 6. Compression ratio comparison



그림 7. 이진화 방법별 인코딩/디코딩 처리시간 비교  
Fig. 7. Processing time comparison of en/decoding a XML document per binarization method

body 이펙트 엘리먼트의 경우 일반적인 이펙트 엘리먼트보다 크기가 크다는 특징이 있다. 예제 2의 압축률은 BiM 94%, gzip 16.7%. MPEG-V 95.7% 순으로 나타났다.

예제1은 각 SEM 샘플의 평균 인코딩/디코딩 시간이다. BiM과 MPEG-V 모두 비슷한 성능을 보였으며 gzip은 낮은 성능을 보였다.

일반적인 SEM인 예제1과 Rigid Body 이펙트 같은 긴 이펙트 엘리먼트인 예제 2 모두 BiM과 MPEG-V가 높은 성능을 보여주었다. 인코딩/디코딩 소모 시간 또한 BiM과 MPEG-V가 우수했다. 그렇기 때문에 SEM 샘플의 이진화는 BiM과 MPEG-V 규칙 기반 이진화가 가장 적절하다고 볼 수 있다.

## VI. 결 론

본 논문의 내용들은 MPEG-A 파트 17(ISO/IEC 23000-17) Multiple Sensorial Media application format 표준화 활동 과정 중 연구한 내용으로, 2014년부터 기고된 기고서 및 표준문서의 내용을 포함한다<sup>[5][6][7][8][9]</sup>.

제안된 다중 감각 미디어 응용 포맷 구조를 정의하고 구현하기 위해 MPEG-4 파트 12(ISO/IEC 14496-12)의 ISO 기본미디어 파일 포맷과 MPEG-V 파트 3(ISO/IEC 23005-3)를 사용하였다. 이 과정에서 MPEG-V 파트 3의 감각효과 메타데이터(SEM)의 샘플화를 위해 XML의 분할과 이진화 방법에 대해 연구하고 본 응용 포맷에 가장 적절한 방법을 제시하였다.

다중 감각 미디어 응용 포맷의 구조는 ISO 기본미디어 파일 포맷의 기본 구조를 재사용하였으며 SEM 샘플의 이진화 방법의 표현을 위해 샘플 디스크립션 박스(stsd)내부에 mpms박스를 새로 정의하였으며 이를 통해 SEM 샘플의 이진화 방법을 명시할 수 있게 하였다. 또한 다중 감각 미디어 응용 포맷의 패커(packager)와 언패커(unpacker)의 구현을 통해 정의한 응용 포맷의 구조대로 구현하였음을 보였다.

SEM 샘플의 이진화를 위해 스키마 기반 이진화 방법인 MPEG-B 파트1(ISO/IEC 23001-1)의 BiM(Binary MPEG for XML)<sup>[2]</sup>, 텍스트 기반 이진화 방법인 gzip<sup>[3]</sup> 마지막으로

MPEG-V 파트3의 감각효과 메타데이터 이진 규칙에 따른 이진화 방법<sup>[4]</sup>을 비교하였다. 이들 이진화 방법을 통해 50,00개의 엘리먼트로 이루어진 SEM 문서와 한 개의 Rigid body 이펙트 엘리먼트로 구성된 SEM 문서의 이진화 압축률을 비교하였다. 이 결과 50,00개의 엘리먼트로 이루어진 SEM문서의 압축률은 BiM(99.34%), gzip(20.71%), MPEG-V 이진 규칙(96%) 순으로 나타났으며, 한 개의 엘리먼트로 구성된 SEM 문서의 압축률은 MPEG-V 이진 규칙(95.7%), BiM(94.0%), gzip(16.7%)순으로 나타났다. SEM 샘플의 경우 반복되는 데이터가 많지 않기 때문에 gzip이 낮은 효율을 보였다.

SEM 데이터의 전송을 위한 지난 연구를 보면, MPEG-2 TS 패킷을 만들 때 SEM 데이터를 시간단위로 자른 후 그것을 다시 PES 패킷으로 만드는 과정을 거친다<sup>[10][11]</sup>. 이는 본 응용 포맷은 경우, 미리 SEM 데이터를 시간단위로 분할하여 SEM 샘플로 저장하기 때문에 이 과정을 생략하고 간단히 PES 패킷으로 만들 수 있을 것으로 기대된다. 향후 본 응용 포맷을 활용한 다중 감각 미디어의 전송과 감각효과 발현기기의 작동까지 가능한 통합 시스템에 대한 연구와 구현이 필요하다.

## 참 고 문 헌 (References)

- [1] Klaus Diepold, Fernando Pereira, Wo Chang, "MPEG-A: Multimedia Application Formats", MultiMedia, IEEE, vol.12, no. 4, pp. 34-41, October-December 2005.
- [2] ISO/IEC 23001-1 Information technology - MPEG systems technologies - Part 1: Binary MPEG format for XML
- [3] Deutsch, L. Peter. "GZIP file format specification version 4.3.", 1996.
- [4] ISO/IEC 23005-3 Information technology - Media context and control - Part 3: Sensory Information.
- [5] Sang-Kyun Kim, Jungyup Oh, Minwoo Kim, Yong Soo Joo, "Implementation of 4D Media Player Application Format", m32402, 2014.01 MPEG 107th San Jose Meeting.
- [6] Jungyup Oh, Minwoo Kim, Yong Soo Joo, Sang-Kyun Kim, "Progress on fragmentation of XML document for multi-sensorial application format", m34156, 2014.07 MPEG 109th Sapporo Meeting.
- [7] Sang-Kyun Kim, Jaejoon Han, Seungju Han, Jungyup Oh, "SEM document fragmentation using Fragment Request Unit (FRU)", m34965, 2014.10 MPEG 110th Strasbourg Meeting.
- [8] Sang-Kyun Kim, Jungyup Oh, Minwoo Kim, In-Su Jang, "Progress on

MulSeMedia Application Format(MSMAF) Reference Software", m36420, 2015.06 MPEG 112th Warsaw Meeting.

[9] ISO/IEC 23000-17:20xx CD Multiple Sensorial Media Application Format, w15873, 2015.10 MPEG 113th Geneva Meeting.

[10] Kyoungro Yoon, Bumsuk Choi, Eun-Seo Lee, Tae-Beom Lim, "4-D

Broadcasting with MPEG-V", Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on, 257 - 262.

[11] Kyoungro "Yoon, End-to-end framework for 4-D broadcasting based on MPEG-V standard", Signal Processing: Image Communication 28 (2013) 127 - 135.

---

## 저 자 소 개



### 오 정 업

- 2014년 : 명지대학교 컴퓨터공학과 학사
- 2016년 : 명지대학교 컴퓨터공학과 석사
- 2015년 10월 ~ 현재 : 한국정보통신기술협회 표준기획부
- ORCID : <http://orcid.org/0000-0002-0859-9383>
- 주관심분야 : metadata encoding, multimedia file format, 4D media, intelligent sensorial effect system, standardization strategy roadmap



### 김 상 균

- 1991년 : 아이오와 대학 (U of Iowa) 전산과학 BS
- 1994년 : 아이오와 대학 (U of Iowa) 전산과학 MS
- 1997년 : 아이오와 대학 (U of Iowa) 전산과학 PhD
- 1997년 3월 ~ 2007년 2월 : 삼성종합기술원 멀티미디어랩
- 2007년 3월 ~ 현재 : 명지대학교 공과대학 컴퓨터공학과 교수
- 2016년 3월 ~ 현재 : IT융합대학 데이터테크놀로지전공 겸임교수
- MPEG 국제표준화 (MPEG-7, MPEG-A, MPEG-V, MIoT) 내 다수의 의장/에디터 수임 중
- ORCID : <http://orcid.org/0000-0002-2359-8709>
- 주관심분야 : digital content (image, video, and music) analysis and management, fast image search and indexing, color adaptation, 4D media, sensors and actuators, VR, Internet of Things, and multimedia standardization