

## 하이브리드 방송 환경기반의 장면 구성 방법

\*조민우 \*박정욱 \*김규현<sup>‡</sup>

\*경희대학교

\*downtorak@khu.ac.kr woogiholic@gmail.com \*kyuheonkim@khu.ac.kr

The method of scene composition  
based on hybrid broadcasting environment\*Jo, Min-Woo Park, \*Jung-Wook \*Kim, Kyu-Heon<sup>‡</sup>

\*Kyung Hee University

## 요약

하이브리드 방송(Hybrid Broadcasting) 환경은 다수의 전송 경로를 통해 다수의 미디어를 동시에 제공할 수 있다. 장면 구성은 다수의 미디어를 동시에 소비하기 위한 기술로 하이브리드 방송의 특징을 효율적으로 활용하는데 적합하다. 장면 구성이 제공해야하는 정보에는 미디어가 표시될 화면상의 위치 정보와 표시되는 시점의 시간 정보 및 사용자 상호작용 정보 등이 있다. 본 논문에서는 미디어의 공간적, 시간적 위치 정보 및 사용자 상호작용 정보를 제공하는 동시에 하이브리드 방송 환경에 적합한 장면 구성 방법을 제안한다.

## 1. 서론

하이브리드 방송(Hybrid Broadcasting)은 다수의 전송 경로를 동시에 이용한 방송 서비스의 일종이다. 현재, 스마트 TV를 필두로 방송망과 IP망을 결합한 하이브리드 방송 환경이 조성되고 있으며, 다양한 미디어의 전달이 가능해졌다. 사용자가 소비할 수 있는 미디어가 증가됨에 따라 장면 구성의 필요성 또한 증가되고 있다.

장면 구성은 비디오, 오디오뿐만이 아닌 다양한 미디어를 소비할 때, 각각의 미디어가 장면에서 차지하는 영역에 대한 공간정보와 어떤 시점에서 소비되어야 하는지에 대한 시간정보를 기술함으로써 다수의 미디어를 동시에 소비할 수 있는 기술이다. 또한, 미디어의 선택적 소비를 제공하기 위해 사용자와의 상호작용을 제공한다. 이러한 장면 구성은 미디어간의 연관관계를 구조적으로 표현하기 위해 주로 마크업 언어(Markup language)를 사용한다.

장면 구성의 필요성은 일찍부터 대두되어 왔으며 그에 따른 표준화 작업도 이루어져 여러 표준이 공표된 상태이다. 대표적인 표준으로는 국제 표준화 기구인 MPEG의 BIFS(ISO/IEC 14496-11:Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine)<sup>[1]</sup>와 LAsER(ISO/IEC 14496-20:Information technology - Coding of audio-visual objects - Part 20: Lightweight Application Scene Representation (LAsER) and Simple

Aggregation Format (SAF))<sup>[2]</sup>, W3C의 SMIL(Synchronized Multimedia Integration Language)<sup>[3]</sup>이 존재한다. 이러한 장면 구성 표준들은 대상으로 하는 환경에 따라 차이를 가지고 있기 때문에 현재 방송을 위해 사용되고 있는 표준은 그림 1과 같이 지상파 DMB(Digital multimedia broadcasting)에 적용된 BIFS가 유일하며, 이 또한 사용이 미비한 실정이다.



그림 1. 지상파 DMB의 BIFS

따라서 본 논문에서는 하이브리드 방송 환경에 적합한 장면 구성 방법을 제안하고 구현 및 검증한다. 본 논문의 구성은 다음과 같다. 2장에서는 장면 구성을 위한 기술인 SMIL에 대하여 소개한다. 3장에서는 하이브리드 환경을 고려한 장면 구성 방법을 제안하고, 4장에서는 3장에서 제안한 장면 구성 방법에 대한 실험 및 검증을 소개한다. 5장에서는 향후 적용 분야를 도

‡교신저자 : 김규현(kyuheonkim@khu.ac.kr)

출하였다.

## 2. Synchronized Multimedia Integration Language

Synchronized Multimedia Integration Language(이하 SMIL)은 W3C가 제시한 장면 구성 방법으로 음향, 동영상, 텍스트 등 동기화된 다수의 미디어를 인터넷 상에 나타내는데 필요한 기준을 정의하고 있다. SMIL의 장점은 XML(eXtensible Markup Language)을 이용하여 구조적이며 간편한 기술 방법으로 장면 구성을 제공하는 것이다. 이러한 SMIL의 장점을 취하고자, 본 논문에서 제안하는 하이브리드 환경을 고려한 장면 구성 방법은 SMIL의 미디어 공간 정보 및 시간 정보 기술 방법을 기반으로 확장하였다.

### 가. 장면 구성 정보 표현

SMIL은 장면 구성을 통해 소비되는 미디어 각각을 하나의 요소로 기술하고 미디어의 공간정보와 시간정보 및 자원의 위치 등을 속성으로 기술한다. 그림 2는 SMIL의 미디어 및 장면 구성 정보의 일반적인 표현 방법을 나타낸 것이다.

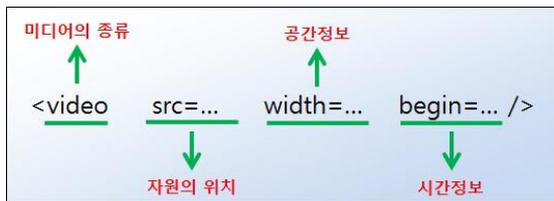


그림 2. SMIL의 미디어 및 장면 구성 정보 표현

미디어를 나타내는 미디어 요소들은 해당 요소의 이름을 통해 미디어의 종류를 나타내며, 요소의 종류에는 ref, animation, audio, img, text, textstream, video가 존재한다. 미디어 자원의 위치는 src 속성의 값으로 기술된다. 미디어의 공간정보는 top, bottom, left, right, width, height, z-index 등의 속성을 통해 기술되며, 시간정보는 begin, end, dur 등의 속성을 통해 기술된다.

### 나. 공간정보 구성

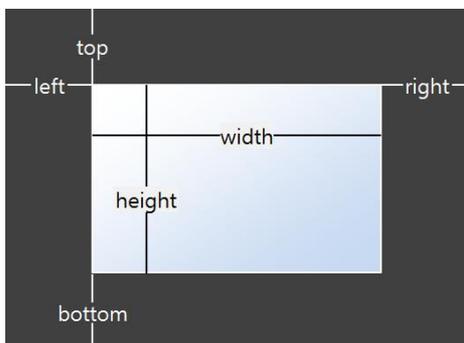


그림 3. 미디어의 공간 구성

SMIL에서 미디어의 공간정보는 주로 top, bottom, left, right, width, height 속성에 의해 기술된다. 전체 화면을 기준으로 top은 미디어가 재생되는 공간의 상단 여백, bottom은 하단 여백, left는 좌측 여백, right는 우측 여백을 나타내며, 화소단위(pixel) 혹은 비율(%)로 표시한다. width는 미디어가 재생되는 공간의 너비, height는 높이를 나타내며, 화소단위(pixel) 혹은 비율(%)로 표시한다. 그림 3는 각 속성을 통해 나타나는 미디어의 화면상 구성을 도식화한 것이다.

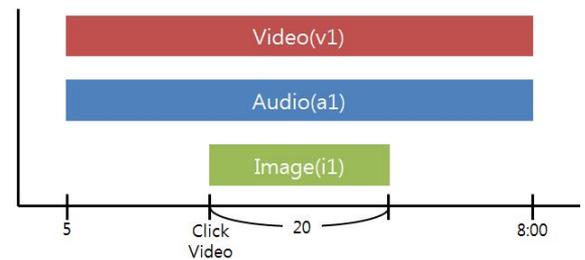
이 외에 z-index 속성은 다수의 미디어가 공간상에서 중첩될 경우, 우선적으로 보여야할 미디어의 선택기준을 제공하며 큰 정수 값을 갖는 미디어가 높은 우선순위를 갖는다.

### 다. 시간정보 구성

SMIL에서 미디어의 시간정보는 주로 begin, dur, end 속성에 의해 기술된다. 해당 미디어에 대해 begin은 재생이 시작되는 시간을, dur은 재생 기간을, end는 재생이 중단되는 시간을 나타낸다. SMIL에서 지원하는 시간에 대한 표현은 상대시간, 절대시간, 이벤트 시점, 동기화 시점 등이 존재한다.

```
<video id="v1" begin="5s" end="wallclock(08:00)" .../>
<audio id="a1" begin="v1.begin" end="v1.end" .../>
<img id="i1" begin="v1.click" dur="20s" .../>
```

(a) SMIL을 이용한 시간 기술



(b) (a)의 시간 구성

그림 4. 미디어의 시간 구성 예

그림 4는 begin, dur, end 속성을 이용한 예를 통해 미디어의 시간 정보 기술과 그에 따른 시간상의 구성을 나타낸 것이다. 장면 구성에 사용되는 미디어는 (a)와 같이 시간정보를 기술하게 된다. (a)에서 정의된 시간정보는 (b)와 같이, 비디오는 서비스를 시작한 후 5초 뒤에 재생되며 오전 8시에 종료되고, 오디오는 비디오와 동기화 되어 재생이 시작되고 종료된다. 이미지는 비디오의 클릭 이벤트에 의해 재생이 시작되고 20초간 재생된 뒤 종료된다.

## 3. 하이브리드 방송 환경기반의 장면 구성 방법

본 논문에서는 하이브리드 방송 환경을 고려한 장면 구성의 XML의 구조를 그림 5와 같이 설계하였다.

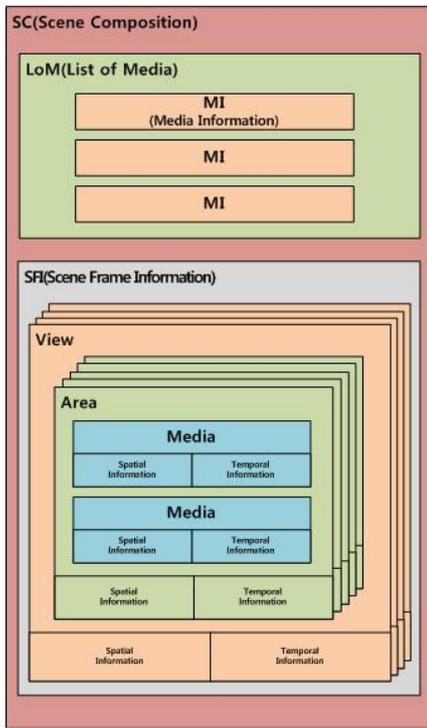


그림 5 하이브리드 방송 환경을 위한 장면 구성의 구조

가. 미디어 속성 정보의 분리

인터넷 환경과 달리 방송환경에서는 유료 서비스가 주를 이루기 때문에 서비스의 품질이 보장되어야 한다. 따라서 서비스 제공 초기에 미디어들의 이용가능여부에 대한 확인이 필요하다. 기존의 장면 구성 방법들은 이러한 미디어 속성 정보가 다른 정보와 혼재되어 있기 때문에 방송환경에서 적용 시, 방송 초기화를 위해 복잡한 절차가 요구된다. 따라서 하이브리드 방송 환경을 위한 장면 구성은 그림 5와 같이 미디어의 속성 정보를 루트 요소인 SC(Scene Composition) 요소의 하위 요소로 LoM(List of Media) 요소를 통해 미디어의 공간정보 및 시간정보와 별도로 기술한다.

LoM 요소는 하위 요소로 MI(Media Information) 요소를 가지며, 각각의 MI 요소는 미디어의 속성 정보를 제공한다. MI의 속성으로는 미디어의 자원 위치를 제공하는 src 속성, 미디어의 종류를 나타내는 type 속성이 존재한다.

미디어의 속성 정보는 LoM 요소를 통해 이미 제공되었기 때문에 미디어 요소에서 다시 기술될 필요가 없다. 그러므로 그림 6의 예제에서와 같이 미디어 요소의 ref속성을 통해 미디어 속성 정보를 참조한다.

나. 다수의 장면 기술

하이브리드 방송 환경을 위한 장면 구성은 하나의 XML 문서로 다수의 장면 기술이 가능하도록 하였다. 기존의 장면 구성은 장면 하나당 하나의 XML파일이 존재하기 때문에 장면 전체의 전환 시, 새로운 XML파일을 이용하므로 이전 장면의 미디어가 이후 장면에서 계속될 경우, 미디어의 연속성을 보장할

수 없다. 따라서 미디어가 연속적으로 재생되어야 하는 방송환경에서 적용할 수 없기 때문이다. 이에 따라 하이브리드 방송 환경을 위한 장면 구성은 그림 5와 같이 미디어의 공간정보 및 시간정보를 제공하는 SFI(Scene Frame Information) 요소의 하위 요소로 하나의 장면을 나타내는 View 요소를 다수 기술할 수 있도록 하였다.

View 요소는 하나의 장면을 나타내므로 장면의 크기를 기술하기 위한 width, height 속성을 제공한다. 다수의 View가 기술됨에 따라 각각의 보여 질 시점이 필요하므로 begin, end, dur 속성 또한 제공한다.

하나의 화면에 여러 장면을 나타낼 수 없으므로 View는 하나만 보이게 되며, 하나의 View가 종료되면 그 다음 View가 이어서 재생된다.

다. 장면의 분할 구성

기존 장면 구성은 장면을 미디어 기반으로 구성하기 때문에 장면의 분할된 일부 영역을 독립적으로 구성할 시, 복잡한 표현이 요구된다. 또한, 특정 영역에 속한 미디어들이 동시에 사라지거나 나타날 때, 제공되어야 할 시간정보의 중복 기술을 피할 수 없다.

이를 위해 하이브리드 방송 환경을 위한 장면 구성은 그림 5와 같이 장면의 분할된 일부 영역을 독립적으로 구성할 수 있는 Area 요소를 제공하며, 미디어는 Area 요소의 하위에 기술하도록 하였다.

Area 요소는 장면 내 일부 영역을 나타내므로 영역의 위치와 크기를 기술할 수 있는 top, bottom, left, right, width, height, z-index 속성을 제공한다. 또한 Area 나타날 시점을 위해 begin, end, dur 속성을 제공한다.

```
<SC>
<LOM>
<MI id="media1" src="video1.avi" type="video"/>
<MI id="media2" src="video2.avi" type="video"/>
<MI id="media3" src="video3.avi" type="video"/>
</LOM>
<SFI>
<View id="view1" width="1920px" height="1080" begin="0s"
dur="10s">
<Area id="area1" top="0px" left="0px" width="1920px"
height="1080px" begin="0s">
<video id="v1" top="0px" left="0px" width="1920px"
height="1080px" begin="0s" ref="media1"/>
</Area>
</View>
<View id="view2" width="1920px" height="1080" dur="30s">
<Area id="area2" top="0px" left="0px" width="960px"
height="1080px" begin="0s">
<video id="v2" top="0px" left="0px" width="960px"
height="1080px" begin="0s" ref="media1"/>
</Area>
<Area id="area3" top="960px" left="0px" width="960px"
height="1080px" begin="0s" end="10s">
<video id="v3" top="0px" left="960px" width="960px"
height="1080px" begin="0s" ref="media2"/>
</Area>
<Area id="area4" top="960px" left="0px" width="960px"
height="1080px" begin="10s">
<video id="v4" top="0px" left="960px" width="960px"
height="540px" begin="0s" ref="media2"/>
<video id="v5" top="0px" left="960px" width="960px"
height="540px" begin="0s" ref="media3"/>
</Area>
</View>
</SFI>
</SC>
```

그림 6 하이브리드 방송 환경을 위한 장면 구성의 기술 예

#### 4. 실험 및 검증

본 논문에서는 하이브리드 방송 환경을 고려한 장면 구성을 제안하였다. 제안한 내용을 검증하기 위하여 PC환경에서 하이브리드 방송 환경을 위한 장면 구성을 통해 미디어를 재생할 수 있는 장면 구성 재생기를 구현하고 실험하였다. 장면 구성 재생기는 XML 문서로 작성된 장면 구성 정보를 분석하여 DOM tree를 구성하고 각 노드의 값을 통해 정해진 공간 및 시간에 미디어를 재생한다. 구현된 재생기는 복잡도를 고려하여 미디어의 공간 정보를 화소단위로만, 시간정보를 초단위로만 분석할 수 있도록 하였다.

실험에 사용된 장면 구성 정보는 그림 6의 예를 사용하였다. 기술된 내용에 따라 4개의 미디어가 지정된 공간과 시간에 재생된다. 재생시작부터 10초간 media1이 전체 화면에 재생되고, 10초부터 20초 사이에 화면 좌측에 media1, 우측에 media2가 재생되며, 20초부터 40초 사이에 좌측에 media1, 우측 상단에 media2, 우측 하단에 media3가 재생될 것이다.

그림 7은 장면 구성 재생기가 본 논문에서 장면 구성 방법에 따라 미디어를 재생하는 장면을 보여준다. 본 그림 7의 결과를 통해 본 논문에서 제안한 하이브리드 환경을 위한 장면 구성을 검증하였다.



(a) 재생 시작 후 5초



(b) 재생 시작 후 15초



(c) 재생 시작 후 25초

그림 7. 장면 구성 재생기의 재생 화면

#### 5. 결론

장면 구성은 다수의 미디어를 동시에 소비하는데 있어 미디어의 공간정보와 시간정보를 제공함으로써 하이브리드 방송과 같이 다수의 미디어를 전송할 수 있는 환경에서 유용하게 사용될 수 있다.

본 논문은 종래의 장면 구성 방법들이 제공하지 못하는 미디어 속성 정보의 분리, 다수의 장면 기술, 장면의 분할 구성을 통해 하이브리드 방송 환경기반의 장면 구성 방법을 소개하였다. 소개한 장면 구성을 통해 다수의 미디어를 방송환경에서 소비하는데 유용하게 활용될 수 있으며, 멀티뷰 서비스나 N스크린 서비스와 같이 다양한 서비스에도 활용될 수 있을 것으로 기대된다. 또한, 장면 구성 기술을 필요로 하는 다른 멀티미디어 서비스 기술 개발의 방향성을 제시할 수 있을 것이다.

#### 감사의 글

이 논문은 지식경제부의 지식경제 기술혁신사업(“스케일러블 데이터기반 Edutainment 핵심기술 개발”)(과제번호: 10035486)의 연구결과와 지식경제부 및 정보통신산업진흥원의 대학IT연구센터육성 지원사업의 연구결과로 수행되었음”(NIPA-2012- H0301-12-1006)

#### 참고문헌

- [1] “ISO/IEC 14496-11:Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine”
- [2] “ISO/IEC 14496-20:Information technology - Coding of audio-visual objects - Part 20: Lightweight Application Scene Representation”
- [3] Synchronized Multimedia Integration Language(SMIL 3.0) - W3C Recommendation 01 December 2008.
- [5] “TTA Hybrid Broadcast and Broadband TV 표준화 동향”