

# 사용자의 관심 영역을 고려한 MPEG-21 디지털 아이템 적응변환

손유미<sup>1)</sup>, 임정연<sup>1)</sup>, 김문철<sup>1)</sup>, 김종남<sup>2)</sup>, 김경수<sup>2)</sup>  
<sup>1)</sup> 한국정보통신대학교 멀티미디어 컴퓨팅, 통신 및 방송 연구실  
<sup>2)</sup> KBS 기술연구소

## MPEG-21 DIA considering user's Region of Interest

Yumi Sohn<sup>1)</sup>, Jeongyeon Lim<sup>1)</sup>, Munchurl Kim<sup>1)</sup>, Jong-Nam Kim<sup>2)</sup>, Kyeongsoo Kim<sup>2)</sup>

<sup>1)</sup> Multimedia Computing, Communication and Broadcasting Laboratory  
Information and Communications University (ICU)

<sup>2)</sup> Technical Research Institute, Korean Broadcasting System (KBS)

E-mail : {yumi, mkim}@icu.ac.kr

### Abstract

멀티미디어 콘텐츠의 트랜스코딩은 디스플레이 사이즈, 컴퓨팅 자원, 네트워크 대역폭이나 QoS 같은 네트워크 자원 등 사용자 터미널의 수용 능력의 관점에서 수행되어 왔다. 오디오비주얼 콘텐츠의 리사이징이나 프레임 드로핑과 같은 제약을 주어 비트 율을 줄여줌으로써 사용자의 터미널 환경에 맞도록 변환 시키는 것이 일반적인 트랜스코딩 방법이었다. 이러한 사용자의 환경에 따른 전통적인 방법 외에, 사용자 선호도에 기반하여 멀티미디어 콘텐츠에서 사용자의 관심 영역(Region Of Interest)을 중심으로 트랜스코딩 하는 의미적인 코딩(segment transcoding)방법을 제안하였다. MPEG-21 DIA 프레임워크 하에서 사용자의 관심 영역을 표현한 의미 정보가 XDI(contexT Digital Item)형식으로 콘텐츠 제공자에게 전달되고 리소스 적응엔진을 통해 적응 변환된 콘텐츠는 다시 사용자에게 전달 된다. 본 논문에서는 사용자의 관심영역 기반 오디오 비주얼 콘텐츠의 의미론적 트랜스코딩 기술을 제안하고 적응 변환의 실험 결과를 설명한다.

### I. 서론

우리는 다양한 종류의 멀티미디어 콘텐츠를 각기 다른 사용자 단말, 가변적 네트워크 등의 멀티미디어 환경에서 이용하고 있다. 이런 멀티미디어 환경은 각기 다른 사용자 터미널, 네트워크 환경에 다른 성향과 소비 형태를 보이기 때문에 매우 복잡하고 다양하다. 이와 같이 다양한

멀티미디어 환경하에서 범용적 멀티미디어 접근을 위하여 전달과 소비를 위한 구성 요소들간의 상호 관련성을 기술하는 표준이 필요하게 되었다. 그런 표준이 MPEG-21 이며 다양한 멀티미디어 콘텐츠를 다양한 환경에서 사용자가 상호 호환적으로 쉽고, 편리하게 생성, 배급, 소비할 수 있는 방법을 정의, 구현 할 수 있도록 하는 '멀티미디어 프레임워크'를 구축하는 것을 목표로 한다.

전통적인 트랜스코딩 방법은 SNR 을 줄이거나 콘텐츠의 시간 또는 공간 축을 변화시켜 양질의 자원을 낮은 질로 변화시키는 것이다. B-picture 을 생략하고 I-picture 만 사용하거나, 터미널 디스플레이 사이즈에 맞도록 영상의 공간적 해상도를 줄이고, 네트워크의 대역폭이 제한되어 있을 경우에는 PSNR 감소시켜 비트 스트림의 양을 줄이는 형태로 트랜스코딩을 수행하였다. 하지만 이 모든 것은 사용자의 환경적인 관점에서 이뤄지고 있어 사용자 자신의 콘텐츠 부분에 대한 선호에 따른 트랜스코딩은 어렵다. 예를 들어 축구장 전경을 내용으로 하는 영상을 사용자 단말의 크기에 따라 비례적으로 공간적 해상도를 축소할 경우, 축구장 내의 움직이는 선수들의 크기 또한 매우 작게 되어 영상의 내용적 의미가 이해되기 어렵게 될 수도 있다. MPEG-21 DIA 내에서 멀티미디어 콘텐츠의 의미론적 트랜스코딩을 가능하게 하기 위해, MPEG-21 의 user characteristics 내에 contents preference, presentation preference, accessibility, mobility 기술구조를 기반으로 멀티미디어 콘텐츠에 대하여 기술하고 있다. 본 논문에서는 사용자의 관심영역 기반 오디오비주얼 콘텐츠의 의미론적 트랜스코딩 기술을 제안하고 적응 변환의 실험 결과를 설명한다.

## II. MPEG-21 Digital Item Adaptation

MPEG-21 멀티미디어 프레임워크에서는 네트워크상에서 디지털 데이터(콘텐츠)를 리소스라고 하며, 디지털 아이템(DI: Digital Item)은 멀티미디어 콘텐츠와 식별자, 그리고 서술자를 포함하는 유통, 처리의 최소 단위이다. MPEG-21 디지털 아이템 적응(DIA: Digital Item Adaptation)은 디지털 아이템 사용자 특성과 환경 정보, 네트워크나 터미널의 특성을 고려하여 다양한 멀티미디어 콘텐츠의 소비를 가능하게 하기 위한 범용적 멀티미디어 접근을 제공하기 위해 사용자 특성, 사용자 환경, 사용자 단말 특성, 네트워크 자원 특성 등에 대한 세부 정보를 체계적으로 기술하는 규격을 정의하고 있다.

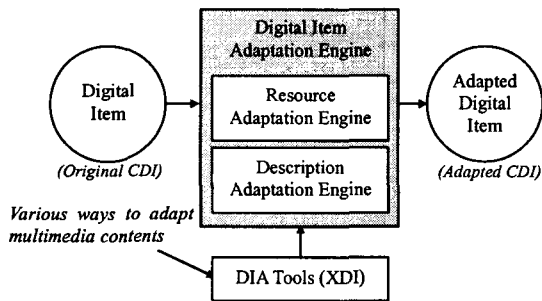


그림 1 디지털 아이템 적응 변환의 개념 [1]

그림 1은 MPEG-21 디지털 아이템 적응 변환의 개념도이다. 처리 과정은 기능적으로 크게 ‘리소스 적응 엔진(resource adaptation engine)’과 ‘서술자 적응 엔진(descriptor adaptation engine)’으로 분류될 수 있는데, 사용자 특성과 사용 터미널 및 네트워크 환경 등의 사용(자) 환경 정보를 기술한 ‘DIA 서술자 툴(DIA description tool)’에 기반하여 입력된 디지털 아이템은 리소스 및 서술자 변환 과정을 거쳐 적응된(adapted) 디지털 아이템으로 출력된다.

## III. 관심 영역에 기반한 의미론적 코딩의 기술 방법

### 3.1 Visual and Text Presentation Preference ROI in the Visual aspect

사용자 단말이 사용자의 터미널 디스플레이 사이즈, 컴퓨팅 자원, 네트워크 대역폭 등으로 제한된 상황일 경우, 원래의 멀티미디어 콘텐츠가 사용자 단말에 제공되어도 원래의 멀티미디어 콘텐츠가 맞춰진 디코딩 환경에 비해 새로 디코딩 될 사용자의 단말의 환경이 더 제한적이라면, 사용자는 그 콘텐츠를 그대로 사용하는 데에는 어려움이

따른다. 기존의 방법은 다음 그림 2와 같이 콘텐츠 자체의 리사이징 혹은 화질을 낮춘다거나, 시간 축에 따라 프레임 드로핑과 같은 조건을 주어 비트 율을 줄이거나 SNR을 줄이는 방법 등이 쓰이고 있다. (a)는 시간 축에 따라 해상도를 줄였으며, (b)는 이미지 크기에 따른 해상도 조절 그리고 (c)는 SNR을 줄여 화질을 줄인 예이다.

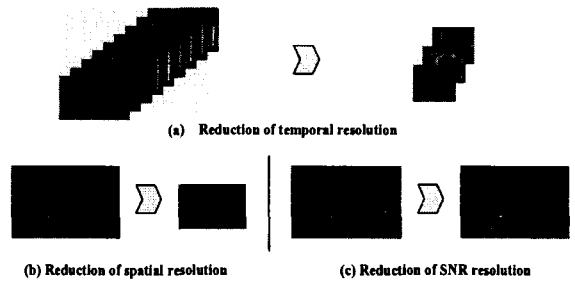


그림 2 전통적인 이미지/비디오 변환 방법

하지만, 기존의 방법과 달리, 농구나 축구경기를 보고 있을 경우, 사용자가 특정 선수의 움직임을 유심히 보고 싶어한다거나 공의 움직임을 놓치고 싶어하지 않을 수 있다. 이러한 경우, 콘텐츠 중에 그의 관심 영역 주위만 해당 사용자의 환경에 맞게 트랜스코딩 해주는 것도 의미 있는 방법이 될 수 있다.

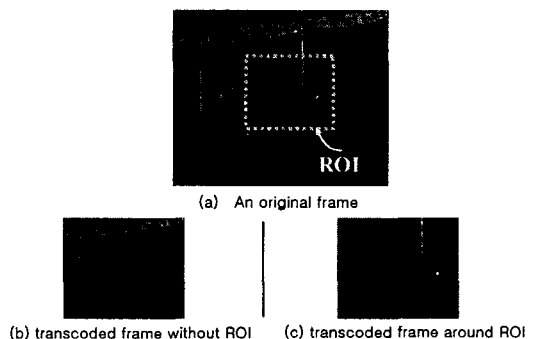


그림 3 관심 영역에 기반한 리소스 적응 변환

그림 3은 축구 경기 장면으로 리사이징만 했을 경우와 관심 영역을 고려하여 트랜스코딩한 경우의 예를 보여준다. (a)와 같이 원래 콘텐츠의 끝인 장면에서 선수가 공을 차는 부분을 보다 자세히 보고 싶을 경우, 단지 리사이징만 해서 보여 주는 경우(b) 보다 공을 주변으로 한 ROI(관심 영역: Region Of Interest)을 중심으로 하여 MPEG-21 DIA을 기반으로 적응 변환한 것을 보여주는 것(c)이 디스플레이 사이즈가 작은 터미널의 경우에는 보다 효과적으로 정보를 전달 할 수 있을 것이다.

사용자의 관심 영역은 그림 4 에서 보는 바와 같이 MPEG-21 DIA 스키마의 *PresentationPreferenceDS* 의 *FocusOfAttention* 이란 요소를 사용하여 XDI 로 기술한다. XDI 는 비디오가 스트리밍 될 때 사용자로부터 서버로 전달되고, 서버는 사용자의 표현 선호 정보가 들어있는 XDI 를 파싱한다. 서버는 사용자의 관심영역을 이해하고 적용 변환된 비디오가 서비스 되도록 한다.

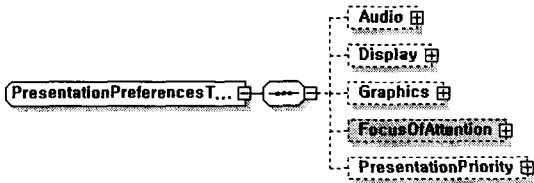


그림 4 PresentationPreference DS 의 구조

*SegmentPresentationPreference Type* 은 MPEG-7 MDS 의 *Segment Type* 으로부터 확장되었고 여기서 *StillRegionROI* 요소를 기술하는 *FocusOnStillRegion* 과 *MovingRegionROI* 을 기술하는 *FocusOnMovingRegion* 이 유도되었다. 이는 각각 MPEG-7 의 *StillRegion* 과 *MovingRegion* 으로부터 유도되었다. 그림 5 는 *FocusOfMovingRegion* 의 구조를 나타내고 있으며, *FocusOfStillRegion* 도 같은 구조에 *StillRegionType* 으로부터 확장된 것이다. *StillRegionROI* 는 *Spatial Mask* 의 *Box* 형태로 4 개의 꼭지점을 지정해주고, *MovingRegionROI* 은 *SpatioTemporal Mask* 의 *Parameter trajectory* 형태로 궤적의 값을 지정하여 사용자의 관심 영역을 표현한다.

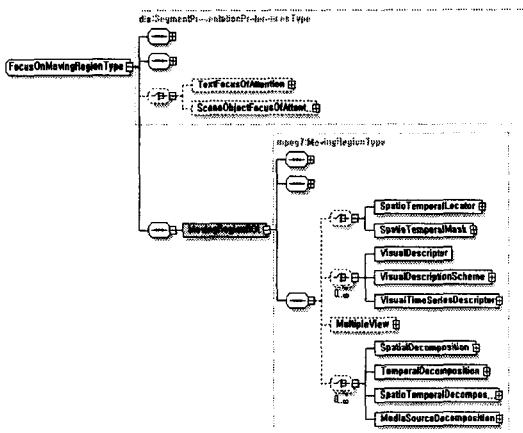


그림 5 PresentationPreference DS 의 구조

**Keyword in the Textual aspect**

스크린에 텍스트를 표시해주는 것은 정보를 전달하는 방법에 있어 매우 효율적인 방법이다. 텍스트 방송이나 텍

스트 기반 메세징 응용 프로그램에서 사용자가 관심 키워드를 위주로 하여 적당한 속도와 폰트로 사용자의 단말에 전달되고 표시된다. 텍스트 콘텐츠는 사용자의 관심 키워드에 기반하여 여러 문단들 중에서 간략하게 요약될 수 있으며 서버는 보다 의미 있고 효율적인 텍스트 정보를 제공할 수 있게 된다.

*TextFocusOfAttention* 은 *FocusOfAttention* 중의 일부분으로 텍스트 타입으로 표현하는 *Keyword* 요소와 폰트 색상, 크기, 형태와 텍스트 표현 속도를 기술할 수 있는 속성으로 이루어져있다.

**3.2 Audio Presentation Preference**

사람들은 보통 빠른 음성을 듣는 것을 어려워한다. 청력에 문제가 있는 사람뿐만 아니라 그렇지 않은 사람들도 음성의 속도에 대한 선호를 가지고 있을 수 있다. 외국어 공부를 한다거나, 노래방에서 노래를 부를 경우 재생 속도를 사용자에게 맞게 조절하는 것은 매우 일반적이다. 그림 6 에서 보는 바와 같이 *SoniferousSpeed* 서술자는 *AudioPresentationPreference DS* 내부에 속해 있으며 5 종류의 빠르기로 표현하고 있다.

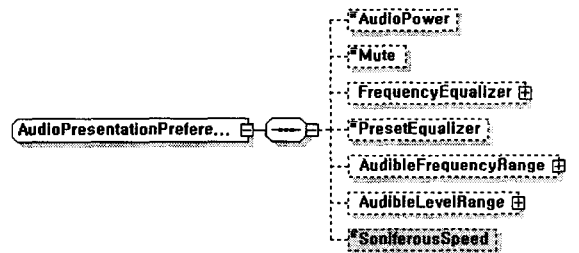


그림 5 AudioPresentationPreference DS 의 구조

**IV. 실험 결과**

그림 7 은 이미지와 비디오에 원하는 ROI 를 지정하여 트랜스코딩 결과를 보여주는 GUI 이다. 응용 프로그램의 왼쪽 부분은 비디오를 오른쪽 부분은 이미지의 예를 보여주고 있다. 왼쪽 윗부분의 원 영상에서 마우스로 관심 있는 영역을 드래킹하여 그 포인트 중심으로 박스 형태로 관심 영역을 나타낸다. 왼쪽 아래 부분에 있는 두 개의 영상 중 왼쪽 것은 ROI 를 고려하여 트랜스코딩 한 것이고, 오른쪽 것은 단순히 리사이징 한 것이다. 디스플레이 사이즈가 작은 터미널이거나 특정 선수에 관심이 있을 경우 영역을 고려하여 서비스하는 것이 훨씬 효율적일 것이다. 오른쪽은 이미지에서 관심 영역을 위주로 하여 트랜스코

당 한 것이다. 이 ROI 부분에 해당하는 디지털 아이템은 그림 8 의 xml file 예제에서 보는 바와 같이 FocusOfAttention 의 FocusOfMovingRegion 또는 FocusOfStillRegion 에 의거하여 표현 할 수 있다.

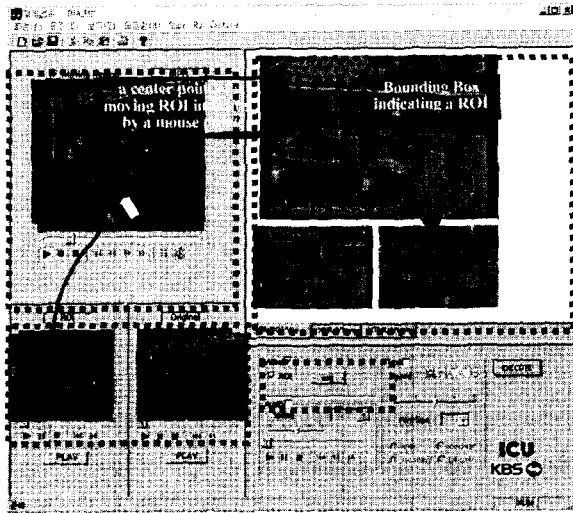


그림 6 FocusOfAttention 의 적용 예

```
<FocusOfAttention xsi:type="FocusOnMovingRegion Type">
  <MovingRegionROI>
    <mpeg7:SpatioTemporalMask>
      <mpeg7:SubRegion>
        <mpeg7:ParameterTrajectory motionModel="translation">
          <mpeg7:InitialRegion>
            <mpeg7:Box mpeg7:dim="2 2"> 0 5 172 149
          </mpeg7:Box>
          </mpeg7:InitialRegion>
          <mpeg7:Params>
            <mpeg7:WholeInterval>
              <mpeg7:MediaDuration>PT14S</mpeg7:MediaDuration>
            </mpeg7:WholeInterval>
            <mpeg7:InterpolationFunctions>
              <mpeg7:KeyValue type="startPoint">84</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">100</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">110</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">121</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">123</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">128</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">128</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">130</mpeg7:KeyValue>
              <mpeg7:KeyValue type="firstOrder">160</mpeg7:KeyValue>
            </mpeg7:InterpolationFunctions>
          </mpeg7:Params>
        </mpeg7:ParameterTrajectory>
      </mpeg7:SubRegion>
    </mpeg7:SpatioTemporalMask>
  </MovingRegionROI>
</FocusOfAttention>
```

그림 8 FocusOnMovingRegion xml file 의 예

그림 9 는 텍스트 프리젠테이션과 오디오의 속도를 사용자의 선호에 맞게 표현 가능하도록 한 응용프로그램의 GUI이다. 오른쪽 패널 윗부분에 사용자가 원하는 키워드와 폰트로 특정 텍스트 콘텐츠의 요약 서비스가 가능하다. 아래쪽 패널의 왼쪽 부분에서 오디오의 속도를 조절할 수 있으며 fastest, faster, normal, slower 또는 slowest의 5가지 레벨 중 하나를 선택하면 해당 오디오 콘텐츠가 그 속도에 맞게 트랜스코딩 되어 전달 된다.

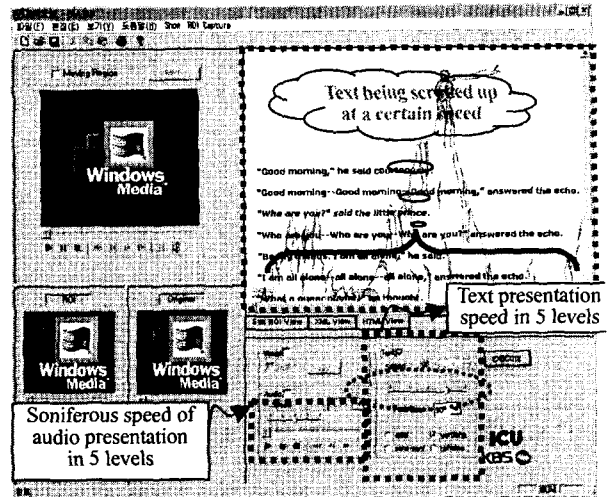


그림 9 TextPresentation & Soniferous Speed 의 적용 예

## V. 결론

MPEG-21 DIA 는 각기 다른 사용자, 네트워크, 단말 등의 환경 하에서의 멀티미디어에 대한 범용적 접근과 완전한 소비 환경을 가능하게 하는 데 있다. MPEG-21 DIA 는 범용적 멀티미디어 접근 환경에 적합한 멀티미디어 콘텐츠를 기술하는 의미적론인 방법이다. 서버가 사용자의 환경을 자동적으로 인식하여 그에 적합한 멀티미디어 콘텐츠를 서비스하기 위해서 꼭 필요한 일이다. 그리하여 비디오, 이미지, 오디오 비주얼과 텍스트에 관하여 사용자 측면에서 멀티미디어 콘텐츠가 표현하기 위한 의미적 적용 변환을 기술할 수 있는 방법을 제안하게 되었다. 향후 과제로는 한층 더 다양한 사용자의 선호를 반영하여 자원에 보다 많은 의미를 부여한 적용 변환 생성하는 형태로 구현되어야 할 것이다.

## 참고문헌

- [1] "MPEG-21 Digital Item Adaptation AM (v.5.0)," ISO/IEC JTC1/SC29/WG11 N5613, Pattaya, Thailand, Mar. 2003.
- [2] "Information Technology - Multimedia Content Description Interface - Part5: Multimedia Description Schemes," ISO/IEC JTC 1/SC 29 N4242
- [3] Munchurl Kim, Jeongyeon Lim, Jong-Nam Kim and Kyeongsoo Kim, "Extensions to Presentation Preference DS of MPEG-21 DIA," ISO/IEC JTC1/SC29/WG11 m8968, Shanghai, China, Oct. 2002.