

대화형 콘텐츠 제작 및 검색을 위한 XMT-A to/from MP4 컨버터 설계 및 구현

이인재, 정원식, 김규현
한국전자통신연구원 방송미디어연구부

Design and Implementation of XMT-A to/from MP4 converter for Authoring and Retrieving Interactive Contents

Injae Lee, Won-sik Cheong, Kyuheon Kim
Broadcasting Media Research Department
Electronics and Telecommunications Research Institute

요 약

대화형 멀티미디어 서비스를 위한 차세대 멀티미디어 콘텐츠의 국제 표준 규격인 MPEG-4에서는 저작권 콘텐츠를 저장하기 위한 포맷으로 XMT (eXtensible MPEG-4 Textual format)와 MPEG-4 파일 포맷 (MP4)를 제공하고 있다. XMT와 MP4는 각각 텍스트 형식과 이진 형식이며 각기 장단점을 가지고 있으므로, 콘텐츠의 효율적인 제작, 관리 및 스트리밍을 위해서는 두 포맷간의 상호 변환을 수행할 수 있는 컨버터가 필요하다. 또한 나날이 콘텐츠의 종류와 수가 증가됨에 따라 MPEG-4 콘텐츠의 용이한 검색도 필요하다 따라서 본 논문에서는 MPEG-4에 기반을 둔 대화형 콘텐츠를 제작하고 검색하기 위한 XMT-A to/from MP4 컨버터를 설계하고 구현하였다.

1. 서론

초고속 정보통신망을 근간으로 실현될 서비스들은 문자, 음성, 영상을 고속으로 처리하는 디지털 단말을 중심으로 한 보고 듣는 멀티미디어형 서비스로 발전할 것이다. 이에 부응하여 현재 다양한 형태의 차세대 오디오/비디오 (Audio-Visual; AV) 서비스를 지원할 수 있도록 컴퓨터의 대화형 기능 (interactivity) 과 통신의 전송 기능을 결합한 국제 표준 규격인 MPEG-4에 대한 표준화 작업이 진행되고 있으며, 많은 부분의 표준화가 완료된 상태이다[1].

MPEG-4에서는 장면을 구성하는 비디오, 오디오, 정지영상, 2차원 그래픽 등을 객체 단위로 취급하

여 콘텐츠를 제작하며, 각 객체에 대해 사용자 대화성(User interactivity)을 부여할 수 있는 특징이 있다. 이러한 특징을 제공하기 위해 MPEG-4 시스템은 장면을 구성하는 미디어 객체들에 대한 정보를 나타내는 객체 기술(Object Descriptor; OD)과, 각 미디어 객체들의 시공간적인 배치 및 사용자 입력이나 시간에 따른 객체의 동작을 나타내는 장면 기술(Scene Description; SD) 정보로 구성된다.

MPEG-4에서는 콘텐츠 저작언어로 텍스트 형식의 XMT[1]를 사용하고 있으며, 이의 이진 형식으로 부호화된 포맷인 MPEG-4 파일 포맷[2]를 제공하고 있다. 여기서, XMT는 MPEG-4의 객체 기술과 장면

기술을 콘텐츠 제작자가 잘 알아볼 수 있도록 텍스트로 나타내기 위한 프레임웍이며, MP4는 스트리밍이 가능한 이진 파일 포맷이다. 그러므로, MPEG-4의 콘텐츠 제작언어인 XMT를 사용하여 제작된 인터랙티브 (interactive) 콘텐츠를 스트리밍하거나, 배포하기 위해서는 이진 포맷인 MP4 파일로 변환하여야 하며, MP4 파일의 재편집을 위해서는 MP4 파일의 XMT 파일로의 변환도 가능하여야 한다.

또한 디지털 방송 서비스가 보편화 되면서 부가 콘텐츠의 종류와 수가 매년 증가함에 따라 부가 콘텐츠의 한 종류인 MPEG-4 콘텐츠의 수도 증가할 것이다. 따라서 MPEG-4 콘텐츠의 검색도 용이하게 이루어질 수 있어야 한다. MPEG-4 시스템에서 MPEG-4 콘텐츠의 용이한 검색을 위해 현재 멀티미디어 데이터의 메타데이터를 표준화 하고 있는 MPEG-7 스트리밍을 활용할 수 있는 방안이 제안되고 있다[2][3].

본 논문에서는 대화형 콘텐츠의 저작 및 검색을 위한 XMT-A to/from MP4 컨버터의 설계 및 구현에 대해 기술하였다. 본 논문의 2장에서는 대화형 콘텐츠를 저작하기 위한 XMT-A to/from MP4 컨버터에 대해 기술하고, 3장에서는 멀티미디어 데이터의 메타데이터를 표준화 하고 있는 MPEG-7을 활용한 XMT-A to/from MP4 컨버터에 대해 기술한다. 4장에서는 구현된 컨버터의 실험결과를 보여주었고, 5장에서는 본 논문의 결론을 맺는다.

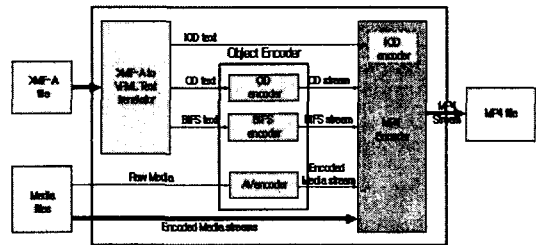
2. 대화형 콘텐츠 저작을 위한 XMT-A to/from MP4 컨버터

MPEG-4의 콘텐츠 제작언어인 XMT-A를 사용하여 제작된 인터랙티브 (interactive) 콘텐츠를 스트리밍하거나, 배포하기 위해서는 이진 포맷인 MP4 파일로 변환하여야 하며, MP4 파일의 재편집을 위해서 MP4 파일을 XMT-A파일로 변환할 수 있어야 한다. 이때, 객체편집 정보를 가지고 있는 XMT-A 파일을 MP4 파일로 변환하는 기능을 수행하는 모듈을 XMT-A to MP4 컨버터라 하고, 반대로 MP4 파일로부터 XMT-A 파일 및 미디어 파일을 생성하는 기능을 수행하는 모듈을 MP4 to XMT-A 컨버터 (혹은 XMT-A from MP4 컨버터)라 한다.

2.1 XMT-A to MP4 컨버터

XMT-A 형식으로 제작된 MPEG-4 콘텐츠는 장면 기술과 객체 기술만을 포함하고 있고, 미디어 객체 자체는 포함하고 있지 않다. 그러므로, XMT-A 파일로부터 MP4 파일을 만들기 위해서는 먼저, XMT-A 파일에 기술되어 있는 장면 기술을 이진 형식의 BIFS 스트림으로, 객체기술을 이진 형식의 OD스트림으로 부호화한 뒤, 부호화된 장면 기술 스트림과 객체 기술 스트림을 부호화된 미디어 스트림과 함께 MP4 파일 규격에 규정되어 있는 box 구조에 맞도록 구성한다.

그림 1은 XMT-A to MP4 컨버터의 구조를 나타낸 것이다. 그림1에서 보여지는 것처럼 XMT-A to MP4 컨버터는 객체 부호화기(object encoder)와 MP4 부호화기 (MP4 encoder)로 구성되어 있으며, 객체 부호화기는 객체 기술자 부호화기, 장면 부호화기 및 미디어 객체 부호화기로 구성된다. 그리고, 미디어 객체 부호화기는 XMT-A에 포함되어 있는 미디어 객체가 부호화 되지 않은 데이터인 경우 객체 기술자에 기술되어 있는 부호화에 필요한 정보를 이용하여 적절히 부호화를 행한다. 그리고, MP4 부호화기에서는 MP4 규격에 따라서, 각 객체 부호화기에서 부호화된 객체 스트림들은 'mdat' atom에 저장하고, 각 객체 스트림에 대한 메타 데이터들을 'moov' atom에 저장한다.



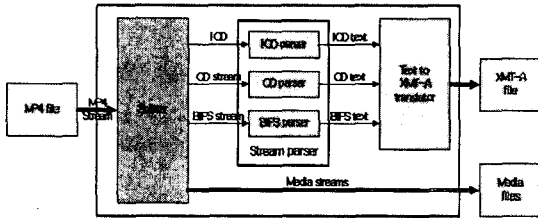
[그림 1] XMT-A to MP4 컨버터의 구조도

2.2 XMT-A from MP4 컨버터

MP4 파일을 XMT-A 파일로 변환하기 위해서는, MP4 파일에 분산되어 저장되어 있는 객체 기술 정보, 장면 기술 스트림 및 미디어 스트림들을 분리해낸 뒤, 객체기술과 장면 기술 스트림을 파싱하여

XMT-A 파일을 생성하고, 분리된 미디어 스트림들은 재사용이 가능하도록 각각의 미디어 파일로 만들어서 저장하여야 한다.

그림 2은 MP4 파일을 XMT-A 파일로 변환하고 MP4 파일 내에 포함되어 있는 미디어 파일들을 분리해 내는 MP4 to XMT-A 컨버터의 구조를 나타낸 것이다. 그림에서 보여지듯이 컨버터는 객체 분리기(object splitter)와 파서(Stream Parser)로 구성된다.



[그림 2] XMT-A from MP4 컨버터의 구조도

3. 대화형 콘텐츠 검색을 위한 XMT-A to/from MP4 컨버터

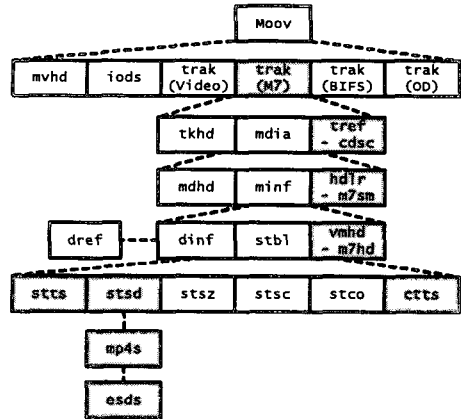
3.1 MPEG-7 스트림

MP4 콘텐츠의 용이한 검색을 위해 MPEG-4 시스템에서는 MPEG-7 메타데이터를 하나의 ES로 간주할 수 있다. 이때 M7 ESD에 의해 기술되는 M7 ES는 같은 OD 내에 존재하는 오디오/비디오/장면 기술 스트림에 대해 유효한 스트림이어야 한다. MPEG-7 스트림이 다른 특정 스트림에 대해 기술하고 있다는 것을 표현하기 위해 DescriptionDescriptor가 사용된다. 이 기술자(descriptor)는 MPEG-7이나 OCI 스트림에 적용될 수 있으며, MPEG-7 이나 OCI 스트림이 기술하고 있는 스트림의 ES_ID를 가지고 있다.

MP4 파일 포맷에서 MPEG-7 ES를 처리하기 위해 'cdsc' (content describes) 라는 참조 타입이 추가되었으며, 이는 DescriptionDescriptor에서 기술된 ES_ID를 가지고 있는 box이다. 그밖에 미디어 핸들러 타입에 'm7sm'이 정의되고, 'minf' box의 header box로 'm7hd'가 추가되었는데 이는 MPEG-7 인코딩 포맷 정보(0인 경우 binary encoding)를 가지고 있다.

MPEG-4에서는 모든 ES를 timed-sequence로 취급하며, MPEG-4에서 AU는 시간 정보가 들어갈 수 있

는 최소 단위로 정의된다. 그러나 MPEG-7의 경우 static sequence로서, MPEG-7에서의 AU은 tree의 segment 단위로 정의될 수 있다. 따라서 본 논문에서는 MPEG-7 스트림을 하나의 AU으로 정의하며, MP4 파일 내에서의 M7 track 구성은 그림 3과 같다.



[그림 3] M7 ES 처리를 위한 MP4 track 구성

3.2 OD/IOD 기술자

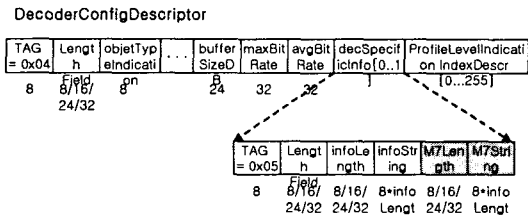
3.2.1 OCI 기술자

MPEG-4 시스템에서 객체에 대한 추가적인 정보를 기술하기 위해 OCI 기술자가 사용될 수 있다. 특정 객체의 추가적인 정보를 기술하기 위해 MPEG-7 스트림을 사용할 경우, MPEG-4 시스템 디코더에는 MPEG-7 스트림 디코딩이 추가적으로 지원되어야 한다. 따라서 객체에 대한 추가적인 기술 정보량이 적은 경우 OCI 기술자를 사용한다면 단말에서의 복잡도를 줄일 수 있어 효과적이다. OCI 기술자는 ContentClassificationDescr, KeywordDescr, RatingDescr, ContentCreatorNameDescr, ContentCreationDateDescr, LanguageDescr 등 15개의 세부 기술자들로 구성될 수 있다. OCI 기술자는 ExpandedTextualDescriptor나 ShortTextualDescriptor를 통해 기술하고자 하는 아이템도 따로 정의해 줄 수 있으므로 기술 범위는 제약적이지 않다. 그러나 IOD/OD 크기에 제약이 있으므로 OCI 기술자를 통한 기술 분량에는 한계가 있다.

3.2. DecoderConfigDescriptor

앞 절에서 객체에 대한 추가적인 정보는 OCI 기

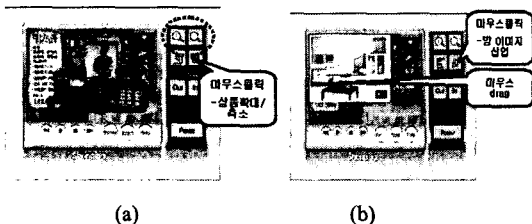
술자를 통해 제공될 수 있었다. 그 밖에 본 논문에서는 영상기반 검색을 위해 DecoderConfigDescriptor에 M7 필드를 추가하는 방법을 제안한다. DecoderConfigDescriptor는 디코더 타입이나 디코딩에 필요한 정보 등을 담고 있는 기술자이다. 예를 들어 JPEG의 경우, 해당 콘텐츠를 만든 사람, 만든 날짜, 분류 및 해당 JPEG에 대한 추가적인 정보 등은 OCI descriptor를 통해 표현될 수 있다. MPEG-7 메타데이터의 경우 JPEG 콘텐츠에 대해 상기 정보 외에 HomogeneousTexture, ScalableColor 등 영상 기반 정보들도 제공해 준다. 따라서 영상 기반 정보들은 DecoderConfigDescriptor에 M7 필드에서 제공해 줌으로써 내용 기반 검색 뿐 아니라 영상 기반 검색도 가능하게끔 제공해 줄 수 있다. 이를 위해 확장된 DecoderConfigDescriptor의 구조는 그림 4와 같다.



[그림 4] 확장된 DecoderConfigDescriptor의 구조

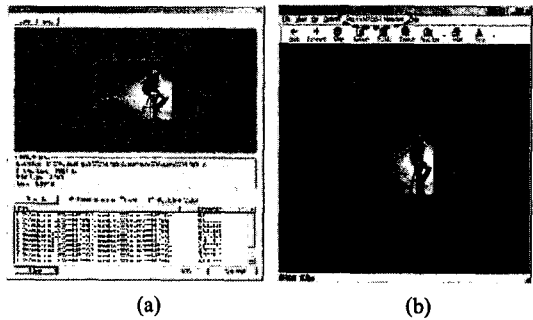
4. 구현 및 고찰

설계된 XMT-A to/from MP4 컨버터는 C/C++, MFC를 이용하여 구현하였다. 그림 5는 구현된 XMT-A to/from MP4 컨버터를 통해 생성된 대화형 MP4 콘텐츠의 예를 보여준다. 해당 콘텐츠는 대화형 홈쇼핑 서비스를 목적으로 제작된 것으로서 상품 확대/축소, 상품 정보 외에 자신의 방 이미지에 제품을 올려보는 등 다양한 대화형 기능을 제공하도록 제작되었다.



[그림 5] 대화형 콘텐츠 제작의 예

그림 6은 DecoderConfigDescriptor의 M7 필드를 이용한 간단한 내용 기반 검색의 결과를 보여준다. 검색하고자 하는 JPEG 영상을 열은 후, MP4 파일들을 검색하면 M7 필드에 특정 JPEG의 ScalableColor, HomogeneousTexture 정보를 담아 인코딩된 MP4 파일 중에서 찾고자 하는 영상과 가장 관련성이 높은 MP4 파일부터 순서대로 리스트 박스에 보여준다. 사용자가 리스트 중에서 보고 싶은 MP4 파일을 선택하면 MPEG-4 시스템 디코더에 해당 MP4 콘텐츠가 플레이되는 것을 볼 수 있다.



[그림 6] MP4 콘텐츠 검색의 실험 결과

5. 결론

본 논문에서는 대화형 콘텐츠의 저작 및 검색을 위한 XMT-A to/from MP4 컨버터의 설계 및 구현에 대해 제안하였다.

텍스트 형식의 파일 포맷인 XMT와 이진 형식의 파일 포맷인 MP4 간의 상호 변환을 행하는 XMT-A to/from MP4 컨버터를 통해 다양한 기능을 가지는 대화형 콘텐츠를 저작해 줄 수 있다. 또한 MPEG-4 콘텐츠에 MPEG-7 메타데이터를 활용함으로써 향후 T-Library, T-School, T-commerce 등에 있어서 보다 풍부한 양방향 서비스를 제공할 수 있을 것이다.

[참고문헌]

- [1] ISO/IEC Information technology, Part1: Systems, ISO/IEC 14496-1:2002, 2002.03. (MPEG-4 systems)
- [2] WD 2.0 of 14496-1:2002/Amd.X2 on the use of MPEG-7 in MPEG-4 Systems, ISO/IEC JTC1/SC29/WG11, N5290, July 2002
- [3] Considerations on Static Meta-data in MP4/ISO Files, ISO/IEC JTC1/SC29/WG11, M8948, Oct. 2002.