

MPEG-4 기반 대화형 복수시점 영상콘텐츠 저작 시스템

이인재, 기명석, 김옥중, 김규현
한국전자통신연구원 방송미디어연구그룹
ninja@etri.re.kr

Interactive Multiview Contents Authoring System based on MPEG-4

Injae Lee, Myungseok Ki, Wook-Joong Kim and Kyuheon Kim
Digital Broadcasting Research Division,
Electronics and Telecommunications Research Institute

요약

This paper introduces interactive multi-view contents authoring system based on MPEG-4. The MPEG-4 standard, which aims to provide an object based audiovisual coding tool, has been developed to address the emerging needs from communications, interactive broadcasting as well as from mixed service models resulting from technological convergence. Due to the feature of object based coding, it has been considered that MPEG-4 is the most suitable for interactive broadcasting content production. This feature is suitable for creation of the content which provides multiple views of object or scene in interactive manner. In this paper, we categorize the multi-view visual content into two types: panoramic multi-view content and object multi-view content. And design and implementation of the authoring system for interactive multi-view contents is presented. We believe that the proposed method can be effectively used for further deployment of MPEG-4 content to various interactive applications.

1. 서론

최근 통신망의 고도화로 인한 네트워크 대역폭 증가와 컴퓨팅 기술의 급속한 성장에 따른 미디어 환경의 변화는 멀티미디어 서비스에 있어서 다양한 변화를 가져왔다. 이로 인해 사용자는 다양한 형태의 고품질 콘텐츠 및 사용자와의 인터랙티브한 기능을 제공할 수 있는 대화형 콘텐츠를 요구하고 있다. 기존의 멀티미디어 콘텐츠는 양방향 서비스보다는 주로 사용자가 수동적으로 콘텐츠를 수용하는 단방향 서비스를 목표로 제작되었다. 하지만 멀티뷰 콘텐츠(MVVC: Multi-View Visual Content)는 사용자와의 양방향 대화형 성격이 두드러져 사용자의 흥미를 유발하기에 충분하다. 따라서 기존의 일반적인 2D 영상 서비스와는 달리 사용자에게 시점의 자유도를 증가시켜 사용자들이 원하는 시점의 영상을 선택할 수 있도록 함으로써 대화형 기능을 제공할 수 있다. 복수시점 영상콘텐츠를 크게 2가지로 분류하면, 다시점 파노라믹 영상콘텐츠(PMVVC: Panoramic multi-view visual content)와 다시점 오브젝트 영상콘텐츠(OMVVC: object multi-view visual content)로 정의할 수 있다. 다시점 파

노라믹 영상콘텐츠는 2장 이상의 영상을 스티칭 처리함으로써 하나의 넓은 시야각을 제공하는데 그 목적이 있다. 다시점 오브젝트 영상콘텐츠는 집중형 방식을 이용하여 대상이 되는 객체를 중심에 놓고 카메라를 물체에 수렴함으로써 객체를 입체감 있게 360도로 관찰할 수 있는 특징이 있다. 현재 이러한 복수시점 영상 관련 기술에 대한 국제 표준은 존재하지 않으며, 단지 복수시점 영상을 이용하는 다양한 제품들이 출시되어 있다[1][2]. 그러나 각 제품의 기술들이 독자적으로 개발되어 왔기 때문에, 각기 다른 제품들 간의 콘텐츠 공유가 이루어지지 못하고 있다. 따라서 복수시점 영상 관련 산업의 발전을 위해서는 표현 및 압축 방법에 대한 기술표준이 필요하다는 인식을 공유하게 되었고, 이러한 배경으로 현재 MPEG 표준화 그룹에서는 3DAV(3D Audio and Visual)에 대한 규격 개발을 진행하고 있다[3][4]. 하지만 현재까지 복수시점 영상 콘텐츠가 MPEG-4 Systems 기술을 이용하여 어떻게 구성될 수 있는가에 대한 충분한 분석이 이루어지지 못하고 있는 실정이다[5][6].

따라서 본 논문에서는 현재의 MPEG-4 Systems 표준 기술을 기반으로 제공될 수 있는 복수시점 영상 콘텐츠 서비스에 대해 기술하고, MPEG-4 Systems를 기반으로 하는 대화형 복수시점 영상콘텐츠 저작시스템을 설계하

고 구현하였다.

2. 복수시점 영상콘텐츠 서비스

일반적으로 복수시점 영상콘텐츠는 크게 다시점 파노라믹 영상콘텐츠와 다시점 오브젝트 영상콘텐츠 2가지 형태로 정의할 수 있다. 본 절에서는 각 유형별로 요구되어지는 기능들에 대해 기술하였다.

가. 다시점 파노라믹 영상콘텐츠

그림 1(a)는 2차원 형태의 파노라믹 영상을 실린더에 매핑하여 3차원의 입체감을 느낄 수 있도록 구성된 파노라믹 영상콘텐츠를 보여준다. 파노라믹 영상은 한 대의 카메라를 이동 시키거나 다수의 카메라를 통해 획득된 여러 장의 이미지를 합성하여 하나의 넓은 시야각의 영상으로 만든 것을 말하는데, 최근에는 파노라마 카메라를 이용하여 실린더 영상을 획득한 후, 매우 긴 길이의 필름에 이미지를 녹화하여 하나의 영상을 만들거나, 어안(fish eye) 렌즈 혹은 반사경을 이용하여 파노라마 영상을 얻는 방법도 사용되고 있다. 한 대의 카메라나 다수의 카메라를 통해 획득된 영상을 이용하여 파노라마 영상을 만들고자 할 경우 다수의 영상을 하나의 영상으로 이어 만드는 파노라마 모자이크라는 영상처리 기법이 필요하다.

파노라마 영상을 이용하면 사용자가 디지털 방송 시청 중에 임의의 시점을 선택할 경우 해당 시점의 정지영상, 동영상 서비스를 제공받을 수 있다. 이를 위해서는 대화형 기능이 부여되어야 하며, 사용자에게 사실감, 입체감을 제공해 주기 위해서는 텍스처 매핑(Texture mapping) 기술이 필요하다. 텍스처 매핑이란 모델링된 3차원 물체에 2차원 영상을 입히는 방식으로서, 이를 통해 평면적인 2차원 영상에 입체감을 부여해 줄 수 있다.

MPEG-4 systems에서는 대화형 파노라믹 영상콘텐츠 생성을 위해서 기본적으로 객체의 3차원 공간 배치 및 오디오/비주얼 객체를 표현하기 위한 노드들이 제공되어야 한다. 그밖에 대화형 기능을 제공하기 위하여 객체간의 상호 동작, 3차원 탐사 기능, 사용자의 시점 제어 기능, 동영상간의 동기 제어 기능, 3차원 물체 모델링 및 텍스처 매핑 기능 등이 필요하다.

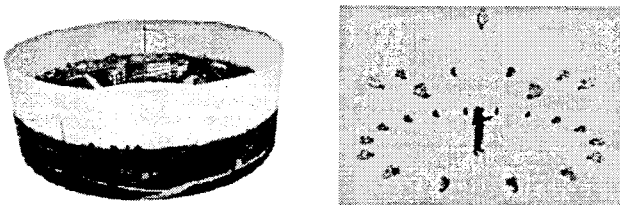


그림 1. 실린더 형태의 파노라믹 영상콘텐츠와 다시점 영상콘텐츠 획득 개념도(25 viewpoint with 3 levels of view-height)

나. 다시점 오브젝트 영상콘텐츠

집중형 방식을 이용한 다시점 오브젝트 영상은 그림 1(b)와 같이 일반적으로 대상이 되는 객체를 중심에 놓고 카메라를 물체에 수렴하는 방식을 말한다. 집중형 방식은 작은 오브젝트를 대상으로 영상 기반 모델링을 할 때 유용하므로 인터넷을 통한 상거래에서 상품을 3차원적으로 보여주는 용도로 사용되고 있으며, 관중에 둘러싸여 이루어지는 스포츠 경기 장면을 제공하는 데도 유용하게 사용될 수 있다.

집중형 방식에서는 각 카메라의 시야가 일치하는 한정된 영역에 속한 객체에 대해서만 제대로 된 3차원 모델링 및 렌더링이 가능하므로, 다시점 비디오 콘텐츠를 획득할 때에도 인터랙티브하게 여러 카메라의 시야(field of view)를 확인한 후 촬영해야 한다. 집중형 방식이 일반 파노라마 영상과 다른 가장 큰 특징은 물체를 입체감 있게 360도로 관찰할 수 있다는 것이다. 모든 임의 중간시점 영상 생성 방법은 물체의 3차원 모델링에 근거한 추정을 전제로 한다.

다시점 오브젝트 영상 서비스를 제공해 주기 위해서는 시점 제어 패널(View Control Panel)과 같이 원하는 영상을 선택할 수 있는 다시점 리스트 정보가 제공되어야 할 것이며, 마우스 관련 이벤트 설정을 통해 대화형 기능이 부여됨으로써 원하는 시점에 해당하는 정지영상/동영상 제공받을 수 있어야 할 것이다. 따라서 기본적으로 오디오/비주얼 객체의 2차원 공간 배치 및 오디오/비주얼 객체를 표현하기 위한 노드들이 제공되어야 할 것이다. 그리고 시점 제어 패널을 통해 영상 선택 기능, 동영상간의 동기 제어 기능 등이 제공되어야 할 것이다.

3. 대화형 복수시점 영상콘텐츠 저작시스템

대화형 복수시점 영상콘텐츠 저작시스템의 구조는 그림 2와 같으며, 사용자 입력 처리와 복수시점 영상콘텐츠 저작시스템의 각 모듈에 대한 전체적인 제어를 담당하는 "사용자 인터페이스부", 편집 정보의 생성/수정/삭제 등을 담당하는 "편집 정보 처리부", 미디어 데이터의 압축과 복원을 담당하는 "미디어 처리부"로 구성된다.

가. 사용자 인터페이스부

사용자 인터페이스부는 보다 용이하게 복수시점 영상 콘텐츠를 생성하고, 다양한 비주얼-오디오 객체를 이용하여 장면을 구성한 후 대화형 기능을 부여하는데 그 목적이 있다. 사용자 인터페이스부에서는 마우스 및 키보드를 통한 사용자의 입력을 해석하고, 적절한 처리를 수행하여 편집/저작 결과를 XMT-A 내부 자료 구조에 저장하며 이를 출력 장치를 통하여 재생하는 기능을 수행한다.

그림 3은 대화형 복수시점 영상콘텐츠 저작시스템의

사용자 인터페이스이다. 그림 3(a)는 파노라믹 영상 생성을 위한 전처리 도구로서, 개별적으로 촬영된 영상들을 영상간 매칭점을 찾아 스티칭하고 컬러 블렌딩 처리하여 한 장의 파노라믹 영상으로 생성시켜 준다. 이렇게 생성된 파노라믹 영상을 그림 3(b)의 메인 윈도우에서 불러들인 후, 다양한 종류의 객체들을 배치시키고 사용자와의 대화적인 조작이 가능한 이벤트들을 설정해 줌으로써 다시점 파노라믹 영상을 이용한 대화형 복수시점 영상콘텐츠를 생성시켜 줄 수 있다. 그림 3(d)는 다시점 오브젝트 영상을 생성하기 위한 사용자 인터페이스로서, 그림 3(c)에 나타난 서브창을 통해 다양한 각도에서 촬영된 영상들을 입력받아 배열하고, 각 영상간의 상하좌우 상관관계 설정한 후 다시점 오브젝트 영상을 컨트롤하기 위해 필요한 컨트롤 버튼 등을 설정해 주면 버튼을 통해 다시점 오브젝트 영상들을 컨트롤할 수 있는 스트립트가 자동으로 생성된다. 그리고 그림 3(d)의 메인창에서는 각 객체들을 배치시키고, 필요하다면 다양한 비디오/오디오 객체를 삽입하며 대화형 조작을 위한 이벤트 설정 등을 통해 다시점 오브젝트 영상을 이용한 대화형 복수시점 영상콘텐츠 서비스를 제공할 수 있다. 이와 같이 다시점 파노라믹/오브젝트 영상 생성 및 대화형 이벤트 설정을 통한 대화형 복수시점 영상콘텐츠 생성 등을 간편하게 저작할 수 있는 구조로 사용자 인터페이스가 설계되었다.

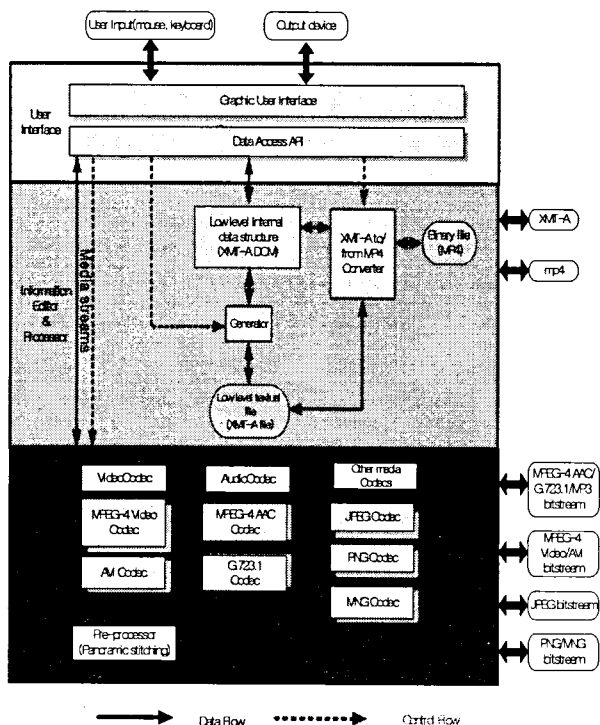
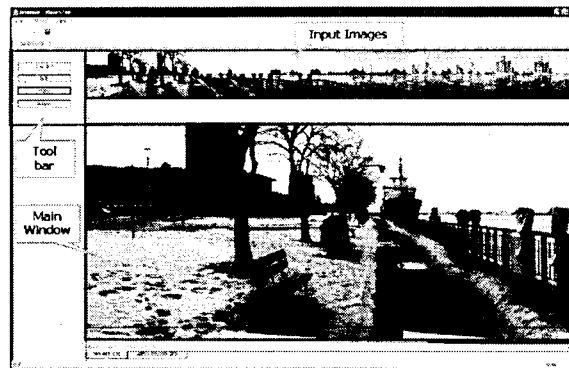
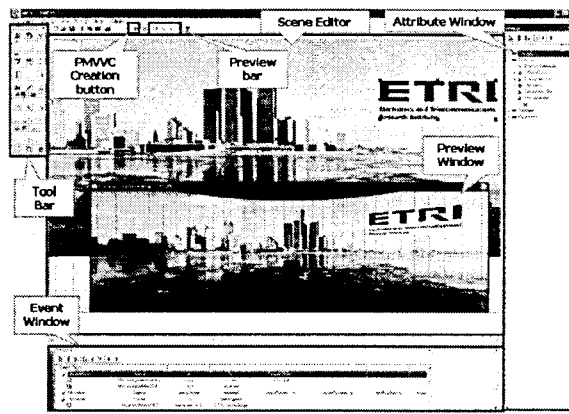


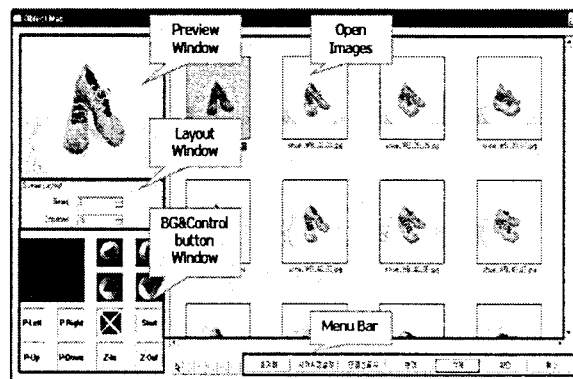
그림 2. 대화형 복수시점 영상콘텐츠 저작시스템의 구조도



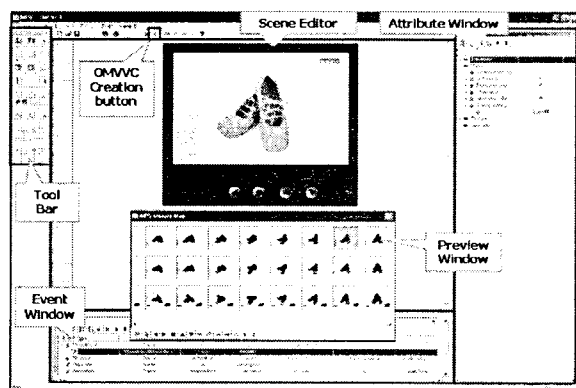
(a)



(b)



(c)



(d)

그림 3. 대화형 복수시점 영상콘텐츠 저작시스템의 사용자 인터페이스

나. 저작정보 처리부

저작정보 처리부는 XMT-A 내부 자료 구조에 저장된 편집 정보를 XMT-A 파일로 저장하고, 필요에 따라 XMT-A 파일과 XMT-A 콘텐츠가 참조하는 외부 미디어 객체들로부터 mp4 파일을 생성하는 기능을 수행한다. XMT는 MPEG-4 장면 기술과 객체 기술을 텍스트 형식으로 기술하기 위한 프레임워크이다. MPEG-4의 콘텐츠 저작언어인 XMT-A를 사용하여 저작된 대화형 콘텐츠를 스트리밍하거나, 배포하기 위해서는 XMT-A 파일을 이진 포맷인 MP4 파일로 변환하여야 한다. 반면에 MP4 파일의 재편집을 위해서 MP4 파일을 XMT-A 파일로 변환할 수 있어야 한다 따라서 대화형 복수시점 영상콘텐츠 저작시스템에서는 이를 처리하기 위한 XMT-A to/from MP4 컨버터 모듈을 구현하였다. 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.

다. 미디어 처리부

미디어 처리부는 동영상, 정지영상, 오디오 및 기타 미디어들을 부호화/복호화 하는 기능을 수행한다. 즉, 저작 시스템에 raw 파일 형태의 미디어가 입력으로 들어올 경우, 해당 미디어를 원하는 형태로 부호화하여 사용할 수 있으며, 부호화된 미디어가 입력으로 들어올 경우 이를 복호화하여 해당 미디어를 미리볼 수 있는 기능을 수행한다.

4. 실험결과 및 결론

현재 복수시점 영상 관련 기술에 대한 국제 표준은 존재하지 않으며, 복수시점 영상을 이용하는 다양한 제품들이 출시되어 있다. 그러나 각 제품의 기술들이 독자적으로 개발되어 왔기 때문에, 각기 다른 제품들 간의 콘텐츠 공유가 이루어지지 못하고 있다. 따라서 복수시점 영상 관련 산업의 발전을 위해서는 표현 및 압축 방법에 대한 기술표준이 필요하다. 따라서 MPEG-4 Systems 규격은 현재까지 개발된 표준들 중에서 복수시점 영상 서비스가 제공하려고 하는 목적에 가장 잘 부합하는 기능을 제공하고 있다.

본 논문에서는 복수시점 영상 서비스를 위한 2가지 유형을 정의하고, 사용자에게 해당 서비스를 제공할 수 있는 대화형 복수시점 영상콘텐츠 생성을 위한 MPEG-4 기반 저작 시스템을 설계하고 구현하였다. 그림 4는 대화형 복수시점 영상콘텐츠 저작 시스템을 통해 구현된 결

과를 보여준다. 그림 4(a), (b)는 파노라믹 배경 위에 정지영상, 비디오, 오디오, 그래픽 객체가 존재하는 대화형 다시점 파노라믹 영상 서비스로서, 마우스 이벤트를 통해 해당 비디오나 정지영상을 삽입/삭제할 수 있다. 그림 4(c), (d), (e)는 대화형 다시점 오브젝트 영상 서비스로서, 특정 객체를 다른 시점에서 바라보았을 때의 영상 및 해당 객체에 대한 상세정보를 마우스 관련 이벤트를 통해 제공받을 수 있다.

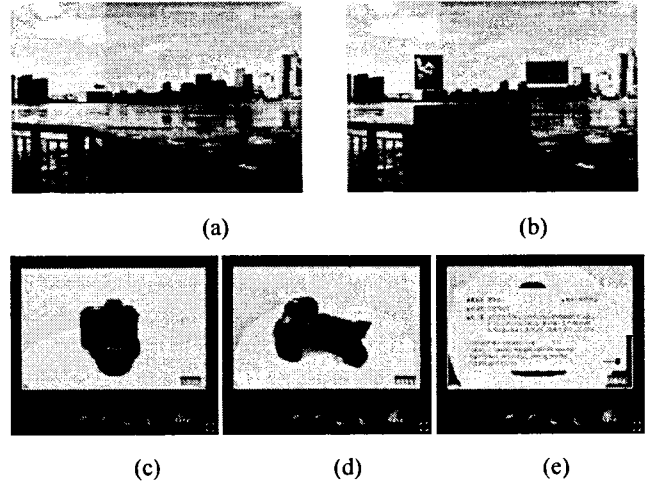


그림 4. 대화형 복수시점 영상콘텐츠 저작 시스템을 통한 저작 결과 영상

대화형 복수시점 영상 서비스는 사용자 시점의 자유도를 증가시키고 보다 넓은 범위의 뷰를 제공함으로써 나날이 증가되는 소비자의 다양한 욕구를 충족시킬 수 있을 것이며, 이는 향후 디지털 방송 분야 및 DMB 데이터 서비스에도 활용될 수 있을 것이다.

참고문헌

- [1] Apple QuickTime VR Authoring Studio. <http://www.apple.com/quicktime/qtvr>
- [2] Photovista Virtual Tour. <http://iseemedia.com>
- [3] S. K. Nayar.: Catadioptric Omnidirectional Camera, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, (1997) 17-19
- [4] Smolic, A., McCutchen, D.: 3DAV Exploration of Video-Based Rendering Technology in MPEG, IEEE trans. on Circuits and Systems for Video Technology, Vol. 14, No. 3, (2004) 348-356
- [5] ISO/IEC JTC1/SC29/WG11 N5877, Applications and Requirements for 3DAV. (2003)
- [6] ISO/IEC JTC1/SC29/WG11 N4668, MPEG-4 Overview. <http://www.chiariglione.org/mpeg/standards/mpeg-4/mp4eg-4.htm>, (2002)
- [6] ISO/IEC 14496-1 Part1: Systems (March 2002)