

특집논문-06-11-4-08

대화형 복수시점 영상콘텐츠 저작시스템 설계 및 구현

이인재^{a)†}, 최진수^{a)}, 기명석^{a)}, 정세윤^{a)}, 문경애^{a)}, 홍진우^{a)}

Design and Implementation of Interactive Multi-view Visual Contents Authoring System

Injae Lee^{a)†}, Jin Soo Choi^{a)}, Myungseok Ki^{a)}, Seyoon Jeong^{a)}, Kyungae Moon^{a)}, and Jinwoo Hong^{a)}

요 약

본 논문에서는 MPEG-4 Systems 기반 복수시점 영상 서비스를 위한 저작 기술에 대해 제안한다. MPEG-4는 이전의 MPEG-1/2 등의 비디오 정보 압축 규격과 달리, 사용자와 콘텐츠 간의 객체 기반 대화형(interactivity) 기능에 대한 다양한 방법을 포함하고 있다. 따라서 MPEG-4 Systems 규격은 현재까지 개발된 표준들 중에서 복수시점 영상 서비스가 제공하려고 하는 목적에 가장 잘 부합하는 기능을 제공하고 있다. 본 논문에서는 MPEG-4 Systems 표준 기술을 기반으로 제공될 수 있는 복수시점 영상 서비스에 대해 기술하였고, 해당 서비스를 제공하기 위한 MPEG-4 장면 기술(scene description) 프로파일을 제안하였다. 또한 제안된 프로파일 에 기반을 둔 대화형 복수시점 영상콘텐츠 저작시스템을 설계하고 구현하였으며, 실험결과를 통해 그 가능성을 보여 주었다.

Abstract

This paper describes issues and consideration on authoring of interactive multi-view visual content based on MPEG-4. The issues include types of multi-view visual content; scene composition for rendering; functionalities for user-interaction; and multi-view visual content file format. 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, the use of MPEG-4 can resolve the format diversity problem of multi-view visual contents while providing high interactivity to users. Throughout this paper, we will present which issues need to be determined and how they can be realized by means of MPEG-4 Systems.

Keywords : 대화형, 복수시점, 저작, MPEG-4, 3D

I. 서 론

복수시점 영상은 한 대 이상의 카메라를 통해 촬영된 영

상들을 기하학적으로 교정하고 공간적인 합성을 통하여 여러 방향의 다양한 시점을 사용자에게 제공할 수 있는 고품질의 다시점 영상 및 파노라믹 영상을 말한다. 기존의 일반적인 영상 서비스는 서비스 제공자의 의도에 따라 시청자가 사전에 고정된 시점의 영상만을 볼 수 있다. 그에 반해, 복수시점 영상서비스는 실생활에서 사람이 보는 것과 유사한 방식으로 사용자의 시점을 자유롭게 선택할 수 있어 시청자에게 효과적으로 실제감과 몰입감을 제공할 수 있다.

a) 한국전자통신연구원 전파방송연구단 방송미디어연구그룹
Broadcasting Research Media Group, Radio & Broadcasting Research Division, ETRI

† 교신저자 : 이인재(ninja@etri.re.kr)

※ 본 연구는 정보통신부의 “지능형 통합정보방송 기술 개발 (SmarTV)” 사업의 일환으로 수행된 연구 결과입니다.

최근 이러한 복수시점 영상을 기반으로 하는 다양한 응용서비스가 출현하고 있다. 한 예로 유럽 최대의 인터넷 서비스 제공사업자(Internet Service Provider)인 T-Online은 ‘T-Online Vision’이라는 IPTV 서비스를 통해 유럽의 축구 경기 슛 장면에서 각기 다른 각도에서 촬영된 4가지 영상을 제공함으로써 시청자가 원하는 시점의 영상을 선택하여 볼 수 있는 멀티트랙 서비스를 제공한 바 있다. 또한 일본 Sony는 일본 여성 그룹의 뮤직비디오를 파노라믹 비디오로 촬영한 후 조이스틱을 이용하여 원하는 시점의 영상을 자유로이 선택하여 볼 수 있는 DVD 타이틀도 제작하였다.

한편 아날로그 TV방송이 디지털 TV방송으로 옮겨감에 따라 고품질의 AV 서비스를 제공할 수 있을 뿐만 아니라, 양방향 채널을 통해 시청자가 AV 방송 프로그램에 직간접적으로 참여할 수 있는 대화형 서비스로 발전하고 있다. 이는 복수시점 영상 서비스의 경우에도 예외가 될 수 없다. 따라서 단순 AV 위주의 복수시점 영상 서비스에서 탈피하여 양방향 대화형 기능을 추가적으로 제공함으로써, 사용자에게 시점 선택의 자유도를 제공해 줄 수 있을 뿐만 아니라 더욱 효과적이고 능동적으로 부가 정보 서비스를 제공할 수 있을 것이다.

그림 1은 대화형 복수시점 영상 서비스를 제공하기 위한

시스템의 형상을 나타낸 것으로서, 크게 5가지로 나누어 볼 수 있다. 먼저 복수시점 영상을 카메라로 획득(acquisition)하는 과정, 획득된 복수시점 영상을 처리하여 부호화(compression)하는 과정, 시점별 영상을 구성하고 부가 콘텐츠를 활용하여 대화형 기능을 제공하기 위해 저작/편집(scene composition)하는 과정, 전송 매체의 특성을 고려하여 전송하는 과정 그리고 사용자의 요구에 따라 원하는 시점의 영상 및 관련 부가 콘텐츠를 제공하는 재생(rendering) 과정으로 이루어진다^{[1][2][3]}.

본 논문은 이상의 과정 중에서 대화형 복수시점 영상 콘텐츠를 저작하고 편집하기 위한 기술을 제안한다. 대화형 복수시점 영상콘텐츠를 저작하기 위해서는 먼저 복수시점 영상을 어떤 형태로 구성할 것인지가 고려되어야 한다. 즉, 3차원 형태인 복수시점 영상을 구성한다는 것은 최종적으로 2차원 영상 장치를 통해 재생될 것을 고려하여 여러 대의 카메라로부터 입력받은 영상들을 합성 처리하여 3차원 공간상의 원통 형태 또는 반구 형태로 영상을 구성하고 여기에 시점 제어 기능을 부여함을 의미한다. 그리고 복수시점 영상 및 부가 콘텐츠에 어떤 방식으로 대화형 기능을 부여할 것인지, 기존 기술과의 호환성은 유지되는지도 중요하게 고려되어야 한다. 여기서 대화형 기능은 복수시점

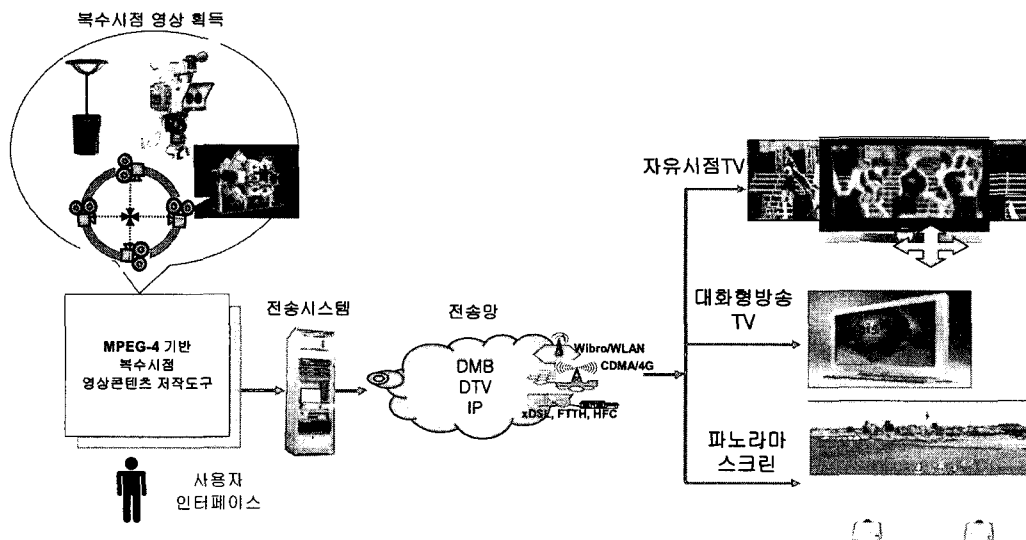


그림 1. 복수시점 영상 서비스를 제공하기 위한 시스템 형상
 Fig. 1. System configuration for interactive multi-view visual service

영상 내에 부가 콘텐츠를 시공간적으로 위치시키고 사용자의 요구에 의해 미리 정해진 동작 절차에 따라 부가 콘텐츠가 동작할 수 있도록 함을 의미한다.

이상의 요구사항을 고려하여 본 논문에서는 MPEG-4 표준을 채용하여 대화형 복수시점 영상콘텐츠 저작 시스템을 설계하고 구현하였다. MPEG-4 표준은 다양한 형태의 오디오/비디오 서비스를 위한 멀티미디어 데이터를 유연성 있게 부호화하기 위한 국제 표준 규격으로서, 사용자와 콘텐츠와의 객체 기반 대화형(interactivity) 기능에 대한 다양한 방법을 포함하고 있다^[4]. 또한 멀티미디어 콘텐츠의 저장, 관리, 배포에 용이한 파일 포맷인 XMT(eXtensible MPEG-4 Textual format)와 MPEG-4 파일 포맷 (MP4)를 제공하고 있다^{[5][6]}. 본 논문에서는 대화형 복수시점 영상서비스의 응용 분야 및 요구사항을 분석함으로써 적합한 프로파일을 제안하고, 이에 따라 구현한 결과를 제시한다.

본 논문의 구성은 다음과 같다. II장에서는 현재의 MPEG-4 Systems 표준 기술을 기반으로 제공될 수 있는 복수시점 영상 서비스에 대해 기술한다. 그리고 이러한 분석을 통하여 현재 규격의 관점에서 향후 추가적으로 제공되어야 할 부분들에 대해 제안한다. III장에서는 제안한 프로파일에 기반을 둔 대화형 복수시점 영상콘텐츠 저작시스템을 설계한다. IV 장에서는 제안한 저작시스템을 통해 제작된 대화형 복수시점 영상콘텐츠를 설명하며, 마지막으로 V장에서는 결론 및 향후 연구 방향을 제안한다.

II. 대화형 복수시점 영상콘텐츠 시스템의 개요

본 장에서는 사용자 시점의 자유도를 증가시켜 주고 실제감을 느낄 수 있는 다시점 영상 및 파노라믹 영상을 이용한 대화형 서비스에 대해 중점적으로 기술하며, 이러한 서비스를 실현시키기 위해 필요한 저작시스템의 요구사항 및 MPEG-4 Systems 활용방안에 대해 살펴본다.

1. 대화형 복수시점 영상 서비스의 정의

1.1 다시점 영상서비스

다시점 영상이란 그림 2 (a)처럼 다수의 카메라를 다양한 위치에 설치하고, 특정 피사체의 다시점 영상을 획득하여 해당 영상을 기반으로 다양한 시점의 영상을 재생할 수 있는 방식이다. 다시점 영상을 활용한다면 사용자가 디지털 방송 시청 중에 임의의 시점을 선택할 경우 해당 시점과 관련된 정지영상, 동영상을 제공해 주는 서비스가 가능하다. 예를 들어 다시점 영상 서비스를 통해 사용자는 특정 물체를 가까이에서 관찰하는 것처럼 다양한 방향에서 살펴볼 수 있다. 즉, 사용자는 특정 물체의 한 단면이 아니라, 앞, 뒤, 옆, 위, 아래 다양한 각도에서 해당 객체를 살펴볼 수 있다. 야구 경기를 시청하는 경우라면 투수, 타자, 포수, 1, 2, 3루 등 다양한 시점 중 원하는 시점을 선택하여 볼 수 있으며, 더 나아가 투수도 앞뒤좌우 다양한 각도에서 지켜볼 수 있을 것이다. 다시점 영상 서비스는 조형물, 전자제

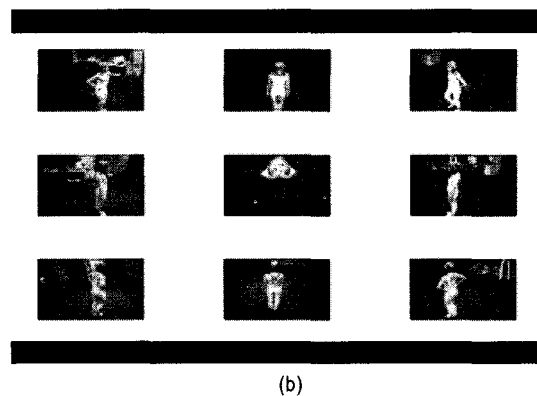
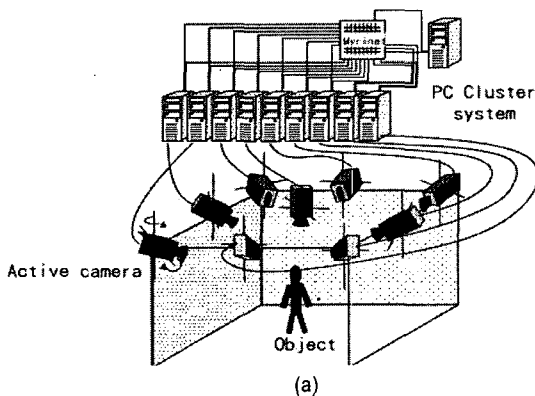


그림 2. 다시점 영상 획득

Fig. 2. Concentric camera setup for OVC(Object Visual Content) capturing

품, 박물관의 유물, 전람회의 조각작품, 자동차, 스포츠 경기, 콘서트 등 다양한 분야에서 유용하게 활용될 수 있다.

1.2. 파노라믹 영상 서비스

파노라믹 영상이란 한 대의 카메라를 이동시키거나 다수의 카메라를 통해 획득된 여러 장의 이미지를 합성하여 하나의 넓은 시야각의 영상으로 만드는 것이다. 파노라믹 영상을 통해 사용자가 디지털 방송 시청 중에 임의의 시점을 선택하게 되면 해당 시점의 정지영상, 동영상 서비스를 제공받을 수 있다. 이는 사용자의 요청에 의해 파노라믹 영상을 적절히 소비할 수 있도록 사용자가 특정 공간에서 주위 배경을 둘러보는 효과를 주는 기술이다. 즉, 최대 좌, 우 360도 혹은 상,하 180도 등 넓은 시야각까지 사용자가 볼 수 있도록 함으로써 특정 장소에 가지 않더라도 그곳에 와 있는 느낌을 제공할 수 있는 서비스이다. 이는 모델하우스, 놀이공원, 호텔, 관공서, 공항, 주요 관광지, 유적지 등 주로 건물의 내부와 외부 그리고 특정 장소나 지역을 설명하는데 효과적으로 쓰일 수 있다.

2. 대화형 복수시점 영상콘텐츠 저작시스템의 요구사항

2.1. 복수시점 영상 처리 기능

대화형 복수시점 영상 서비스를 지원해 주기 위해서는 사용자들이 복수시점 영상을 손쉽게 다룰 수 있도록 지원해 주어야 한다. 따라서 저작시스템에서는 다양한 각도에

서 촬영된 복수개의 영상을 입력받아 하나의 파노라믹 영상으로 스티칭해 주는 파노라믹 영상 처리 기능을 제공하여야 한다. 또한 입체감을 느낄 수 있도록 파노라믹 영상을 기하객체에 텍스처 매핑하는 기능도 지원해 주어야 한다. 다시점 영상의 경우 각각의 다시점 영상들을 자유롭게 선택할 수 있도록 영상간의 자동 전환 기능(play), 영상 자동 전환 멈춤 기능(stop), 영상 수동 전환 기능(backward, forward) 등이 제공되어야 한다.

2.2. 사용자 대화형 기능

다시점 및 파노라믹 영상은 기본적으로 사용자가 원하는 시점의 영상을 선택하여 볼 수 있는 기능을 제공하여야 한다. 또한 복수시점 영상 외에 사용자에게 더욱 풍부한 서비스를 지원해 주기 위하여 사용자와의 대화적인 조작성이 가능한 부가콘텐츠를 제공해 주어야 한다. 부가콘텐츠에 대한 대화형 기능은 특정 객체를 선택했을 경우 관련된 웹페이지로 이동하는 웹링크 기능, 시간에 따라 객체가 나타나거나 사라지는 타임 이벤트 기능, 외부 입력장치를 통한 사용자 입력에 의해 특정 객체가 나타나거나 사라지는 마우스 이벤트 기능, 동영상/오디오 객체를 play, stop, pause 할 수 있는 미디어 제어 기능, 특정 객체에 대한 회전, 위치 변환, 크기 변환 등의 애니메이션 기능이 있으며, 상기 기술한 이벤트들은 특정 객체에 두가지 이상의 이벤트가 동시에 혹은 순차적으로 적용될 수 있다. 본 저작시스템은 표 1에 기술된 바와 같이 다양한 멀티미디어 포맷 및 대화형

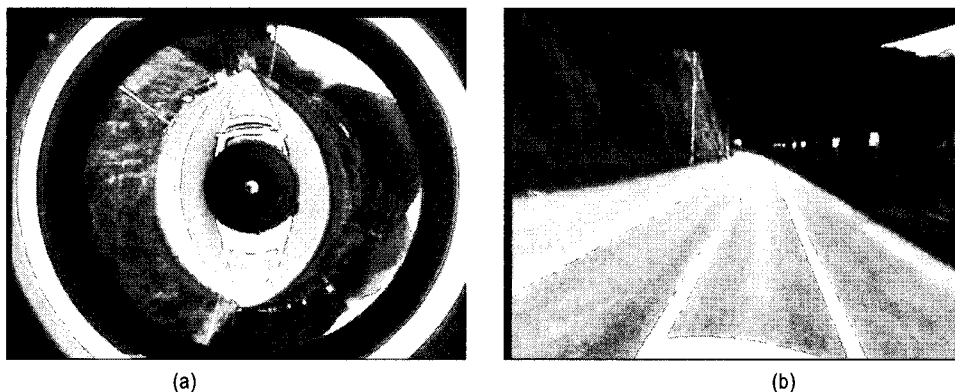


그림 3. 파노라믹 영상 획득 (a) 옴니 카메라로 촬영된 파노라믹 영상 (b) (a)로부터 획득된 임의시점 파노라믹 비디오
 Fig. 3. Panoramic video acquisition

기능을 지원하고자 한다.

표 1. 객체 기반 대화형 기능 제공을 위한 미디어 객체 타입
Table 1. Media object types for object-based interaction

종 류	내 용
복수시점 영상	<ul style="list-style-type: none"> ○ 파노라믹 정지영상 (JPEG) ○ 파노라믹 동영상 (AVI) ○ 다시점 정지영상 (JPEG)
부가 콘텐츠	<ul style="list-style-type: none"> ○ 동영상 (MPEG-4 simple video) ○ 정지영상 (JPEG, PNG) ○ 오디오 (MPEG-4 AAC) ○ 텍스트 (영문/한글 지원) ○ 그래픽(rectangle, circle, box, cylinder, sphere, etc)
콘텐츠 제어 및 이벤트 기능	<ul style="list-style-type: none"> ○ 사용자 시점 전환 ○ 웹 링크 ○ 타임 이벤트 ○ 마우스 이벤트 (객체 삽입/삭제) ○ 미디어 제어 (play/stop/pause) ○ 애니메이션 (회전, 위치 변환, 색깔 변환, 크기 변환 등)
출력	<ul style="list-style-type: none"> ○ 텍스트 형식 (XMT-A) ○ 이진 형식 (MP4)

2.3. 장면 표현 및 압축 기능

앞서 기술한 바와 같이 대화형 기능을 제공하는 복수시점 영상콘텐츠를 지원하기 위해서는 각 미디어 객체 정보 및 관련 이벤트에 대한 장면 표현 및 압축 기술이 요구되어 진다.

대화형 복수시점 영상콘텐츠 저작시스템에서는 저작 및 재편집이 용이하도록 저장된 장면이 텍스트 형식의 파일로 출력될 수 있어야 하며, 콘텐츠 관리와 전송이 용이하도록 이진 형식의 파일로도 출력될 수 있어야 한다. 국제 표준 규격인 MPEG-4는 장면 표현을 텍스트 형식으로 기술하기 위한 XMT 파일 포맷이 존재하며, 콘텐츠 배분 및 스트리밍 등을 고려한 이진 형식의 MP4 파일 포맷이 존재하므로 본 저작시스템에서는 장면 저작에 대한 결과로서 XMT와 MP4 파일 포맷을 지원하고자 한다.

3. 대화형 복수시점 영상 서비스를 위한 MPEG-4 장면 구성 기술

국제 표준 규격인 MPEG-4는 복수시점 영상콘텐츠 및

부가콘텐츠에 대화형 기능을 부여할 수 있는 다양한 방법들을 포함하고 있다. 따라서 MPEG-4는 현재까지 개발된 표준들 중에서 복수시점 영상 서비스가 제공하려고 하는 목적에 가장 부합하는 기능을 제공하고 있으며, 이를 잘 활용한다면 사용자에게 대화형 기능을 제공하는 복수시점 영상콘텐츠를 기술표준에 맞게 제공할 수 있을 것이다. 따라서 본 절에서는 대화형 복수시점 영상 서비스를 실현시키기 위해 필요한 MPEG-4 Systems 활용방안에 대해 살펴본다.

3.1. 다시점 영상 서비스를 위한 MPEG-4 장면 구성 기술

MPEG-4 Systems에서는 시스템 설계자와 콘텐츠 제공자 간의 호환성을 보장해 주기 위해 프로파일과 레벨을 정의한다. 프로파일은 각 응용 서비스 분야 별로 해당 분야에 유용한 노드들의 서브셋을 정의해 둔 것이며, 레벨은 해당 프로파일에 있어 오디오/비주얼 객체의 수 및 복잡도의 정도 등에 대해 제약을 둔 것이다.

장면 구성 관련 프로파일은 크게 장면 그래프 프로파일과 그래픽스 프로파일로 나눌 수 있다. 장면 그래프 프로파일은 장면을 구성하는 오디오/비주얼 객체의 시공간적 위치 및 객체간의 동작 등 제공 기능에 따라 필요한 노드들을 분류해 둔 것이다. 그래픽스 프로파일은 그래픽 표현과 관련된 BIFS 노드들을 분류해 둔 것이다.

그림 4에서 보여지는 것처럼 다시점 영상 서비스를 제공해 주기 위해서는 사용자의 시점을 자유롭게 전환시킬 수 있는 시점 제어 패널(View Control Panel)이 제공되어야 하며, 사용자 입력을 받아들일 수 있도록 대화형 기능을 부여함으로써 사용자의 요청이 있을 때 원하는 시점에 해당하는 정지영상/동영상을 제공해 주어야 한다. 따라서 사용자 시점의 자유 전환을 위한 노드들이 지원되어야 한다. 또한 오디오/비주얼 객체의 2차원 공간 배치 및 오디오/비주얼 객체를 표현하기 위한 노드들이 제공되어야 하며, 동영상 서비스를 지원할 경우 각 동영상 간의 동기를 맞추기 위해 필요한 노드들도 제공되어야 한다.

기존의 장면 그래프 프로파일에는 복수시점 영상 서비스에 적합한 프로파일이 존재하지 않으므로 다시점 영상 서비스를 위해 필요한 프로파일을 표 2와 같이 제안하였다.

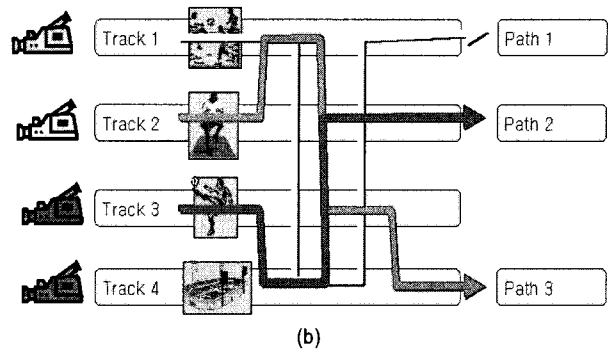
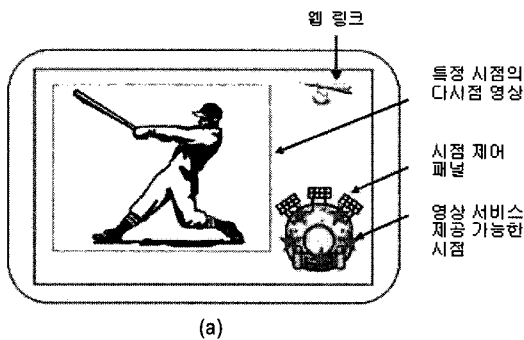


그림 4. 다시점 영상 서비스의 예
Fig. 4. Examples of OVC

그래픽스 프로파일의 경우 다시점 영상 서비스를 위해 기존의 'simple2D+Text' 혹은 'core2D' 프로파일을 이용해도 무관하다. 제안된 'simplified Core 2D' 프로파일은 앞서 설명한 것처럼 2차원 장면 배치, 미디어 표현, 객체간의 상호동작, 동영상 동기 제어 기능 등을 제공해 줄 수 있으며, 이는 다시점 영상 서비스를 목표로 한다.

3.2 파노라믹 영상 서비스를 위한 MPEG-4 장면 구성 기술
그림 5와 같이 파노라믹 영상 서비스를 제공하기 위해서는 파노라믹 영상의 특정 선택영역에 대한 정지영상/동영상을 제공할 수 있어야 하며, 이를 위한 대화형 기능이 부여되어야 한다. 또한 사용자에게 실제감, 입체감을 제공해주기 위해서는 텍스처 매핑 기술이 필요하다.

표 2. 다시점 영상 서비스를 위해 제안된 프로파일
Table 2. Proposed list of nodes for OVC scene composition

프로파일	기능	노드
장면 그래픽스 프로파일 Simplified Core 2D	2차원 장면 배치	Group, Transform2D
	객체간 상호 동작	InputSensor, TimeSensor, PlaneSenser2D, Valuator, Conditional, ROUTE
	영상 선택	TouchSensor, Switch
	동영상 동기 제어	MediaControl, MediaSensor, Script
	웹링크 접속	Anchor
	미디어 표현	Sound2D, AudioSource, Bitmap, ImageTexture, MovieTexture
그래픽스 프로파일 Simple2D+Text	2차원 기하객체 및 텍스트	Background2D, Appearance, Material2D, Shape, Bitmap, Rectangle, Text, fontStyle

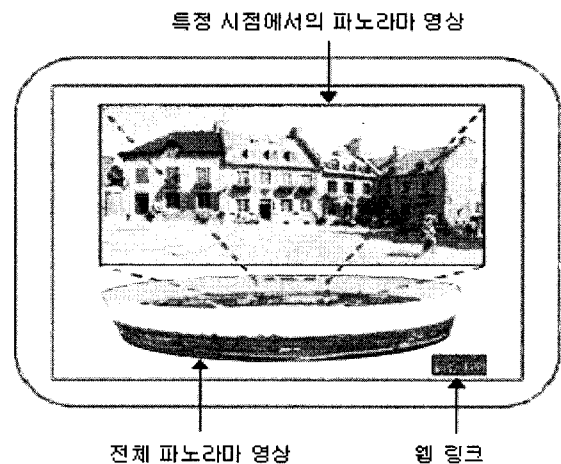


그림 5. 파노라믹 비디오 서비스의 예
Fig. 5. Examples of PVC (Panoramic Visual Content)

텍스처 매핑이란 모델링된 3차원 물체에 2차원 영상을 입히는 방식으로서, 이를 통해 평면적인 2차원 영상에 입체감을 부여해 줄 수 있다. 평면적인 2차원 영상을 3차원 물체에 매핑시키기 위해서는 정육면체 매핑, 원기둥 매핑, 구 매핑 등 다양한 방법이 존재하며, 그림 5의 경우 원기둥 매

핑에 해당된다.

파노라믹 영상 서비스를 위해서는 기본적으로 객체의 3차원 공간 배치 및 오디오/비주얼 객체를 표현하기 위한 노드들이 제공되어야 한다. 그리고 3차원 공간상에서 자유롭게 사용자 시점을 전환하는 기능을 지원해 주기 위한 노드들이 제공되어야 한다. 이때, 파노라믹 장면을 구성하는 영상이 단일 동영상이라 아니라 다수의 패치들로 구성된 동영상의 집합일 경우, 각 동영상 간의 동기를 맞춰줄 수 있는 노드들이 제공되어야 한다. 그 밖에 입체감을 제공해 주기 위하여 3차원 기하객체 모델링 및 텍스처 매핑 기능을 지원해 주는 노드들이 필요하다. 기존의 MPEG-4 Systems에서는 2차원 장면 구성과 관련하여 다양한 프로파일이 정의되어 있으나, 3차원의 경우 '3D audio'와 'Complete' 프로파일만이 정의되어 있다. 즉, 3차원 장면 구성의 경우 특정 응용 분야를 목표로 하는 프로파일에 대한 정의가 미흡한

실정이다. 따라서 본 논문에서는 표 3과 같이 파노라믹 영상 서비스를 목표로 장면 그래프 프로파일에는 'Simplified Core 3D' 프로파일을, 그래픽스 프로파일에는 'Simplified Advanced 3D' 프로파일을 제안하였다. 'Simplified Core 3D' 프로파일은 앞서 설명한 것처럼 3차원 장면 배치, 미디어 표현, 객체간의 상호 동작, 동영상 동기 제어 기능 및 자유 시점 전환 기능 등을 제공해 줄 수 있고, 'Simplified Advanced 3D' 프로파일은 3차원 물체 모델링 및 텍스처 매핑 기능을 제공해 줄 수 있다.

Ⅲ. 대화형 복수시점 영상콘텐츠 저작시스템 설계 및 구현

대화형 복수시점 영상콘텐츠 저작시스템은 복수시점 영상을 통해 다양한 시점의 영상을 사용자에게 제공함으로써 기존의 한정되고 지협적인 뷰의 제공에서 벗어나 사용자 시점의 자유도를 증가시키고 더욱 넓은 범위의 뷰를 제공해 줄 수 있는 저작시스템이다. 또한 사용자의 요구에 따라

표 3. 파노라믹 영상 서비스를 위해 제안된 프로파일
Table 3. Proposed list of nodes for PVC scene composition

프로파일		기능	노드
장면 그래프 프로파일	Simplified Core 3D	3차원 장면 배치	Group, Transform
		객체간 상호 동작	InputSensor, TouchSensor, TimeSensor, VisibilitySensor, Conditional, ROUTE
		자유 시점 전환	Viewpoint, NavigationInfo
		동영상 동기 제어	MediaControl, MediaSensor, Script
		웹링크 접속	Anchor
		미디어 표현	Sound, AudioSource, ImageTexture, MovieTexture
그래픽스 프로파일	Simplified Advanced 3D	3차원 기하객체 모델링	Shape, Appearance, Material, IndexedFaceSet, Coordinate
		텍스처 매핑	TextureCoordinate
		3차원 기하객체 및 텍스트	Box, Cylinder, Sphere, Text, fontStyle

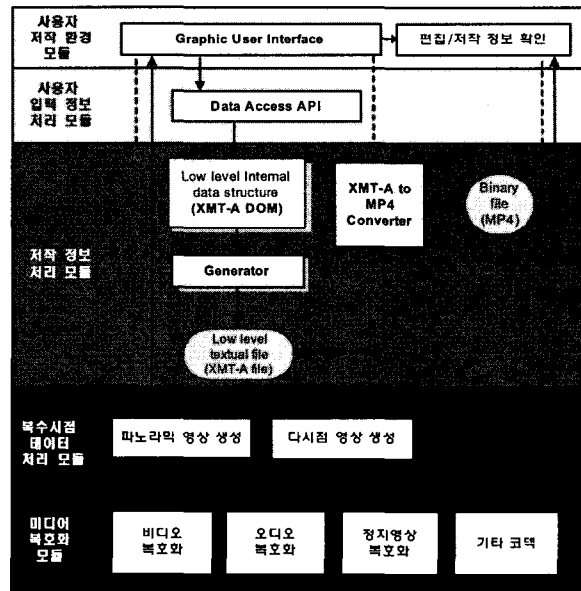


그림 6. 대화형 복수시점 영상콘텐츠 저작시스템의 구조도
Fig. 6. Architecture of the interactive multi-view visual contents authoring system

복수시점 영상 뿐 아니라 이와 관련된 부가 콘텐츠 및 독립적인 미디어 콘텐츠들도 추가 제공해 줄 수 있다. 그림 6은 대화형 복수시점 영상콘텐츠 저작시스템의 구조를 나타낸 것으로서, 사용자 입력 정보 처리와 대화형 복수시점 영상 콘텐츠 저작시스템의 각 모듈에 대한 전체적인 제어를 담당하는 "사용자 저작 환경 모듈", 사용자 저작 정보를 저작 정보 처리 모듈의 규격에 맞게 변환하여 전달하는 "사용자 입력 정보 처리 모듈", 저작 정보의 생성/수정/삭제 등을 담당하는 "저작 정보 처리 모듈", 파노라믹 및 다시점 영상 생성을 위한 전처리 기능을 수행하는 "복수시점 데이터 처리 모듈"과 미디어 데이터의 복원을 담당하는 "미디어 복호화 모듈"로 구성된다.

1. 사용자 저작 환경 모듈

사용자 저작 환경 모듈은 사용자가 손쉽게 복수시점 영상 콘텐츠를 생성하고, 다양한 비주얼-오디오 객체를 이용하여 장면을 구성한 후 대화형 기능을 부여하는데 그 목적이 있다. 사용자 저작 환경 모듈에서는 마우스 및 키보드를 통한 사용자 입력을 해석하고, 적절한 처리를 수행하여 저작 정보를 사용자 입력 정보 처리 모듈로 전달하는 기능을 수행한다. 또한 저작 중인 장면 저작 결과를 출력장치를 통해 미리볼 수 있는 기능을 제공한다.

그림 7은 사용자 저작 환경 모듈의 사용자 인터페이스로서, 복수시점 영상 생성 및 다양한 미디어 객체를 입력받기 위한 메뉴 및 툴바, 공간적인 객체 단위 편집/저작을 위해 장면 편집창, 각 객체에 대한 노드 속성을 편집하기 위한 속성창, 마우스/애니메이션/웹링크 등 다양한 이벤트들을 설정하고 편집해 주는 이벤트창, 저작 결과를 미리보기 위한 미리보기창 등으로 구성되어 있다.

복수시점 영상은 사용자 인터페이스의 복수시점(다시점/파노라믹) 영상 생성 버튼을 통해 활성화되는 복수시점 데이터 처리 모듈을 통해 생성된다. 해당 모듈을 통해 생성된 복수시점 영상을 장면 편집창으로 불러들인 후, 동영상/정지영상/그래픽/텍스트 등 다양한 객체들을 배치시키고 사용자와의 대화적인 조작이 가능한 이벤트들을 설정해 줌으로써 다시점 영상 및 파노라믹 영상을 이용한 대화형 복수

시점 영상콘텐츠를 저작할 수 있다.

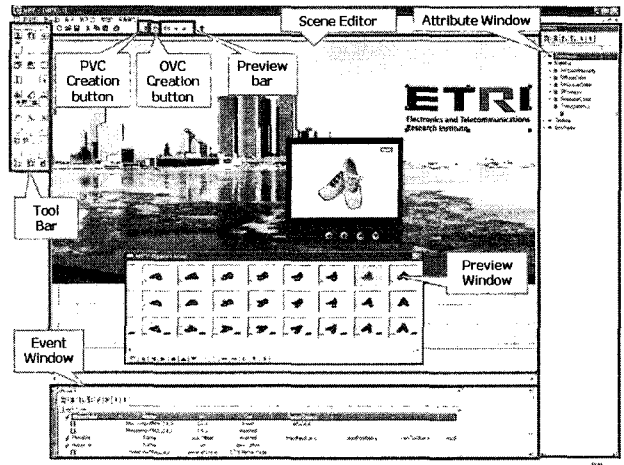


그림 7. 사용자 저작 환경 모듈의 사용자 인터페이스
Fig. 7. GUI of the authoring environment module

2. 사용자 입력 정보 처리 모듈

사용자 입력 정보 처리 모듈은 XML DOM 규격을 따르는 XMT-A 내부 자료구조를 생성하기 위해 사용자 저작 환경 모듈의 사용자 인터페이스를 통해 전달받은 콘텐츠 저작 정보를 XMT-A 내부 자료구조에 맞는 저작 정보로 변환하여 저작 정보 처리 모듈에 전달하는 기능을 제공한다. 이는 전달받은 콘텐츠 저작 정보로부터 사용자 입력에 의해 생성된 각 미디어 객체들을 이용해 장면을 구성하는데 필요한 노드 및 해당 노드에 필요한 필드값들을 생성하여 Data access API를 통해 저작 정보 처리 모듈로 전달한다.

3. 저작 정보 처리 모듈

저작 정보 처리 모듈은 XMT-A 내부 자료 구조를 생성하고 내부 자료 구조에 저장된 편집 정보를 XMT-A 파일로 저장하는 기능을 수행한다. 그리고 사용자 요청에 따라 XMT-A 파일과 XMT-A 콘텐츠가 참조하는 외부 미디어 객체들로부터 mp4 파일을 생성하는 기능을 수행한다. XMT는 MPEG-4 장면 기술과 객체 기술을 텍스트 형식의

로 기술하기 위한 프레임워크로서, MPEG-4의 콘텐츠 저작언어인 XMT-A를 사용하여 저작된 대화형 콘텐츠를 스트리밍하거나 배포하기 위해서는 XMT-A 파일을 이진 포맷인 MP4 파일로 변환하여야 한다. 따라서 대화형 복수시점 영상콘텐츠 저작시스템에서는 XMT-A 파일을 MP4 파일로 변환하는 기능을 수행하는 XMT-A to MP4 컨버터를 구현하였다.

4. 복수시점 데이터 처리 모듈

복수시점 데이터 처리 모듈은 다양한 각도에서 촬영된 복수개의 영상을 입력받아 하나의 파노라믹 영상을 생성해주는 기능을 제공하며, 이렇게 내부적으로 생성된 파노라믹 영상이나 혹은 외부로부터 별도로 입력받은 파노라믹 영상을 텍스처 매핑하는 기능을 제공한다. 또한 다시점 영상의 경우 각 다시점 영상 간의 상하좌우 물리적인 관계를 설정하고, 다시점 영상 간의 영상 전환 기능을 제공한다. 단, 해당 모듈에서는 다시점 영상 간의 중간 영상 생성 기술은 포함하지 않는다.

그림 8(a)는 파노라믹 영상 생성을 위한 전처리 도구로서, 입력 영상들을 보여주는 입력 영상창과 단계별 영상처리 결과를 보여주는 메인창, 파일 열기/스티칭/컬러 블렌딩/결과영상 저장 기능을 지원해 주는 메뉴로 구성된다. 먼저 개별적으로 촬영된 영상들을 입력받아 영상간 매칭점을 찾

아 스티칭하고 컬러 블렌딩 처리하는 과정을 거쳐 한 장의 파노라믹 영상이 생성된다.

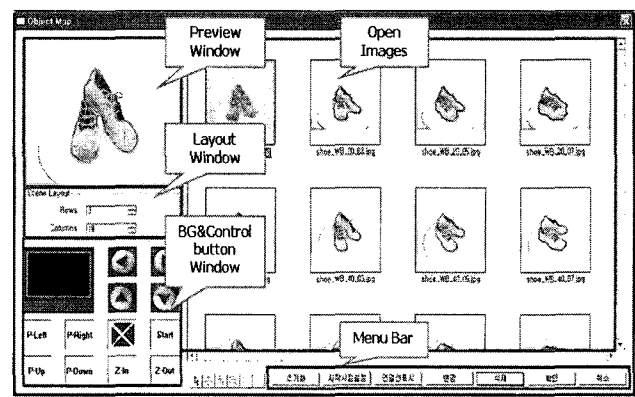
그림 8(b)는 다시점 영상을 위한 사용자 인터페이스로서, 입력 영상을 보여주고 각 영상간의 물리적인 관계를 설정할 수 있는 입력 영상창, 다시점 영상의 물리적인 개수를 설정해 주는 레이아웃창, 다시점 영상 자동 전환 기능을 제어해 주기 위한 제어버튼 설정창, 입력 영상 열기/추가/제거 등의 편집을 위한 메뉴 등으로 구성된다. 다시점 영상을 생성하기 위해서는 먼저 다양한 각도에서 촬영된 영상들을 입력받아 배열하고, 각 영상간의 상하좌우 물리적인 관계를 설정한 후 다시점 영상들을 제어하기 위한 시점 전환 버튼을 설정해 주면 이를 통해 다시점 영상들을 자유롭게 전환할 수 있는 장면 기술 정보가 자동으로 생성된다. 이렇게 생성된 복수시점 영상을 그림 7의 장면 편집창에 배치시키고, 다양한 객체를 삽입하여 대화형 조작이 가능한 이벤트 설정을 통해 복수시점 영상을 이용한 대화형 서비스를 제공해 줄 수 있다.

5. 미디어 복호화 모듈

미디어 복호화 모듈은 동영상, 정지영상, 오디오 등 다양한 미디어 객체를 복호화 하는 기능을 수행한다. 즉, 저작시스템에 부호화된 미디어 객체가 입력으로 들어올 경우 미디어 복호화 모듈을 활성화하여 관련 코덱을 호출하고 이



(a)



(b)

그림 8. 복수시점 데이터 처리 모듈의 사용자 인터페이스
Fig. 8. GUI of the multi-view data processor module

를 복호화한 후 해당 미디어의 복호화 데이터를 사용자 저작 환경 모듈로 넘겨줌으로써 해당 객체를 미리볼 수 있도록 한다. 미디어 복호화 모듈은 표 1에서 보여지는 것과 같이 다양한 동영상, 정지영상, 오디오 파일 포맷들을 지원해 준다.

IV. 실험결과

본 장에서는 앞서 III장에서 제안한 저작시스템을 이용해 저작된 대화형 복수시점 영상콘텐츠들을 보여준다. 먼저 저작시스템의 입력으로 사용된 파노라믹 정지영상의 경우 디지털 카메라를 이용해 다양한 각도에서 획득된 영상들을 파노라믹 스티칭 처리한 것으로 2,000 X 1,000 크기의 영상을 실험영상으로 사용하였다. 파노라믹 동영상의 경우 파노라믹 동영상 획득을 위해 Point Grey사의 Ladybug 카메라를 사용한 것으로 초당 15프레임의 2,000 X 1,000 크기의 영상을 실험영상으로 사용하였다. 그밖에 부가콘텐츠로 사용된 정지영상, 동영상은 상용화된 그래픽틀을 이용하여 독립적으로 생성하였다. 이렇게 생성된 복수시점 영상 및 부가콘텐츠를 대화형 복수시점 영상콘텐츠 저작시스템의 입력으로 하여 각 객체를 자유로이 배치시키고 원하는 대화형 기능을 부여해 준 후, 저작된 결과를 MPEG-4 Systems 규격을 따르는 XMT-A 파일로 생성해 준다. 이렇게 생성된 XMT-A 파일은 전송/저장 관리를 위해 바이너리 MP4 파일로 변환되어 대화형 복수시점 영상콘텐츠 플레이어를 통해 저작된 결과를 확인할 수 있

었다. 개발된 대화형 복수시점 영상콘텐츠 저작시스템 및 플레이어의 개발 환경은 Microsoft Visual C++를 사용한 것으로, 해당 시스템은 III장에서 제안된 장면 구성 프로파일을 지원한다.

그림 9는 파노라믹 정지영상을 배경으로 동영상, 정지영상, 오디오, 그래픽 등 다양한 객체가 존재하는 대화형 복수시점 영상콘텐츠로서, 그림 9(a)-(c)에서 보이는 다양한 객체들은 외부 입력 장치를 통해 사용자 요청이 발생했을 때 해당 객체가 장면에 삽입되거나 삭제된다. 그림 9(a)는 카메라 다시점 객체로서 사용자가 외부 입력장치를 통해 카메라 객체를 선택하면 시점 전환이 가능한 카메라의 다시점 영상 및 카메라 관련 상세정보를 제공받을 수 있다. 그림 9(b)는 사용자가 액자를 선택하면 액자 내의 그림과 관련된 화가의 상세정보를 제공받을 수 있고 액자 내의 그림도 다른 것으로 교체할 수 있다. 또한 원통에 특정 영상을 텍스처 매핑시켜 그래픽적인 효과도 줄 수 있다. 그림 9(c)는 부가동영상 서비스로서, 사용자가 리모콘 객체를 선택하면 특정 동영상을 TV 화면에 보여 준다.

그림 10은 파노라믹 동영상을 배경으로 하여 다양한 객체들이 존재하는 대화형 복수시점 영상콘텐츠로서, 카페, 빌딩, 박물관, 산 등 각 장소와 관련된 상세정보를 제공해 준다. 그림 10(a)는 사용자가 요청할 경우, 사용자에게 카페에 관한 위치 정보나 할인 쿠폰 다운로드 서비스를 제공해 준다. 그림 10(b), (c)는 빌딩 및 박물관에 대한 규모와 시설 안내 등의 부가 정보를 제공해 주며, 그림 10(d)의 경우 특정 산의 위치 및 소개, 등산코스에 관한 상세한 부가

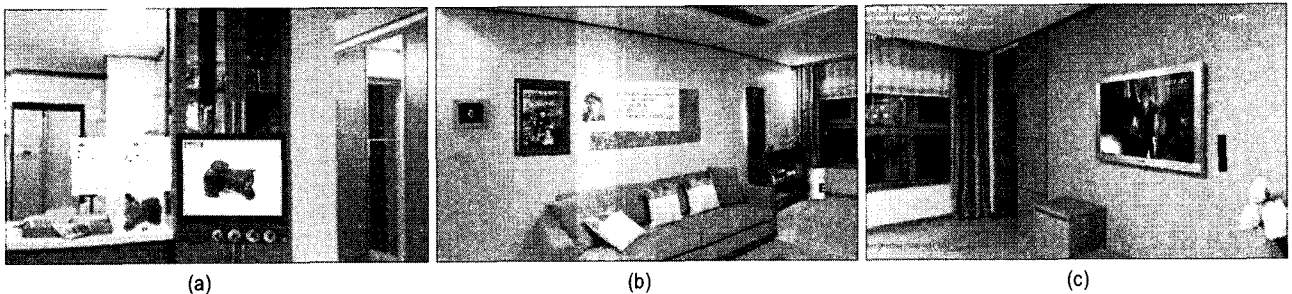


그림 9. 파노라믹 정지영상 기반 대화형 복수시점 영상콘텐츠의 예
 Fig. 9. Screen shots of interactive multi-view visual content based on the panoramic image

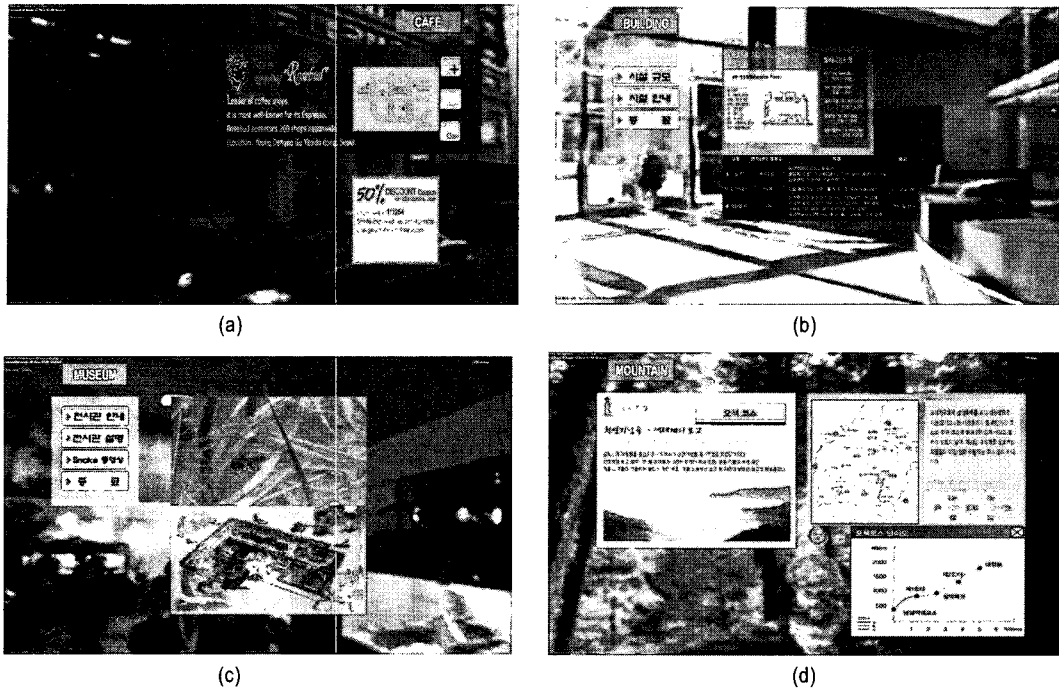


그림 10. 파노라믹 동영상 기반 대화형 복수시점 영상콘텐츠의 예
 Fig. 10. Screen shots of interactive multi-view visual content based on the panoramic video

정보를 제공해 준다.

현재 복수시점 영상을 이용한 다양한 제품들이 출시되어 왔다^{[7][8]}. 하지만 각 제품들의 기술은 순수하게 복수시점 정지영상을 생성하는 영상처리를 수준으로 복수시점 정지영상 외에 별도의 부가콘텐츠를 지원해 주지 않고 있다. 하지만 본 저작 시스템의 경우 실험결과에서 보여지는 것처럼 복수시점 정지영상 뿐 아니라 복수시점 동영상까지 지원하며, 다양한 대화형 기능을 가지는 부가콘텐츠까지 지원해 준다. 또한 상용 제품들의 출력 파일 포맷은 비표준 규격으로 다운로드 서비스에 국한되는데 반해, 본 저작시스템은 스트리밍이 가능한 MP4 파일을 지원해 줌으로써 효율적인 콘텐츠 전송도 가능하다.

V. 결 론

현재 복수시점 영상 관련 기술에 대한 국제 표준 또는 사실상 표준(de-facto standard)이 존재하지 않고 있다. 따

라서 본 논문에서는 MPEG-4 Systems를 통해 이를 실현해 보고자 하였다. 이를 위해 복수시점 영상 서비스를 크게 다 시점 영상 서비스와 파노라믹 영상 서비스로 분류하여 각 서비스별 MPEG-4 장면 구성 기술 방법에 대해 살펴보았 으며, 복수시점 영상 서비스를 위한 MPEG-4 장면 그래프 프로파일 및 그래픽스 프로파일을 제안하였다. 또한 본 논문에서는 표준 파일 포맷을 지원하면서 콘텐츠를 편리하게 저작할 수 있는 대화형 복수시점 영상콘텐츠 저작시스템을 설계하고 구현하였다. IV장의 실험결과에서 보여지는 것처럼 본 저작 시스템을 통해 복수시점 정지영상 뿐 아니라 복수시점 동영상 그리고 대화형 기능을 포함하는 다양한 부가콘텐츠까지 지원해 줌으로써 풍부한 복수시점 영상 서비스가 가능하다. 또한 바이너리 포맷의 MP4 표준 규격까지 지원해 줌으로써 다운로드 서비스 뿐만 아니라 스트리밍 서비스까지 가능하다.

복수시점 영상 서비스는 다양한 분야에서 서서히 두각을 나타내며 이를 이용한 서비스가 시작되고 있다. 이에 단순히 복수시점 A/V 서비스 제공에만 머무르지 않고 대화형

기능이 가능한 복수시점 영상 서비스로의 기술적 발전 모색을 통해 사용자 시점의 자유도를 증가시키고 보다 넓은 범위의 뷰 및 대화형 부가콘텐츠를 제공함으로써 나날이 증가되는 소비자의 다양한 욕구를 충족시킬 수 있을 것이다. 이는 향후 디지털 방송 분야 및 DMB 데이터 서비스에도 활용될 수 있을 것이다.

참 고 문 헌

[1] ISO/IEC JTC1/SC29/WG11 N5877, "Applications and Requirements for 3DAV," July 2003.

[2] ISO/IEC JTC1/SC29/WG11 N5878, "Report on 3DAV Exploration,"

July 2003.

[3] 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

[4] ISO/IEC Information technology, Part1: Systems, ISO/IEC 14496-1:2002, 2002.03. (MPEG-4 systems).

[5] ISO/IEC JTC1/SC29/WG11 N3573, "Working Draft 3.0 of ISO/IEC 14496-1/Amd3: Extensible MPEG-4 Texture Format (XMT)," July 2000.

[6] ISO/IEC JTC1/SC29/WG11 N5298, 14496-14 : MP4 File Format, "Text of ISO/IEC FDIS 14496-14," April 2003.

[7] Apple QuickTime VR Authoring Studio. <http://www.apple.com/quicktime/qtvr>

[8] Photovista Virtual Tour. <http://iseemedia.com>[1] S. K. Nayar.: Catadioptric Omnidirectional Camera, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition,(1997) 17-19

저 자 소 개



이 인 재

- 1999년 : 성균관대학교 전자공학과 학사
- 2001년 : 성균관대학교 전기전자및컴퓨터공학부 석사
- 2001년 ~ 현재 : 한국전자통신연구원 방송미디어연그룹 연구원
- 주관심분야 : 영상처리, 디지털 방송, 컴퓨터비전



최 진 수

- 1990년 : 경북대학교 전자공학과 학사
- 1992년 : 경북대학교 전자공학과 석사
- 1996년 : 경북대학교 전자공학과 박사
- 1996년 ~ 현재 : 한국전자통신연구원 선임연구원
- 2001년 2월 ~ 2005년 3월 : 한국전자통신연구원 데이터방송연구팀장
- 2004년 10월 ~ 2006년 3월 : TTA 데이터방송프로젝트그룹(PG312) 의장
- 주관심분야 : 멀티미디어방송, 영상통신



기 명 석

- 1999년 : 전남대학교 컴퓨터공학과 학사
- 2001년 : 전남대학교 컴퓨터공학과 석사
- 2001년 ~ 현재 : 한국전자통신연구원 방송미디어연그룹 연구원
- 주관심분야 : 영상처리, 컴퓨터비전

 저 자 소 개

**정 세 윤**

- 1995년 : 인하대학교 전자공학과 학사
- 1997년 : 인하대학교 전자공학과 석사
- 1996년 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹 선임연구원
- 주관심분야 : 스케일러블비디오코딩, 멀티미디어통신, 디지털 대화형 방송

**문 경 애**

- 1985년 : 충남대학교 계산통계학과 학사
- 1988년 : 충남대학교 대학원 전산학과 석사
- 1997년 : 충남대학교 대학원 전산학과 박사
- 1988년 ~ 1990년 : 충남대학교 전산학과 조교 및 시간강사
- 1990년 ~ 현재 : 한국전자통신연구원 대화형미디어연구팀 팀장 (책임연구원)
- 주관심분야 : 영상처리, 컴퓨터비전, 디지털 멀티미디어방송

**홍 진 우**

- 1982년 : 광운대학교 응용전자공학과 학사
- 1984년 : 광운대학교 전자공학과 석사
- 1993년 : 광운대학교 전자계산기공학과 박사
- 1998년 ~ 1999년 : 독일 프라운호퍼연구소 파견연구원
- 1984년 ~ 현재 : 한국전자통신연구원 방송미디어연구그룹장 (책임연구원)
- 2000년 ~ 현재 : 한국방송공학회 학술위원 및 편집위원
- 주관심분야 : 오디오 신호처리 및 부호화, 디지털 콘텐츠 보호 및 관리, 디지털 콘텐츠 보호 및 관리, 통방융합 프레임워크 기술