특집논문 (Special Paper)
방송공학회논문지 제17권 제4호, 2012년 7월 (JBE Vol. 17, No. 4, July 2012)
http://dx.doi.org/10.5909/JBE.2012.17.4.584

3D 비디오의 HTTP 적응적 스트리밍을 위한 전송규격 설계

이 장 원[a)], 김 규 헌[a)‡]

Design of 3D Video Delivery Format for HTTP Adaptive Streaming Service

Jangwon Lee[a)] and Kyuheon Kim[a)‡]

요 약

최근 멀티미디어 서비스 관련 업계 및 시장에서는 콘텐츠로서의 3D 입체영상과, 전송기술로서의 HTTP 적응적 스트리밍 기법이 큰 관심을 받고 있다. 주목받고 있는 두 기술을 통해 유망하고 시장성 있는 멀티미디어 서비스 모델을 창출할 수 있을 것으로 기대되나, 현재의 표준기술에서는 두 기술의 유기적인 결합 방법을 제공하고 있지 않다. 3D 콘텐츠 서비스는 대역폭이 보장되거나 구애받지 않는 방송 및 저장 서비스 위주로 편중되어 있고, 적응적 스트리밍 서비스는 특정 코덱에 의존적이며, 단순한 영상구성 정보만을 제공하는 기초적 형태의 3D 서비스 제공 방법만을 제공하고 있다. 이에, 본 논문은 인터넷망을 기반으로 안정적이고 끊김 없는 스트리밍 서비스를 다양한 형태의 스테레오스코픽 비디오를 통해 호환성 있는 형태로 제공하기 위한 3D 비디오의 적응적 스트리밍 전송규격을 제안하고자한다. 제안하는 기술은 스테레오스코픽 비디오의 저장에 대한 서비스 지향적 표준기술인 스테레오스코픽 비디오 어플리케이션 포맷을 기반으로, 적응적 스트리밍에 대한 대표적 표준기술인 DASH를 통해 직접 적용 가능하도록 설계되었다. 본 논문의 제안 기술은 MPEG에서 작업표준안으로 채택되어 표준기술로 고려되고 있어, 향후 관련 분야에서의 실용적인 산업화를 위한 기틀을 제공할 수 있을 것으로 기대된다.

Abstract

Recently, 3D stereoscopic video and HTTP adaptive streaming technologies have received a lot of attention from relevant industrial fields and markets in terms of multimedia contents and delivery services, respectively. It is expected that promising and marketable service models can be created by means of these noticeable two technologies. However, current standard specifications do not provide a method for organized connection between those two technologies. 3D stereoscopic video services are weighted in broadcasting and storage services that are only available under environments in which the network bandwidth is guaranteed or free. Also, HTTP adaptive streaming technologies only provide plain 3D service methods that are dependent on particular Codec. Therefore, this paper proposes 3D video delivery format for HTTP adaptive streaming service which enables stable and seamless display for various stereoscopic video sequences over internet networks. The proposed technology is designed on the basis of Stereoscopic Video Application Format which is a service-oriented standard specification for storing stereoscopic video sequences. Also, this delivery format is directly applicable over DASH that is the representative standard technology for HTTP adaptive streaming services. The delivery format proposed in this paper has been submitted to MPEG and it has been accepted as a working draft, thus it expected to pave the way for practical industrialization in relevant fields from now on.

Keyword : 3D stereoscopic video, HTTP adaptive streaming, Delivery format, Stereoscopic Video Application Format, DASH

Ⅰ. 서 론

현재 방송, TV, 영화 모바일 서비스를 비롯한 모든 멀티미디어 관련 업계에서 3D 콘텐츠 서비스는 가장 유망하고 시장성 있는 분야로 평가받고 있다. 이러한 관련 업계 및 소비자의 관심은 최근 수년 사이 급격하게 증대되어 영화, 방송뿐만 아니라, 인터넷, 모바일 서비스에 이르기까지 3D 콘텐츠 서비스는 그 저변을 넓혀가고 있으며, 이에 따라 TV, 디스플레이, 휴대폰, 게임기 등 3D 영상을 지원하는 다양한 디지털 디바이스들이 시장에 등장하게 되었다. 3D 콘텐츠 서비스를 위한 기반 기술에 대한 연구 또한 빠르게 진행되고 있다. 특히, MPEG(Moving Picture Experts Group), DVB(Digital Video Broadcasting), SMPTE (Society of Motion Picture and Television Engineers) 등 국제적 표준화 기구에서는 3D 비디오의 압축, 전송, 저장 등 다양한 기술요소에 대한 표준규격을 제공하고 있다[1]-[3].

멀티미디어 콘텐츠 제공 측면에서 3D 입체영상 서비스가 주목을 받고 있다면, 멀티미디어 전송 측면의 기술로서는 HTTP 적응적 스트리밍(HTTP adaptive streaming) 기법이 각광을 받고 있다. 본 기술은 기존의 프로그레시브 파일 다운로드(Progressive file download)와 멀티미디어 스트리밍(Multimedia streaming)을 접목한 기술로, 인터넷을 이용한 멀티미디어 전송 서비스의 단점으로 지적되었던 초기 지연시간과 대역폭 변화에 따른 끊김 현상을 극복하기 위한 대안기술로 여겨지고 있다. 적응적 스트리밍 기법에 대한 MPEG, 3GPP(3rd Generation Partnership Project) 등의 표준화 기구의 움직임 또한 활발하게 이루어지고 있다[4]-[6].

이처럼 3D 콘텐츠 서비스와 적응적 스트리밍 서비스가 개별적으로는 시장과 업계로부터 큰 관심을 받고 있으나, 두 기술의 유기적인 결합은 기대만큼 이루어지고 있지 않다. 3D 콘텐츠 서비스는 안정적 대역폭 확보가 가능한 방송 또는 대역폭에 구애받지 않는 파일기반 서비스를 위주로 치우쳐 있으며, 적응적 스트리밍에서는 특정 코텍(Codec)을 기반으로 기초적인 수준의 3D 비디오 서비스 방법만을 정의하고 있다.

이에, 본 논문은 인터넷망을 기반으로 안정적이고 끊김 없는 스트리밍 서비스를 다양한 형태의 스테레오스코픽 비디오(Stereoscopic video)를 통해 기존 기술들과의 호환성을 유지하여 제공하기 위한 3D 비디오의 적응적 스트리밍 전송규격을 제안하고자한다. 제안하는 기술은 3D 콘텐츠 저장에 대한 서비스 지향적 표준인 스테레오스코픽 비디오 어플리케이션 포맷(SVAF, ISO/IEC 23000-11 Stereoscopic Video Application Format)[7][8]을 기반으로 고안되었다. 또한, 적응적 스트리밍 기법의 대표적 표준 기술인 MPEG-B DASH(Dynamic Adaptive Streaming over HTTP, ISO/IEC 23001-6)[4][5]를 통해 직접 전송 가능하도록 설계되었다.

본 논문은 다음과 같은 구성으로 3D 비디오의 적응적 스트리밍 전송규격을 제안하고자 한다. Ⅱ장에서는 전송규격 설계를 위한 기반기술로서의 스테레오스코픽 비디오 어플리케이션 포맷에 대한 전반을 소개하고, Ⅲ에서는 적응적 스트리밍 서비스의 전송규격 설계를 위한 파일의 구조화 과정을 소개한다. Ⅳ장에서는 제안하는 전송규격의 상세 설계과정을 설명하며, Ⅴ장에서는 설계된 전송규격 기반의 3D 적응적 스트리밍 송수신 시스템의 구현 및 시연을 통해 전송규격으로의 적합성 및 기능성을 검증한다. 마지막으로, Ⅵ장에서는 제안된 기술의 향후 연구방향 및 활용방안을 도출한다.

Ⅱ. 3D 입체영상 서비스를 위한 스테레오스코픽 비디오 어플리케이션 포맷

스테레오스코픽 비디오는 현재 방송, 영화 콘텐츠 및 스트리밍 서비스 등 멀티미디어 관련업계 전반에 걸쳐 가장 폭넓게 사용되는 3D 입체영상 구성기술로, 양안식 좌, 우영상을 통해 입체감을 제공한다. 본 기술은 2D 기반의 비디오 부호화, 전송, 저장 등 기존의 기술요소들을 채용하여 기존 시스템의 큰 변경 없이 3D 입체영상 서비스를 구현할 수

a) 경희대학교 전자정보대학 (College of Electronics and Information, Kyung Hee University)
‡ 교신저자 : 김규헌 (Kyuheon Kim)
E-mail: kyuheonkim@khu.ac.kr
Tel: +82-31-201-3810 Fax: +82-31-203-0146
※ 이 논문은 지식경제부의 지원을 받는 지식경제 기술혁신사업(“스케일러블 데이터기반 Edutainment 핵심기술 개발”)(과제번호 : 10035486)의 연구결과와 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2012- H0301-12-1006)의 연구결과로 수행되었음.
・접수일(2012년5월15일),수정일(2012년6월19일),게재확정일(2012년6월19일)

있는 장점을 갖고 있다. 이러한 스테레오스코픽 비디오의 장점을 활용하여, MPEG에서는 3D 입체영상 타겟 서비스를 위한 표준규격으로 스테레오스코픽 비디오 어플리케이션 포맷을 제정하였다. 본 장에서는 제안하고자 하는 전송규격 설계를 위한 기반 기술로서의 SVAF의 기술 전반에 대해 소개하고자 한다.

1. 스테레오스코픽 비디오 어플리케이션 포맷 파일 기본구조

스테레오스코픽 비디오 어플리케이션 포맷은 파일 저장에 기반한 방법으로 스테레오스코픽 비디오 서비스를 위한 기본 프레임워크(Framework)를 제공한다. 스테레오스코픽 비디오라는 특정 어플리케이션을 타켓으로 기존 요소기술의 통합 운용체계를 제공하여 3D 입체영상 서비스의 실용적인 산업화를 목적으로 제정된 표준인 만큼, 기존 기술들과의 호환성(Compatibility) 유지는 필수적이다. 이를 위하여 대표적 파일규격 표준인 ISO base media file format (ISO BMFF, ISO/IEC 14496-12)[9]을 기반으로 개발되었다.

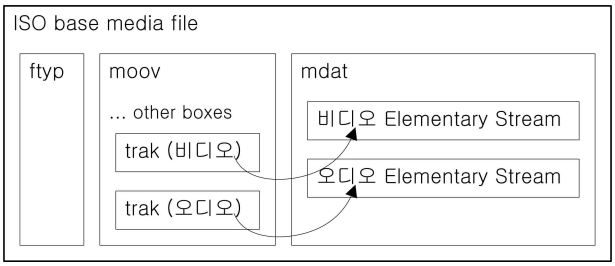


그림 1. 로컬 ISO base media file format의 기본 구조
Fig. 1. Local ISO base media file format structure

그림 1은 로컬(Local) 서비스를 위한 ISO BMFF의 기본 구조를 보여준다. ISO BMFF는 4바이트의 고유식별코드를 갖는 박스(Box)라는 계층적 구조체의 집합으로 이루어지며, 개괄적으로 파일의 형식 식별을 목적으로 하는 파일타입박스(File Type Box, 'ftyp')와 프리젠테이션(Presentation)을 위한 메타데이터들을 기술하는 무비박스(Movie Box, 'moov') 그리고 부호화된 ES(Elementary Stream)들을 담는 미디어박스(Media Box, 'mdat')로 구성된다. 또한 개별 ES의 프리젠테이션을 위한 메타데이터는 무비박스 내에 개별 트랙박스(Track Box, 'trak')로 대응되어 저장된다.

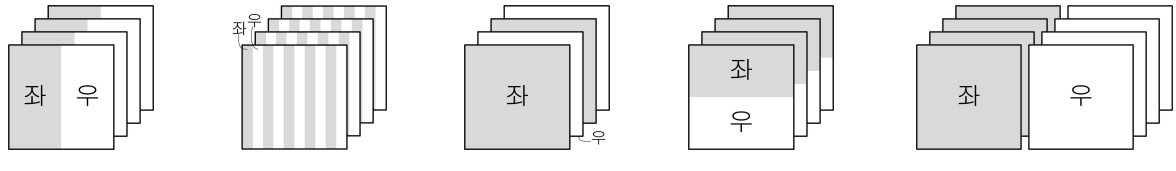
스테레오스코픽 비디오 어플리케이션 포맷은 상기의 ISO BMFF와 같은 방식으로, 하나 또는 두 개의 스테레오스코픽 비디오 ES를 오디오, 스피치(Speech), 정지영상(Image) 등의 부가 스트림의 ES와 함께 미디어박스에 저장한다. 여기서 스테레오스코픽 비디오의 ES는 MPEG-4 Visual(ISO/IEC 14496-2), MPEG-4 AVC(Advanced Video Coding, ISO/IEC 14496-10) 등의 2D 기반 방식으로 부호화되기 때문에, 기존의 단말 및 시스템과 호환성을 확보할 수 있다. 또한, 3D 재생장치에서 다양한 형태의 스테레오스코픽 비디오의 ES를 올바르게 재생시킬 수 있도록, 다음절에서 설명하는 신규박스들을 트랙박스(Track Box, 'trak') 내에 정의하고 있다.

2 스테레오스코픽 비디오의 특성 기술을 위한 신규 박스

2.1 스테레오스코픽 비디오 미디어 정보 박스

SVAF는 스테레오스코픽 비디오 트랙의 샘플테이블박스(Sample Table Box, 'stbl') 내에 스테레오스코픽 비디오 미디어 정보 박스(Stereoscopic Video Media Information Box, 'svmi')를 정의하고 있다. 표 1과 같은 구문으로 표현되는 이 박스는 3D 재생을 위한 필수(Mandatory) 박스로서, 스테레오스코픽 비디오의 시각구성 정보 및 2D/3D 구간 정보를 포함한다.

첫 번째 필드인 ‘stereoscopic_composition_type'은 사람의 좌안, 우안을 위한 두 개의 비디오 시퀀스(Sequence)로 이루어지는 스테레오스코픽 비디오의 구성방식(Composition type)을 나타낸다. 현재 시장에서 널리 사용되는 구성방식에는, 그림 2의 (가)-(라)와 같이 기존 하나의 비디오 프레임(Frame)에 좌/우영상의 이미지를 분할하여 구성하는 프래임 패킹 (Frame packing) 방식과, 그림 2 (마)와 같이 좌/우 영상이 독립적으로 부호화되는 듀얼 스트림 (Dual stream) 방식이 있다. 상기 필드는 이러한 구성방식을 각각 하나의 정수 값으로 할당하여 표현하며, 'is_left_first'를 통



(가) 좌우분할 (Side-by-side) (나) 수직격렬주사 (Vertical line interleaved) (다) 순차 프레임 (Frame sequential) (라) 상하분할 (Top-bottom) (마) 듀얼 스트림 (Dual stream)

그림 2. 스테레오스코픽 구성방식
Fig. 2. Stereoscopic composition types

해 개별 구성방식에서의 좌/우영상 위치를 표시한다.

SVAF는 2D와 3D가 혼용된 다양한 시나리오에 유연하게 대응하기 위해, 스테레오스코픽 비디오 트랙에 2D 시퀀스를 혼용하여 부호화하는 형식을 지원하고 있다[8]. 'stereo_mono_change_count'는 이처럼 2D/3D로 혼용된 시퀀스에서, 2D/3D로 구분되는 개별 구간의 개수를 의미하며, ‘sample_count'는 각 구간의 길이를 프레임수로 표현하고, 'stereo_flag'는 각 구간의 2D/3D 여부를 나타낸다.

표 1. 스테레오스코픽 비디오 미디어 정보 박스의 구문
Table 1. Syntax of Stereoscopic Video Media Information Box

구 문
StereoscopicVideoMediaInformationBox {
stereoscopic_composition_type
is_left_first
stereo_mono_change_count
for(i=1; i<=stereo_mono_change_count; i++) {
sample_count
stereo_flag
}
}

2.2 스테레오스코픽 비디오 카메라 디스플레이 정보 박스

SVAF는 스테레오스코픽 비디오 미디어 정보 박스를 통해 표현되는 3D 재생을 위한 필수정보 외에, 3D 재생시 입체효과를 증진시키기 위한 부가정보를 저장하기 위한 방법을 제공한다. 이러한 부가정보에는 폭주거리(Convergence Distance), 초점거리(Focal length), 두 카메라의 상대위치 벡터(Vector) 등의 카메라 파라미터(Camera Parameters)와, 최대/최소 시차 (Disparity), 적정화면 크기 및 시청거리 등의 디스플레이 파라미터(Display Parameters)가 포함된다. SVAF는 이러한 카메라, 디스플레이 파라미터 지원을 위해, 스테레오스코픽 비디오 트랙의 메타박스(Meta Box, 'meta') 내에 표 2와 같이 스테레오스코픽 카메라 디스플레이 정보 박스(Stereoscopic Camera and Display Information Box, ‘scdi’)라는 선택적(Optional) 신규 박스를 정의하여 카메라, 디스플레이 파라미터를 지원한다.

또한, 상기의 카메라, 디스플레이 파라미터들은 시간에 따라 가변적으로 적용될 수 있어, SVAF는 복수의 파라미터 셋(Set)을 구간별로 적용하기 위한 메커니즘(Mechanism)을 제공하고 있다. 표 2에 나타난 바와 같이, 스테레오스코픽 카메라 디스플레이 정보박스는 내부의 음영 처리된 부분에 개별 파라미터 셋을 복수로 저장하며, 구간별 식별

표 2. 스테레오스코픽 카메라 디스플레이 정보 박스의 구문
Table 2. Syntax of Stereoscopic Camera and Display Information Box

구 문
StereoscopicCameraAndDisplayInformationBox {
item_count
for(i=0; i<item_count; i++){
item_ID
is_item_ID_ref
if(is_item_ID_ref){
ref_item_ID
}
else{
카메라, 디스플레이 세부 파라미터 ...
}
}
}

자 ('item_ID') 또는 참조 식별자('ref_item')를 부여한다. 개별 구간의 위치정보는 메타박스 안에 존재하는 아이템로케이션박스(Item Location Box, 'iloc')를 통해 기술하여, 구간별로 다르게 지정된 파라미터 셋을 적용한 재생이 가능하다.

3. 신규박스를 포함하는 스테레오스코픽 비디오 어플리케이션 포맷 파일 세부구조

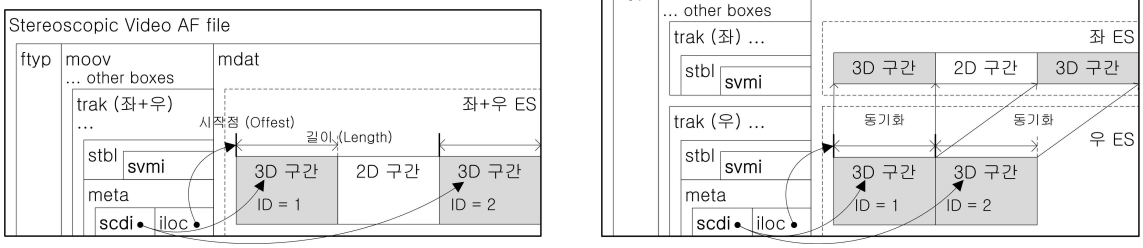
본 절에서는 앞서 설명한 ISO BMFF 기반의 SVAF 파일 기본구조에, 스테레오스코픽 비디오 특성 기술을 위한 신규박스를 포함하는 파일의 실제 구성 예에 대해 설명한다.

2.1절에서 설명하였듯, 스테레오스코픽 비디오는 하나의 시퀀스를 사용하는 프레임 패킹 방식, 두 개의 시퀀스를 사용하는 듀얼 스트림방식으로 크게 구분될 수 있으며, 이에 따라 파일 구조 또한 그림 3과 같이 하나 또는 두 개의 트랙으로 구성될 수 있다.

그림 3 (가)와 같이 프레임 패킹 방식의 스테레오스코픽 비디오를 포함하는 파일 구조에서는 좌/우 영상을 모두 포함하는 하나의 스테레오스코픽 비디오 트랙이 존재하게 된다. 상기 예에서 ‘svmi' 박스는 ’stbl' 박스 내에 위치하여 3D 시각정보 및 3개로 구성되는 2D/3D구간을 표현하며, 두 개의 파라미터 셋을 기술하는 ‘scdi’ 박스가 'iloc' 박스와 함께 'meta' 박스에 저장된다. 특히, 2 절에서 설명된 바와 같이, 파라미터 셋이 적용되는 개별 스테레오스코픽 구간은 'iloc' 박스를 통해 표현되는데, 그림 3에 표현된 바와 같이 구간의 시작점(Offset)과 그 길이(Length)를 바이트 단위로 기술하는 형식을 취한다. 'iloc'를 통해 표현된 구간 위치정보와 ‘scdi’에서 표현되는 개별 파라미터 셋은 식별자(‘item_ID')를 통해 링크(Link)되어 재생시 적용된다.

그림 3 (나)와 같이 듀얼 스트림 방식의 스테레오스코픽 비디오를 포함하는 파일 구조에서는 각각 독립된 좌, 우영상을 포함하는 두 개의 스테레오스코픽 비디오 트랙이 존재하게 된다. 기본 트랙의 구성은 그림 3 (가)에서와 같으나, 중복되는 정보 기술을 피하기 위해 하나의 트랙에서는 ‘scdi’ 및 ‘iloc' 박스가 생략될 수 있다. 또한, 중간에 삽입된 2D 구간의 경우는 좌, 우영상중 하나의 시퀀스만을 통해 재생될 수 있으므로, 하나의 ES만 저장되어도 무방하다. 특히, 이러한 파일 구조에서는 두 개의 트랙이 동기화(Synchronization)되어 재생되어야하는데, SVAF는 ISO BMFF의 프레임별 부호화 시간(Decoding Time) 화면재생 시간(Composition Time) 제공에 의한 동기화 메커니즘을 채용하여 좌/우 영상간의 동기화를 지원한다.

이처럼 SVAF는 현재 시장에서 사용되는 다양한 형태의 스테레오스코픽 비디오를 특정 코덱에 구애받지 않고 호환성 있게 서비스할 수 있는 방안을 제공한다. 또한, 다양한 서비스 시나리오와 카메라 디스플레이 세부 파라미터들을 지원하여 3D 입체영상 서비스를 위한 풍족한 환경을 제공해 준다. 그러나 SVAF는 적응적 스트리밍 서비스에 직접 적용할 수 없다는 한계점을 갖고 있다. 본 장에서의 설명을 통해 알 수 있듯이, 로컬 서비스를 목적으로 개발된 SVAF의 파일 구조는 시퀀스만을 기준으로 구조화 되어있다. 적



(가) 프레임 패킹 방식 (Frame packing type)

(나) 듀얼 스트림 방식 (Dual stream type)

그림 3. 스테레오스코픽 비디오 어플리케이션 포맷 파일구조
Fig. 3. File structure of Stereoscopic Video AF

응적 스트리밍 서비스의 전송규격으로 활용되기 위해서는 시퀀스 기준 뿐만 아니라 시간 기준에 의한 파일의 구조화가 필요하다.

III. HTTP 적응적 스트리밍을 위한 파일의 구조화

본 장에서는 앞서 SVAF의 한계점으로 제기되었던 적응적 스트리밍 서비스에서의 전송규격으로서의 기능 확보를 위한 시간 기준에 의한 파일의 구조화 방안에 대해 소개하고, 구조화된 파일을 이용한 HTTP 적응적 스트리밍 서비스에서의 전송규격 구성 방법을 설명하고자 한다.

적응적 스트리밍은 HTTP를 이용한 프로그레시브 파일 다운로드 기법을 기반으로 한다. 여기에 추가적으로, QoS (Quality of Service)를 보장할 수 없는 가변 대역폭의 인터넷 기반 전송 환경에서 끊김 없는 미디어 스트리밍 서비스를 제공하기 위해, 하나의 미디어 콘텐츠를 다양한 품질의 스트림으로 부호화 한다. 개별 스트림은 일정 시간간격으로 나누어져 여러 개의 작은 파일 형태로 서버에 저장되며, 단말은 서버와의 송수신과정에 지속적으로 가용 대역폭을 측정하여, 대역폭 이내의 품질을 갖는 스트림의 파일을 선택적으로 다운로드하여 실시간으로 재생한다.

MPEG-B DASH는 적응적 스트리밍 제공을 목적으로 하는 대표적 표준기술로, 개별 스트림을 일정 시간간격으로 나누어 전송하기 위해 ISO BMFF 또는 MPEG-2 TS (Transport Stream) 기반의 전송규격인 세그먼트(Segment)를 정의하고 있다. 특히, ISO BMFF 기반 세그먼트는 프래그먼티드(Fragmented) 파일구조를 기본으로, 간결하고 압축적인 형태의 적응적 스트리밍 전송규격을 제공하고 있다.

그림 4는 프래그먼티드 ISO BMFF의 기본 형태를 보여준다. 그림 1에서 설명한 로컬(Local) 파일 구조와 비교해 볼 때, 부호화된 ES를 저장하는 ‘mdat' 박스가 시간간격에 의해 여러 개로 세분화 되어 있고, 각각의 'mdat' 박스 앞에는 무비프래그먼트 박스 (Movie Fragment Box, 'moof')가 위치함을 볼 수 있다.

‘moof' 박스는 뒤따르는 'mdat' 박스에 저장된 부분 ES포함되는 개별 프레임의 시간 및 색인 정보를 포함하고 있다. 따라서 이러한 구조의 'moov' 박스에서 ES의 개별 프레임 정보 기술을 포함하는 각 트랙의 'stbl' 박스는 개별 프레임 정보를 포함하지 않는다. 또한 'moov' 박스 뒤에 위치하는 무비 확장 박스(Movie Extends Box, 'mvex')는 'moov'에 뒤따르는 'moof' 박스가 존재함을 알려주고, 'moof'에서 사용되는 파라미터들의 디폴트(Default) 값을 기술하는 역할을 한다.

DASH에서는 그림 4에 표시된바와 같이, 상기 프래그먼티드 ISO BMFF의 'ftyp', 'moov', 'mvex' 박스를 초기화 세그먼트(Initialization segment)로 파일화하여, 개별 품질의 시퀀스 재생을 위한 초기화 정보로 서비스 시작시점에 전송한다. 또한 ‘moof'와 ’mdat' 박스의 쌍으로 구성되는 프래그먼트(Fragment)를 하나 혹은 여러 개를 하나의 파일에 담아 미디어 세그먼트(Media segment)로 구성하여, 적응적 미디어 스트리밍 서비스를 위한 미디어 전송규격으로 사용한다.

본 논문에서 제안하는 3D 비디오 전송규격은 위와 같은 프래그먼티드 ISO BMFF 및 세그먼트 규격을 토대로 설계

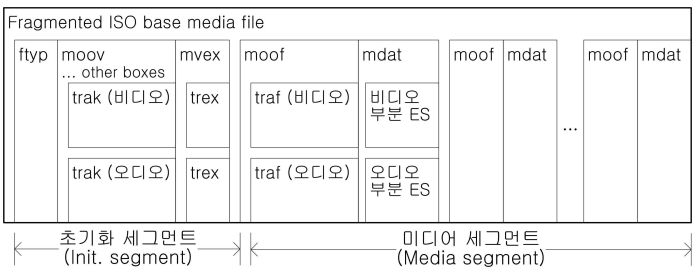


그림 4. 프래그먼티드 ISO base media file 구조
Fig. 4. Fragmented ISO base media file structure

되었다. Ⅱ장에서 설명하였듯, SVAF는 로컬 ISO BMFF를 기반으로 스테레오스코픽 비디오의 특성 기술을 위한 신규 박스를 포함하는 형태로 설계되었기 때문에, SVAF 기반의 3D 적응적 스트리밍 전송규격 또한 상기구조를 바탕으로 신규박스의 적용 방법을 고려하여 간결하게 설계될 수 있을 것이다. 또한 코덱으로 부터의 독립성, 호환성, 다양한 서비스 시나리오 및 파라미터 지원 등 SVAF의 장점을 그대로 갖는 적응적 스트리밍 전송규격 설계가 가능할 것이다.

Ⅳ. 스테레오스코픽 비디오 어플리케이션 포맷 기반 3D 비디오 전송규격의 설계

본 장에서는 앞서 Ⅱ장에서 소개한 스테레오스코픽 비디오 어플리케이션 포맷을 기반으로, 스테레오스코픽 비디오를 적응적 스트리밍 서비스에 적용시키기 위해 본 논문에서 제안하고자 하는 3D 비디오 전송규격의 상세 설계 과정에 대해 설명한다. 설계 과정은 Ⅲ장에서 소개한 시간 기준의 파일 구조화 과정을 토대로 설명된다.

앞서 Ⅱ장 2.1절에서 설명된 바와 같이, 스테레오스코픽 비디오 미디어 정보 박스는 스테레오스코픽 비디오의 시각구성 정보와, 2D/3D 구간 정보를 포함하고 있다. 여기서 시각구성 정보는 시퀀스 전체에 적용되는 정보이며, 2D/3D 구간 정보는 개별 프레임에 대한 정보이다. 프래그먼티드 SVAF를 설계하기 위하여, Ⅲ장에서 소개된 프래그먼티드 ISO BMFF에서와 마찬가지로, 전자의 정보는 시퀀스 초기화 정보로서 초기화 세그먼트의 ‘moov' 박스에 위치시키고, 후자의 정보는 ’moof' 박스에 위치시켜야 한다.

표 1의 'svmi' 박스에서 'stereoscopic_composition_type'와 'is_left_first'는 시각구성 정보로서 기존 SVAF와 동일하게 기술될 수 있다. 다만, 이 박스에 2D/3D 구간 정보는 포함되지 않으므로, 'stereo_mono_change_count'는 0이 될 것이다. 또한 ‘moof'에 2D/3D 구간 정보 기술하기 위한 새로운 박스가 설계되어야 한다.

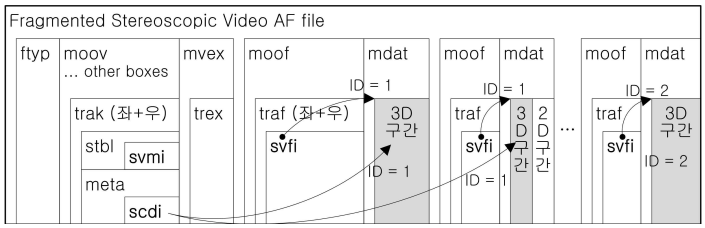
표 3. 스테레오스코픽 비디오 프래그먼트 정보 박스의 구문
Table 3. Syntax of Stereoscopic Video Fragment Information Box

구 문
StereoscopicVideoMediaInformationBox {
stereo_mono_change_count
for(i=1; i<=stereo_mono_change_count; i++) {
sample_count
stereo_flag
scdi_flag
if(stereo_flag && scdi_flag) {
scdi_item_ID
}
}
}

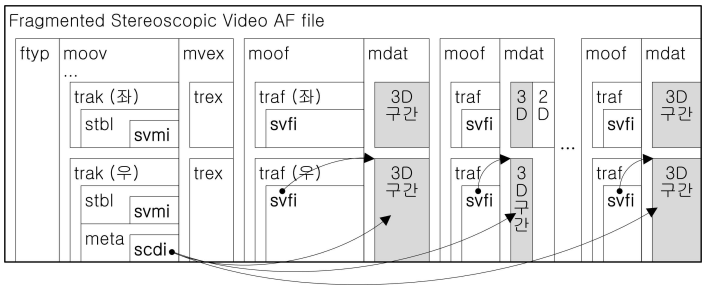
표 3은 본 논문에서 제안하는 스테레오스코픽 비디오 프래그먼트 정보 박스(Stereoscopic Video Fragment Information Box, 'svfi')의 구문을 보여준다. 이 박스는 'moof'내 트랙 프래그먼트 박스(Track Fragment Box, 'traf')에 위치하며, 기본적으로 'svmi' 박스와 같이 ‘stereo_mono_change_count', 'sample_count', 'stereo_flag' 등의 2D/3D 구간 표현을 위한 필드들을 포함하여, 개별 프래그먼트 내부에서의 2D/3D 구간을 표현한다.

또한, 'scdi' 박스는 프래그먼티드 SVAF 구조에서도 기존 SVAF에서와 마찬가지로 'moov' 내부에 위치할 수 있다. 다만, 'iloc'를 통해 하나의 파일 내부에서 'scdi'에서 기술된 파라미터 셋의 적용 위치를 표현했던 메커니즘은 제안하는 프래그먼티드 SVAF에서 적용될 수 없다. 그 이유는 프래그먼트 구조에서는 동일한 스테레오스코픽 구간이 하나의 파일내부에서 여러 개로 나뉘어져 기술상의 복잡성을 유발하고, 세그먼트 구조에서는 하나의 시퀀스를 포함하는 파일이 여러 개로 나뉘어져, 기존 방법대로의 기술이 원천적으로 불가능하기 때문이다.

그러므로, 스테레오스코픽 파라미터 셋 적용 구간을 표현하기 위해 표 3에 나타난 바와 같이, ‘scdi_flag'와 ’scde_item_ID' 필드를, 제안하는 'svfi' 박스에 추가적으로 구성하였다. 'scdi_flag'는 'moov'박스에 'scdi'박스가 포함되어 있음을 나타낸다. 'scdi_item_ID'는 프래그먼트 내부의 스테



(가) 프레임 패킹 방식 (Frame packing type)



(나) 듀얼 스트림 방식 (Dual stream type)

그림 5. 프래그먼티드 스테레오스코픽 비디오 어플리케이션 파일 포맷
Fig. 5. Fragmented file structure of Stereoscopic Video AF

레오스코픽 구간에 부여되는 식별자로써, 재생시 'scdi'와의 링크를 통해 구간별로 각기 다른 파라미터 셋을 적용시키는 역할을 한다.

그림 5는 제안하는 'svfi' 박스 및 'svmi', 'scdi' 박스가 포함된 프래그먼티드 SVAF 구조의 구성 예를 보여주고 있다. 그림 3에서 소개되었던 로컬 SVAF 구조와 마찬가지로, 상기의 예에서는 스테레오-모노-스테레오로 구성되는 3개의 구간을 포함하는 스테레오스코픽 비디오를 사용하고 있으며, 두 개의 스테레오 구간은 각기 다른 'scdi' 적용 파라미터가 적용되고 있다.

그림 5의 (가)는 하나의 트랙으로 구성되는 프레임 패킹 방식의 프래그먼티드 SVAF 구조를 보여주고 있다. 스테레오스코픽 비디오 트랙에는 로컬 SVAF에서와 같이 'svmi'와 'scdi' 박스가 포함되는데, 여기서 'svmi'는 시퀀스 전체에 해당하는 스테레오스코픽 시각구성 정보만을 포함한다. 2D/3D 구간 정보는 각 프래그먼트 'moof'의 'svfi'에 기록되며, 스테레오구간의 경우 파라미터 셋의 식별자를 함께 기술한다. 그림에서와 같이, 앞서 'moov'에 저장된 'scdi'의 파라미터 셋과 부여된 식별자가 일치하는 구간에 해당 파라미터 셋을 링크함으로써, 각기 다른 파라미터 셋을 이용한 재생이 가능하다.

그림 5의 (나)는 두 개의 트랙으로 구성되는 듀얼 스트림 방식의 프래그먼티드 SVAF 구조를 보여준다. 로컬 SVAF에서와 마찬가지로, 'scdi' 박스는 중복성을 피하기 위해 하나의 트랙에만 포함될 수 있으며, 두 번째 프래그먼트에서와 같이 2D 구간에서는 좌/우 영상중 하나의 영상의 ES만 구성하여 저장 및 전송 효율을 높일 수 있다. 또한 이와 같은 듀얼 스트림으로 구성된 전송규격은 2D 장치에서는 하나의 스트림만을 이용한 2D 모드 재생이 가능하여, 3D 장치뿐 아니라 2D장치에서도 호환성 확보가 가능하다는 장점이 있다.

본 논문에서는 제안한 프래그먼티드 SVAF 구조를 DASH 기반의 적응적 스트리밍 서비스에 적용하기 위해, Ⅲ장에서 소개한 프래그먼티드 ISO BMFF에서의 경우와 같은 방법으로, 'ftyp', 'moov', 'mvex' 박스를 포함하는 초기화 세그먼트와, 하나 또는 둘 이상의 프래그먼트를 포함하

는 미디어 세그먼트 구성을 제안한다.

다양한 품질로 부호화된 스테레오스코픽 비디오 스트림을 제안한 방법으로 구성하여, 세그먼트 구성 목록 기술을 위한 DASH의 MPD(Media Presentation Description)[10]와 서버와 단말의 연속적인 요청(Request)과 응답(Response)에 의한 기존의 HTTP 메시지 송수신 방법에 기반하여[11], 호환성있는 3D 비디오의 적응적 스트리밍 서비스 구현이 가능하다.

Ⅴ. 실험 및 검증

본 장에서는 Ⅳ장에서 제안한 3D 비디오의 적응적 스트리밍을 위한 전송규격의 기능성 및 적합성을 검증하기 위한 실험과정 및 검증결과에 대해 소개한다. 본 논문에서는 상기 실험을 위해 그림 6과 같은 3D 적응적 스트리밍 송수신 시스템을 구현하였다. 3D 적응적 스트리밍 송수신 시스템은 Ⅳ장에서 제안한 프래그먼티드 SVAF 및 세그먼트를 미디어 전송규격으로 사용하여 콘텐츠 생성에서부터 단말에서 영상이 제생되기까지의 전 과정을 구현한 엔드투엔드(End-to-end) 시스템으로 그림 6과 같은 구조를 갖는다.

시스템은 그림과 같이 크게 콘텐츠 생성부, 서버부와 단말부로 이루어진다. 콘텐츠 생성부는 Ⅱ장에서 소개한 로컬 SVAF 파일을 입력으로 하는데, 적응적 스트리밍 서비스를 위해 하나의 콘텐츠를 다양한 품질로 부호화한 복수개의 SVAF 파일을 사용한다. 콘텐츠 생성부 내부 SVAF 파일 분석기는 로컬 SVAF 파일을 분석(Parsing)하여, Ⅲ장에서 제안된 'svmi', 'scdi' 박스를 포함하는 초기화 세그먼트와, 'svfi'를 포함하고 일정 시간간격의 스테레오스코픽 비디오의 ES를 담는 미디어 세그먼트로 분할 및 변환하는 역할을 한다. MPD는 분할된 미디어 세그먼트와 초기화 세그먼트들은 품질별, 시간별로 분류하여 목록화한 디렉터리(Directory) 정보로서, XML(eXtensible Mark-up Language) 형식의 텍스트로 저장된다[10].

생성된 MPD와 세그먼트들은 서버부에 저장되며, 서버부는 단말부와의 HTTP 메시지 응답/요청에 의해 세그먼트들을 전송한다. 단말부는 최초 MPD를 요청하여 품질별로 요구되는 대역폭의 크기를 파악하고, 가용대역폭에 따라 선택적으로 세그먼트를 요청한다. 초기화 세그먼트는 해당 품질의 시퀀스를 처음 재생할 때 요청되어 미디어 세그먼트 재생을 위한 초기화 정보로 활용된다. 단말부 내부의 세그먼트 분석기는 전송된 세그먼트에서 스테레오스코픽 비디오 ES와 ES 복호화를 위한 기본 정보들과 3D 특성정보들을 추출하여 각각 비디오 복호화기와 2D/3D 렌더러(Renderer)로 전달한다. 2D/3D 렌더러는 복호화된 프레임들을 3D 특성정보를 기반으로 화면표시장치에 재생하는 역할을 한다.

상기와 같이 설계된 3D 적응적 스트리밍 송수신 시스템을 통해 제안된 전송규격을 검증하기 위하여, 표 4에 명시된 사양의 실험 콘텐트를 구성하였다. 실험 콘텐츠는 MPEG-4 AVC를 이용하여 각각 512, 1024, 3548, 7096 Kbps의 평균 비트율(Bit-rate)로 부호화된 4개의 스테레오

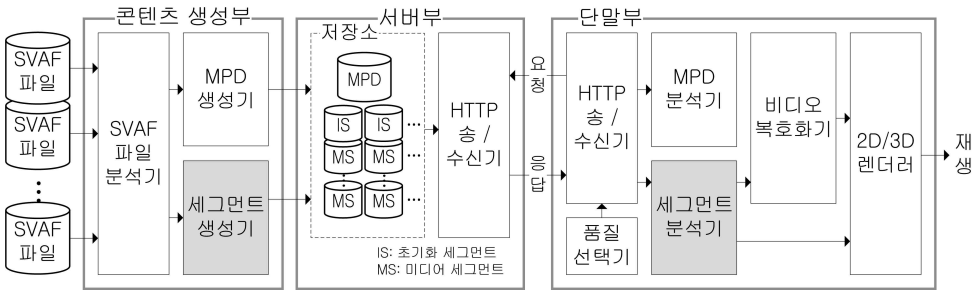


그림 6. 3D 적응적 스트리밍 송수신 시스템
Fig. 6. 3D Adaptive Streaming Service System

스코픽 비디오 시퀀스를 포함한다. 첫 번째와 세 번째 시퀀스는 프레임 패킹방식, 두 번째와 네 번째 시퀀스는 듀얼 스트림 방식으로 스테레오스코픽 비디오가 포함되고, 각각 15초 간격으로 3D 구간과 2D 구간이 반복되는 형태이며, 제안한 'svfi' 박스를 통해 상기 정보를 기록하였다. 또한, 전체 120초의 시퀀스를 10초 단위로 등분하여 개별 미디어 세그먼트에 저장하였다.

그림 7은 상기 구성된 실험 콘텐츠를 위한 실험환경 구성 및 결과를 보여준다. 그림과 같이 실험을 위해 3D HD (High-definition), 2D HD, 3D SD(Standard-definition), 2D SD, 4종류로 각기 다른 특성을 갖는 단말을 사용하였다.

그림 7 상단의 3D HD 단말에서의 실험은 제안된 전송규격을 통해 하나의 단말에서 다양한 품질로 부호화된 스테레오스코픽 비디오가 끊김 없이, 2D/3D 전환 기능을 포함하여 올바르게 재생되는지를 확인하기 위한 것으로, 30초 단위로 차등적으로 가용대역폭을 각각 1000, 2000, 5000, 9000 Kbps로 제한한 환경에서 실시하였다. 제한된 가용대역폭에서 가장 높은 품질의 스테레오스코픽 비디오가 끊김 없이 재생되었으며, 특히 01:32, 01:42 시점의 영상에서 3D에서 2D로 재생모드의 전환이 이루어 졌음을 확인할 수 있다. 이 결과를 통해, 제안된 전송규격을 통해 하나의 단말에서 스테레오스코픽 비디오를 통한 적응적 스트리밍 서비스가 가능함을 확인할 수 있다. 뿐만 아니라, 실험에서와 같이 미디어 세그먼트의 경계와 2D/3D 전환지점의 경계가 일치

표 4. 실험 콘텐츠의 구성 및 세부사항
Table 4. Test content configuration and specifications

구분	시퀀스 1	시퀀스 2	시퀀스 3	시퀀스 4
평균 비트율	512 Kbps	1024 Kbps	3548 Kbps	7096 Kbps
스테레오스코픽 비디오 구성방식	프레임 패킹 (좌우분할)	듀얼 스트림	프레임 패킹 (좌우분할)	듀얼 스트림
영상 크기	640*480	640*480*2	1920*1080	1920*1080*2
2D/3D 구간 구성	2D구간(15초) / 3D구간(15초) 반복 구성			
시퀀스 길이	120 초			
부호화 방식	MPEG-4 AVC Main Profile			
미디어 세그먼트 시간간격	10 초			

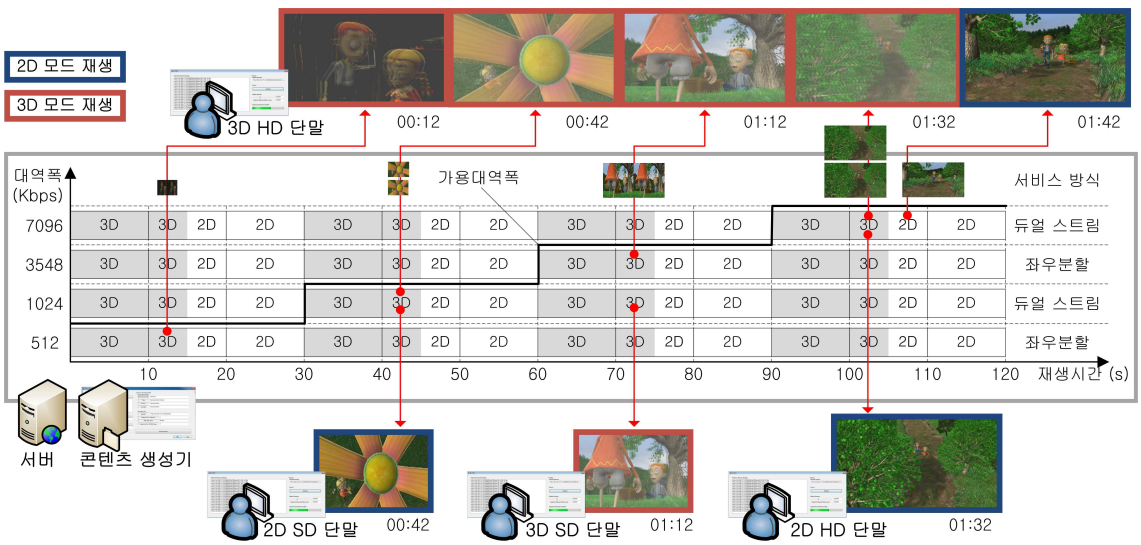


그림 7. 실험환경 구성 및 결과
Fig. 7. Experimental configuration and results

하지 않더라도 2D/3D 모드전환을 할 수 있는 기능을 제공하여, 제안된 전송규격을 통해 다양한 3D 영상 서비스 시나리오에 유연하게 대처할 수 있을 것이다.

또한, 그림 7 하단에 표시된 바와 같이, 3D HD 단말에서 뿐만 아니라, 2D SD, 3D SD, 2D HD 등의 다양한 단말들에서도 실험 콘텐츠를 통한 서비스가 가능함을 확인하였다. 이 결과는, 제안하는 전송 규격을 통해 기존의 단말들과 호환성을 보장하며, 원 소스 멀티 유즈(One source multi-use) 서비스 시나리오의 범주를 넓혀, 2D/3D 선택 기능을 포함하여 적용시킬 수 있는 가능성을 보여준다.

Ⅵ. 결 론

본 논문에서는 최근 멀티미디어 서비스에서 각광받고 있는 콘텐츠 측면의 스테레오스코픽 비디오와, 전송 측면의 적응적 스트리밍 기법을 결합하여 3D 비디오의 적응적 스트리밍 서비스를 위한 전송규격을 제안하였다. 본 규격은 ISO BMFF, DASH, SVAF 등 기존의 기반기술들과 호환성을 고려하여 설계되어, 기존 시스템의 최소한의 변경을 통해 3D 비디오의 적응적 스트리밍 서비스가 구현될 수 있도록 하였으며, 3D 뿐만 아니라 2D 지원 단말과의 호환성을 유지한 서비스가 가능하다. 또한, 현재 시장에서 널리 사용되는 다양한 스테레오스코픽 비디오를 지원하고, 카메라, 디스플레이 파라미터 지원을 통해 고품질의 3D 비디오 재생을 가능하게 하며, 2D/3D 전환 기능 제공을 통해 다양한 3D 기반 서비스 시나리오에 유연하게 대처할 수 있도록 설계 되었다. 본 논문에서 제안하는 기술은 MPEG에서 작업 표준안(Working draft)로 채택되어 표준기술로서도 고려되고 있어[12], 3D 비디오의 적응적 스트리밍 전송방안의 규격화를 통해 관련 산업에서의 실용적인 산업화를 위한 기틀을 제공할 수 있을 것으로 기대된다.

참 고 문 헌

[1] A. Vetro, A.M. Tourapis, K. Muller, and T. Chen, “3D-TV Contents Storage and Transmission”, IEEE Transactions on Broadcasting, Vol. 57, No. 2, pp 384-394, Jun. 2011.

[2] SMPTE, “Stereoscopic 3D Full Resolution Contribution Link Based on MPEG-2 TS”, SMPTE ST 2063-2012, 2012.

[3] DVB, “Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV)”, DVB Document A154, Feb. 2011.

[4] T. Stockhammer, “Dynamic adaptive streaming over HTTP – standards and design principles,” in Proceedings of the Second Annual ACM Conference on Multimedia Systems, New York, NY, USA, pp. 133–144, Feb. 2011.

[5] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC 2nd DIS 23009-1 Dynamic Adaptive Streaming over HTTP”, N12166, Torino, Italy, Jul. 2011.

[6] 3GPP, “Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)”, 3GPP TS 26.247 v10.0.0, Jun. 2011.

[7] ISO/IEC, “Information technology – Multimedia application format (MPEG-A) – Part 11: Stereoscopic video application format”, ISO/IEC 23000-11:2009 First edition, Nov. 2009.

[8] 이장원, 김규헌 “3D 입체영상 서비스를 위한 저장규격 표준 : 스테레오스코픽 비디오 응용포맷”, 제6회 한국정보통신협회 정보통신표준화 우수논문집, pp.46 - 59, 2010. 12.

[9] ISO/IEC, “Information technology – Coding of audio-visual objects – Part 12: ISO base media file format”, ISO/IEC 14496-12:2008 Third edition, Oct. 2008.

[10] 박기준, 이길복, 이장원, 김규헌, “MPEG DASH 기반 service-compatible 3D 콘텐츠 대상 HTTP adapptive streaming 적용방안”, 방송공학회논문지, 제17권 제2호, pp.207-222, 2012. 3.

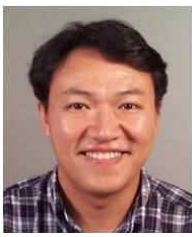
[11] 김정한, 이장원, 김규헌, 서덕영, “HTTP Streaming 환경에서 User Profile 기반 Seamless Framework 제공방법”, 방송공학회논문지, 제16권 제1호, pp.155 - 173, 2011. 1.

[12] ISO/IEC JTC1/SC29/WG11, “WD of ISO/IEC 23000-11 Amd. 3 DASH support”, N12290, Geneva, Switzerland, Nov. 2011.

저 자 소 개

이 장 원

- 2007년 2월 : 경희대학교 전자공학과 공학사
- 2007년 3월 ~ 현재 : 경희대학교 전자전파공학과 석·박사 과정
- 주관심분야 : 멀티미디어 서비스, 디지털 방송



김 규 헌

- 1989년 2월 : 한양대학교 전자공학과 공학사
- 1992년 9월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학석사
- 1996년 7월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학박사
- 1996년 ~ 1997년 : 영국 University of Sheffield, Research Fellow
- 1997년 ~ 2006년 : 한국전자통신연구원 대화형미디어연구팀장
- 2006년 ~ 현재 : 경희대학교 전자정보대학 교수
- 주관심분야 : 영상처리, 멀티미디어 통신, 디지털 대화형 방송