www.osia.or.kr

OSIA S&TR Journal

ISSN 1738-9887 Vol. 30, No. 3, September 2017
• 발행인 : 장영민 회장/OSIA, 국민대 • 편집위원 : 김규헌 교수/경희대

30th OSIA
(사)개방형컴퓨터통신연구회
OPEN STANDARDS & INTERNET ASSOCIATION

MPEG 표준화 관련 최근 동향 소개

ETRI
차지훈 그룹장

많은 ICT 기술들이 그러하듯 미디어 기술도 하루가 다르게 진화하고 있다. 10년 전만해도 모바일 단말에서 동영상을 자유롭게 소비하는 것이 쉽지 않았으나, 이제는 누구나 일상에서 즐기는 보편화된 서비스가 되었다. MPEG (Moving Picture Experts Group: ISO/IEC 산하 동영상 전문가 그룹)은 지난 25년간 다양한 멀티미디어 서비스를 위해 기반이 되는 표준기술을 제공하여 전세계 멀티미디어 산업 발전에 혁혁한 공을 세웠다.

최근 MPEG은 기존의 2차원적 서비스 패러다임을 넘어서 진일보한 고품질 몰입형 멀티미디어 서비스를 지원하기 위해 2016년 10월 회의에서 처음으로 향후 5년간의 표준기술 개발 로드맵을 제시하였다. 산업계와 관련 전문가의 의견을 수렴하여 2017년 7월에 업데이트된 로드맵에는 대표적인 서비스로 이종 플랫폼간의 미디어 분배, VR360, 6자유도 몰입형 미디어 등이 포함되어 있다. 여기에 제시된 대표적인 서비스와 기술을 살펴보면, 차세대 미디어 서비스를 수식하는 키워드로 360, 전방향, 몰입형 등을 꼽을 수 있다.

이러한 기술 및 표준 개발 동향은 산학연을 망라하여 한동안 지속이 될 것으로 전망된다. 이에 맞추어 본 호에서는 활발한 표준기술 개발이 진행되고 있는 고품질 몰입형 미디어 서비스를 위한 기반 기술로서 차세대 비디오 부호화, VR 서비스를 위한 3D 모델링 등에 많은 활용이 예상되는 포인트 클라우드 코딩, OMAF & 몰입형 미디어 전송 표준, 3D 오디오 표준기술을 소개하고자 한다.

첫 번째 차세대 비디오 부호화 표준화 동향 논문에서는 HEVC (High Efficiency Video Coding) 이후 차세대 비디오 부호화 표준을 위한 기술 및 표준화 동향에 대해 기술한다. 특히 새로운 비디오 부호화 기술의 탐구를 위해 결성된 JVET (Joint Video Exploration Team)에서의 주요 기술 동향과, 차세대 비디오 부호화 표준을 위한 CfE (Call for Evidence), CfP (Call for Proposals) 등의 표준화 동향에 대해 소개한다.

두 번째 논문은 최근 MPEG에서 시작하고 있는 PCC (Point Cloud Coding/Compression) 기술의 표준화에 대한 동향을 소개하였다. 포인트 클라우드의 획득기술로 시작하여 포인트 클라우드 기반의 서비스 시나리오를 알아보고, 이를 바탕으로 포인트 클라우드 압축 기술에 대한 요구사항도 소개하고 있다. 기술 표준화의 현황 부분에서는 CfP를 중심으로 평가 조건, 테스트 데이터, 주요 경쟁 기술 등을 요약하고 향후에 전개될 PCC 표준화에 대해 저자의 의견을 제시하고 있는 부분이 매우 흥미롭다.

세 번째 논문은 MPEG 에서 진행중인 OMAF (Omnidirectional MediA Format)과 DASH (Dynamic Adaptive Streaming over HTTP), MMT (MPEG Media Transport) 기반의 몰입형 미디어 전송 기술에 대해 소개한다. 3DoF (Degrees of Freedom)를 지원하는 전방향 미디어의 획득, 공유 및 전송을 위한 본 표준들의 기술요소 전반에 대해 소개하고 관련 표준화 동향과 향후 전망에 대해 논의한다.

마지막으로 네 번째 논문은 최신 오디오 표준 기술인 MPEG-H 3D Audio에 대해 소개한 논문이다. 다채널, 다객체, 장면 오디오 신호에 대한 부호화 기술과 함께 헤드폰이나 다양한 스피커 배치에 따른 최적의 오디오 재생을 위한 렌더링 기술을 포함하고 있는 MPEG-H 3D Audio의 표준화 현황, 기술 요소 및 서비스에 대해 제시하고 있다.

바쁜 중에도 훌륭한 원고를 준비해 주신 저자들께 감사를 드리며, MPEG 표준기술에 관심을 갖고 계시는 독자들에게 관련 자료들이 도움이 되기를 기대한다.

차세대 비디오 부호화 표준화 동향

백아람, 최상규, 최해철
한밭대학교 멀티미디어공학과

Abstract

2013년 1월에 표준화가 완료된 High Efficiency Video Coding (HEVC) 이후 ISO/IEC MPEG (JTC 1/SC 29/WG 11)과 ITU-T VCEG (Q6/16)은 차세대 비디오 부호화 표준 기술 개발을 위해 Joint Video Exploration Team (JVET)을 결성하고 2015년 10월 1차 회의를 시작으로 현재까지 7번의 JVET 회의를 진행하며 기술 탐색을 위한 논의를 활발히 진행하고 있다. 본 고에서는 JVET에서 다루고 있는 주요 기술 동향에 대해 살펴보고, 차세대 비디오 부호화 기술의 전망과 함께 차세대 비디오 부호화 표준을 위한 Call for Evidence, Call for Proposals 등의 표준화 동향에 대해 소개한다.

I. 서론

높은 부호화 효율을 기반으로 다양한 응용 분야에서 폭넓게 사용되고 있는 High Efficiency Video Coding (HEVC) [1]와 확장 표준들은 앞으로도 그 활용도가 더욱 증가할 것으로 예상되고 있다. 하지만 여전히 제한된 네트워크 대역을 통해 전송 또는 저장이 필수적이므로 더 높은 비디오 부호화 효율을 제공할 차세대 비디오 부호화 기술이 요구되고 있다. 또한 사용자 소비 단말 소유에 따른 다음과 같은 새로운 비디오 서비스 및 응용 분야가 생성되며 고효율 비디오 압축 기술에 대한 산업 수요도 증가할 것으로 예상되며, 새로운 비디오 압축 부호화 기술을 가장 잘 받아들일 수 있는 분야로도 전망되고 있다.

- 순수한 소프트웨어 기반 응용 분야로서 압축 비트스트림에 대한 저장 필요성이 없는 서비스
- 소프트웨어 복호화 기반의 응용 분야로서 이미 압축된 비트스트림을 서버에서 단말 사용자에게 제공하는 서비스
- 하드웨어 복호화 기반의 응용 분야로서 스마트폰과 같이 빠른 사용자 단말 교체 주기를 가지고 있는 분야
- 새로운 비디오 압축 부호화 기술을 빠른 주기의 교체가 어려운 분야로서 하드웨어 복호기 적용이 필요한 응용 분야

〈그림 1. 차세대 비디오 부호화 기술이 적용 가능한 신흥 시장의 예〉

기존 시장과 향후 시장에서의 고효율 비디오 압축 기술 적용 시나리오 또한 긍정적으로 전망되고 있다. 기존 시장으로는 지상파 및 위성 방송, 케이블, IPTV, 스튜디오 콘텐츠, 홈시네마 및 패키지 미디어, 감시 비디오, 이동 통신망을 통한 VoD 서비스 등이 있으며, 비디오 품질, 포맷, 네트워크 환경 등을 고려하였을 때 새로운 비디오 압축 부호화 기술의 적용 시나리오가 긍정적으로 전망되었다. 더불어 기존 시장에 비해 신흥 시장에서 새로운 비디오 압축 부호화 기술의 적용 가능성이 더욱 높은 것으로 전망되었고, 신흥 시장으로는 증강 및 가상 현실 그리고 비디오 센서를 사용하는 자동차 응용, 미디어 중심 사물 인터넷(Internet of Things, IoT) 등이 있다[2].

이처럼 다양한 분야와 시장에서 지속적으로 요구되고 있는 차세대 비디오부호화의 현재 동향을 본고를 통해 살펴볼 것이다. 본고는 다음과 같이 구성되어 있다. 2 장에서는 JVET에서 수행한 차세대 비디오 부호화 표준을 위한 기술 탐구(Exploration) 동향을 소개하며, 3 장에서는 본격적으로 진행되고 있는 차세대 비디오 부호화 표준화 동향을 기술한다. 마지막으로 4 장에서 결론으로 마무리 짓는다.

II. JVET의 차세대 비디오 부호화 기술 탐구

2.1 JVET 회의 동향

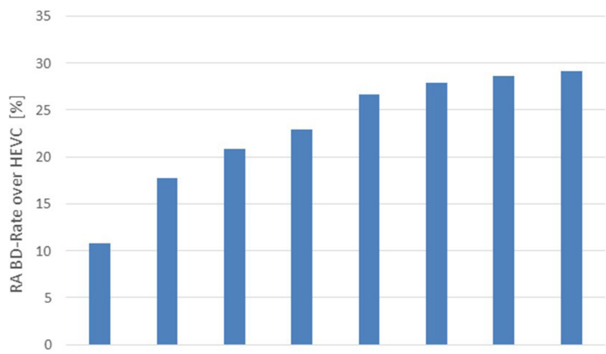
2013년 HEVC의 표준 최종안[1]이 확정된 이후 ISO/IEC MPEG (JTC 1/SC 29/WG 11)은 2014년 10월 제110차 MPEG 표준화 회의를 시작으로 차세대 비디오 압축 부호화 기술의 표준 재정을 위한 차세대 비디오 부호화(Future Video Coding, FVC) 표준 개발 논의를 시작하였다. 제110차 MPEG 표준화 회의에서 초청된 7 개 기관의 차세대 비디오 부호화 표준 제정에 대한 입장을 기반으로 FVC 기술 탐색을 위한 그룹을 결성하며 논의 준비를 시작하였고, 제111차 회의에서는 FVC 표준 개발을 위한 요구사항 초안 문서를 도출하였다. 제112차 회의에서는 FVC 기술에 대한 2차 워크숍을 개최하였고, 제출된 기술 및 의견 기고서에 대한 토의를 거쳐 FVC 요구사항에 대한 첫 번째 문서와 함께 FVC 표준 개발을 위한 테스트 영상 제안 요청서를 발표하였다. 제113차 회의에서는 FVC 표준 개발에 앞서 사전 기술 탐색을 수행하기 위해 Joint Video Exploration Team (JVET)을 ITU-T VCEG (Q6/16)과 함께 결성하였고, FVC 기술 탐색을 위한 공통의 플랫폼으로 Joint Exploration Model (JEM) 첫 번째 버전[3]을 발표하였다. 이후 표 1과 같이 JVET은 2015년 10월 제네바 회의를 시작으로 2017년 9월 현재까지 7번의 회의를 통해 Standard Dynamic Range (SDR), High Dynamic Range (HDR), 360° 비디오 분야에 대한 기술 탐색과 FVC 표준 개발을 위한 테스트 영상에 대한 조사 활동을 함께 진행하고 있다. 또한 JVET은 현재까지 진행된 FVC 표준을 위한 기술 성숙도를 살펴보기 위해 2017년 4월 Call for Evidence (CfE) [4]를 발행하였고, 2017년 7월 이탈리아 토리노에서 개최된 제119차 MPEG 표준화 회의에서 Qualcomm, InterDigital, Samsung 등 여러 기업 및 기관에서 제출한 총 8건의 CfE response에 대해 논의하였다.

<표 1. 차세대 비디오 부호화 기술 탐구를 수행한 JVET 회의 현황>

MPEG Meeting	Date	City	Released JEM Version
MPEG 113	2015.10.19~10.21	Geneva	JEM 1.0
MPEG 114	2016.2.20~2.26	San Diego	JEM 2.0
MPEG 115	2016.5.26~6.01	Geneva	JEM 3.0
MPEG 116	2016.10.15~10.21	Chengdu	JEM 4.0
MPEG 117	2017.1.12~1.20	Geneva	JEM 5.0
MPEG 118	2017.3.31~4.7	Hobart	JEM 6.0
MPEG 119	2017.7.13~7.21	Torino	JEM 7.0

2.2 JVET 기술 동향

MPEG과 VCEG이 JVET을 구성하기 이전에 VCEG에서는 비디오 부호화 효율 향상을 위한 기술적 논의가 이루어지고 있었으며, 2015년 2월 ITU-T SG 16 VCEG 미팅에서 HEVC를 기반으로 추가적인 부호화 툴을 이용하여 부호화 효율을 향상시킨 기고서가 발표되었다[5]. 기고서에서는 Large block & transform, Advanced Temporal Motion Vector Prediction (ATMVP), Overlapped Block Motion Compensation (OBMC), Linear Model Cross-Component Prediction (LMCCP), Adaptive Multiple Transform (AMT), Adaptive Loop Filter (ALF)의 추가적인 툴을 HEVC에 추가하여 All intra (AI), Random access (RA), Low-delay B (LDB) 조건에서 각각 Y-BD Rate 기준으로 6.8%, 10.4%, 8.9%의 부호화 효율 향상을 얻었으며, 이 기술들을 조합하여 HMKTA 1.0 플랫폼으로 공개되었다. 2015년 8월에는 HMKTA 2.0이 공개되었고 Finer granularity arbitrary directional intra prediction (67 intra modes), Multi-Parameter Intra (MPI), Intra Coding Improvement, Adaptive Motion Vector Resolution (AMVR), Frame Rate Up-Conversion (FRUC) mode, Bi-directional Optical Flow (BIO), Local Illumination Compensation (LIC), Adaptive Secondary Transform, Enhanced CABAC (eCABAC) 등의 부호화 툴이 추가되어 AI, RA, LDB에서 각각 10.9%, 17.9%, 14.8%의 부호화 효율 향상을 획득하였다.

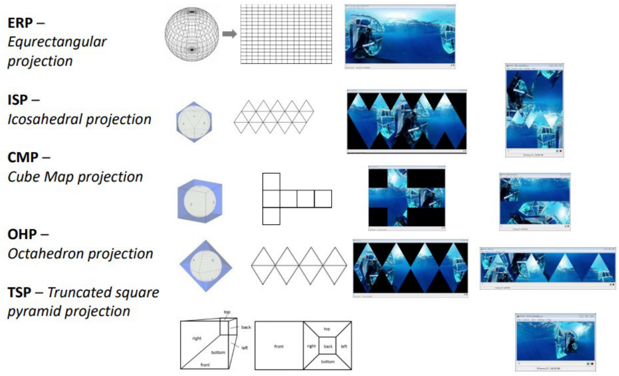


<그림 2. 차세대 비디오 부호화를 위한 JEM 버전별 기존 HEVC 대비 Random Access 부호화 효율과 복잡도 현황 [6]>

이후 VCEG과 MPEG이 JVET을 결성하여 2015년 10월 JEM 1.0이 공개되었으며, 이는 Position Dependent Intra Prediction Combination (PDPC), Adaptive Reference Sample Smoothing (ARSS), Mode Dependent Non-separable Secondary Transform (NSST), Affine Transform Prediction (AFFINE), Quadtree plus Binary tree (QTBT) block structure 등의 툴을 포함시켜 AI, RA, LDB, Low-delay P (LDP) 에서 각각 14.2%, 20.8%, 16.7%, 19.9%의 성능 향상을 보였다. 2016년 3월에 공개된 JEM 2.0에서는 AFFINE, OBMC, De-blocking Filter (DBF) 툴 간의 조율 기술과 Non-MPM Intra Mode Coding, 1/16 pel Motion Vector Storage Accuracy에 대한 기술 기고를 채택하였으며, AI, RA, LDB, LDP에서 각각 15%, 22.9%, 16.9%, 20.0%의 부호화 효율 향상을 이루었다. 2016년 6월에 공개된 JEM 3.0에서는 Exploration Experiments (EE)로 진행되었던 QTBT Block Structure가 HEVC의 Quadtree Structure를 대신하여 적용되었다. 추가적으로 AFFINE, BIO, Intra Prediction의 MPM 유도, ATMVP, ALF, NSST에 대한 간소화 및 개선과 함께 64x64 Transform Block에 대한 Transform Skip (TS) 툴이 적용되어, AI, RA, LDB, LDP에서 각각 18.2%, 26.1%, 20.7%, 24.1%의 부호화 효율 향상을 보였다. 2016년 11월에 공개된 JEM 4.0은 QTBT에 대한 Parameter 조정과 NSST에서의 불필요한 중복 과정 제거, NSST 개선 등에 대한 기술이 적용되어, AI, RA, LDB, LDP에서 각각 19.3%, 27.9%, 21.7%, 25.2%의 성능 향상이 이루어졌다. 2017년 2월에 공개된 JEM 5.0에서는 Decoder Side Motion Vector Refinement, FRUC, Motion Vector Difference (MVD) coding, ALF에 대한 개선, Multiple Direct Modes for Chroma Intra Coding, Enhanced Cross-component Linear Model Intra Prediction 등의 기술을 채택하여, AI, RA, LDB, LDP에서 각각 19.7%, 28.6%, 22.1%, 25.6%의 부호화 성능 향상을 이루었다. 한편 그림 2와 같이 JEM 5.0부터는 이전까지 증가하고 있던 복잡도에 대해 고려하기 시작하여 복잡도를 낮추기 위한 노력을 수행하였다. 2017년 4월에 공개된 JEM 6.0 [7]은 BIO, FRUC, Transform Skip (TS), In-loop Filter의 개선 등에 대한 기술을 바탕으로 AI, RA, LDB, LDP에서 각각 20.2%, 29.1%, 22.4%, 26.1%의 부호화 효율 향상이 이루어졌다.

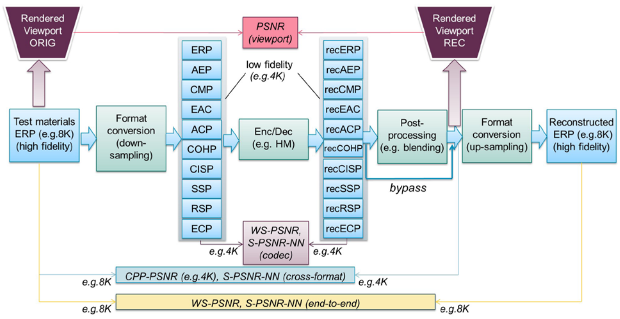
2.3 JVET 360° 비디오 기술 동향

JVET의 FVC 표준 개발을 위한 기술 탐색 활동에서 360° 비디오 부호화 기술 역시 중요 사안으로 진행되고 있다. 360° 비디오 부호화 기술의 요구사항으로 30~50%의 비트율 감소, VGA 부터 8K×4K 영상 포맷 지원, HDR과 Wide Colour Gamut (WCG) 등의 다양한 비디오 콘텐츠 특성 지원, 적응적 스트리밍 지원 등이 있으며, 가상 현실과 기존과 다른 프로젝션 방법 등도 고려되고 있다. JVET은 360° 비디오 부호화 기술을 위해 다양한 프로젝션 포맷을 지원하는 360Lib



〈그림 3. JVET의 360° 비디오 부호화를 위한 프로젝션 방법 예〉

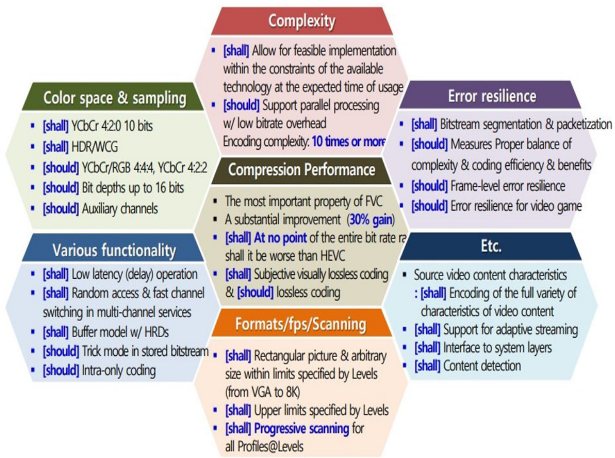
Software (SW)를 발표하고 Common Test Condition (CTC)를 정의하였다. 그림 3은 360Lib가 지원하는 대표적인 프로젝션 포맷들을 나타낸다. 또한 그림 4와 같이 입력 360° 비디오에 대한 부호화 프로젝션 포맷 변환 등을 지원하는 프레임워크를 구성하였고, 성능 측정을 위한 기준도 정의하였다. 현재까지 360Lib은 3.0 버전[8]까지 공개되었다.



〈그림 4. JVET의 360° 비디오부호화 기술을 위한 프레임워크 [9]〉

III. 차세대 비디오 부호화 표준화 동향

3.1 차세대 비디오 부호화 표준 요구사항



〈그림 5. 차세대 비디오 부호화 표준화를 위한 요구사항 [10]〉

MPEG에서는 FVC 표준을 위한 본격적인 첫 걸음으로 FVC의 요구사항을 제시하였다. 그림 5는 핵심 요구사항들을 간략하게 나타내고 있으며, 가장 중요히 요구되고 있는 압축 효율은 이전 HEVC과 동일한 인지 화질 대비 30%의 비트율 개선을 목표로 하고 있으며, 객관적 무손실 압축을 지원과 함께 주관적 무손실 압축도 고려하고 있다. 상기 요구사항에서, 포맷은 직사각형 화면과 VGA부터 8K×4K까지의 공간 해상도를 지원해야 하며, 레벨에 의해 규정된 수치 내에서 임의 크기의 해상도도 지원 가능해야 한다. 프레임율은 레벨에 의해 최대 프레임 율이 명시 가능해야 하며, 스캐닝 방법으로 순차 주사 지원을 모든 프로파일과 레벨에서 지원해야 한다. 컬러 공간 및 컬러 샘플링에 대해서 4:2:0 샘플링 YCbCr 컬러 공간과 컴포넌트 별 10 비트 입력을 지원해야 하며, YCbCr/RGB 4:4:4 및 YCbCr 4:2:2 포맷 지원, 컬러 컴포넌트 별 16비트 입력 지원, 부가 채널 지원 등을 포함해야 한다. 소스 비디오 콘텐츠 특성에 대한 요구사항으로는 범위가 매우 넓은 비디오 특성을 포괄하는 압축 기능 지원과 적응적 스트리밍 지원, 시스템 계층으로의 인터페이스, 콘텐츠 검출 방법 지원 등이 고려되었다. 인코더와 디코더 복잡도는 새로운 표준이 이용되는 시점에 예상 사용 시간이 허용 가능한 범위 이내여야 하며, 비트율 오버헤드 비용 관점에서 병렬 처리 지원도 고려되었다. 부호화 복잡도는 기존 HEVC보다 10배 또는 그 이상도 수용 가능하다. 오류 강건성과 관련하여 네트워크 전송을 위해 비트스트림 분할과 패키지화가 지원되며, 복잡도와 코딩 효율 등의 적절한 균형, 프레임 단위에서의 오류 강건성 지원, 비디오 게임을 위한 오류 강건성 지원 등도 고려되었다. 마지막으로 다양한 기능들과 관련하여 저지연 동작, 임의 접근과 다중 채널 서비스에서의 빠른 채널 전환, 가상 디코더 버퍼 모델(Hypothetical reference decoders, HRDs)을 고려한 버퍼 모델을 지원하며, 저장된 비트스트림에서 트릭 모드와 모든 프레임의 인트라 부호화 지원도 함께 요구사항으로 고려되었다.

3.2 Call for Evidence

JVET의 기술 탐구 활동 동안 부호화 효율 향상 수치가 요구사항에서 정의한 목표치와 근접함에 따라 제118차 MPEG 표준화 회의의 6차 JVET 회의 기간 중인 2017년 4월 7일 최종 CfE[4]가 발행되었다. CfE에서는 SDR, HDR, 360° 비디오에 대한 평가 조건 및 항목에 대해 정의하고 있으며, 각각의 평가 조건에 대한 실험 영상은 SDR의 경우 8 개의 4K 영상과 5 개의 2K 영상, HDR은 5 개의 2K 영상, 360°은 5 개의 8K 영상으로 구성하였다. 6차 JVET 회의의 CfE에 대해 2017년 7월에 개최된 제119차 MPEG 및 7차 JVET 회의에서 SDR, HDR, 360° 각각 2, 2, 4 개, 총 8개의 CfE response에 대한 평가를 진행하였다.

• SDR 비디오 부호화

InterDigital은 BIO 기술의 복잡도 간소화를 통해 표 2와 같이 JEM 6.0 버전과 비슷한 부호화 효율을 획득하면서 부/복호화 복잡도 상당히 감소시킨 코덱을 제안하였다. 삼성의 제안 코덱은 context model 초기화, BIO, 화면내 부호화의 Four-tap Interpolation Filter, PDPC, ARSS 기술 등의 개선을 통해 표 2와 같은 부호화 효율 향상을 보였다.

〈표 2. JVET Call for Evidence의 SDR Category Response에 대한 기존 HEVC 대비 Random Access 부호화 효율 및 복잡도 비교〉

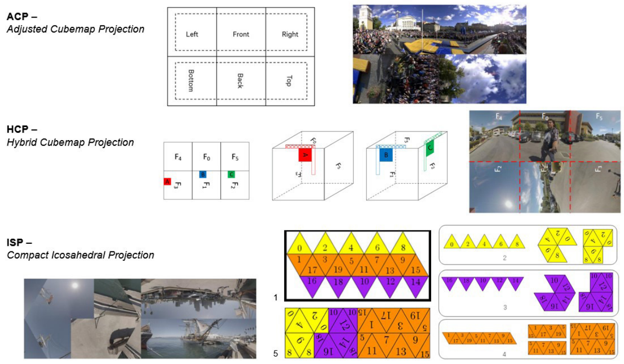
	BD-rate			Time	
	Y	U	V	Encoding	Decoding
JEM 6.0 CfE bit-matching [6]	-34.3%	-46.2%	-47.7%	100%	100%
InterDigital's SDR CfE Response [11]	-34.2%	-46.2%	-47.7%	92%	75%
Samsung's SDR CfE Response [12]	-36.2%	-42.4%	-44.0%	-	-

• HDR 비디오 부호화

FastVDO는 입력 영상의 Y'CbCr 4:2:0 포맷 변환 방법을 개선하였고, Technicolor는 Luma-based Cross Component Scaling 기술을 바탕으로 Y'CbCr의 신호 특성을 개선하여 표 3과 같이 부호화 효율을 향상시켰다.

〈표 3. JVET Call for Evidence의 HDR Category Response에 대한 기존 HEVC 대비 Random Access 부호화 효율 비교〉

MPEG Meeting	tPNSR Y	DE100	PSNRL 100	wPSNR Y	wPSNR U	wPSNR V
FastVDO's HDR CfE Response [13]	-24.0%	-41.3%	-30.0%	-30.0%	-54.4%	-43.0%
Technicolor's HDR CfE Response [14]	-27.9%	-39.6%	-30.3%	-30.4%	-55.8%	-60.7%



〈그림 6. JVET CfE 360˚ Category의 Response들에서 사용된 Projection Format〉

• 360˚ 비디오 부호화

Qualcomm은 기존 HEVC의 플랫폼인 HM 16.15과 360Lib 3.0 기반으로 Adjusted Cubemap Projection (ACP) 포맷을 이용하였고, InterDigital은 JEM 6.0과 360Lib 3.0 기반으로 Hybrid Cubemap Projection (HCP) 포맷을 이용하였다. 삼성은 SDR 비디오 response에서 소개한 JEM 6.0 기반의 코덱 엔진에서 Compact Icosahedral Projection (ISP) 포맷을 이용하였고, 각각의 Response에서 이용된 포맷은 그림 6과 같다. TNO는 각 프레임을 subsampling하여 4개의 서브프레임을 생성하고 4개의 서브프레임에 대해 부/복호화를 진행한 후 원 영상으로 복원하는 Polypahse Subsampling을 소개하였다. 360˚ 비디오 부호화의 Response에 대한 부호화 효율 결과는 표 4와 같다.

〈표 4. JVET Call for Evidence의 360˚ Video Category Response에 대한 기존 HEVC 대비 Random Access 부호화 효율 비교〉

	WS-PSNR (End to End)		
	Y	U	V
Qualcomm's 360˚ CfE Response [15]	-10.7%	-6.1%	-6.2%
InterDigital's 360˚ CfE Response [16]	-31.22%	-57.29%	-60.20%
Samsung's 360˚ CfE Response [17]	-29.8%	-42.6%	-47.3%

3.3 Call for Proposals

Call for Evidence의 response 결과를 통해 HEVC 이후 개발된 비디오 부호화 기술에 대한 성숙도가 새로운 비디오 부호화 표준화를 시작하기에 충분하다고 판단한 MPEG 및 ITU-T SG 16은 2017년 9월 제119차 MPEG 회의 및 제7차 JVET 회의에서 Call for Proposals (CfP)의 draft를 공개하였다[18]. 이 CfP draft에서 기술하고 있는 새로운 FVC 표준화 일정은 다음과 같다.

• 최종 CfP 발행: 2017년 10월 31일
• 정식 등록 기간: 2017년 11월 1일 ~ 2018년 1월 15일
• CfP에 대응하는 제안 기술에 대한 문서 등록: 2018년 3월 중
• CfP에 대응하는 제안 기술에 대한 평가: 2018년 4월 중
• Test Model 선택 기간: 2018년 4월 ~ 2018년 10월
• 최종 FVC 표준화 종료: 2020년 10월

IV. 결론

본고에서는 HEVC 이후 새로운 비디오 부호화 기술과 관련된 ITU-T SG.16과 MPEG이 결성한 JVET의 기술 탐구 동향과 차세대 비디오 부호화 표준을 위한 표준화 동향에 대해 살펴보았다. 차세대 비디오 부화화 기술은 주로 가상 및 증강현실, IoT, 8K 비디오, 대형 파노라마 비디오와 같은 신흥 서비스 및 시장에서 그 필요성이 제기되었으며, 이에 따라 JVET은 SDR과 HDR 뿐만 아니라 360˚ 비디오에 대해서도 표준화를 준비하고 있다. 현재까지 JVET은 SDR과 HDR, 360˚ 비디오 부호화에 대한 기술 탐구 진행과 함께 새로운 비디오 부호화 표준을 위한 CfE에 대한 Response를 통해 비디오 부호화 기술의 발전 성숙도를 확인하였기에, 차세대 비디오 부호화를 위한 CfP를 2017년 10월 31일에 발행하고 본격적인 표준화를 시작할 예정이다. 그러므로 신흥 시장을 위한 영상 기술을 포함하는 차세대 비디오 표준에 대한 기술력 경쟁은 지금 가장 중요한 단계로 접어들고 있으며, 최종 표준화 종료 시점인 2020년 10월까지 전 세계의 관련 기업과 연구기관, 대학들의 기술력 확보를 위한 노력은 지속될 것으로 판단된다. 이에 앞으로 펼쳐질 차세대 비디오 부호화 표준화에서의 국내 산학연의 선전을 기대한다.

References

[1] ITU-T Rec. H.165 and ISO/IEC 23008-2, "High efficiency video coding," Final draft approval Jan 2013.

[2] Donggyu Sim, Jong-Ki Han, Munchurl Kim, "Future Video Coding 기술추적 보고서," 2015년 MPEG 포럼 기술 추적 보고서, Dec 2015.

[3] "Algorithm Description of Joint Exploration Test Model 1," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 1nd Meeting, Geneva, Oct 2015, Doc. JVET-A1001.

[4] "Joint Call for Evidence on Video Compression with Capability beyond HEVC," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1002.

[5] J. Chen, Y. Chen, M. Karczewicz, X. Li, H. Liu, L. Zhang, and X. Zhao, "Coding tools investigation for next generation video coding," ITU-T SG 16, doc. COM-16-C806, Jan 2015.

[6] "JVET AHG report: Tool evaluation (AHG1)," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0001.

[7] "Report of AHG3 on JEM software development," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0003.

[8] "JVET AHG Report: 360 video conversion software development," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0006.

[9] "JVET common test conditions and evaluation procedures for 360° video," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G1030.

[10] ISO/IEC JTC 1/SC 29/WG 11 document N15340, "Requirements for a Future Video Coding Standard v1," Jun 2015, Warsaw

[11] "InterDigital's Response to the SDR Category in Joint Call for Evidence on Video Compression with Capability beyond HEVC," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0028.

[12] "Samsung's response to Joint CfE on Video Compression with Capability beyond HEVC (SDR category)," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0029.

[13] "FastVDO Response to JVET CfE for HDR," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0021.

[14] "CfE response to the HDR category from Technicolor," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0022.

[15] "Qualcomm's response to Joint CfE in 360-degree video category," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0023.

[16] "InterDigital's Response to the 360° Video Category in Joint Call for Evidence on Video Compression with Capability beyond HEVC," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0024.

[17] "Samsung's response to Joint CfE on Video Compression with Capability beyond HEVC (360° category)," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G0025.

[18] "Draft Joint Call for Proposals on Video Compression with Capability beyond HEVC," Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/ WG11), 7th Meeting, Torino, July 2017, Doc. JVET-G1002.

Biographies

백아람

2012년: 한밭대학교 멀티미디어공학과 학사
2014년: 한밭대학교 멀티미디어공학과 석사
주요 관심분야: 비디오 부호화, 병렬 프로그래밍
e-mail : aram98123@naver.com

최상규

1975년: 단국대학교 수학과 학사
1981년: 단국대학교 수학과 석사
1989년: 인하대학교 수학과 박사
주요 관심분야: 멀티미디어 신호처리
e-mail : sgchoi@hanbat.ac.kr

최해철

1997년: 경북대학교 전자공학과 학사
1999년: 한국과학기술원 전기및전자공학과 석사
2004년: 한국과학기술원 전기및전자공학과 박사
주요 관심분야: 영상통신, 비디오 부호화, 패턴 인식
e-mail : choihc@hanbat.ac.kr

MPEG-I Point Cloud Coding 표준화

장의선
한양대학교

Abstract

MPEG-I (Immersive) 표준의 하나로 포인트 클라우드 압축 (Point Cloud Coding/Compression) 기술의 표준화가 시작단계에 있다. 비디오와 그래픽스 부분의 접점이 AR/VR 부문에서 확대되고 있는 가운데, 공간상의 화소로 볼 수 있는 점 구름 데이터의 압축 기술에 대한 표준화에 대한 요구 또한 과거 그 어느 때보다 높다. 본 고에서는 점 구름 압축 표준화에 대한 현재의 동향과 향후 전개 방향에 대하여 소개하고자 한다.

I. 서론

점 구름이라고도 부르는 포인트 클라우드는 말 그대로 점들이 구름처럼 모인 형태를 의미한다. 포인트 클라우드는 원래는 3D 모델의 획득과정에서 추출된 샘플들의 모임을 의미하는 말이다. 정교한 3D 모델일수록 더 많은 3D 포인트들이 발생하고 이 과정에서 많은 포인트를 빠른 시간 안에 저장, 전송, 계산하는 과정은 전통적으로 그래픽스 영역에서 매우 중요한 문제이다.

전통적으로 포인트 클라우드는 모델의 획득 과정에서 주로 사용되었다. 저장, 전송, 렌더링 과정에 직접 사용하기에는 포인트의 수가 많아 빠른 처리에 부담이 되었기 때문이다. 이러한 문제를 해결하기 위해 등장한 모델이 3D 메쉬(Polygonal Mesh)다. 하나의 면을 공유하는 여러 개의 포인트를 한 개의 다각형 (대개의 경우 삼각형)으로 표시하는 3D 메쉬 표현은 포인트 개수 자체의 감소에 손쉬운 해결책이 되었다. 그리고 추가적인 압축 기술을 MPEG에서 3D Mesh Coding이란 이름으로 표준화하기도 하였다.

지난 수십 년간 끊임없지 진화하고 있는 컴퓨팅 성능은 그래픽 데이터 처리 부문에서도 많은 변화를 가져왔다. 그 중 하나가 메쉬 기반의 표현에서 포인트 클라우드를 직접 사용하는 것이다. 과거에는 처리과정의 복잡도로 인해 다수의 포인트 클라우드를 사용하는 대신 메쉬 기반의 표현을 사용해왔다. 그러나 키넥트 등 사용자 측면의 포인트 클라우드 획득 장치의 대중화와 함께 포인트 클라우드 자체를 서비스단에서 직접 사용하는 사례가 대량으로 증가하게 되었다. 특히 실시간 포인트 클라우드 처리는 메쉬 표현 변환과정에서 야기되는 문제들(예를 들면, Manifold 문제, topology 변환, boundary edge 처리 등)이 원천적으로 차단되어 복잡도가 문제가 되지 않는다면 매력적인 표현으로 인식되고 있다. 최근 포인트 클라우드 서비스가 전통적인 그래픽, 게임, AR/VR 등의 영역을 넘어서서 자동차나 드론 등에 탑재된 LIDAR등과 같은 레이저 센서를 이용한 위치기반 서비스의 영역으로 확대되면서 공간상의 점 기반 서비스의 중요성이 더욱 확대되어 가고 있다.

앞서 언급한대로, 포인트 클라우드 자체의 데이터 용량을 줄이는 좋은 방법은 Mesh 형태로 변환하여 처리하는 것이다. 그리고, 이러한 변환은 게임 등의 기존 서비스에서는 매우 효과적인 전략이기도 하다. 그러나, 포인트 클라우드의 사용범위가 넓어지면서, 포인트 자체의 위치가 서비스의 중요한 일부가 되는 경우, 포인트 클라우드 데이터의 압축은 선택이 아닌 필수 요건이 되어가고 있다.

본 기고에서는 포인트 클라우드 압축 표준화에 관련하여 최근 MPEG에서 진행 중인 표준화 활동을 소개하고, 향후 기술의 발전 동향을 살펴보고자 한다. 2절에서는 먼저 주요 포인트 클라우드 기반 응용 및 서비스를 소개한다. 3절에서는 포인트 클라우드의 기본 데이터 구조 및 파일 포맷을 소개한다. 4절에서는 CfP를 중심으로 MPEG에서 진행 중인 표준화 활동에 대해 소개한다. 그리고 5절에 결론으로 맺음하였다.

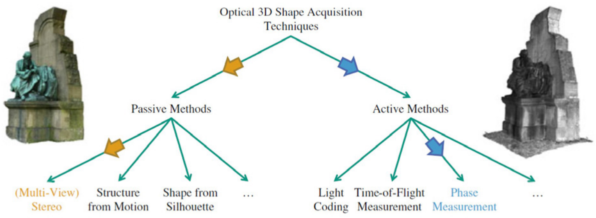
II. 포인트 클라우드 기반 응용 및 서비스

2.1 포인트 클라우드 획득

포인트 클라우드는 원래 3D 그래픽 모델링의 한 분야로 시작했다. 3D 그래픽 모델을 만들어 내는 과정은 그래픽 아티스트의 순수 창작 아이디어인 경우도 있지만, 많은 경우 실물에서 직접 3D 기하(Geometry) 정보를 추출하여 이를 다양한 응용해 활용하여 왔다. 실물에서부터 포인트 클라우드를 획득할 수 있는 방법은 그림 1과 같이 크게 능동(Active)적인 방법

과 피동(Passive)적인 방법으로 구분된다. [1]

피동적인 방식은 스테레오 영상, 비디오, 음영 등과 같은 주어진 영상 속의 정보를 분석하여 영상 속 3D 기하 정보를 추출하는 방식이다. 우리 눈이 3D를 인식하는 것도 피동적인 획득 방식이라고 볼 수 있다. 피동적인 획득 방식은 부가적인 장비가 없어도 주어진 영상에서 기하 정보를 추출할 수 있다는 장점이 있지만, 주어진 정보로부터 획득불가능한 영역이 Occlusion과 같은 영상 간 간섭으로 인해 획득에 제약이 존재한다. 아울러, 상대적인 기하 정보는 추출 가능하나 실제 크기나 거리 같은 절대적인 기하 정보를 추출하는 데에는 오차가 클 수 있다는 단점도 존재한다.



〈그림 1. 포인트 클라우드 획득 방식의 분류 예 [1]〉

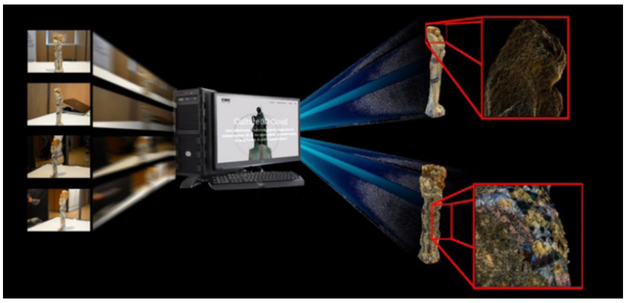
능동적 획득 방법은 레이저 광선을 쏘고 그 광선이 반사되는 시간을 계산하는 이른 바 이동 시간 (Time-of-Flight)을 계산하여 실물과 센서간의 실측 거리를 정확하게 측정하는 방법이다. 센서의 특징에 따라, 이동 시간을 계산하는 방식, 측정 오차, 측정 가능 거리 등이 결정된다. 정확한 거리를 측정할 수 있다는 장점이 있지만, 센서를 탑재한 장비가 기존의 카메라 같이 저렴하지만은 않다는 점에서 그동안 전문 영역에서만 사용되어 왔다. 마이크로소프트에서 게임기 Xbox의 3D 센서로 2010년 발표된 키넥트 등 저렴한 3D 획득 장치의 도입으로 대중 미디어 영역으로 확장되고 있다. 아울러, Lidar로 불리는 빛과 레이더를 결합한 형태의 능동 센서가 인공위성, 드론, 차량 등의 영역에 광범위하게 응용되면서 대량의 정밀한 포인트 클라우드 데이터가 만들어지고 있다.

2.2 포인트 클라우드 기반 서비스 시나리오

획득 장치의 다양성만큼이나 획득된 포인트 클라우드가 사용될 수 있는 응용분야 또한 다양하다. 2016년 MPEG에서 정리한 문서에 따르면 다음과 같은 서비스로 정리할 수 있다. [2]

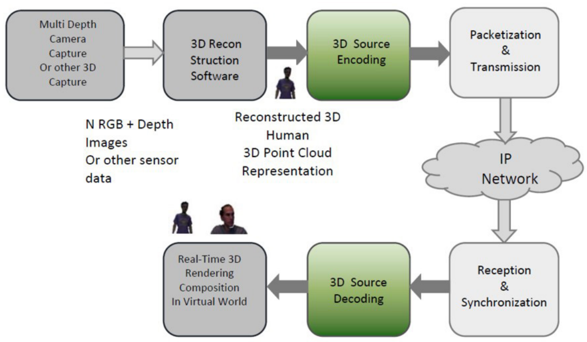
- Real-Time 3D immersive telepresence
- Content VR viewing with Interactive Parallax
- 3D Free viewpoint Sport Replays Broadcasting
- Geographic Information Systems
- Cultural heritage
- Autonomous Navigation Based on Large-Scale 3D Dynamic Maps

MPEG에서 정리한 서비스 시나리오는 다시 크게 문화 유산을 위한 포인트 클라우드 서비스 (그림 2), Interactive/immersive 포인트 클라우드 서비스 (그림 3), GIS/Navigation을 위한 포인트 클라우드 서비스(그림 4)의 세 분야로 나눌 수 있다.


〈그림 2. 문화 유산을 위한 포인트 클라우드 서비스의 예 [2]〉

2.3 포인트 클라우드 압축 기술의 요구사항

포인트 클라우드를 이용한 서비스는 그 용도와 사용 범위가 다양하지만, 몇 가지 공통점을 추릴 수 있다. 첫번째는 무손실 및 손실 압축의 필요성이다. 좋은 화질의 정지영상 한 장의 해상도가 8백만 화소 (4Kx2K)를 넘어간다. 8백만개의 정지영상 한 장에는 공간상의 기하 정보는 포함되어 있지 않다. 이미 정의된 2차원 평면 안에 스캔 오더를 따라 8백만 화소들이 정렬되어 있기 때문이다. 그런데, 각 화소의 공간 정보가 추가적으로 압축되어야 하고, 화소와 화소간의 연관성도 없어진 상태이므로 기존의 비디오 압축과는 달리 압축에 대한 더 많은 연구가 필요한 상황이다.



〈그림 3. Interactive/immersive 포인트 클라우드 서비스의 예 [2]〉

- 3D Point Cloud Representation
  - X, Y, Z: 3D 공간 좌표를 이른다. 공간좌표는 Voxel화 할 경우, 최소 좌표 간격이 실제 공간의 해상도를 지칭

하게 된다. 그러나, 현재의 표현 영역에서는 실제 공간 상의 해상도를 표현하지 않고, 수치로 주어진 가상 공간의 영역 내에서 표현되는 값으로 표현되어 있다.

- Attributes (예: colour, reflectance, normal vectors and transparency): 각각의 포인트에 연결되는 색, 반사도, 법선, 투명도 등의 정보 등이 공간정보외에 추가 될 수 있다.
- View-dependent attributes: 대용량 포인트 클라우드가 경량의 디바이스에서 렌더링되어야 하는 상황에 필요한 기능으로, 관찰자의 시점, 터미널의 렌더링 능력 등을 고려하여 필요한 추가 기능을 제공할 수 있도록 하였다.
- Time-varying: 움직임이 없는 정적인(static) 포인트 클라우드와 시간적으로 변화는 (dynamic) 포인트 클라우드를 모두 포함하여 표현하도록 하였다.

〈그림 4. GIS/Navigation을 위한 포인트 클라우드 서비스의 예 [3]〉

다음으로, 압축에 대한 요구사항을 정리하면 다음과 같다.

- 3D Point Cloud Compression
- 손실 (Lossy) 압축: 원격 채팅, 게임 등의 약간의 손실이 주어져도 미디어의 화질에 크게 영향을 주지 않는다고 판단되는 경우에 기하정보와 색 등 속성값(Attribute)이 원본과 다르게 압축되거나 생략될 수 있다.
  - 무손실(Lossless) 압축: 원본 데이터의 최대 품질의 재현을 가능하게 하고자, 원본의 데이터 정밀도(Precision)을 최대한 잃지 않고 유지하며, 또 포인트의 수도 원본과 같이 유지하는 것이 무손실 압축의 목표이다. 이 때, 부동소수점으로 입력된 값의 경우, 일정 범위의 임계치 (Tolerance value)를 설정하고 해당 임계치 이내의 오차만 가능하도록 하는 유사 무손실 (Near Lossless) 압축도 무손실의 범위로 간주하고 있다.
  - Progressive and/or scalable 압축: 전체 비트스트림이 다 전송되지 않거나, 일부만 전송되어도 포인트 클라우드의 복원이 가능하도록 점진적인 (Progressive) 복원 또는 가변 해상도를 적용한 (Scalable) 복원이 가능하도록 하는 기능이다.
  - View-dependent coding: 데이터의 크기, 데이터 전송속도, 단말기의 계산 용량 등의 제약에 따라, 단말기에서 한꺼번에 포인트 클라우드를 복원하기 어려운 경우, 비트스트림의 일부에 접근하여 부분적인 전송, 복호화, 복원이 가능하도록 하는 기능이다.
  - Random Access: 다이나믹 포인트 클라우드와 같이 시간에 따라 변화하는 포인트 클라우드의 경우 특정 시간의 포인트 클라우드에 비트스트림의 구조에 따른 순차적 접근이 아닌 임의의 접근이 가능하도록 하는 기능이다.
  - Error resilience: 데이터 전송 시 오류를 최소화하는 기능이다.

포인트 클라우드의 표현과 압축을 위한 요구사항을 앞서 살펴보았다. 매우 많은 요구사항들이 있지만, 여기서 가장 중요한 요구사항은 압축률이다. 현재, MPEG에서 10월에 있을 제안서 성능 평가 때 중요하게 평가하는 기준 중 하나는 압축 비트율이다. 목표하는 비트율은 5 ~ 40 Mbps 구간이다. 이것은 포인트 클라우드 압축에 특화된 값이라기 보다는 비디오 압축에서 사용하는 비트율을 참고한 값이다.

앞서 언급한 5 ~ 40 Mbps 서비스가 어느 정도의 압축 성능일까? 간단한 예를 들어보자. 백 만개 정도의 포인트를 가진 포인트 클라우드가 기하정보로만 표현되고 있다고 가정하자. 초당 30 프레임을 서비스해야 하고 각 축 별로 10비트씩 표현한다면 한 포인트는 30비트, 한 프레임은 30 Mb, 그리고 초당 비트율은 900 Mbps가 된다. 900 Mbps를 5 ~ 40 Mbps 사이로 추가 압축하려면 180 ~ 22.5:1의 압축이 되어야 한다. 현재 MPEG에서 사용하고 있는 참조 기술은 대략 10:1 정도의 압축 기능을 제공한다. 따라서, 향후 만들어질 표준 압축 기술은 이러한 기대 성능을 맞추어야 할 것이다.

III. 기술 표준화 현황

본 절에서는 MPEG에서 진행중인 CFP를 중심으로 현재의 기술 표준화 상황을 설명하겠다.

3.1 MPEG 포인트 클라우드 압축 표준화 일정

MPEG에서는 표 1과 같이 포인트 클라우드 압축 기술의 표준일정을 예정하고 있다. 올 10월에 Call for Proposals[5]에

대한 제안들에 대한 평가를 진행하고, 작업표준안 (WD)를 만들고 2018년에는 표준을 최종 완성 짓는다는 계획이다. 일정은 표준화의 진행에 따라 좀 더 긴 시간 동안 진행될 확률도 높을 것으로 예상된다.

〈표 1. 포인트 클라우드 압축 표준화 일정〉

Year	Month	Day	MPEG meeting	City	Country	Stage
2017	01	20	117	Geneva	CH	Approval of CFP
2017	10	27	120	Macao	CN	Approval of WD
2018	04	20	122	San Diego	US	Approval of CD
	07	20	123	Ljubljana	SI	Approval of DIS
	10		124	??	CA	Approval of FDIS

3.2 Test Data

앞에서 논의한 요구사항에 기반하여 MPEG PCC 표준화를 위해서 선정된 테스트용 데이터는 표 2와 같다. 크게 Static, Dynamic, Dynamic acquisition 등 세가지로 분류하고 있다. 그림 5, 그림 6, 그림 7에 각각의 범주에 따른 데이터의 유형을 확인할 수 있다. PCC test data set은 MPEG website [6]에서 구할 수 있다.

3.3 평가 조건 및 방법 – 객관적 방법

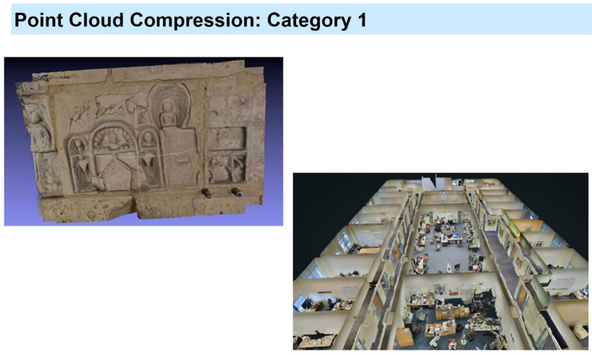
MPEG PCC CFP에서 제시하는 제안 평가 방법은 크게 객관적 방법과 주관적 방법으로 나누어 진행될 예정이다. 먼저 객관적 방법을 설명하도록 하겠다.

Test Data에서 제시된 세개의 범주에 따라 표 2와 같이 다양한 조건으로 테스트를 진행하고 그 결과를 제안서에 보고하도록 하고 있다. 범주 별로 다양한 테스트를 실시하고 있지만, 크게 세가지 범주의 테스트로 다시 분류해 볼 수 있다. 첫번째는 무손실 기하정보 압축, 두번째는 손실 기하정보 압축, 세번째는 손실 기하정보 및 색정보 압축이다.

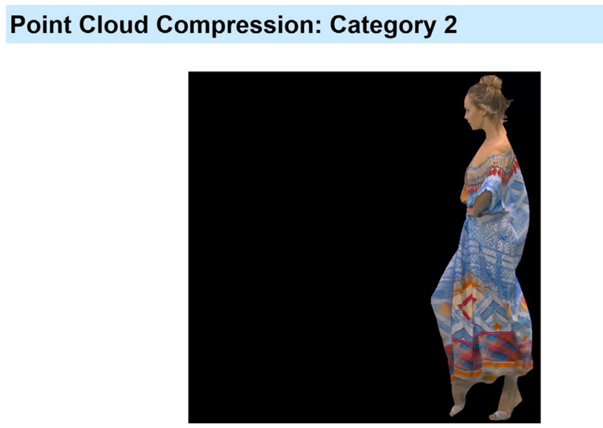
〈표 2. Category 별 실험 조건〉

Test Sub-Condition	Code for bitstream naming convention
Category 1: lossless geometry – no attributes	00
Category 1: lossless geometry – lossy color	01
Category 1: lossy geometry – lossy color	02
Category 2: lossy geometry – lossy color / All Intra	03
Category 2: lossy geometry – lossy color / Random Access	04
Category 3: lossless geometry – no attributes	05
Category 3: lossless geometry – lossy color	06
Category 3: lossless geometry – lossy reflectance	07
Category 3: lossy geometry – no attributes / All Intra	08
Category 3: lossy geometry – no attributes / Random Access	09
Category 3: lossy geometry – lossy color & reflectance / All Intra	10
Category 3: lossy geometry – lossy color & reflectance / Random Access	11
Category 3: fused – lossy geometry	12
Category 3: fused – lossy geometry, lossy color & reflectance	13

기하 정보의 압축은 간단히 보면 원래 점의 수가 압축 후에 몇 개의 점으로 줄어드는 가 하는 것과 압축된 점들이 원래의 점들에 얼마나 가까운가 하는 두 가지로 평가할 수 있다. 압축 전후의 점의 수가 같고, 압축 전후의 가장 가까운 두 점들 (즉, 압축전 점과 압축후 점) 사이의 거리가 허용오차 이내로 모두 계산되는 경우는 무손실 압축으로 평가한다.



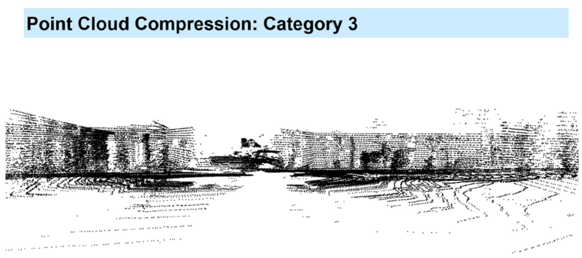
〈그림 5. MPEG PCC Test Data Category 1의 예〉



〈그림 6. MPEG PCC Test Data Category 2의 예〉

점의 수가 바뀌는 경우와 허용오차보다 점간 거리가 커지는 경우는 모두 손실 압축에 해당된다. 압축 전후의 두 점간 거리를 계산할 때 크게 'Point-to-point' (점간) 거리를 측정하는 방법과 'Point-to-plane' (점면간) 거리를 측정하는 방법을 다 사용하여 측정한다.

색 정보의 경우 RGB, YUV 등 칼라 정보와 상관없이 YUV로 변환하고 압축 전후에 최단거리로 쌍을 이루는 점들의 색을 서로 비교하는 것으로 색의 왜곡도를 평가한다. 기존의 비디오에서 사용하는 MSE, PSNR을 그대로 사용한다. 자세한 평가 도구 및 방법에 대한 설명는 웹사이트를 잠조하면 된다. [7]



〈그림 7. MPEG PCC Test Data Category 3의 예〉

3.4 평가조건 및 방법 – 주관적 방법

주관적인 평가는 객관적이지 않은 평가를 의미하는 것이 아니다. 다만, 평가자의 평가가 사람마다 달라질 수 있다는 것을 의미한다. 객관적 방법이 있음에도 주관적 화질 평가를 도입하는 것은 객관적 방법으로만 확인하기에는 부족하다는 판단 때문이다. 이미 비디오나 오디오에도 이 같은 방식이 도입되어 있기도 하다. 다만, 무손실 압축의 경우는 주관적 방법 없이 객관적 방법으로만 평가한다.

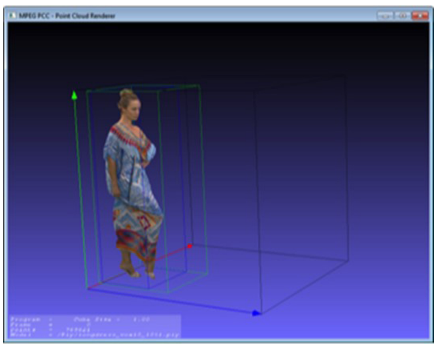
표 3은 주관적 화질 평가를 위한 테스트 데이터와 평가 기준이다. 표에서 처럼, 모든 테스트 데이터를 사용하지 않고, Class A와 Class B의 일부 경우만 가려 뽑아서 시행한다. 각 데이터 별로 모두 4개의 압축 포인트를 설정하고 각 포인트에 가까운 기술을 모아서 참조 기술과의 성능을 비교하게 된다. 참조 기술은 다음 절에서 자세히 설명하기로 한다.

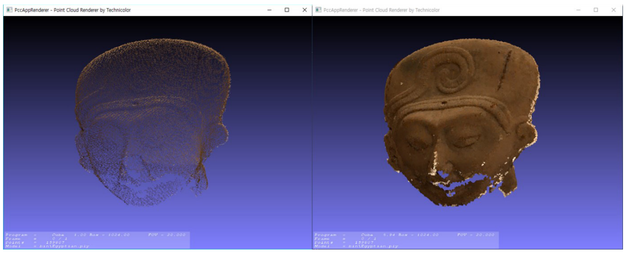
〈표 3. 주관적 화질평가를 위한 Test Data와 평가 기준〉

Test Data Set	Category	Operating Point	Point Size*
Egyptian_mask (Class A)	1	R1*	16.9
		R2*	8.33
		R3*	5.2
		R4*	4.62
Façade9 (Class A)	1	R1*	7.4
		R2*	3.65
		R3*	2.03
		R4*	1
Longdress_vox10_1300 (Class B)	1	R1*	7.7
		R2*	3.65
		R3*	2.03
		R4*	1.8
8i VFB – Long_dress (Class B)	2	R1	
		R2*	8.32
		R3	
		R4*	2.02
8i VFB – Red_and_Black (Class A)	2	R1*	15.61
		R2	
		R3*	4.12
		R4*	2.02
8i VFB – Soldier (Class A)	2	R1	
		R2*	8.32
		R3	
		R4*	2.02

표 3에서 Point Size는 참조기술의 압축 결과물의 렌더링시 사용한 포인트의 (각 축당) 크기이다. 포인트 클라우드 결과물의 렌더링은 사용하는 렌더러의 성능에 따라 화질이 크게 달라질 수 있다. 이 점을 고려하여, MPEG에서는 그림 8과 같이 모든 압축 기술의 결과물을 렌더링할 렌더러를 별도로 개발하여 사용하고 있다. [8]

포인트 사이즈의 결정은 해당 포인트 클라우드 압축 성능에 많은 영향을 줄 수 있다. 그림 9에서처럼, 같은 포인트 클라우드라 할 지라도 포인트의 크기가 너무 작으면 포인트와 포인트 사이의 공간이 노출되고, 너무 크면 포인트와 포인트간의 간섭이 커져 화질이 나빠 질 수 있으므로 해당 압축기술에 가장 최적의 포인트 사이즈를 제안자가 제안기술에 대하여 제시하도록 하였다.


〈그림 8. MPEG PCC Player〉


〈그림 9. Point Size를 달리했을 때의 복원된 결과〉

3.5 주요 경쟁 기술

포인트 클라우드 압축 기술이 MPEG에서 표준화 하기 이전에도 이미 존재하고는 있다. 대표적인 기술이 LASzip이라는 기술로 주로 대량의 포인트 클라우드 데이터를 Zip과 유사한 형태로 압축하는 기술이다. [9] 이 기술이 ZIP에 비해 더 유리한 것은 스트리밍에 더 유리하다는 것이다. 따라서, 대용량의 포인트 클라우드 데이터를 압축하고 복원하는데 유리하다. 보고된 자료에 근거하면, 현재 LASzip의 압축률은 10:1 수준이다. 물론, 이 정도의 압축률보다 높은 압축률을 목표로 현재 MPEG PCC가 진행중이다.

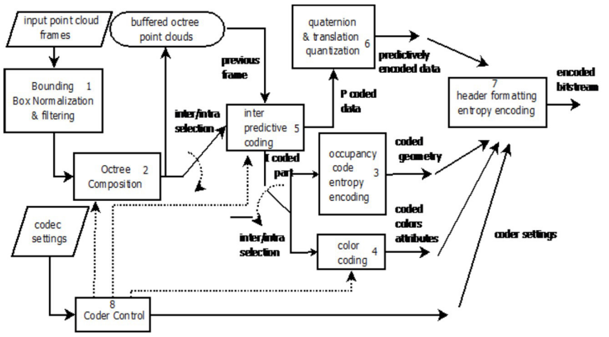
최근 Google의 Draco라는 메쉬 압축기술이 발표되었다. 이 기술은 메쉬 객체를 압축하는 것이 주요 목표이지만, 포인트 클라우드도 압축 가능하다고 알려져 있다. 자세한 성능이나 정보는 아직 많이 공개되어 있지 않아 정확한 성능 비교를 논하기는 어렵다.

MPEG에서 Anchor로 선택하여 사용하는 참조 기술[10]은 네덜란드의 CWI에서 제안한 기술로, 그림 10과 같이 Intra/Inter 구조를 가지고 있다. Intra 압축의 경우 입력된 포인트 클라우드를 바운딩 박스를 기반으로 Octree로 재구성하고 옥트리의 노드정보를 엔트로피 압축을 하는 방식이다. Inter 압축은 이미 압축된 참조 프레임과의 유사도를 비교하여 Quaternion과 Translation으로 참조 프레임으로부터의 변화치를 압축하는 P coding과 유사도가 적다고 판단하여 I coding하는 두가지 방식이 있다. 색 정보의 경우, 64개의 색을 모아 8x8 블록화하고 이를 JPEG으로 압축하는 방식을 채택하고 있다.

참조 기술의 경우 현재 대략 10:1 정도에서 무리 없는 압축 기술로 판단되고 있다. 그러나, 현재 PCC가 추구하는 목표에는 미치지 못하는 만큼 10월에 있을 제안서 리뷰에서 더 좋은 기술들이 제안될거라는 기대감이 있다.

3.6 향후 기술적 쟁점

10월에 있을 제안서 평가에서 어느 정도 성능의 기술들이 발표될 지 미리 예측하는 것은 어려울 것 같다. 다만, 10개의 기관이 제안서를 발표하겠다는 의향서를 지난 7월 MPEG에 전했다. MPEG의 다른 표준 코덱의 개발 역사를 고려해보면, 우수하지만 비슷한 성능을 가진 코덱 몇 개가 최종 경쟁을 거쳐 검증 모델로 선택되고, 약 1년간 이 모델의 압축률을 개선하는 실험을 거쳐 최종 표준으로 발전할 것으로 예상된다.



〈그림 10 MPEG PCC 참조 기술의 Encoder diagram [11]〉

PCC CfP에 얽힌 이야기를 하나 소개하면, 내용이 회차를 거듭하면서 많이 바뀌었다는 것이다. 더 정확히는 내용이 더 많이 늘어났다. 그 중에 하나가 무손실 압축에 대한 논의다. 무손실의 개념을 영상처럼 주어진 비트표현에 충실한 무손실로 정의했다가, 입력 데이터의 반 이상이 부동 소수점 표현을 쓰고 있어서 As Is 무손실의 개념이 무색해졌다. 그래서, 지금은 Near-lossless까지를 무손실에 가깝게 생각하는 개념가지 확장이 되었다. 향후 정확한 무손실의 개념은 표준화와 함께 정립되어 갈 것으로 보인다.

아직 많이 논의가 되지 않은 부분 가운데 하나가 바로 기하정보와 색정보간의 비트율 배분문제이다. CfP에서는 이 부분을 제안자의 임의 선택 영역으로 분류해서 문제삼고 있지 않지만, 향후 효율적인 PCC 기반 서비스를 생각한다면 꼭 표준화 내에서 연구되어야 할 내용이다.

논의가 많이 이루어지지 않은 또 하나는 포인트의 저장 순서 보존에 대한 것이다. 많은 응용 서비스에서 포인트 클라우드 내의 포인트는 저장순서에 상관이 없다. 그러나 일부 서비스는 저장 순서가 중요한 경우가 있을 수 있다. 만약 포인트 순서를 지켜서 압축해야 한다면 이러한 기능을 표준이 제공할 수 있을까? 제공할 수 있다면 그 비용은 추가로 얼마나 늘어날 것인가? 이러한 물음 역시 PCC 표준화를 통해 그 해답을 제시해야 할 것이다.

PCC가 기하정보와 색정보를 압축하는 기능을 제공하게 되면 비디오 압축에 쓰였던 많은 기술들이 응용 가능 할 것으로 보인다. 그래서, 기술의 발전 속도가 기대된다. 충분한 기술의 숙성이 이루어 진다면, 실감형 미디어의 전달 기술의 하나로 충분한 자리매김을 할 것으로 보인다.

IV. 결론

본 고에서는 최근 MPEG에서 시작하고 있는 포인트 클라우드 압축 기술의 표준화에 대한 동향을 소개하였다. 포인트 클라우드의 획득기술로 시작하여 포인트 클라우드 기반의 서비스 시나리오를 알아보고, 이를 바탕으로 포인트 클라우드 압축 기술에 대한 요구사항도 살펴보았다. 기술 표준화의 현황 부분에서는 CFP를 중심으로 평가 조건, 테스트 데이터, 주요 경쟁 기술 등을 요약하고 향후에 전개될 PCC 표준화에 대해 나름의 의견을 제시하였다.

Acknowledgement

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신.방송 연구개발사업의 일환으로 하였음 [2017000000001493, 자율비행을 위한 포인트 클라우드 부호화 및 렌더링 기술 연구]

References

[1] Weinmann, Martin. Reconstruction and Analysis of 3D Scenes: From Irregularly Distributed 3D Points to Object Classes. Springer, 2016.
[2] Use Cases for Point Cloud Compression (PCC), ISO/IEC JTC1/SC29/WG11 MPEG2016/N16331, Geneve, June 2016.
[3] ISO/IEC JTC1/SC29/WG11/M40495, “Mobile Mapping System Point Cloud Data from Mitsubishi Electric”, Hobart, AU, April, 2017.
[4] ISO/IEC JTC1/SC29/WG11 MPEG2016/N16330, “Requirements for Point Cloud Compression,” Geneve, June 2016.
[5] ISO/IEC JTC1/SC29/WG11 MPEG2016/N16330, “Call for Proposals for Point Cloud Compression,” Geneve, January 2017.
[6] Point Cloud Compression: Data set and anchors, http://mpegfs.int-evry.fr/MPEG/PCC/DataSets/pointCloud/CfP.

[7] Point Cloud Compression: objective evaluation, http://wg11.sc29.org/svn/repos/MPEG-04/Part16-Animation_Framework_eXtension_(AFX)/trunk/3Dgraphics/3DG-PCC/tags/CfP/distortion/
[8] PCC player, http://wg11.sc29.org/svn/repos/MPEG-04/Part16-Animation_Framework_eXtension_(AFX)/trunk/3Dgraphics/3DG-PCC/tags/CfP/player/
[9] Isenburg, Martin. "Laszip." Photogrammetric Engineering & Remote Sensing 79.2 (2013): 209-217.
[10] PCC Anchor Software, http://wg11.sc29.org/svn/repos/MPEG-04/Part16-Animation_Framework_eXtension(AFX)/trunk/3Dgraphics/3DG-PCC/tags/cfp
[11] Mekuria, Rufael, and Pablo Cesar. "MP3DG-PCC, open source software framework for implementation and evaluation of point cloud compression." Proceedings of the 2016 ACM on Multimedia Conference. ACM, 2016.

Biographies

장의선

1991년: 전북대학교 전자계산기공학과 학사
1996년: SUNY at Buffalo, 전기 및 컴퓨터공학 박사
2002년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학부 교수
주요 관심분야: Image/Video/Graphics Media Compression
e-mail : esjang@hanyang.ac.kr

Omnidirectional MediA Format 및 몰입형 미디어 전송 표준 동향

이장원
LG Electronics

Abstract

VR (Virtual Reality) 콘텐츠 획득과 소비가 가능한 사용자 디바이스들이 널리 보급되고 있는 가운데, MPEG (Moving Picture Experts Group)에서는 전방향 미디어 (omnidirectional media)의 획득, 공유 및 전송을 위한 OMAF (Omnidirectional MediA Format)와, 몰입형 미디어 전송 기술 표준화를 활발히 진행하고 있다. 본고에서는 OMAF와 몰입형 미디어 전송 표준의 기술 전반과 표준 동향, 그리고 향후 전망에 대해 소개하고자 한다.

I. 서론

최근 HMD (Head Mount Display)와 360도 카메라의 사용이 보편화 되면서, 전방향 미디어를 이용한 몰입형 미디어 서비스 기술에 대한 관심이 높아지고 있다. CG (Computer Graphics)에서 주로 사용되던 VR 콘텐츠 처리 기술들이 실제 촬영, 획득 가능한 실사 영상에까지 확대되어 사용되기 시작하면서, 사용자들은 전방향 미디어를 손쉽게 획득, 소비할 수 있게 되었다.

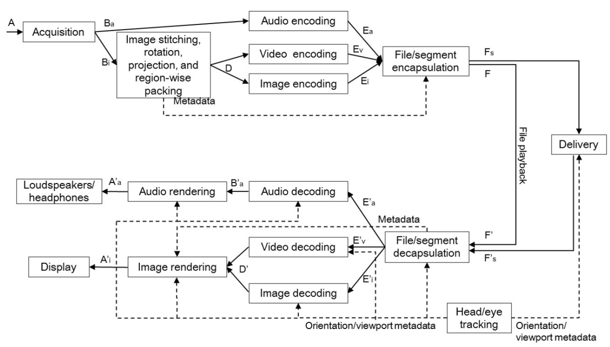
이러한 미디어 시장과 기술의 움직임 속에 MPEG은 다양한 디바이스와 플랫폼 간에 전방향 미디어의 공유, 유통을 원활히 위해, 전방향 미디어 저장 포맷 기술인 OMAF (Omnidirectional MediA Format)와 DASH (Dynamic Adaptive Streaming over HTTP), MMT (MPEG Media Transport) 기반 전방향 미디어 전송 기술 표준화 작업을 진행하고 있다.

본고는 최근 활발히 진행중인 OMAF와 몰입형 미디어 전송 표준의 기술 전반과 표준화 동향에 대해 소개하고자 한다. II장에서는 본 표준들의 요소기술에 대해 전반적으로 살펴볼 것이다. III장에서는 표준화 동향과 향후 전망에 대해 소개하며, IV장에서 결론을 맺는다.

II. OMAF 및 몰입형 미디어 전송 표준 기술 개요

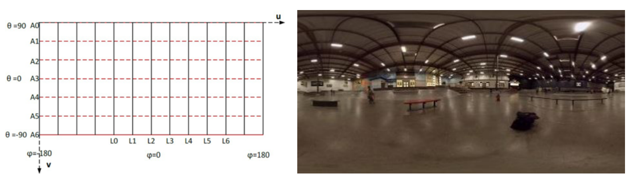
2.1 OMAF 표준 기술

OMAF는 전방향 미디어 서비스를 위한 저장 포맷 표준으로, 기존의 2D (2-dimensional) 미디어 처리를 위한 획득, 압축, 저장, 전송을 위한 요소기술들을 기반으로 설계되고 있다. 또한, 전방향 이미지, 비디오 처리를 위한 새로운 프로세싱 단계들과 여기에 필요한 메타데이터들을 기술 범주로 포함한다. 그림 1은 OMAF 콘텐트의 동작흐름을 end-to-end 관점에서 도시하여 보여주고 있다.

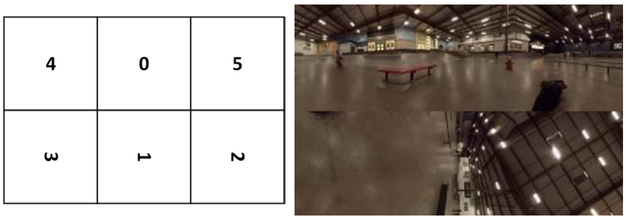


<그림 1. OMAF 콘텐트 동작 흐름도 [1]>

전방향 미디어는 오디오 센서와 카메라들로부터 획득된다. 일반적으로 전방향 비디오, 이미지는 여러 대의 카메라로부터 획득되며, 획득된 이미지를 하나로 연결하는 스티칭 (stitching), 3D 공간상의 전 방향 이미지를 정해진 포맷의 사각형 픽처에 투영하는 프로젝션 (projection)과정을 거치게 된다. 다양한 프로젝션 포맷들이 연구/개발되고 있으며, OMAF는 현재 시장에서 보편적으로 사용되는 Equi-Rectangular Projection (ERP), Cube Map Projection (CMP)을 지원하고 있다. 그림 2, 그림 3은 각각 ERP, CMP



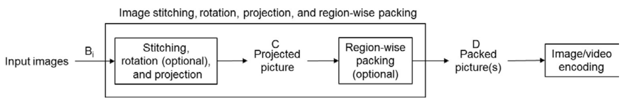
<그림 2. Equi-Rectangular Projection (ERP) [2]>



〈그림 3. Cube Map Projection (CMP) [2]〉

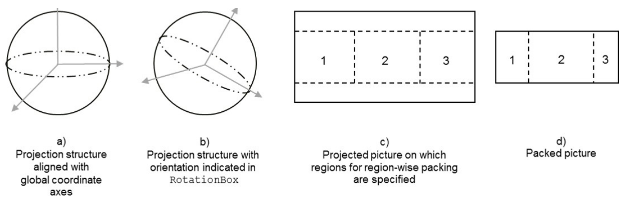
에서의 프로젝션 방법을 도시하여 보여주고 있다 [2].

그림 4.에서와 같이 프로젝션의 결과물인 프로젝티드 픽처 (projected picture)는 선택적으로 리전와이즈 패킹 (region-wise packing) 과정을 거칠 수 있다. 이 과정에서 프로젝티드 픽처를 구성하는 영역들의 위치, 크기 및 방향을 변형시킬 수 있으며, 전체적인 압축 효율을 높이거나, 특정 뷰포트 (viewport)에 해당하는 영역의 해상도를 다른 영역에 비해 높게 처리함으로서 사용자 시점 기반 프로세싱에 사용될 수 있도록 픽처의 구성을 가공하여 팩트픽처 (packed picture)를 생성한다.



〈그림 4. 리전와이즈 패킹 (region-wise packing) [1]〉

이외에도 OMAF는 그림 5.와 같이 스티칭, 프로젝션, 리전와이즈 패킹에 걸친 팩트픽처 생성 과정에 있어서, 3차원 공간의 회전 (rotation) 정도와 전방향 콘텐트의 시야범위 (coverage)를 시그널링 할 수 있는 메타데이터를 제공한다. 이러한 메타데이터를 이용하여, 중심점 (center point) 위치 이동을 통한 압축 효율 향상, 360도 이하의 비디오 콘텐트 지원 등 다양한 콘텐트 형태와 목적을 갖는 전방향 미디어 어플리케이션 서비스가 가능하도록 하고 있다.



〈그림 5. 팩트픽처 생성을 위한 프로세싱 단계 [1]〉

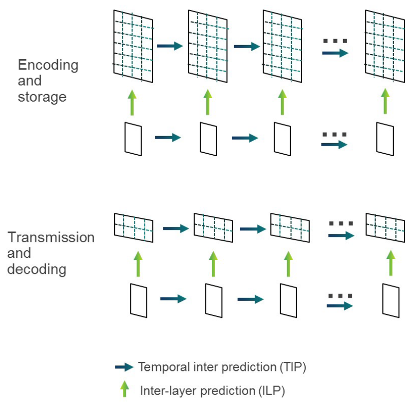
OMAF는 미디어 데이터의 프로세싱과 관련된 메타데이터 뿐만 아니라, 전방향 어플리케이션 서비스에서의 사용자 경험 향상을 위한 타임드 메타데이터 (timed-metadata)를 제공하고 있다. 초기 시점 (initial viewpoint), 추천 뷰포트 (recommend viewport)와 같이 전방향 미디어의 재생 시간 흐름에 따라 변화하는 시점, 뷰포트 관련 데이터들을 제공하고 있다.

전방향 미디어는 HEVC, MPEG-H Audio 등의 규격으로 부호화 되며, 앞서 설명한 메타데이터들과 함께 ISO base media file format [3] 기반의 파일 포맷에 저장된다. 저장된 파일은 파일 기반 서비스, II.2절에서 설명될 DASH, MMT 기반의 전방향 미디어 전송 및 스트리밍을 통해 사용자에게 전달된다.

사용자 디바이스에서는 전달된 파일을 파싱 (parsing)하여 미디어 데이터와 메타데이터를 추출하고, 미디어 데이터를 복호화한다. HMD 등의 출력 장치를 통해 복호화된 미디어 데이터와 메타데이터를 통해 전방향 미디어를 3D 공간에 복원하고, head/eye tracking을 통해 사용자 시점에 해당되는 전방향 미디어의 부분을 몰입감있는 형태로 사용자에게 제공하게 된다.

전방향 비디오의 압축과 전송에 있어서 많은 관심을 받고 또 하나의 기술요소는 사용자 시점 기반 프로세싱 (viewport dependent processing)이다. 전방향 비디오의 특성상 사용자는 항상 전체 비디오의 일부에 해당하는 뷰포트만을 시청하고 있으므로, 뷰포트에 해당하는 일부분에 해상도, 대역폭 등의 자원을 집중하여 사용한다면, 프로세싱 효율을 향상 시킬 수 있을 것이다.

그림 6.은 MCTS (Motion-Constrained Tile Set)와 SHVC (Scalable extension of HEVC)을 이용한 사용자 시점 기반 프로세싱의 예를 보여준다. 저화질의 base layer는 전체 비디오를 부호화하고, 고화질의 enhancement layer는 전체 비디오를 MCTS로 분할하여 부호화 한다. Enhancement layer는 사용자의 뷰포트에 따라 해당 MCTS만 전송, 복호화된다. 사용자 시점에 상관없이 항상 제공되는 저화질의 base layer는 사용자 시점 이동과 실제 뷰포트 이미지가 사용자에게 보여지기까지의 지연시간 (Motion-to-photon latency)에 사용자에게 제공될 수 있어, 사용자 경험 악화를 최소화 할 수 있다.



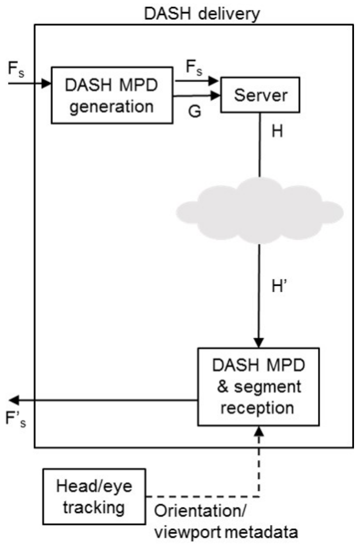
〈그림 6. MCTS와 SHVC를 이용한 사용자 시점 기반 프로세싱의 예 [1]〉

2.2 몰입형 미디어 전송 표준 기술

MPEG에서는 OMAF 표준과 연계하여, 몰입형 미디어의 전송 및 스트리밍을 지원하기 위한 기존 전송기술과 OMAF의 접목 방안이 논의되고 있다.

DASH AhG (Add-hoc Group) 에서는 IP/TCP 환경에서 사용자 시점기반 화질/대역폭 적응적 전방향 미디어 서비스를 제공하기 위한 DASH-VR 표준화를 수행 중이다. 현재 DASH-VR은 3 DoF (Degrees of Freedom) 를 지원하는 전방향 미디어의 전송 표준화에 집중하고 있으며, 관련 요구사항을 토대로 고화질 VR전송에 대한 Core Experiment (CE)가 진행 중이다.

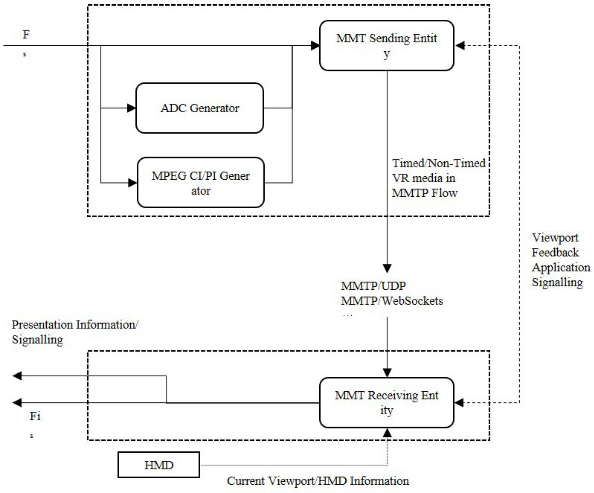
그림 6.에 도시된 바와 같이, DASH를 통한 OMAF 전송을 위해 전방향 미디어 전송 관련 메타데이터 지원을 위한 DASH MPD (Media Presentation Description) 확장 방안과, head/eye tracking을 통해 사용자 시점에 해당하는 DASH segment의 선택적 송수신 방안이 논의되고 있다.



<그림 7. DASH를 통한 OMAF 콘텐트 전송>

MMT AhG에서는 IP/UDP 환경에서 전방향 미디어를 사용자 시점, 제작자 시점기반으로 효율적으로 전송하기 위한 MMT-VR 표준화를 수행 중이다. 모바일 환경에서의 FoV (Field of View) 스트리밍 및 웹 환경에서의 타일드 미디어 스트리밍을 MMT 기반으로 지원하기 위한 세션 컨트롤 등 MMT 개선사항에 대한 CE가 진행 중에 있다.

그림 7.에 도시된 바와 같이, MMT를 통한 OMAF 전송을 위해 전방향 미디어 전송 관련 메타데이터 생성과, head/eye tracking을 통해 사용자 시점에 해당하는 MMTP Flow의 효율적 송수신 방안이 논의되고 있다.



<그림 8. MMT를 통한 OMAF 콘텐트 전송>

III. 기술 표준화 현황

OMAF 표준화는 Oculus Rift, 삼성 Gear VR 등의 HMD가 대중화 되기 시작했던 2015년 10월, 스위스 제네바에서 열린 115차 MPEG 회의에서 시작되었다. 당시 이미 시장에 전방향 미디어 서비스 및 디바이스가 보급되어 있었기 때문에 이들을 위한 공통의 어플리케이션 포맷 제정을 위한 MPEG-A MAF (Media Application Format)의 부속 표준으로 OMAF를 표준화하는 것으로 참여사들의 의견이 모아졌다.

이후 3 DoF 기반의 전방향 미디어 저장 및 전송 규격에 대한 요구사항과 기술규격 설계를 위한 논의들이 본격적으로 이어졌고, 2016년 10월 중국 청두에서 개최된 116차 MPEG 회의에서는 3DoF 기반의 저장, 전송 규격 표준뿐만이 아니라, 6 DoF를 지원하는 몰입형 미디어를 위한 Systems, Visual, Audio, Graphics를 포괄하는 MPEG-I 표준 제정의 장기적 계획이 수립되었다. MPEG-I는 현재 OMAF의 3 DoF를 시작으로, 3 DoF+, Omnidirectional 6 DoF, Windowed 6 DoF, 6 DoF와 같이 사용자가 경험할 수 있는 시각적 자유도에 대한 개념을 세분화하여 2021년 까지의 표준화 작업을 계획하고 있다 [4].

OMAF는 이후 MPEG-I의 부속표준으로 이동하게 되었고, 2017년 말까지 3 DoF 표준화 완료라는 MPEG-I의 계획에 따라 표준화 작업의 속도를 높여, 2017년 1월에 Committee Draft (CD), 4월에 DIS (Draft International Standard), 7월에는 Study of DIS를 발간하였으며, 10월 마카오에서 개최될 120차 MPEG 회의에서 FDIS (Final Draft International Standard) 발간을 앞두고 있다.

최근의 MPEG 회의에서는 3 DoF 기반의 현재의 OMAF를 OMAF version 1으로 규정하였고, 2017년 이후에도

MPEG-I의 계획에 맞춰 3 DoF+ 등을 지원하는 OMAF version 2 표준화 작업이 계속될 것이다. OMAF version 2 에서는 사용자 시점 자유도 확장과 사용자 시점 기반 프로세싱 등 OMAF version 1의 성능 개선 및 기능 확장 등의 주제가 논의될 것으로 전망된다.

IV. 결론

본고에서는 최근 MPEG에서 활발히 작업중인 OMAF와 몰입형 미디어 전송 기술 표준의 기술 전반과 표준화 동향에 대해 소개하였다. 2015년 시작된 OMAF 표준화 작업은 3 DoF 지원을 위한 version 1 표준화 작업이 마무리 단계에 이르고 있다. OMAF는 시장의 요구에 따라 현존하는 몰입형 미디어 서비스 기술을 표준에 담아내고, 다양한 어플리케이션 서비스 시나리오 지원, 전방향 미디어 프로세싱 효율성 증대 등 향후 몰입형 미디어 서비스 기술의 방향에 대한 전망도 제시하고 있다. 소개한 표준 기술들은 MPEG-I의 표준화 계획에 발맞춰 6 DoF 기반 몰입형 미디어 서비스 기술 제공에 이르기까지 지속적으로 표준화가 진행될 것으로 전망되며, 앞으로도 관련 시장의 요구와 기술 동향을 충실히 반영하는 표준 기술이 될 것을 기대해 본다.

References

[1] ISO/IEC JTC1/SC29/WG11/N16950, “Study of ISO/IEC DIS 23000-20 Omnidirectional Media Format”, MPEG 119, July 2017.

[2] Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, “360Lib Software Manual”, MPEG 117, Jan. 2017

[3] ISO/IEC JTC1/SC29/WG11/N15177, “Text of ISO/IEC 14496-12 5th edition”, MPEG 111, Feb. 2015.

[4] ISO/IEC JTC1/SC29/WG11/N16541, “New Work Item Proposal on Coded Representation of Immersive Media”, MPEG 117, Jan. 2017

Biographies

이장원

2007년: 경희대학교 전자공학과 학사

2013년: 경희대학교 전자전파공학과 박사

2013년 ~ 현재 : LG전자 CTO부문 선임연구원

주요 관심분야: image processing, media streaming, broadcasting

e-mail : jangw.lee@lge.com

최신 MPEG 오디오 표준 기술: MPEG-H 3D Audio

이태진, 백승권, 이미숙
Electronics and Telecommunications Research Institute (ETRI)

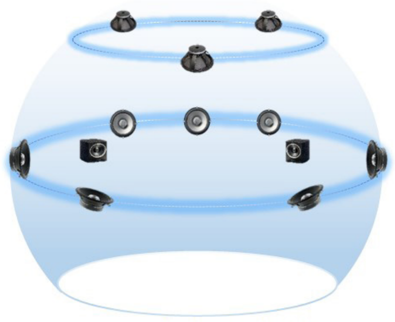
Abstract

디지털 오디오 신호의 음질 손실을 최소화 하면서 데이터 량을 줄이기 위한 오디오 압축 기술은 1980년대 후반부터 MPEG을 중심으로 기술개발과 표준화를 진행하고 있다. MPEG-H 3D Audio는 MPEG의 최신 오디오 압축 표준으로 다채널, 다객체, 장면 오디오 신호에 대한 부호화 기술과 함께 헤드폰이나 다양한 스피커 배치에 따른 최적의 오디오 재생을 위한 렌더링 기술 까지를 포함하고 있다. 본 논문에서는 MPEG 오디오 압축 표준의 기본 원리 및 역사와 함께 MPEG-H 3D Audio 표준에 대해 기술한다.

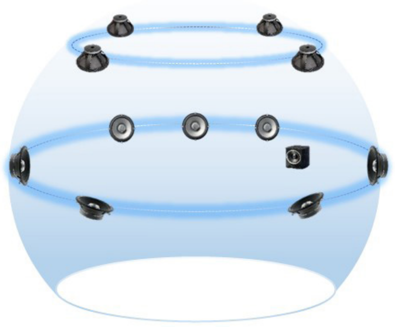
I. 서론

가정과 극장에서 4K와 같은 고화질 영상 서비스의 확대에 따라 기존 5.1채널 서비스를 넘어서는 새로운 오디오 서비스에 대한 시장과 사용자의 요구가 증대되고 있다. UHDTV와 같은 고화질 영상을 위한 UHD 오디오 서비스는 몰입감(Immersive)과 개인화(Personalization)라는 키워드로 정의할 수 있다. 몰입감은 그림 1과 같이 수직면까지 음상을 제공하는 다채널(10.2채널/11.1채널) 오디오 기술, 헤드폰 환경에서 3D 입체음향을 제공하는 바이노럴 오디오 기술 등을 활용하여 기존 스테레오(2.0)나 5.1채널 서비스를 넘어서는 현장감과 사실감을 제공해 주는 것을 의미한다[1].

개인화는 개인의 취향에 따른 오디오 제어 및 선택이 가능한 서비스로 그림 2와 같이 스포츠 중계에서 아나운서/해설자의 목소리만을 크게 하여 좀 더 명료한 방송을 청취하거나, 관중과 현장의 소리를 크게 하여 좀 더 현장감 있는 방송을 청취

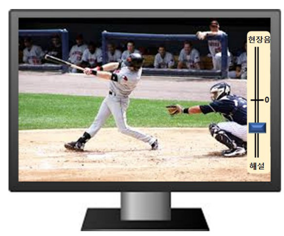


a) 10.2 채널 스피커 구성



b) 11.1 채널 스피커 구성

〈그림 1. 다채널 오디오 스피커 구성〉



a) Dialogue Enhancement

b) 편파 방송

c) 오디오 객체 위치 제어

〈그림 2. MPEG-H 3D Audio 서비스〉

할 수 있다. 또한 다양한 해설/아나운서 음성을 제공하여 자신이 응원하는 팀의 해설/아나운서 음성과 관중의 응원을 선택하여 청취하거나 음악 서비스에서 오디오 객체(가수, 악기 등)의 위치나 레벨을 자신의 취향에 따라 변경할 수 있다[2][3].

5.1채널을 넘어서는 몰입감과 객체기반의 개인화된 오디오 서비스를 제공하는 UHD 오디오 서비스를 위한 표준 기술은 MPEG의 MPEG-H 3D Audio와 Dolby의 AC-4등이 있다. 두 기술 모두 몰입감과 개인화라는 UHD 오디오 서비스를 제공할 수 있지만, 본 논문에서는 국내 지상파 UHDTV를 위한 오디오 표준인 MPEG-H 3D Audio에 대해 소개하고자 한다.

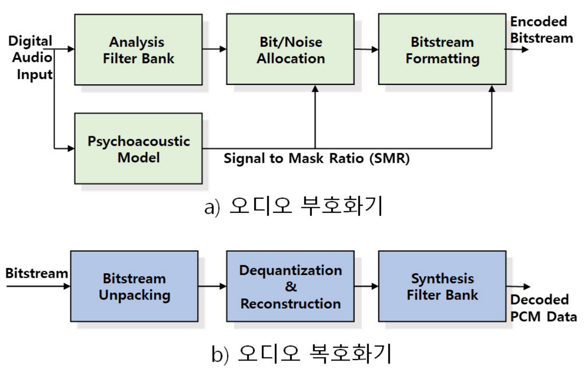
II. MPEG 오디오 표준

오디오 신호의 데이터 량은 초당 획득한 샘플의 수인 샘플링 율(Sampling rate)과 채널 수(No. of channel), 획득한 샘플을 디지털로 표현하기 위한 비트 해상도(Bit resolution)를 이용하여 계산할 수 있다. 예를 들어 스테레오 신호인 경우 44,100(샘플링 율) × 2(채널 수) × 16(비트 해상도) = 1,411,200 bits/sec의 데이터 량을 가지며, UHDTV 를 위한 다채널 오디오인 경우 48,000(샘플링 율) × 16(채널 수) × 24(비트 해상도) = 18,432,000 bits/sec의 데이터 량을 가진다. 데이터 량을 줄이기 위해서 샘플링 율이나 비트 해상도를 낮추거나 채널 수를 줄이는 방법 있을 수 있다. 그러나 채널 수를 줄이면 다채널 오디오에 의한 스테레오 이미지가 약해져서 몰입감이 저하되고, 샘플링 율을 줄이면 고주파 신호가 없는 대역 제한된 신호가 생성되며, 비트 해상도를 줄이면 양자화 잡음이 발생한다. 따라서 최대한 원음의 음질을 유지하면서 데이터 량을 줄이기 위한 오디오 압축 기술이 필요하다.

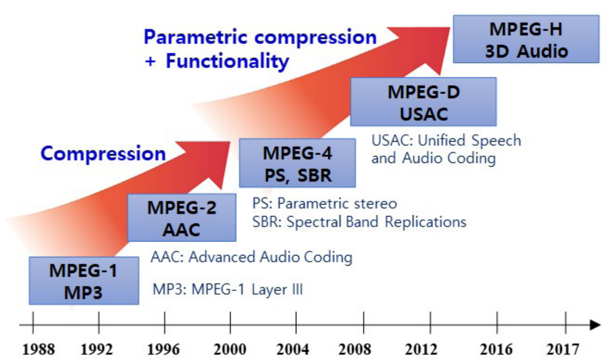
데이터 량을 최소화 하면서 음질을 최대화 하기 위한 오디오 압축 기술의 기본 구조는 그림 3과 같은 형태를 가지고 있다. 오디오 부호화기의 경우 입력 오디오 신호를 주파수 영역으로 변환하고, 심리 음향 모델을 이용하여 인간이 들을 수 있는 신호와 들을 수 없는 신호로 구별하고, 들을 수 있는 신호에 비트를 할당하여 압축된 비트스트림을 생성한다. 오디오 복호화기는 부호화기의 역 과정을 거쳐서 압축된 비트스트림 입력신호를 이용하여 복원된 오디오 신호를 생성한다. 오디오 압축에 대한 본격적인 연구를 시작한 1980년대 후반부터 다양한 오디오 압축 기술이 발전해 오고 있지만, 그림 3과 같은 기본적인 구조에 다채널, 다객체 지원이나 저 비트율 지원을 위한 툴을 추가하는 구조를 가지고 있다.

MPEG에서는 1992년 MPEG-1 Audio 표준(ISO/IEC 11172-3)[4]을 시작으로 2017년 하반기 표준화가 예정되어 있는 MPEG-H 3D Audio(ISO/IEC 23008-3)[5]까지 다양한 오디오 압축 기술에 대한 표준화를 진행해 왔다.

그림 4는 MPEG의 오디오 압축 기술에 대한 역사를 간략하게 정리한 것이다. 1.41Mbps의 정보량을 가지는 CD의 스테레오 오디오 신호를 192kbps로 압축하기 위한 MPEG-1 오디오 표준은 우리에게 MP3(MPEG-1 Layer 3)라는 표준으로 잘 알려져 있으며, DTV 방송과 모바일 오디오 재생장치 등 MPEG 오디오 표준 중 가장 성공적으로 시장에서 서비스 되고 있다. AAC(Advanced Audio Coding)라는 표준 명으로 잘 알려져 있는 MPEG-2 오디오 표준(ISO/IEC 13818-7)은 128kbps로 스테레오 오디오 신호를 압축하기 위한 기술로 DAB, DRM 등에서 오디오 표준으로 서비스 되고 있다[6]. 고주파 대역이나, 스테레오에 대한 정보를 파라미터로 표현하여 64kbps까지 스테레오 오디오 신호를 압축하기 위한 MPEG-4 오디오 표준(ISO/IEC 14496-3)은 HE-AAC(High-Efficiency AAC) v1/v2로 알려져 있으며 DVB, DAB 등에서 사용되고 있다[7]. 5.1채널과 같은 다채널 오디오 서비스가 확장되어 가면서, 다채널 오디오에 대한 효율적인 압축 기술에 대한 시장의 요구에 따라 MPEG-D MPEG-Surround(ISO/IEC 23003-1) 표준이 개발되었다[8]. 또한 다객체 오디오 신호에 대한 압축 기술인 MPEG-D SAOC(ISO/IEC 23003-2)가 2010년에 표준화 되었다[9].



〈그림 3. 오디오 부/복호화기 기본 구조〉



〈그림 4. MPEG 오디오 압축 기술〉

음성과 오디오 신호 모두에 대해 우수한 음질을 제공하는 오디오 압축 기술의 개발을 위해 MPEG에서는 2007년 CfP를 시작으로 표준화를 진행하여 2012년 MPEG-D USAC(Unified Speech and Audio Coding) 표준을 승인하였다(ISO/IEC 23003-3)[10].

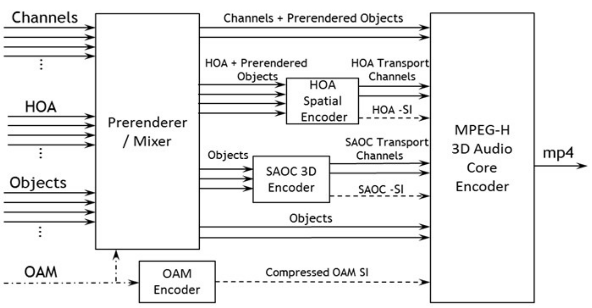
III. MPEG-H 3D Audio

MPEG의 최신 오디오 압축 표준인 MPEG-H 3D Audio는 고차 다채널 오디오 신호, 객체기반 오디오 신호와 함께 장면기반 오디오 신호를 지원한다.

- 고차 다채널 오디오(High-order Multichannel Audio): 채널 오디오 신호는 각각의 스피커로 전달되는 음향 신호를 의미하며, MPEG-H 3D Audio에서는 기존 5.1채널뿐 아니라 10.2/11.1채널과 같은 고차 다채널 오디오 신호를 지원한다.

- 객체기반 오디오(Object-based Audio): 채널 신호와 함께 전송되는 음향 신호로 수신단에서 특수 취급 또는 사용자 상호작용을 요구하는 음원을 의미한다. 예를 들어, 별개의 음성 대화나 악기음과 같은 대화형(interactive) 객체뿐만 아니라, 움직이는 비행기, 자동차 등 비대화형(non-interactive) 특수효과 객체 등을 포함한다.

- 장면기반 오디오 신호(Scene-based Audio): 음향 장면(Sound Scene)을 공간상에서 직교하는 데이터로 표현한 신호를 의미한다. HOA(Higher Order Ambisonics) 오디오 신호는 하나의 Eigen 마이크를 이용하여 음향 장면을 획득하고, 스테레오, 5.1/10.2/11.1채널 등 재생 환경에 따른 최적의 입체음향 신호를 재현할 수 있는 특징이 있다.



〈그림 5. MPEG-H 3D Audio 부호화기 구조〉

MPEG에서는 이와 같은 다양한 형태의 입력신호를 처리하기 위해 그림 5와 같이 기존의 MPEG 오디오 표준 기술인 USAC, MPS, SAOC 기술을 활용하고 장면기반 오디오 신호의 처리를 위한 HOA와 다양한 재현 환경에 최적화된 오디오 재생을 위한 렌더링(Rendering) 기술을 포함하여 MPEG-H 3D Audio 기술을 표준화하였다[11].

3.1 MPEG-H 3D Audio 압축 기술

MPEG-H 3D Audio 압축 기술의 주요 부호화 모듈(Core encoder)은 USAC 표준에 새로운 툴을 추가한 형태를 가지고 있다. 기존 USAC 과 같이 오디오 신호의 특성에 따라 음성 특성 신호인 경우 LPD(Linear Prediction Domain) 모드, 음악 특성 신호인 경우 FD(Frequency Domain) 모드로 구분하여 동작하면서 다채널 오디오 신호에 대한 고음질/저 비트율 부호화를 위한 MCT(Multi-Channel Coding) 툴, 저 비트율에서 고대역 신호를 생성하기 위한 IGF(Intelligent Gap Filling) 툴 등을 추가하여 다채널/다객체 오디오 신호에 대한 저 비트율/고음질 압축을 제공한다.

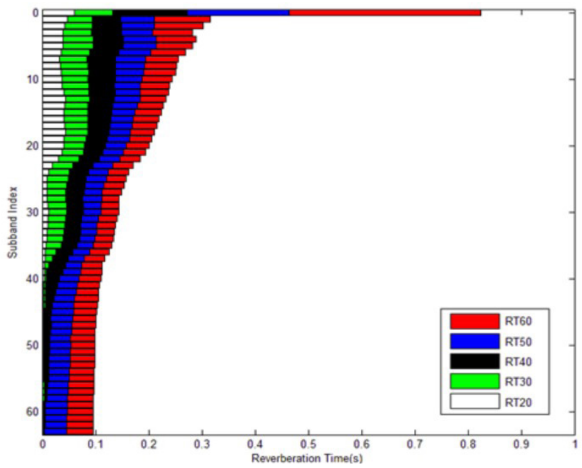
3.2 MPEG-H 3D Audio 렌더링 기술

MPEG-H 3D Audio를 활용하여 10.2 채널과 같은 다채널 오디오 콘텐츠를 제공하더라도 모든 시청자가 10.2 채널의 스피커를 정해진 표준 위치에 설치하는 것이 아니기 때문에 스테레오, 5.1채널 스피커나 다른 위치에 배치된 스피커 또는 헤드폰 사용자에게도 원 제작자의 의도에 충실한 오디오 장면을 제공할 수 있어야 한다. MPEG-H 3D Audio의 주요 특징 중 하나는 다양한 스피커 배치 환경과 헤드폰에서도 최적화된 오디오를 재현할 수 있는 렌더링 기술이다.

MPEG-H 3D Audio에서 제공하는 플렉서블 렌더러(Flexible Renderer)에는 채널, 객체, 그리고 장면 오디오 신호 각각을 렌더링하는 모듈이 포함되어 있다. 포맷 변환기(Format Converter)는 입력 채널과 출력 채널 구성 사이의 변환 장치로 원 콘텐츠의 효과를 최대한 반영하기 위한 능동적 믹싱(Dynamic mixing)을 지원한다. 객체 오디오 신호는 객체 렌더러를 통해 특정 스피커 재생 환경에 맞게 렌더링되고, 장면 오디오 신호는 복호화된 채널 오디오 신호와 HOA 메타데이타를 사용하여 HOA 렌더러를 통해서 특정 스피커 재생 환경에 맞게 생성된다[11].

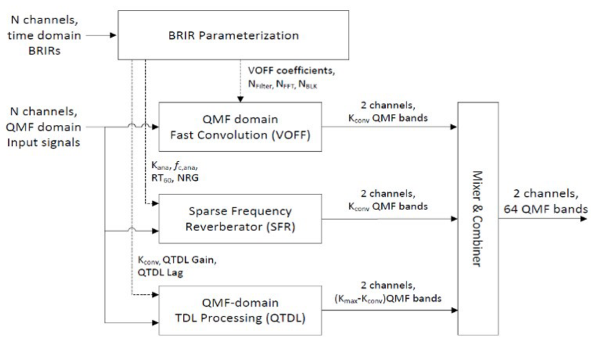
사용자가 헤드폰 환경을 선호하게 된다면 MPEG-H 3D Audio는 원 콘텐츠의 효과를 최대한 반영할 수 있도록 헤드폰을 위한 바이노럴 기술 기반의 스테레오 오디오 신호를 생성할 수 있다. 고차 다채널 오디오 신호를 헤드폰을 통해 재생하기 위한 바이노럴 재현(Binaural Rendering)기술은 복호화

기의 최종 출력 신호인 다채널 오디오 신호를 BRIR(Binaural Room Impulse Response) 필터 와 Convolution 연산을 통해 이루어진다. 음질과 3차원 음장감에 영향을 미치는 BRIR의 주파수별 잔향시간(Reverberation Time) 특성은 저주파 대역이 고주파 대역에 비해 상대적으로 길며, 주파수 대역별로 길이가 다른 특징을 가지고 있다(그림 6). 따라서 주파수 대역별로 다른 길이의 BRIR 필터를 사용하게 되면, 음질을 최적으로 유지하면서도 BRIR 필터에 필요한 연산량을 최소가 되도록 설계할 수 있다.



〈그림 6. BRIR의 주파수대역별 잔향시간〉

MPEG-H 3D Audio의 바이노럴 오디오 기술은 그림 7과 같이 부호화기의 출력 신호인 N채널 오디오 신호를 주파수 대역별로 가변적인 길이를 가지는 QMF(Quadrature Mirror Filter) 대역의 가변 차수 필터링(Variable Order Filtering in Frequency-domain) 방법으로 BRIR 필터를 적용한다. 특히 상대적으로 청감상 둔감한 고주파 대역에 대해서는 하나의 이득 값과 지연 값으로만 표현되는 TDL(Tapped Delay Line)을 QMF 밴드에서 적용(QMF-domain TDL)하여 지각되는 음질의 손상을 최소화하면서 연산량을 줄일 수 있다.

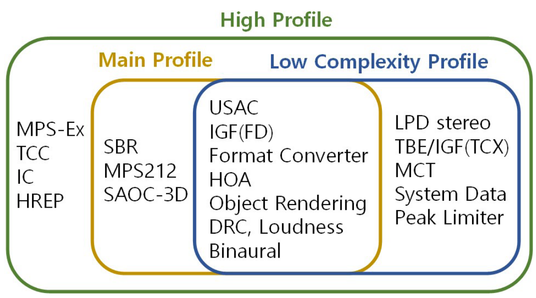


〈그림 7. MPEG-H 3D Audio 바이노럴 오디오 기술 블록도〉

마지막으로, MPEG-H 3D Audio에서 지원하는 기능으로 DRC(Dynamic Range Control) 모듈이 있다. MPEG-H 3D Audio는 기본적으로 다양한 포맷의 오디오 신호를 처리한다. 따라서 상호간에 발생하는 오디오 신호의 레벨 차가 존재할 수 있다. DRC의 첫 번째 요구 기능은 다양한 출력 오디오 포맷 사이의 레벨 차를 맞추는 것이다. 따라서 프로그램별로 다양한 형태의 오디오 포맷이 존재하더라도 최종 출력 레벨의 차이를 느낄 수 없도록 DRC가 오디오 포맷 별로 출력 레벨을 조절할 수 있어야 한다. 기본적으로 DRC 기능은 출력 신호의 레벨을 자동적으로 일정 레벨로 맞추어 줄 수 있어야 한다. 예를 들어, 방송 콘텐츠 별로 상이한 오디오 레벨로 제작되었다 하더라도, 사용자 단말에서는 사용자가 개별적으로 볼륨을 조절하지 않아도 동일한 라우드니스(loudness)를 가질 수 있어야 한다. 이는 최근 많은 이슈가 되어왔던 사항으로, 사용자가 채널 별로 혹은 콘텐츠 별로 인지하는 라우드니스 레벨 차이가 상이하여 시청 시, 오디오 볼륨을 상시 조절해야 하는 번거로움을 줄이기 위한 기술이다. MPEG-H 3D Audio는 코덱단에서 DRC를 통해 라우드니스를 제어하여 사용자의 볼륨 조절의 번거로움을 최소화 시킬 수 있다[11].

3.3 MPEG-H 3D Audio 프로파일

MPEG 표준은 RM(Reference Model) 선정 이후에도 성능향상을 위해 다양한 툴들을 추가하기 때문에 활용 분야에 따라 채널 수, 비트율, 사용 툴 등에 대해 프로파일로 정의한다. 각 프로파일은 서비스 목적에 따라 제공하는 기능 및 처리 가능한 채널 수를 명시하고 있다. MPEG-H 3D Audio는 그림 8과 같이 세 개의 프로파일을 지원하며 각각의 특징은 아래와 같다[2].



〈그림 8. MPEG-H 3D Audio 프로파일〉

- Main 프로파일: Phase 1에서 개발된 기술들에 대한 프로파일로, 복잡도에 대한 제약사항 없이 높은 몰입감과 여러 기능을 지원하기 위해 필요한 기술들을 포함하고 있으며, 지원 가능한 채널 수 및 객체 수도 제약을 두지 않고 있다.

- LC(Low-Complexity) 프로파일: MPEG-H 3D Audio가 높은 몰입감을 제공하고 다양한 기능을 지원함으로써 새로운 오디오 코덱 형상을 제시하였으나, 높은 복잡도로 인해 Main 프로파일을 실제 서비스 단말에서 지원하기는 다소 부담이 될 수 있다. 이를 해결하기 위해서 Phase 2 기술 개발 기간 동안에 복잡도가 낮은 기술, 또는 기존 기술의 복잡도를 낮춘 기술들을 표준에 반영하였다. LC 프로파일은 이러한 기술들을 조합하여 만든 프로파일이지만 몰입감의 수준과 제공하는 기능은 Main 프로파일과 크게 다르지 않다.

- High 프로파일: LC 프로파일이 저-복잡도를 위한 프로파일 이기는 하지만 Main 프로파일에는 기술되지 않은 새로운 기술들을 포함하게 되었다. 이에 따라 모든 기술들이 포함된 High 프로파일을 새롭게 정의하여 MPEG-H 3D Audio Phase 1과 Phase 2의 모든 기술들을 반영하였다.

일본, 미국, 유럽 등에서 UHDTV에 대한 기술개발과 표준화가 활발하게 진행 중에 있다. 국내에서는 2015년 지상파 UHDTV 방송을 위한 700MHz 주파수 대역을 각 방송사가 할당 받음으로써, 본격적인 UHDTV 방송을 위한 작업에 착수하였다. 국내 지상파 UHDTV를 위한 오디오의 경우 Dolby의 AC-4, MPEG-H 3D Audio, DTS 기술이 경합하여, UHDTV 오디오 서비스를 실현하기 위한 렌더링 기능을 포함하고 있는 MPEG-H 3D Audio의 LC 프로파일을 최종 오디오 표준으로 선정하였다.

IV. 결론

본 논문에서는 MPEG을 중심으로 한 오디오 압축 기술의 역사와 함께 5.1채널을 넘어서는 몰입감과 개인형 오디오 서비스를 제공하기 위한 MPEG-H 3D Audio의 기술의 표준화 현황 및 기술의 개요에 대해 살펴보았다.

국내에서는 MPEG-H 3D Audio LC 프로파일을 기반으로 최대 16트랙을 지원하는 UHDTV 오디오 표준을 제정하였고, 국내 지상파 UHDTV 서비스가 시작된 2017년 5월 말부터 MPEG-H 3D Audio 복호화기를 탑재한 TV가 판매 중이다.

2018년부터 객체기반의 개인화 방송을 먼저 시작하고, 가정 내에 다채널 오디오 스피커 환경이나 사운드 바 등이 보급될 예정인 2019년 이후부터 다채널 오디오 서비스를 제공하는 단계적 접근 방법을 통해 머지 않아 UHD 오디오 서비스를 가정에서 제공 받을 수 있기를 기대해 본다.

Acknowledgement

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

References

[1] 이태진 외, “UHDTV를 위한 10.2채널 기반 다채널 오디오 재현 기술”, 방송공학회 논문지 제 17권 제 5호, 2012년

[2] 이미숙, 백승권, 이태진, “UHDTV 방송 서비스를 위한 MPEG-H 3D Audio 기술”, TTA Journal 167호, 2016년

[3] 백승권, “MPEG-H 3D Audio 배경 및 진행과정”, 방송과 기술, 5월호, 2016년

[4] ISO/IEC 11172-3:1993, Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio

[5] ISO/IEC 23008-3:2015, MPEG-H (High efficiency coding and media deliver in heterogeneous environments), Part 3: 3D Audio

[6] ISO/IEC 13818-7:2006, Information Technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding(AAC)

[7] ISO/IEC 14496-3:2009, Information Technology - Coding of audio-visual objects - Part 3: Audio

[8] ISO/IEC 23003-1:2007, Information Technology - MPEG Audio Technologies - Part 1: MPEG Surround

[9] ISO/IEC 23003-2:2010, Information Technology - MPEG Audio Technologies - Part 2: Spatial Audio Object Coding(SAOC)

[10] ISO/IEC 23003-3:2012, Information Technology - MPEG Audio Technologies - Part 3: Unified Speech and Audio Coding

[11] 이태진, 백승권, “UHDTV를 위한 차세대 오디오 표준: MPEG-H 3D Audio”, 한국방송공학회지 21권, 3호, 2016년

Biographies

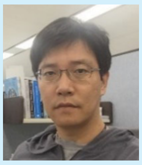
이태진

2000년 ~ 현재: 한국전자통신연구원 미디어연구본부 전문위원

2015년 ~ 현재: 과학기술연합대학원대학교(UST) 부교수

주요 관심분야: 오디오 신호처리, 부호화, 음향 신호 인식

e-mail : tjlee@etri.re.kr



백승권

2005년 ~ 현재: 한국전자통신연구원 미디어연구본부 책임연구원

주요 관심분야: 오디오 신호처리, 부호화

e-mail : skbeack@etri.re.kr

이미숙

2004년 ~ 현재: 한국전자통신연구원 미디어연구본부 책임연구원

주요 관심분야: 음성·오디오 신호처리, 부호화

e-mail : lms@etri.re.kr