이머시브미디어를 3DoF+ 비디오 부호화 표준 동향

Standardization Trend of 3DoF+ Video for Immersive Media

이광순 (G.S. Lee, gslee@etri.re.kr) 실감미디어연구실 책임연구원
정준영 (J.Y. Jeong, jyj0120@etri.re.kr) 실감미디어연구실 연구원
신홍창 (H.C. Shin, hcshin@etri.re.kr) 실감미디어연구실 선임연구원
서정일 (J.I. Seo, seoji@etri.re.kr) 실감미디어연구실 책임연구원/실장

ABSTRACT

As a primitive immersive video technology, a three degrees of freedom (3DoF) 360° video can currently render viewport images that are dependent on the rotational movements of the viewer. However, rendering a flat 360° video, that is supporting head rotations only, may generate visual discomfort especially when objects close to the viewer are rendered. 3DoF+ enables head movements for a seated person adding horizontal, vertical, and depth translations. The 3DoF+ 360° video is positioned between 3DoF and six degrees of freedom, which can realize the motion parallax with relatively simple virtual reality software in head-mounted displays. This article introduces the standardization trends for the 3DoF+ video in the MPEG-I visual group.

KEYWORDS 이머시브비디오, 6DoF, 3DoF+, 운동시차, 비디오 압축

Ⅰ. 서론

몰입(Immersion)이란 가상의 세계에 본인이 빠져 현실과 실제가 불분명하게 되는 현상으로 정의할 수 있다. 이머시브미디어(Immersive Media)는 사용자에게 몰입감을 경험하게 하는 오디오 및 비디오 등을 포함한다. PC로부터 스마트폰 등으로 미디어를 소비하는 환경이 변화하는 것을 고려할 때 이머시브미디어는 HMD 등의 다양한 VR기기로부터 멀티 TV로 구성되는 다양한 형태의 대형 디스플레이 환경에서 몰입감을 증대시키는 미디어라고 볼 수 있다. 특히 이머시브비디오 관점에서 보면 사용자의 자유로운 움직임에 대해 완전한 6DoF의 자유도를 제공하는 것이 완전한 몰입감을 제공할 수 있는 기본적인 요건이라고 여겨져 관련 기술이 개발되고 있다.

* DOI: https://doi.org/10.22648/ETRI.2019.J.340614
* 본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 정보 통신기술진흥센터의 지원을 받아 수행된 연구임[No. 2018-0-00207, 이머시브전문연구실].

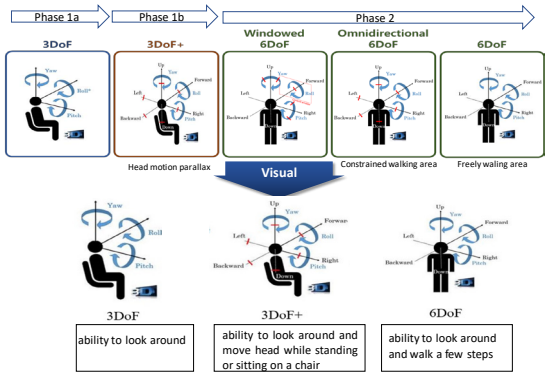


그림 1 MPEG-I에서 이머시브비디오 표준화

MPEG에서의 이머시브비디오 표준화 논의는 2016년 10월 청두 회의 때 시작된 후, visual, system 그룹 등에서 MPEG-I 프로젝트명으로 현재 활발히 진행 중이다[1,2]. MPEG-I 프로젝트의 궁극적인 목표는 사용자에게 최대 6자유도(DoF)를 제공하는 몰입형 미디어 서비스에 필요한 비디오, 오디오 그리고 시스템 기술을 표준화하는 것이다. MPEG-I 프로젝트는 그림 1과 같이 고정된 중심을 기준으로 3축 회전 운동(Rotational movement)을 지원하는 Phase 1a(3DoF) 단계, 전방위 장면을 포함한 복수 개의 영상과 이를 기반으로 합성(synthesize)한 다수의 가상 영상을 기반으로 약간의 머리 움직임과 같이 매우 제한된 병진 운동(Translational movement)을 지원하는 Phase 1b(3DoF+) 단계, 그리고 실제 생활하는 것과 같이 미디어 환경 내에서 자유롭게 이동이 가능한 Phase 2(6DoF) 단계의 총 세 단계로 구성된 타임라인을 기반으로 진행되고 있다.

2017년 3월 호바트 회의 때 Phase 2 단계는 병진 운동의 지원 범위에 따라서 Windowed 6DoF, Omnidirectional 6DoF, 그리고 6DoF 단계로 세분화되었으며, 각 단계에 대한 설명은 다음과 같다.

- Windowed 6DoF: 사용자가 HMD와 같은 특수 디스플레이 장치를 착용한 상태가 아닌 정면에 위치한 디스플레이를 기준으로 한정된 범위 내에서 병진 운동을 체험할 수 있는 단계를 의미함. 여러 대의 2D 카메라로 구성된 카메라 어레이를 통해 획득한 멀티뷰(multiview) 비디오가 Windowed 6DoF 콘텐츠의 예라 할 수 있음
- Omnidirectional 6DoF: Phase 1b 단계에서 병진 운동의 범위를 조금 더 확장하여 사용자가 미디어 환경 내에서 몇 걸음 걷는 것이 가능한 단계를 의미
- 6DoF: 미디어 환경 내에서 특별한 제약 없이 자유롭게 이동이 가능한 단계를 의미

하지만 MPEG-I visual 그룹에서는 2019년부터 3DoF+ 비디오의 표준화를 본격적으로 시작하고 6DoF의 EE를 진행하면서, 6DoF의 범위를 시청자가 몇 걸음 이동하는 것에 대한 자유도를 제공하는 것으로 범위를 한정하면서 Omnidirectional이라는 용어는 사용하지 않고 있다. 또한 3DoF+ 및 6DoF의 비디오 포맷 내에 perspective 프로젝션 포맷을 포함하면서 windowed 6DoF는 3DoF+ 및 6DoF에 포함되었다.

II. 3DoF+ 비디오 부호화 아키텍쳐

MV-HEVC, 3D-HEVC 등의 전통적인 다시점 영상 압축 기술은 많은 시점 영상 간의 중복성을 제거하여 압축 효율을 증가시키는 데 관점을 두었었다. 하지만 MPEG-I 비디오 그룹은 HEVC, VVC 등 2D 비디오 코덱 표준과의 호환성을 유지하면서 어떻게 360 비디오에 운동시차(Motion parallax)를 부여하느냐는 관점에서 출발하

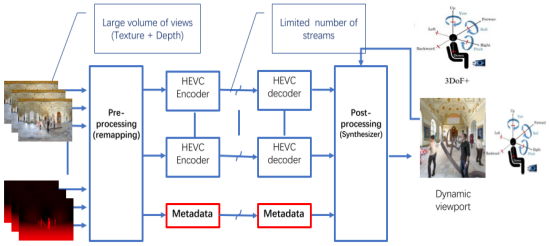


그림 2 3DoF+ 비디오 부호화 아키텍처

였다. 이외 2D 비디오 코덱과 차별화되는 요구사항으로 픽셀율(Pixel rate)의 최소화를 들 수 있다. 즉 기존의 MV-HEVC, 3D-HEVC는 비교적 작은 단방향의 시청 공간에 필요한 다시점 영상의 압축만은 고려하였지만, 이머시브비디오는 전방위로 확장된 시청 공간에 필요한 시점 영상을 고려하므로 시점 영상의 개수가 증가하고 시야의 증가에 따라 해상도가 증가한다. 이에 따라 수반되는 입출력 인터페이스, 압축 처리를 위한 데이터량 등이 반영된 픽셀율 증가 문제를 해결하고자 하는 것이다. 따라서 MPEG-I Visual에서는 깊이맵 추출, 시점 영상 합성 등의 3차원 영상처리를 통한 전후처리 과정을 통해, 상기 두 가지 요구사항을 만족할 수 있는 이머시브비디오 압축 기술을 표준화하고 있다. 즉, 그림 2에서와 같이 전방위 공간을 획득하여 추출되는 텍스처(texture) 및 깊이맵(depth map)을 이용하여 3차원 공간상에서 중복성을 제거할 수 있는 부호화 전처리 기술(pre-processing)을 표준화함으로써 픽셀율을 줄임과 동시에 적은 수의 2D 비디오 코덱을 통해 압축 처리하도록 하고자 한다. 디코더단에서는 다시 후처리(post-processing)와 시점합성을 통해 시청자의 움직임에 따라 동적으로 끊어짐 없이 뷰포터(Viewport) 영상을 재생하도록 하는 것이다. 이 모든 과정을 제어하기 위해서는 메타데이터가 필요한데, 이는 현재 "Coded Representation of Immersive Media-part 7: Metadata for Immersive Media(Video)"라는 이름으로 표준화 중에 있다.

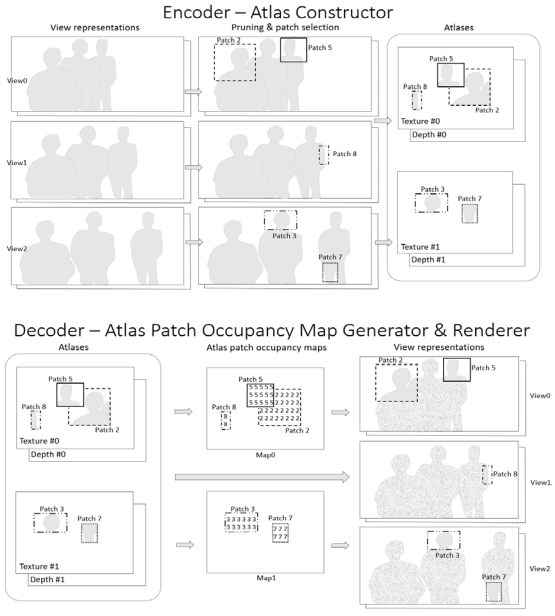
III. 3DoF+ Test Model

2019년 3월 제네바 회의 때 CfP 대응을 통해 제안된 5개의 기술은 기술 분류별로 분류된 후 객관적 주관적 평가 결과를 바탕으로 TMIV(Test Model for Immersive Video) 버전 1.0이라는 명칭으로 테스트 모델로 통합되었으며, 각 참여기관들이 지원하여 다섯 개의 CE 아이템을 도출하였다. 2019년 5월에는 TMIV 1.0을 바탕으로 참조소프트웨어를 구현하였으며, MIV라는 이름으로 WD 문서의 초안을 발간하였다[3].

1. 3DoF+ TMIV 개념

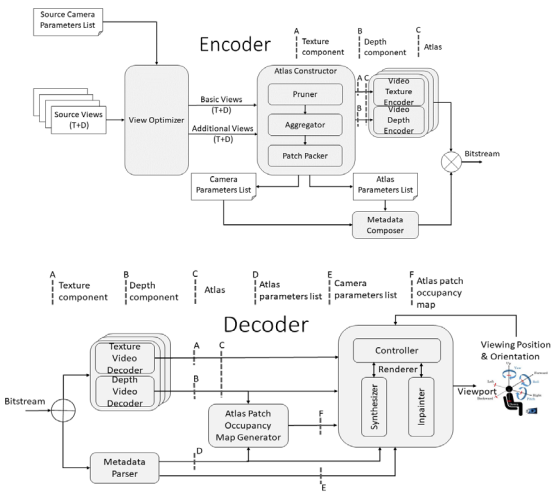
3DoF+ TMIV의 개념은 그림 3에서와 같이 다중의 원본 이머시브비디오가 입력되면, TMIV 인코더는 기본시점(Base view) 비디오와 부가시점(Additional view) 비디오로부터 프루닝(Pruning) 과정을 통해 패치(Patch)를 추출하고, 패킹(Packing) 과정을 통해 텍스처와 깊이맵으로 구성된 아틀라스(Atlas) 쌍을 생성한다. 이 아틀라스 쌍이 다중계층의 HEVC Main profile 10으로 인코딩 및 디코딩되므로 호환성이 보장되는 것이다[4,5]. 단말측에서는 메타데이터를 파싱하여 Occupancy Map을 생성하는데, 이는 픽셀별로의 패치정보를 기록한 것으로서 패치가 서로 겹쳤을 때 픽셀별로 분리하는 데 이용된다. TMIV 2.0에서는 패치의 경계를 정확히 표시하기 위해 깊이맵 내에 Occupancy Map 데이터를 전송하도록 개정되었다. 이 Occupancy Map 정보와 카메라파라미터를 이용하여 원래의 시점 영상을 복원하고 중간시점 영상을 합성하게 된다.

3DoF+ TM 블록도는 그림 4와 같이 크게 부호화 단계와 복호화 단계로 나눠져 있으며, 주요한 기



출처 B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019. 복제 금지.

그림 3 3DoF+ TMIV 인코더/디코더 개념



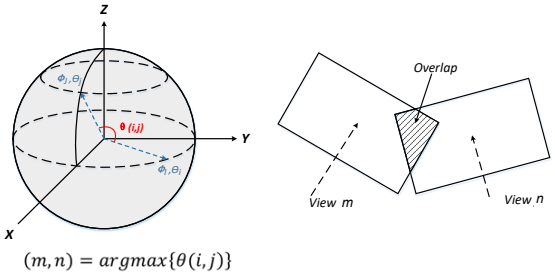
출처 B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019. 복제 금지.

그림 4 3DoF+ TMIV 인코더/디코더 블록도

능적 모듈로는 "View Optimizer", "Pruner", "Aggregator", "Patch Packer", "Metadata Composer", "Atlas Patch Occupancy Map Generator", "Synthesizer" 등으로 구성되어 있다. TMIV 디코더의 최종 출력은 시청 공간 내에서 운동시차를 지원하도록 시청자의 위치 및 각도에 부합되게 프로젝션된 perspective 뷰포트 혹은 전방위 영상이다. 전방위 영상은 HMD에 입력되어 360 비디오를 재생하며, perspective 뷰포트영상은 일반적인 2D 모니터에도 직접 재생될 수 있다.

2. View Optimizer

View Optimizer는 비디오 코덱에서 처리할 픽셀 율을 줄이기 위한 첫 번째 단계로서, 기본시점 영상과 부가시점 영상을 선택하는 기능을 수행한다. 기본시점은 원본시점을 그대로 선택하거나 특정 시점으로 영상 합성을 통해 프로젝션하여 생성할 수 있다. 모든 부가시점 영상은 pruner를 통해 기본시점 및 부가시점, 부가시점들 상호 간의 공간적 중복성이 제거된다. 입력되는 원본시점 중에서 기본시점 영상을 결정하는 척도는 그림 5에서와 같이 카메라 간 방향의 이격(deviation), FoV, 거리 및 투영되는 영상의 중첩 정도 등이 있으며, 기본시점



출처 B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019. 복제 금지.

그림 5 View Optimizer 알고리즘 개념

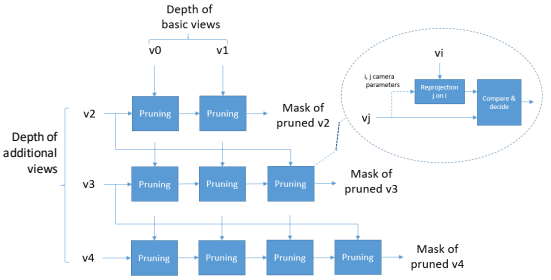
영상은 여러 개일 수 있다. 기본시점의 개수는 시청 공간이 클수록 증가할 것이며, 깊이맵의 품질이 좋지 않으면 기본시점의 개수를 증가시켜야 하지만, 공간상의 중복성 제거 정도가 작아 픽셀율이 증가하는 단점으로 작용할 수 있다.

3. Atlas Constructor

Atlas Constructor는 view optimizer에서 선택된 기준 시점과 부가 시점 영상의 깊이정보 및 카메라 파라미터를 기반으로 중복영역을 제거하고 패킹함으로써 아틀라스를 생성하고, 아틀라스 구성정보 리스트 및 카메라 파라미터 리스트를 출력한다. 여기서 아틀라스란 기본시점 영상, 프루닝된 부가시점 영상 또는 기본시점과 프루닝된 부가시점 영상으로 패킹된 영상포맷을 의미한다. 그림 4와 같이 Atlas Constructor는 크게 Pruner, Aggregator, Patch Packer로 구성된다. 각 구성요소에서의 영상처리는 깊이 정보만으로 진행되며 그 결과를 바탕으로 텍스처와 깊이에 대한 아틀라스가 생성된다. 프루닝은 프레임 단위로, Aggregation과 Clustering은 일정 시간구간(Intra Period) 단위로 진행되고, 이를 바탕으로 최종 아틀라스는 프레임 단위로 생성된다.

4. Pruner

시점 영상 간 중복성을 제거하기 위한 핵심 알고리즘은 기본시점 영상에서 차폐되었지만 부가시점 영상에서 보이는 영역을 추출함으로써 구현되는데, 그 역할을 pruner가 수행한다. 중복성 제거 효과를 극대화하기 위해 pruner는 그림 6에서와 같이 첫 번째 단계에서는 기본시점 영상과 모든 부가시점 영상 간의 프루닝을 수행하고, 두 번째 단계에서는 각 부가시점 위치에서 서로 다른 부가시점



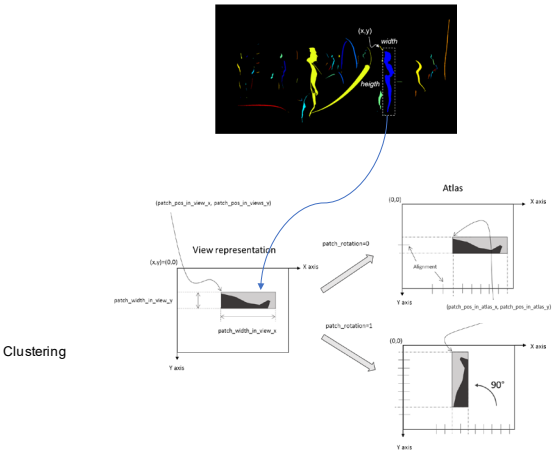
출처 B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019. 복제 금지.

그림 6 Pruner 모듈 블록도

들 간의 프루닝을 수행한다. 그림 6에서와 같이 프루닝 과정은 타겟 시점 영상(Vj)의 깊이픽셀들을 기준 시점 영상(Vi)으로 re-projection 후, 깊이값을 비교하여 invalidate pixel과 validate 픽셀로 구분하여 마스크(Mask)를 생성한다. 이 마스크 정보는 Intra period를 시작하면서 업데이트되고 Intra period 내에서 프레임 간 OR 연산을 통해 누적(aggregation)된다. Intra period 동안 마스크를 누적하는 목적은 메타데이터를 인트라 기간별로 전송하여 메타데이터의 총 양을 줄이고, 패치들이 시간축 상으로 일관성을 유지하도록 하여 압축률을 높이는 데 있다.

5. Patch Packer

Pruner에서 생성된 마스크는 Patch Packer를 통해 최종 패킹 정보를 생성하게 된다. 이를 위해 클러스터링(Clustering)이 선행되며, 각 cluster는 마스크에서 1의 값을 가진 픽셀이 연결된 영역으로서 그림 7에서와 같이 사각형 영역에 해당한다. Pruner에서 구해진 마스크는 1값을 가진 validate pixel이 분산되어 끊긴 형태인데, region growing 과정을 거쳐 최대한 연결하여 cluster의 개수를 줄여야만 패치의 개수를 줄여 픽셀율 및 메타데이터량을 줄일 수 있다. 최종적으로 패치 패킹은 각 cluster를 아



<Packing and related parameters>

출처 B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019. 복제 금지.

그림 7 패치 패킹 개념

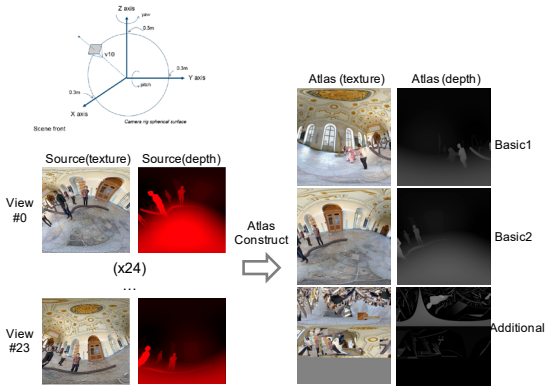


그림 8 Atlas 생성 예

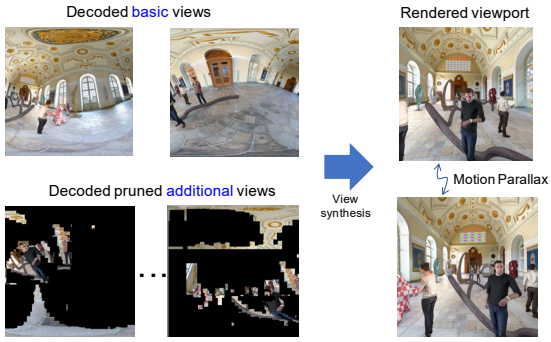


그림 9 TMIV 레더링 예

플라스에 순차적으로 패킹하는 것으로, MaxRect algorithm[6]을 기반으로 수행된다. 이 알고리즘은 단순한 방법으로 MinPatchSize 값을 기반으로 split, PiP(patch in patch), 회전, flip 등을 통하여 아틀라스 영역 내에 최대한 밀접하여 패킹한다. 하지만, 과도한 split 및 PiP는 단말에서 렌더링 품질에 영향을 미치는 것으로 실험적으로 드러났다. 따라서 단말 렌더링 품질을 최상으로 유지하면서 적절한 프루닝, 패치 패킹을 통해 픽셀율을 줄이는 것을 CE를 통해 실험하고 있다.

최종적으로, 마스크로만 이루어져 있는 패치패킹 정보를 이용하여 소스영상으로부터 해당하는 텍스처 및 깊이맵을 추출하여 아틀라스를 생성하며, 그 예는 그림 8에서와 같다.

6. TMIV Decoder

텍스처 및 깊이맵 아틀라스는 독립적으로 디코딩되고, 메타데이터는 파싱된 후 camera parameter list와 atlas parameters list로 분리된다. Atlas parameters list는 Atlas Patch Occupancy Map Generator에 사용되고 Camera parameter list는 렌더러에서 시점합성을 통해 뷰포트 영상을 재생하는 데 사용된다. TMIV는 다중의 아틀라스 혹은 시점 영상으로 부터 직접 중간 시점 영상을 합성하여 뷰포트 영상을 재생할 수 있다. 픽셀율이 낮으므로 다중시점에 비해 비교적 작은 메모리의 사용과 연산으로 뷰포트 영상을 재생할 수 있다는 장점이 있다. 시점합성 과정에서는 뷰포트에 인접한 시점으로부터 단계적으로 합성할 수 있는 multi-pass synthesis 방식도 지원하고 있으며, RVS(Reference View Synthesizer)를 참조 소프트웨어로 해서 여러 가지 알고리즘들이 적용되

고 있다.

그림 9는 디코딩된 기본시점 및 부가시점 영상들을 이용하여, 시청자의 시청위치 및 각도에 따라 가상시점을 합성하고 이를 통해 운동시차가 지원되는 뷰포트 영상을 렌더링하는 개념을 설명하고 있다.

Ⅳ. WD에서의 메타데이터

3DoF+ 비디오 압축 표준을 위한 메타데이터의 표준은 "Coded Representation of Immersive Media—Part 7: Metadata for Immersive Media(Video)"이란 표준명으로 제정되고 있으며, 현재 이를 위한 WD 문서가 작성 중에 있다. 이 WD는 3DoF+ 비디오 압축 비트스트림에 대해 정의하고 있으며 아틀라스를 압축한 HEVC 시퀀스 파라미터(Sequency Parameter), 카메라파라미터에 대한 메타데이터 및 아틀라스 관련 파라미터 등이 포함되고 있다. 3DoF+ 압축 비트스트림은 한 개 이상의 아틀라스 쌍(텍스처 및 깊이맵)으로 구성되므로 MIV 비트스트림은 독립적으로 인코딩된 다중 layer의 HEVC 비트스트림이라고 볼 수 있다. MIV 시퀀스파라미터에는 프로파일, tier, level 등의 기본적인 파라미터들이 정의 중이며, Camera parameters list, Atlas parameters list, Atlas parameters syntax 등의 새로운 메타데이터가 대표적으로 정의하고 있다. Camera parameters는 전통적으로 다시점 비디오를 이용한 영상합성에 필요한 내외부 파라미터들을 포함하고 있다. 특히 전방위 영상을 표현하기 위한 이머시브비디오는 OMAF에서 정의한 sphere 좌표 체계를 준수하므로, 360° 공간상에 yaw, pitch, roll로 표현되는 카메라의 회전각과 X, Y, Z로 표현되는 카메라의 위치로 표현된다. 콘텐츠 생성과정에서 16비트 depth가 추출된다면 이를 10비트의 압축표준으로 변환하기 위한 depth quantization도 정의되었다. Atlas parameter는 access unit의 기본적인 단위로서 아틀라스별로 id가 부여되며, 이는 텍스처와 깊이맵을 동시에 지칭한다. 아틀라스 파라미터의 요소에는 크기를 지정할 수 있는데, 아틀라스별로 서로 다른 해상도가 인코딩될 수도 있다. 아틀라스파라미터 요소에는 아틀라스를 구성하는 모든 패치에 정보가 기술되는데, 프루닝될 때 패치의 소스 view id, 소스 시점의 위치, 크기, 아틀라스에 패킹될 때의 위치, 회전각도 등으로 구성된다.

Ⅴ. CE 현황 및 계획

2019년 3월 MPEG 제네바 회의 때 결정된 5개의 CE를 진행한 결과를 바탕으로 10월 회의에서 레퍼런스 소프트웨어를 TMIV 3.0으로 업그레이드하였다. 10월 회의에서 추가된 것은 시점 순위 기반의 pruning, 깊이기반 occupancy map, group-based TMIV, object-based application, generalized viewing space signaling 등이다. 이를 기반으로 향후 표준화 진행을 위하여 그림 10에서와 같이 3개의 CE를 지속적으로 추진하고, 차기 회의에서 CD를 진행할 예정이다[6]. 이에 따라 3DoF+ 비디오 부호화 표준화는 CD(2020.01), DIS(2020.07), FDIS(2020.10)

Core Experiments for MIV

- CE1: Metadata (Intel, Interdigital, ZJU, Nokia, ETRI)
  - V-PCC alignment
  - Pre-screening of patches
  - Study of view_id
  - Viewing space
- CE2: Pixel pruning (Philips, ETRI, Interdigital)
- CE3: Atlas preparation (Interdigital, ZJU, KAU, Intel, PUT, ETRI)

그림 10 3DoF+ 표준화 CE 계획 및 수행기관

의 일정으로 진행될 예정이다.

한편 V-PCC를 표준화 중인 MPEG 3DG그룹과의 공동회의를 통해 V-PCC와 공통되는 기술을 프로파일링하여 MIV를 V-PCC의 annex에 포함하는 작업을 시작하였으며, 동시에 3DoF+는 "Coded Representation of Immersive Media-Part 12: Immersive Video(ISO/IEC 23090-12)"라는 프로젝트명으로 별도의 표준으로 제정될 예정이다.

따라서 3DoF+ 비디오 표준은 3D-HEVC와 같은 복잡한 3D 비디오 코덱과는 달리 산업계의 요구사항을 충실히 반영함으로써 향후 포인터 클라우드, 6DoF 등의 다양한 이머시브미디어의 표준을 위한 기본적인 프레임워크로서의 역할을 수행할 것으로 전망된다.

약어 정리

CD	Certificate Document
CE	Core Experiment
CfP	Call for Proposal
DoF	Degree of Freedom
EE	Experiment explorer
HMD	Head Mounter Display
MIV	Metadata for Immersive Video
MPEG	Moving Picture Experts Group
MV-HEVC	Multiview High Efficiency Video Coding
TM	Test Model
TMIV	Test Model for Immersive Video
V-PCC	Video codec based Point Cloud Compression
VVC	Versatile Video Coding
WD	Working Draft

참고문헌

[1] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," ISO/IEC JTC1/SC29/WG11, w16768, Apr. 2017.

[2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, "Standardization Status of Immersive Video Coding," IEEE Jour. Emerg. Select. Topics Circuits Syst., vol. 9, no. 1, pp. 5-17, Mar. 2019.

[3] J. Boyce, R. Dore, V. Vadakital, "Working Draft2 of ImmersiveVideo," ISO/IEC JTC1/SC29/WG11, w18576, Jul. 2018.

[4] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, Jul. 2019.

[5] HM reference software, [Online]. Available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[6] B. Kroon, "Report of the BOG on Immersive Video," ISO/IEC JTC1/SC29/WG11, m49228, Oct. 2019.