

## 10 MIV 기술 표준화 - MPEG뉴미디어포럼

### □ 개요

○ 기술 정의 : MIV (MPEG Immersive Video) 부호화 표준

▶ 이머시브 비디오는 시청자가 공간상에서 완전한 6DoF(Degree of Freedom) 자유도로 움직일 때 보여져야 할 영상을 재현함으로써 사실감과 몰입감을 동시에 제공할 수 있는 비디오이며, 이를 부호화 및 렌더링하기 위한 비디오 압축 표준이 MIV 부호화 표준임

※ OSI 7 Layer 중 Presentation layer(layer 6)에 해당

○ 기술에 대한 부연 설명

- MPEG-I에서 MIV 비디오 표준화 과정

- 3DoF+ 단계: 전방위 장면을 포함한 복수 개의 영상과 이를 기반으로 합성한 다수의 가상 시점 영상을 기반으로 착석환경에서 약간의 머리 움직임과 같이 매우 제한된 운동에 대한 운동시차 지원
- 6DoF 단계: 시청자의 6방향(좌우이동, 상하이동, 전후이동, 전후 기울기(pitch), 좌우 기울기(roll), 좌우회전(yaw)) 움직임에 대한 자유도를 지원할 수 있는 시청 환경을 의미하며, 시청자가 한두 발짝의 움직임에 대해 운동시차 지원

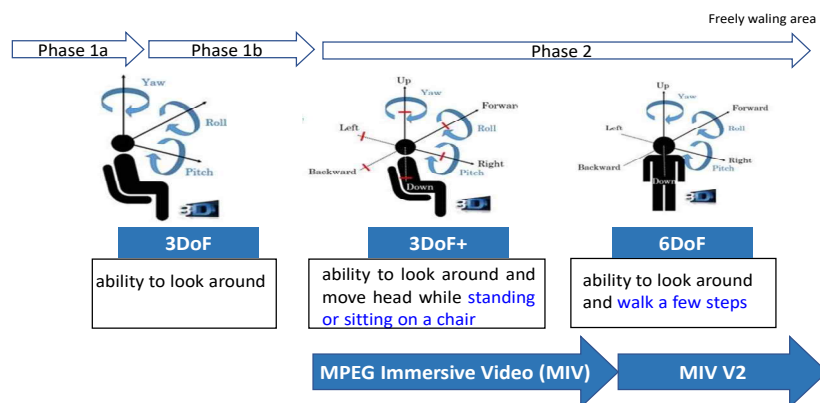


그림. MPEG-I 이머시브 비디오 압축 표준화 과정

- MIV 부호화 표준 기술 특징

- MIV 부호화 표준 기술은 전통적인 비디오 코딩에서의 인코딩 효율 향상과 더불어 입출력 인터페이스, 압축 처리를 위한 코덱의 개수 등을 나타내는 픽셀율(pixel rate)의 최소화를 기본적인 요구사항으로 함
- 비교적 적은 수의 2D 비디오 코덱을 통해 사용자의 움직임에 따른 끊김 없는 운동시차의 제공과 고품질 뷰포트(viewport) 영상의 재현을 목표로 함

- HEVC, VVC, LEVC 등 일반적인 2D 비디오 코덱과 조합 가능한(agnostic) 비디오 표준
- V-PCC 표준과 공통 규격(ISO/IEC 23090-5, V3C)을 공유하며 2021년 7월에 FDIS를 발간(ISO/IEC 23090-12:2021(E))

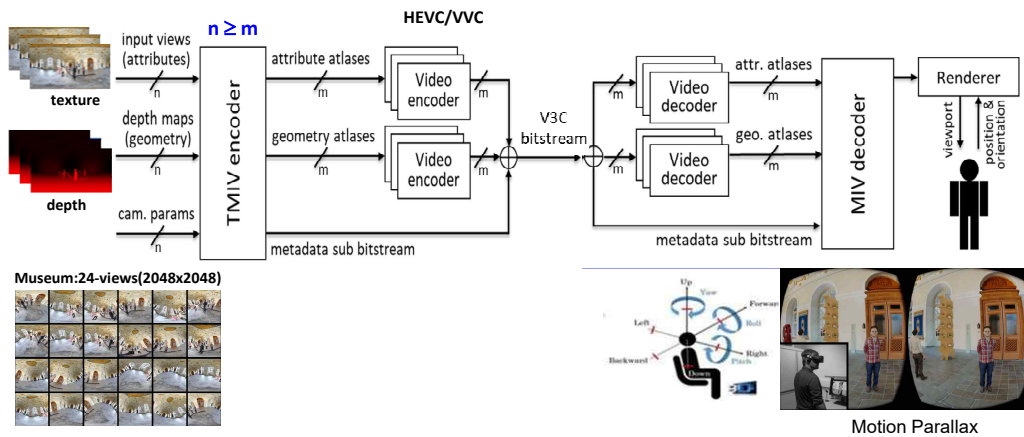


그림. MIV 부복호화 기술 아키텍처

- MIV 부호화에서 TMIV(Test Model for MIV) 인코더는 핵심적인 전처리 기능을 수행하며, 일반적인 비디오 인코더가 부호화 가능하도록 다수의 시점영상들로부터 소수의 아틀라스 영상을 생성
- 입력되는 다중시점 형태의 전방위 비디오로부터 3D 공간상에서 대응되는 픽셀들의 중복성을 줄이기 위한 푸루닝(pruning)과 패치 패킹(patch packing)이 주요 기능임

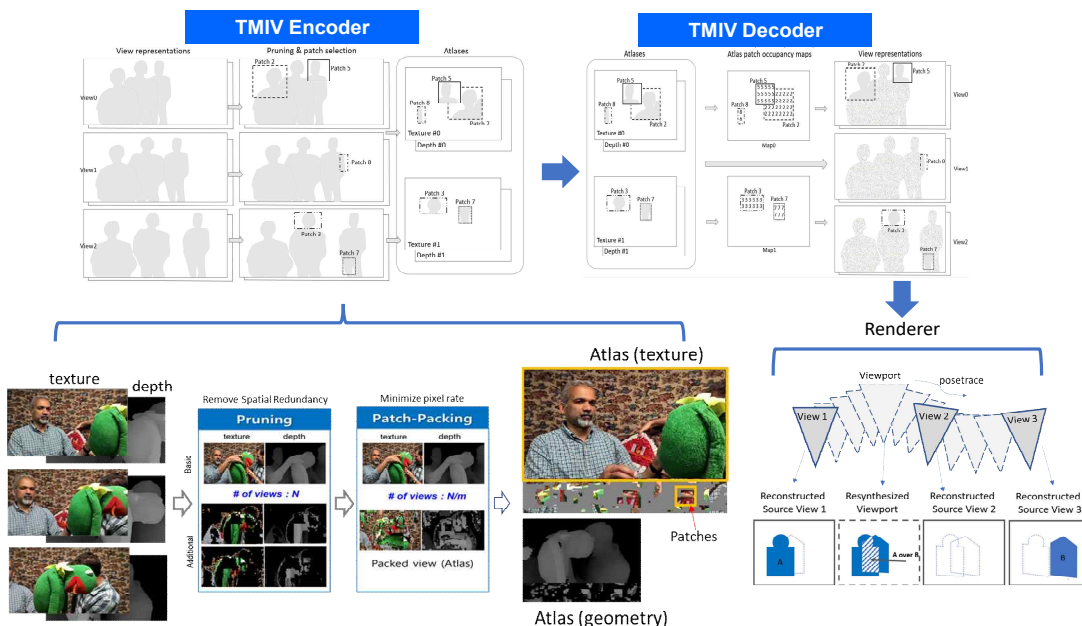


그림. MIV 이머시브 비디오 부복호화 개념

- MIV 디코더는 MIV 표준 메타데이터를 파싱하여 푸루닝된 원본시점 영상을 복원
- MIV 렌더러(renderer)는 디코딩되어 분리된 패치들로부터 시청자가 바라보는 방향에 해당하는 중간시점 영상을 합성하며, 시청 경로에 해당하는 포스트레이스(posetrace)에 따라 연속적으로 뷰포트 영상을 렌더링
- 푸루닝은 기하정보를 이용해서 모든 소스 시점의 픽셀들을 3D 공간상으로 역투영하고 상호 대응되는 픽셀들의 깊이 및 칼라값을 비교하여 중복성 제거
- 비교과정을 통한 중복성 제거는 계층적인 구조로 모든 시점들간의 조합으로 수행되며, 이 계층 구조를 푸루닝 그래프로 표시하여 시그널링함

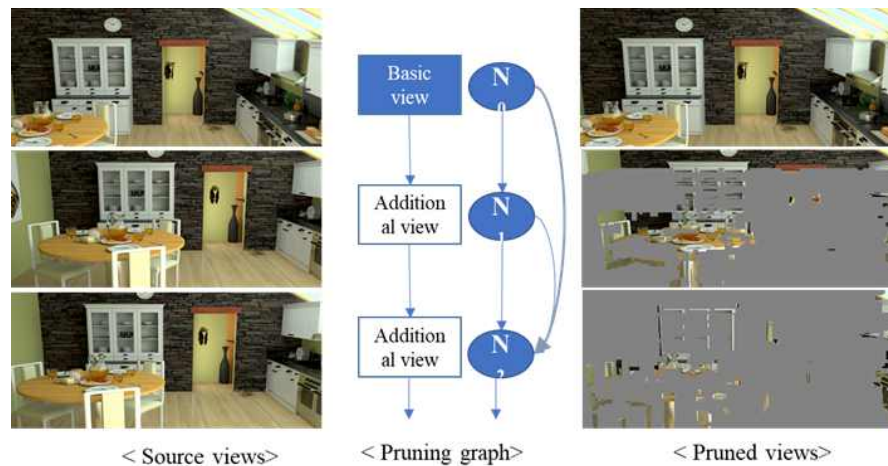
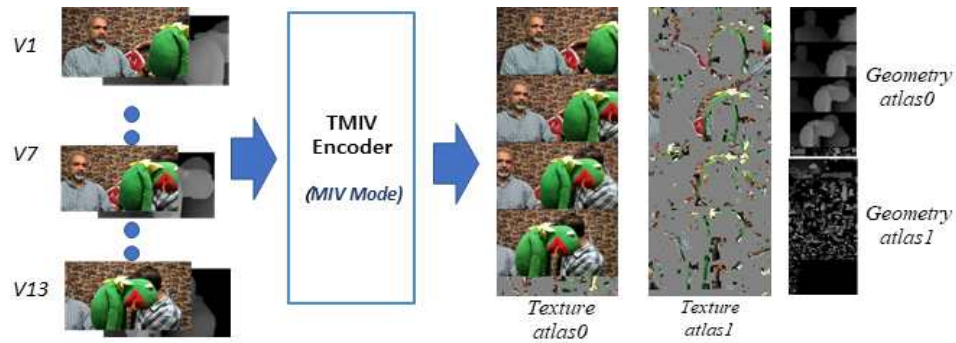
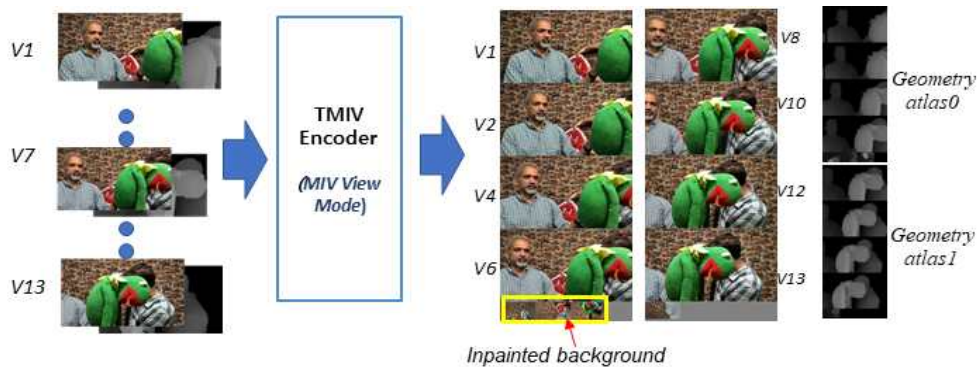


그림. MIV 부호화에서 푸루닝 개념

- TMIV 인코더는 MIV 모드 및 MIV 뷰(view) 모드로 동작함
  - MIV 모드는 푸루닝을 하지 않은 기본시점 영상과 푸루닝을 통한 패치들로 구성된 아틀라스 영상을 생성
  - MIV 뷰 모드는 모든 입력 시점영상으로부터 허용된 픽셀레이트에 따른 몇 개의 기본시점을 선택하고 인페인팅된 배경과 함께 패킹하여 아틀라스 영상 생성 (푸루닝 과정 없음)
  - MIV 모드는 기하정보가 정확한 소스영상으로부터 푸루닝에 필요한 공간정보를 비교적 정확히 추출할 수 있는 환경에서 비실시간 인코딩에 적합하며, MIV 뷰 모드는 그렇지 않은 환경에서 고속 인코딩에 적합함



(a) MIV 모드



(b) MIV View 모드

그림. MIV 부호화에서의 TMIV 인코더 동작 모드

- MIV 부호화에서는 기하정보 즉 깊이정보를 인코딩할 때, 데이터양 특히 픽셀들의 최소화를 위해 기하 아틀라스 영상의 크기를 일정 비율만큼 축소 가능함
  - 깊이맵 영상은 텍스처 영상에 비해 낮은 공간 주파수 성분으로 구성되어 있어, 원래의 크기로 복원했을 때 상대적으로 정보 손실이 덜 발생함
  - TMIV 인코더는 축소 작업에  $2 \times 2$  크기의 “최대 풀링(max pooling)” 필터를 사용하여 렌더링 품질에 중요한 전경 객체의 정보를 최대한 보존함

## □ 현황

### ○ 기술개발 현황 및 전망

- (국제)21년 1월 표준화 미팅에서 인텔(Intel)은 자사 GPU 칩이 탑재된 노트북을 통해 MIV extended 프로파일로 압축된 비트스트림의 실시간 디코딩과 렌더링 기술을 선보임
- 그림과 같이 다시점 영상 형태의 이머시브 비디오를 TMIV 인코더를 통해 인코딩한 후, 노트북에서 듀얼 HEVC 디코더 칩으로 디코딩함



- 딥러닝 기반의 헤드 트래킹 툴을 이용하여 시청자의 시청방향을 추적하고 그 방향으로의 시점영상을 실시간으로 렌더링함으로써, 일반적인 모니터에서도 이머시브 비디오 서비스가 가능함을 선보임

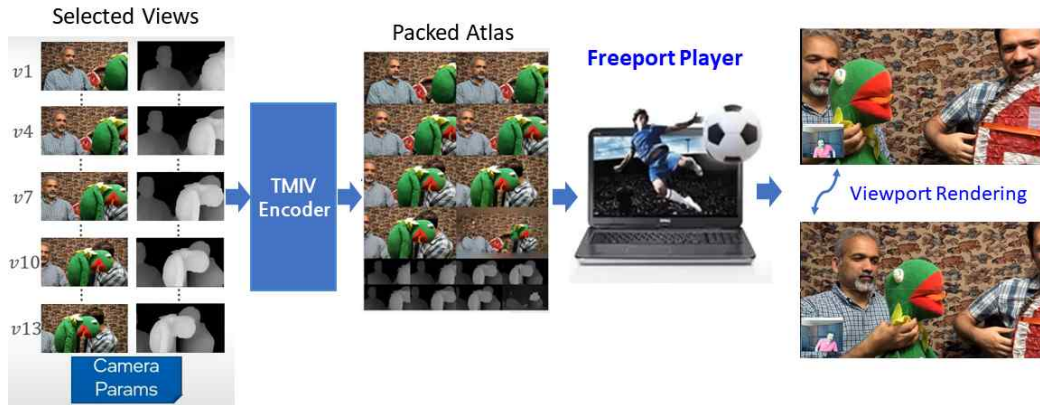


그림. Intel의 MIV기반 Freeport player

- (국제)21년 표준화 4월 미팅에서 인터디지탈은 MIV MPI(Multiple Plane Image)가 스마트 단말에서와 같이 low-end 디바이스에서 실시간으로 구현 가능함을 선보임
- MPI는 단말에서 가장 간단하게 렌더러를 구현할 수 있는 MIV 프로파일이기 때문에, 테블릿에서 실시간으로 MV 콘텐츠의 렌더링이 가능함을 검증함
- MIV 디코딩 기능은 미디어코덱을 이용한 듀얼 HEVC 디코더에서 구현되었고 GPU로 렌더링 기능을 구현함. 최종적으로 사용자의 터치 조작에 의해 렌더링되는 시점이 변하는 것을 선보임

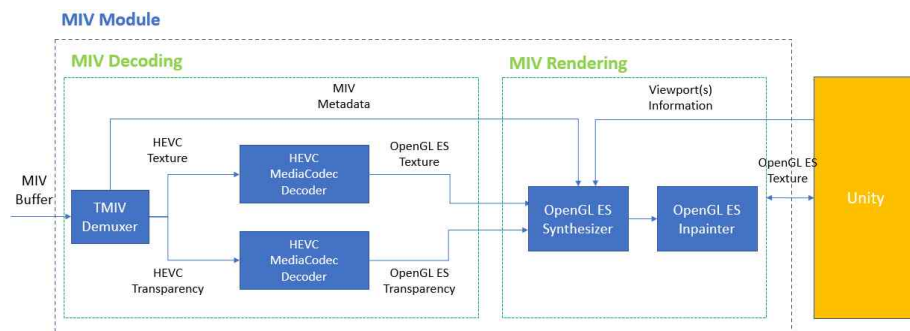


그림. Interdigital의 MIV player

- (국제) 22년 MIV 표준화 그룹의 에디터는 Vimmerse라는 벤처를 창업하고, MIV 기반의 3D 비디오 획득, 3D 콘텐츠 생성, 재생 플레이어, SDK 관련 솔루션을 제공 중
- 사용자가 MIV 콘텐츠 획득을 위한 툴은 다운 받고 이를 통해 콘텐츠를 획득하여 업로드 하면 MIV 뷰 모드의 MIV 콘텐츠로 생성하여 다시 제공함. 이에 사용자는 스마트폰, PC 등에서 재생할 수 있음

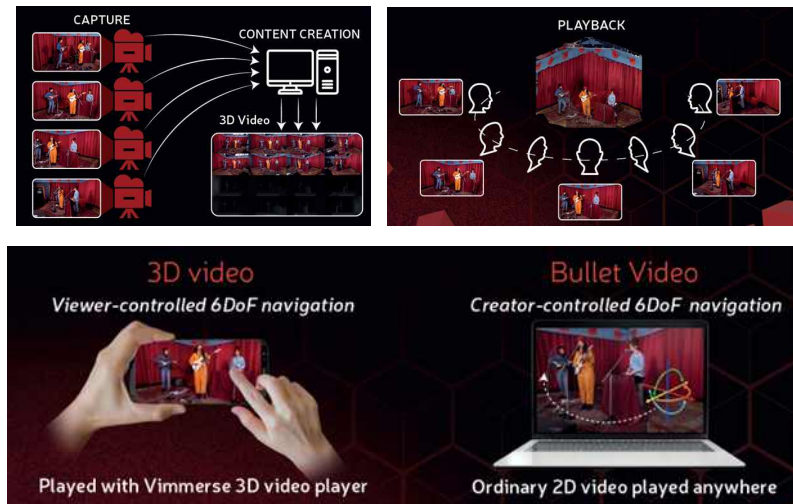


그림. Vimmerse사의 MIV기반 3D 비디오 솔루션

- (국내) ETRI는 MIV 표준화에 참여하여 다수의 표준 IPR 확보와 더불어 표준기술 검증을 위해 2021년에 MIV 기반의 3DoF+ 비디오 생성 및 재현 시제품을 개발함
- MIV 표준 부합 3DoF+ 비디오 인코딩 검증 소프트웨어 개발
- 시청 위치에서의 시점영상을 재현하기 위한 딥러닝 기반(MobilNet) 사용자 헤드 추적 모듈 개발 및 MIV 재현 SW 모듈 연동
- CUDA 기반으로 TMIV 렌더링 모듈의 고속화 구현을 통한 실시간(30fps) 3DoF+ 비디오 재현 가능



그림. ETRI의 MIV기반 3DoF+ 비디오 생성(좌) 및 재현 실험시제품(우)

### ○ 시장 및 산업체 현황 및 전망

- (Technicolor) 스위스 제네바에서 열린 제128차 MPEG 회의에서 패치 기반 3DoF+ 비디오 재현 개념을 시연하고, 21년 MPEG 표준화 4월 미팅에서 MIV MPI가 스마트패드와 같은 저성능 디바이스에서도 실시간 구현이 가능함을 선보임
- 제한된 사용자 움직임 범위에 대한 운동시차를 제공할 수 있는 3DoF+ 비디오 기술은 패치기반으로 경량화된 시스템에서 재현 가능함을 최초로 시연



그림. Technicolor사의 MIV 기반 3DoF+ 비디오 개념 시연

- (구글) 이머시브 비디오의 주요한 특징 중 하나인 실제감(Presence)을 표현하기 위한 수단으로 입체영상 기술에 주목하고 있음
- 구글은 46개 반구형 카메라 어레이와 딥러닝기반 메쉬·깊이 모델링을 이용한 라이트필드 영상 획득 및 3차원 공간 표현 기법인 이머시브 라이트필드 기술을 공개('20.7)

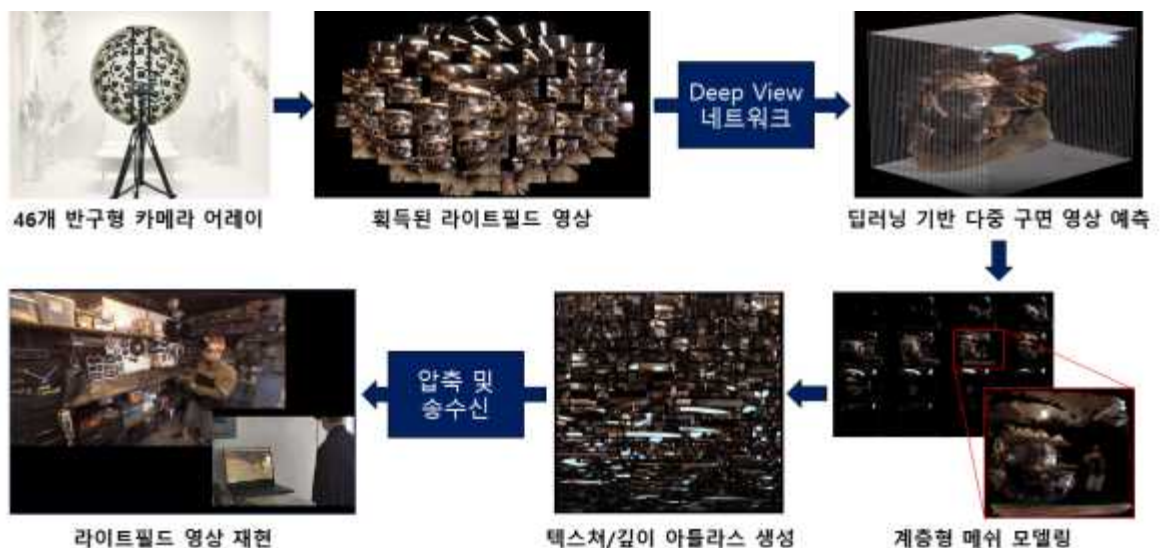


그림. 구글의 이머시브 라이트 필드 기술 흐름도

- (4D리플레이) 4D리플레이는 다시점 영상 솔루션 포드리플레이(4DReplay)와 최대 360도까지 돌려볼 수 있는 4D 인터랙티브 기술을 보유하고 있으며, 4D리플



레이 애플리케이션을 이용하여 시청자가 화면 줌인·줌아웃, 360도 회전, 되감기, 다시 보기를 제공.

- MLB, NBA, LPGA, UFC 등의 스포츠 중계영상 제작과 평창동계올림픽에서는 인텔과 경쟁하여 OBS 입찰에 성공하여 경기 장면을 중계함
- 도쿄올림픽에서도 다양한 종목의 스포츠 비디오 판독 등에도 활용. 2021년 4차 원 인터랙티브 웹드라마 ‘미스터LEE’ 제작을 진행함

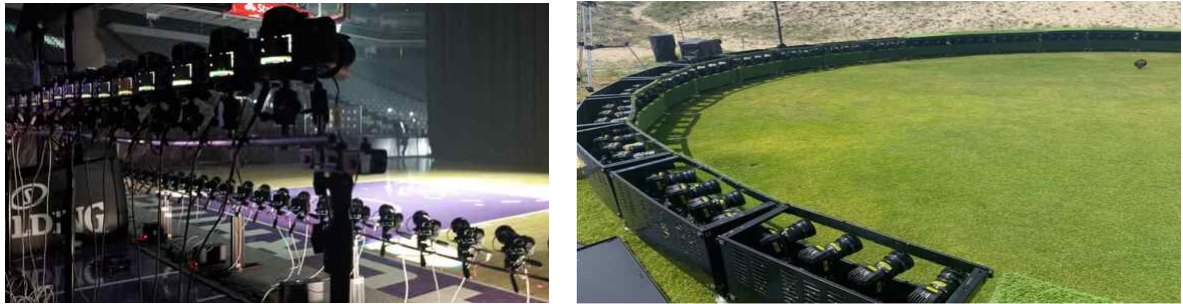


그림. (좌) NBA 새크라멘토 및 (b) 4DReplay: 다시점 리플레이 영상 솔루션

#### ○ 표준화 현황 및 전망

- (국제) ‘19년 3월 회의 때 CfP(Call for Proposal)에 대응하여 제안된 다섯 개의 기술을 기반으로 테스트 모델(TMIV) 소프트웨어를 통해 본격적인 표준 기술을 개발하였고 22년 7월 버전 14.0이 배포됨

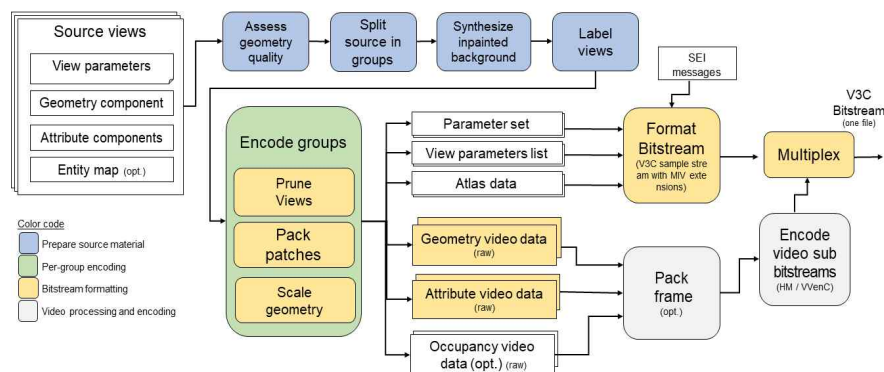


그림. TMIV 인코더 블록도

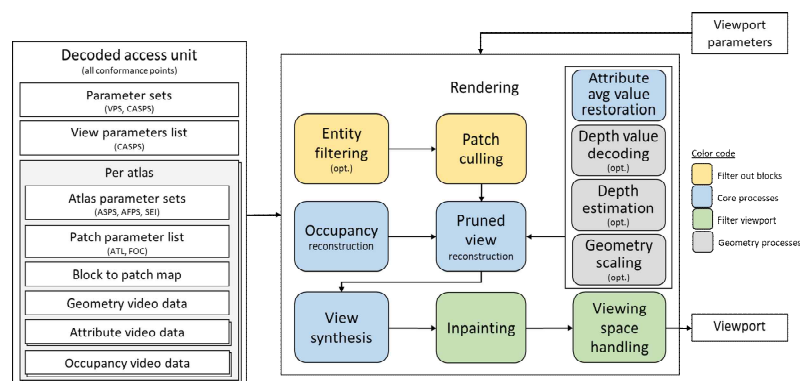


그림. TMIV 디코더 및 렌더러 블록도



- ‘21년 1월에 DIS(Draft International Standard)를 발간하고 DIS ballot에 대한 77개의 comment를 모두 처리한 후, 7월 미팅에서 MIV FDIS를 발간하고 10월초에 최종적으로 SC29에 공식적으로 접수함
- MIV 표준은 “ISO/IEC 23090-12:2021(E), Information technology - Coded representation of immersive media - Part 12: MPEG Immersive video”이란 표준명으로 제정됨
- MIV 표준의 참조 규격은 ISO/IEC 23090-5, Information technology - Coded Representation of Immersive Media - part 5: V3C(Visual Volumetric Video-based Coding) and V-PCC(Video-based Point Cloud Compression)
- 즉, 패치방식의 아틀라스 포맷은 포인트클라우드 표준 중의 하나인 V-PCC와 공통적인 규격으로 표준화되었고, 이러한 공통적인 규격은 V3C에 규정하고 있으며 part12에는 MIV 특화된 규격만 정의함

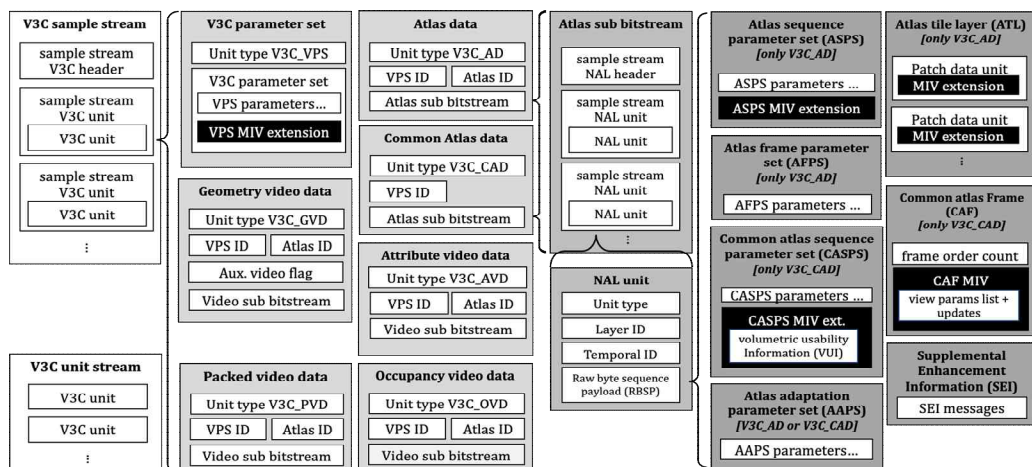


그림. MIV 부호화 표준에서 선택스 구조

- MIV 표준에서 규정한 프로파일과 이의 대표적인 선택스는 표와 같음. MIV Extended 프로파일은 아틀라스 프레임 패킹 기능, 별도의 점유지도 인코딩, 2개의 속성 정보 인코딩 기능 등을 포함

표. MIV 부호화 표준 프로파일과 대표적인 선택스

Key Syntax element	MIV Main	MIV Extended	MIV Extended Restricted Geometry(MPI)	MIV Geometry Absent
vps_packing_information_present_flag	0	0, 1	0, 1	
vps_occupancy_video_present_flag	0	0, 1	0	
vps_geometry_video_present_flag	1	0, 1	0	
vme_embedded_occupancy_enabled_flag	1	0, 1	0	
al_attribute_count	0,1	0, 1, 2	2	
al_attribute_type_id	ATTR-TEXTURE	ATTR_TEXTURE, ATTR_TRANSPARENCY	ATTR_TEXTURE, ATTR_TRANSPARENCY	ATTR_TEXTURE

- MIV에서의 아틀라스 프레임 패킹(frame packing)은 텍스처, 기하 및 점유지도 아틀라스들을 단일 아틀라스로 패킹함으로써, 디코더에서의 동기화를 용이하게 하고 실행되는 디코더의 개수를 줄일 수 있음

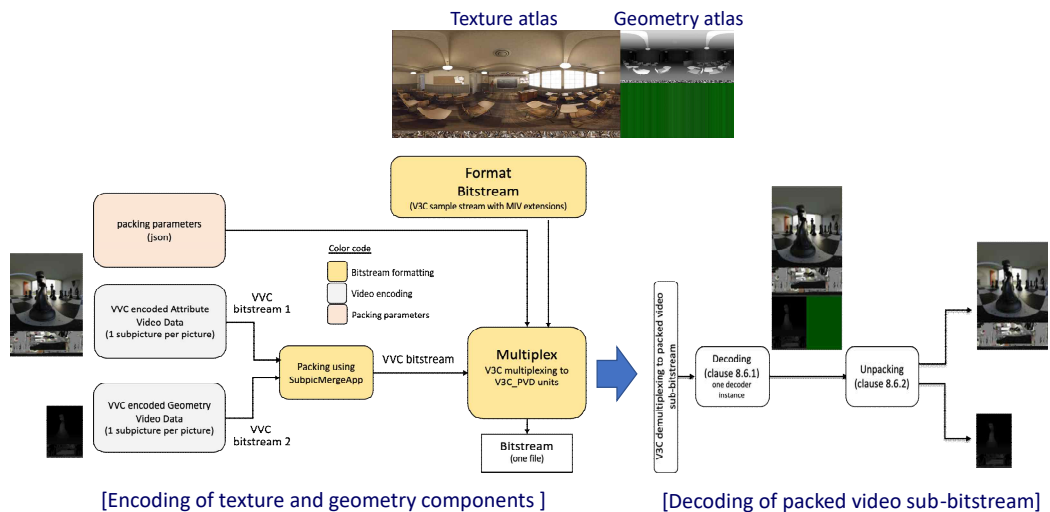


그림. MIV 부호화 표준에서의 아틀라스 프레임 패킹

- MIV Extended Restricted Geometry 프로파일은 MPI 포맷 형태의 이머시브 비디오를 부호화하기 위한 것임
- MPI는 3D 공간을 깊이 방향으로 계층화하는데 이는 중앙 카메라를 기준으로 일정 간격의 계층화된 면에 텍스처를 재투영하여 계층화된 깊이 영상(layered depth image)으로 표현하는 것임
- 이 계층화된 텍스처 및 투명 정보를 MIV의 아틀라스로 패킹하여 압축 전송하면 단말에서는 비교적 간단한 연산을 통해 렌더링 가능
- 인코딩되는 속성으로서 투명정보(transparency)를 포함하여 2개로 고정하며, 불 연속적인 기하 값(restricted geometry)을 통해 계층 정보를 인코딩함

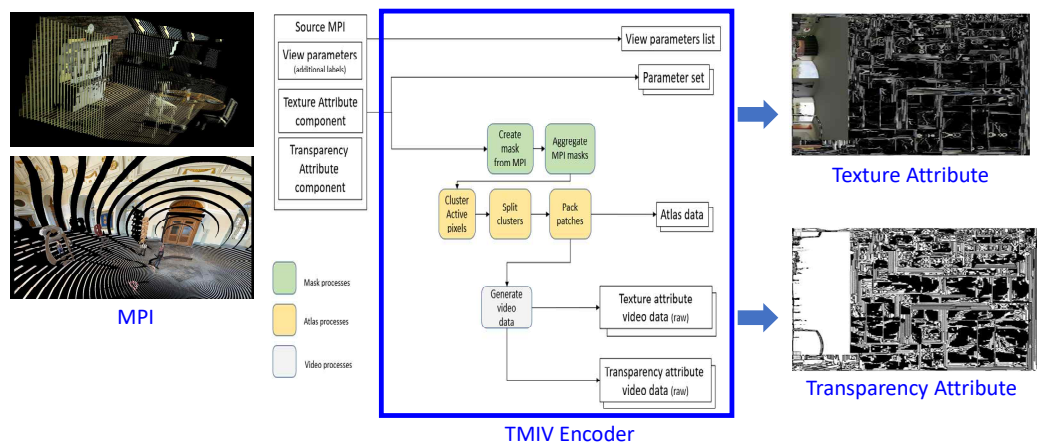


그림. MIV 부호화 표준에서의 MPI 프로파일

- MIV Geometry Absent 프로파일은 중간 시점영상 합성에 필요한 기하 정보를 인코더로부터 전송하지 않고 디코더에서 직접 추출하게 하는 기술들을 포함
- 이때 인코딩하는 아틀라스는 기본시점 영상만으로 패킹된 MIV 뷰 모드를 사용함으로써 저사양의 인코더를 사용할 수 있지만, 디코더를 상당히 복잡하게 하는 단점이 있음
- 따라서 디코더의 복잡도를 줄이기 위해, 깊이 추출에 필요한 특징들을 미리 서버에서 추출하여 기하 보조 (Geometry assistance) SEI를 통해 전송함
- 21년 10월 미팅에서 MIV 비트스트림의 conformance test를 위해 발췌된 17개의 포인터를 테스트하도록 7개의 비트스트림을 생성하였고, 이를 테스트한 결과는 WD 문서(WD3 of ISO/IEC 23090-23:2021(E))로 발간함
- MIV 압축 비트스트림 : V3C 파라미터 셋, 공통 아틀라스 데이터(common atlas data), 아틀라스 데이터(atlas data) 그리고 실제 렌더링되는 시퀀스인 속성 비디오, 기하 비디오 및 점유지도(옵션)로 구성

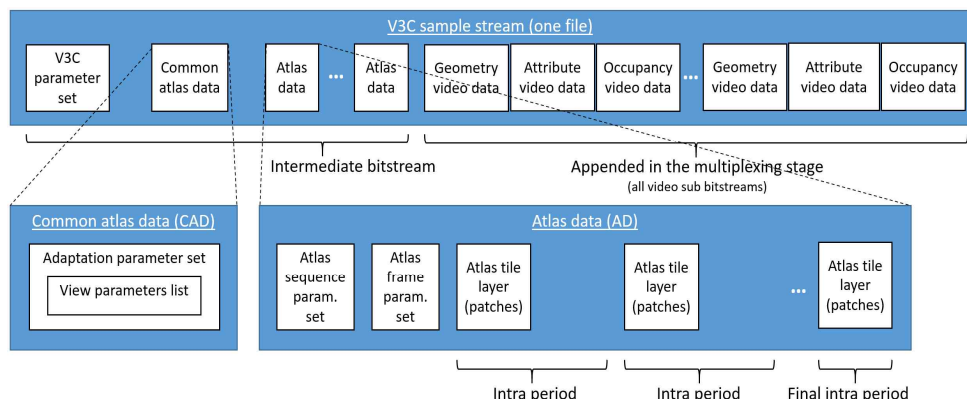


그림. MIV 부호화 비트스트림 구조

- 22년에 MIV IS의 발간을 목표로 함과 동시에 MIV 표준기술의 우수성을 검증하고 홍보하기 위해 verification test를 진행 중임
- MIV와의 성능 비교를 위한 앵커는 표준화 시작 전의 다시점 코덱인 MV-HEVC이며, 전제 조건으로서 인코딩하기 위한 영상의 픽셀률을 동일하게 설정
- 정해진 픽셀률을 맞추기 위해 시점 선택에 의해 공간정보가 유실되는 MV-HEVC와 공간정보의 중복성을 제거하고 렌더링에 필요한 픽셀만 처리하는 MIV와의 성능 비교 실시
- 객관적인 평가 결과 대부분의 비교 콘텐츠에서 MIV의 성능이 우수하였으나, 포스트 트레이스 파일 재생을 통한 주관적인 평가에서는 객관적인 성능 대비 우수성이 미비하여 지속적으로 TMIV 소프트웨어의 성능을 업그레이드 중임

H	BabyUnicorn	Natural content
S	CBABasketball	Natural content
Y	Breakfast	Natural content
Z	Barn	Natural content
F	Guitarist	Computer generated
W	Dancing	Computer generated
X	Cyberpunk	Computer generated

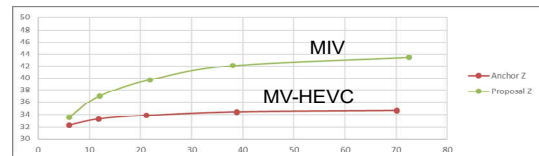
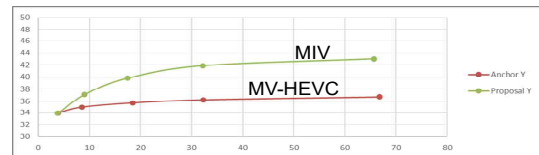
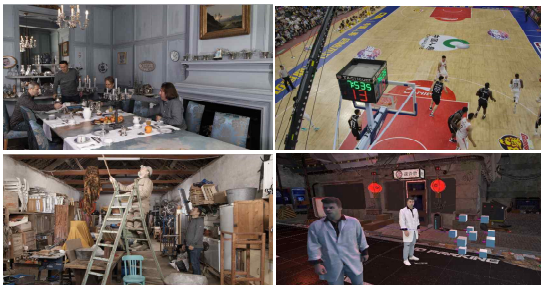
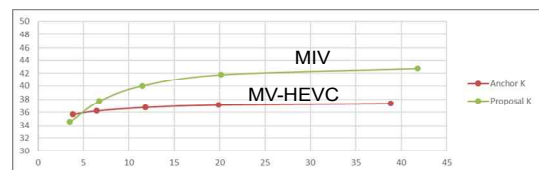


그림. MIV verification test 용 콘텐츠(좌) 및 부호화 성능(우)

- 21년 4월 회의 때부터 MIV 표준의 확장을 목표로 MIV edition-2를 위한 유저 케이스(usecase) 및 요구사항(requirement) 문서를 준비하였으며 22년 1월 회의에서 최종 발간함



## □ 주요이슈 및 대응방안

### ○ (주요이슈)

- 현실과 가상이 공존하는 메타버스를 통한 다양한 신규 서비스 모델이 기대됨에 따라 사실감과 몰입감을 동시에 지원하기 위해서, 전방위 실감 콘텐츠의 획득, 압축전송, 렌더링을 위한 표준 및 서비스 플랫폼이 필요
- 이머시브 비디오 압축 전송 표준(MIV edition-2, V3C edition-3)의 확장에 따라 지속적인 표준기술 개발 및 IPR 확보 필요

### ○ (현황 및 문제점)

- 기존의 다시점 비디오 부호화 코덱(MV-HEVC, 3D-HEVC 등)은 부복호화 시스템이 복잡하여 상용화가 부진함
- 현재의 메타버스 서비스는 CG 콘텐츠, 2D 평면 콘텐츠 위주로 제한된 환경에서만 실감형 콘텐츠를 제공하고 있어, 사실감 재현 및 몰입감 제공이 부족하여 서비스 활성화에 걸림돌로 작용

### ○ (대응방안)

- MIV 부호화의 요구사항이었던 ‘낮은 픽셀레이트를 통한 끊김 없는 운동시차의 지원’이라는 특징을 부각하여 경량 MIV 코덱 시스템 상용화를 통한 6DoF 기반의 이머시브 비디오 서비스 제공
- MIV 기반 이머시브 비디오 시스템 개발 및 서비스 검증을 위한 지속적인 연구 개발, 상용화 지원을 통해 관련 시장 활성화 및 기술 국제 경쟁력 확보
- MIV 기반 이머시브 비디오, 체적형 비디오 객체 등의 생성, 부호화 및 실감/몰입감 재현 관련 핵심기술을 고도화하여 관련 산업에서의 경쟁력을 제고하고 상용화를 선도할 수 있는 기회 확보

[약어표]

3DoF	3 Degree Of Freedom	3DoF+	3 Degree Of Freedom Plus
6DoF	6 Degree Of Freedom	HEVC	High Efficiency Video Coding
MIV	MPEG Immersive Video	TMIV	Test Model for MIV
V3C	Multiview High Efficiency Video Coding	V-PCC	Point Cloud Compression/Coding
MPI	Multiple Plane Image	LEVC	Low Complexity Enhancement Video Coding
VVC	Versatile Video Coding	CfP	Call for Proposal
DIS	Draft International Standard	FDIS	Final Draft International Standard

[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, Mar. 2019, pp.5-17.
- [3] ISO/IEC 23090-2, Information technology - Coded Representation of Immersive Media - Part 2: Omnidirectional Media Format (OMAF) 2nd Edition.
- [4] Broxton, Michael, et al., “Immersive light field video with a layered mesh representation,” ACM Transactions on Graphics (TOG) 39.4 (2020): 86-1.
- [5] J.Boyce, V.K.Malamalkital, B.Chupeau, “Information technology - Coded representation of immersive media - Part 12: MPEG Immersive video“, ISO/IEC JTC1/SC29/WG4, N00111, Oct. 2021.
- [6] 이광순, 서정일, “MPEG 이머시브 비디오 표준기술 동향과 전망”, IITP, 주간기술 동향 1969호, 2020. 10. 21.
- [7] Information Technology - Coded Representation of Immersive Media - Part 5: Visual Volumetric Video-Based Coding(V3C) and Video-Based Point Cloud Compression(V-PCC), Standard ISO/IEC23090-5, N19329, May 2020.

- 
- [8] Call for Proposals on 3DoF+ Visual, ISO/IEC JTC1/SC29/WG11 MPEG/N18145, Jan. 2019, Marrakesh, Morocco.
  - [9] B. Salahieh, B. Kroon, J. Jung, M. Doma-ski (Eds.), “Test model 11 for Immersive Video,” ISO/IEC JTC1/SC29/WG11, N0142, Oct. 2021
  - [10] ISO/IEC 23008-2: Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 2: High efficiency video coding.
  - [11] Boyce, Jill M., et al., “MPEG Immersive Video Coding Standard,” Proceedings of the IEEE 2021.
  - [12] J.Jung, B.Kroon, “Common Test Conditions for MPEG Immersive Video,” ISO/IEC JTC1/SC29/WG4, N0143, Oct. 2021, Online
  - [13] V.K.Malamalkital, B.Kroon, “Draft use-cases and requirements for MIV - edition 2,” ISO/IEC JTC1/SC29/WG4, N0143, Oct. 2021.