

특집논문 (Special Paper)

방송공학회논문지 제27권 제6호, 2022년 11월 (JBE Vol.27, No.6, November 2022)

<https://doi.org/10.5909/JBE.2022.27.6.861>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

MPEG 몰입형 비디오를 위한 Geometry Packing 구현

정 중 범^{a)}, 이 순 빈^{a)}, 류 은 석^{a)‡}

Implementing Geometry Packing for MPEG Immersive Video

Jong-Beom Jeong^{a)}, Soonbin Lee^{a)}, and Eun-Seok Ryu^{a)‡}

요 약

실사 및 컴퓨터 그래픽을 표현하는 다수의 몰입형 영상을 효율적으로 부호화하는 표준으로 moving picture experts group (MPEG)에서는 MPEG immersive video (MIV) 표준을 개발하였다. MIV 표준은 다수의 몰입형 영상을 압축하여 다수의 출력 영상인 아틀라스를 생성하나, 다수의 아틀라스를 부호화 후 복호화 시 저사양 장비에서 복호기 간 동기화 문제가 발생할 수 있다. 본 논문은 저사양 및 고사양 장비에 모두 대응하여 적응적 복호기 개수 조절을 위한 geometry packing 기법을 제안하고 구현한다. 제안하는 방법은 MIV의 최신 참조 소프트웨어에서 문제없이 동작함을 확인하였다.

Abstract

The moving picture experts group (MPEG) developed the MPEG immersive video (MIV) standard for efficient compression of multiple immersive videos representing natural contents and computer graphics. The MIV compresses multiple immersive videos and generates multiple output videos which are defined as atlases. However, there is a synchronization issue of multiple decoders in a legacy device when decoding multiple encoded atlases. This paper proposes and implements the geometry packing method for adaptive control of decoder instances for low-end and high-end devices. The proposed method on the recent version of the MIV reference software worked correctly.

Keyword : Virtual reality, MIV, 6DoF, Frame packing, Metaverse

a) 성균관대학교 컴퓨터교육학과(Department of Computer Science Education, Sungkyunkwan University)

‡ Corresponding Author : 류은석(Eun-Seok Ryu)

E-mail: esryu@skku.edu

Tel: +82-2-760-0677

ORCID: <https://orcid.org/0000-0003-4894-6105>

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-00167169, Development of Moving Robot-based Immersive Video Acquisition and Processing System in Metaverse)

· Manuscript September 7, 2022; Revised November 4, 2022; Accepted November 4, 2022

1. 서론

가상 현실 (virtual reality) 및 증강 현실 (augmented reality)의 발전에 따라, 고품질의 실사 및 컴퓨터 그래픽 영상을 가상 공간에서 감상이 가능한 메타버스 기술이 주목받고 있다. 외부 환경과는 다른 가상의 영역을 생성하여 콘텐츠를 표현하는 head-mounted display (HMD) 사용 시 고품질의 몰입형 영상을 감상할 수 있으나, 기존 2-D 영상 스트리밍 실시예에서 요구되는 영상 대비 높은 품질의 영상 전송을 요구한다. 현재 널리 사용되는 HMD인 Meta의 Oculus Quest 2의 경우 양안에 각각 2K×2K의 영상을 표현할 수 있으며, 이는 전체 영상의 일부만을 렌더링하는 HMD 특성 상 2쌍의 8K×4K 영상 전송이 필요함을 의미하여 기존 2-D 영상 스트리밍 대비 높은 수준의 기술을 요구한다^[1]. 한편, 실사를 촬영한 다수의 텍스처 (색상) 및 지오메트리 (깊이) 로 이루어진 2-D 몰입형 영상을 통해 현실 세계를 가상 세계에서 표현할 수 있고, 이를 위해 다수의 고품질 영상 전송이 선행되어야 하나 종래의 영상 압축 표준인 high-efficiency video coding (HEVC/H.265)^[2] 및 versatile video coding (VVC/H.266)^[3] 로는 대역폭 및 복호기

복잡도 측면에서의 어려움이 존재한다.

따라서, 영상 압축 표준 단체인 moving picture experts group (MPEG) 은 영상 간 중복성을 제거하여 비트율을 절감하고 잔차 영상을 추출 및 병합하여 복호기 개수를 절감하는, 기존 영상 압축 표준을 사용 가능한 MPEG immersive video (MIV) 표준화를 진행하였다^[4].

MIV 표준 부호기는 텍스처와 지오메트리로 구성된 다수의 몰입형 영상 및 카메라 매개변수를 입력받아 각 시점의 영상을 중복성 제거의 기준이 되는 기본 시점 (basic view) 및 그 외의 추가 시점 (additional view)으로 분류한다. 그림 1은 MIV 부호기에 의한 15개의 시점을 가진 몰입형 영상 부호화 과정 예시를 나타내며, 기본 시점은 빨강, 추가 시점은 파랑으로 도시되었다. 영상 간 중복되는 영역은 오직 추가 시점에서 제거되며, 상기 중복되는 영역은 기본 시점에 의해 표현된다. 이후 추가 시점에서의 잔차 영상은 직사각형의 패치 (patch) 단위로 추출되고, 아틀라스 (atlas) 라 정의되는 출력 영상에 저장된다. 기본 시점은 그 자체로 커다란 단일 패치로 분류되어 아틀라스에 저장되며, 각 패치의 위치는 visual volumetric video coding (V3C) 표준을 따르는 메타데이터로써 비트스트림의 형태로 저장된다^[5]. 그림

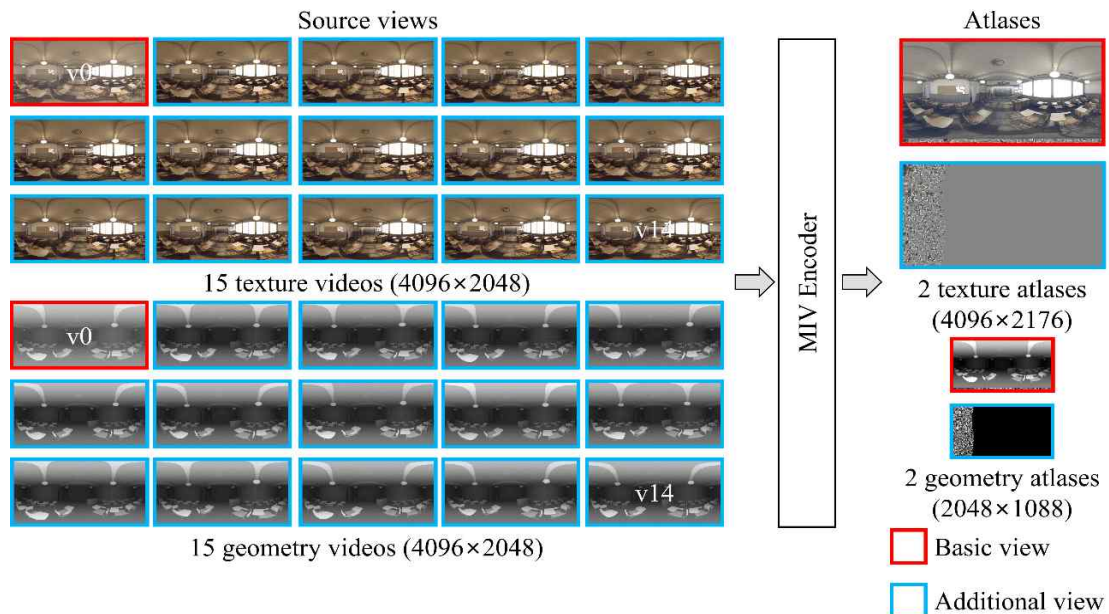


그림 1. MIV 부호기에서의 몰입형 영상 부호화 과정
Fig. 1. MIV encoder-side processes of immersive videos

1에서 아틀라스는 2개의 텍스처 지오메트리 영상으로 구성되어 총 4개의 아틀라스가 생성되고, 지오메트리 아틀라스는 픽셀 레이트 (pixel rate) 절감 및 비트율 향상을 위해 텍스처 아틀라스 대비 2×2 비율로 다운샘플링 (down-sampling) 되어 저장된다. 이후 이 아틀라스는 기존의 HEVC 또는 VVC 영상 부호기에 의해 부호화된다. 따라서 4개의 아틀라스 비트스트림이 생성되고 이는 서버에서 클라이언트로 전송되어 복호화 과정을 거치나, 실시간화를 위한 병렬 처리를 거칠 경우 복호기 간 동기화 문제가, 순차 처리를 거칠 경우 지연 시간 증가의 문제가 생기므로, 복호기 개수를 적응적으로 조절할 수 있는 기술이 필요하다. 이를 해결하기 위해 360도 영상 처리 시 사용되었던 프레임 패킹 (frame packing) 기술과 유사하게 텍스처 및 지오메트리 아틀라스에 적용하여 동일 인덱스 내 아틀라스 구성요소들을 하나로 패킹하는 프레임 패킹 기법이 제안되었다^{6,7}. 그러나, 상기 프레임 패킹 기법은 일부 HEVC 및 VVC 구현체에서 지원되지 않는 기능을 요구하여 MIV 표준화 그룹 내에서 널리 논의되고 있지 않다. 따라서, 현존하는 영상 부호화 표준 구현체와 호환되는 프레임 패킹 기법이 논의될 필요가 있다.

Abbreviations: Attr, attribute (texture); Geo, geometry.

본 논문은 MIV의 적응적 복호기 조절을 위한 geometry packing 기법을 제안하고 구현한다. 제안하는 기법은 지오메트리 영상에 대해 프레임 패킹 기법을 적용하여 HEVC의 타일 (tile), VVC의 서브픽처 (subpicture) 기반으로 부호화를 진행하고, 복호기 사양에 따라 선택적으로 텍스처 및 지오메트리 비트스트림 병합을 진행하여 복호기 개수를 적응적으로 조절한다. Geometry packing 적용 시 일부 패치의 위치 정보가 변동되므로 V3C 비트스트림 수정 기능도 필요하며, 본 논문에서 제안된 상기 프레임 패킹 및 비트스트림 수정 기능은 MIV 표준에 채택되었으며⁸⁻¹⁰, MIV의 참조 소

프트웨어인 test model for immersive video (TMIV)¹¹ 최신 버전에서 구현하여 정상적으로 동작함을 확인하였고, 검증 을 거쳐 TMIV v15.0에 최종 반영되었다.

본 논문의 구성은 다음과 같다. 2 절에서는 관련 연구로 MIV의 기존 프레임 패킹 기법을 소개한다. 3 절에서는 제안하는 geometry packing 기법을, 4 절에서는 TMIV 소프트웨어에서의 구현 내용을, 마지막으로 5 절에서는 본 논문의 결론을 서술한다.

II. 관련 연구

MIV 표준 이전에도 몰입형 영상 압축 기법에 대한 연구는 지속되어왔다. multi-view HEVC (MV-HEVC) 표준은 HEVC의 확장 표준으로 픽처 내, 픽처 간 압축 외에도 영상 간 압축을 진행하여 다시점 영상에서 기존 HEVC 대비 높은 압축률을 달성하였다¹². 그러나 종래의 HEVC 기반 가속화 기술 적용이 어려운 점, 그로 인해 실시간 구현이 어렵고 별도의 복호기가 필요하여 호환성 문제로 인해 널리 사용되기 어려운 문제가 있었다. 따라서 몰입형 영상 압축 표준의 활성화를 위해서는 기존 표준과의 호환성을 확보할 필요가 있었으며 이를 위해 MIV는 시점 간 압축을 전처리 단계에서 진행하고 영상 압축은 HEVC, VVC 등의 기존 압축 표준을 사용하여 호환성 문제를 해결하고자 하였다. 특히 MIV는 픽셀 레이트를 크게 줄여 복호기 복잡도를 감소시키고자 하였는데, MIV의 common test conditions (CTC) 는 픽셀 레이트에 따른 프로파일 (profile)을 정의하여 통상의 복호기에서의 효율을 검증하고자 하였다¹³. 표 1은 MIV CTC가 권고하는 픽셀 레이트 제한 조건을 나타낸다. Low 및 high 픽셀 레이트는 각각 VVC에서의 Main

표 1. MIV 픽셀 레이트 제한 조건
 Table 1. MIV pixel rate test condition constraints

Test conditions	Max. number of luma samples	Max. resolution of luma picture	Max. number of decoder instances
Low pixel rate constraints (Main10@Level 5.2)	1,069,547,520 (e.g., 32MP@30FPS)	8,912,896 (e.g., 4K, 4096×2048)	4
High pixel rate constraints (Main10@Level 6.2)	4,278,190,080 (e.g., 128MP@30FPS)	35,651,584 (e.g., 8K, 8192×4096)	4

10 프로파일에서의 레벨 5.2, 6.2를 준수하며, 약 4K, 8K의 아틀라스를 생성할 수 있고, 4개의 복호기를 사용 가능하다. MIV의 성능 평가는 주로 low 픽셀 레이트에서 이루어지며, 이는 생성 가능한 최대 아틀라스의 크기와 개수가 각각 4K, 4개로 제한됨을 의미한다. 그러나 4개의 아틀라스 비트스트림을 생성하여 4개의 복호기를 사용한다고 하더라도, 실시간 시스템 구현 시 다수의 복호기를 사용할 경우 복호기 간 동기화 문제가 발생함이 보고되었다^[14,15]. 따라서 MIV는 복호기 개수를 절감하기 위한 프레임 패킹 기법에 대해 논의하였고, Intel이 제안한 프레임 패킹 기법이 채택되어 TMIV에서 구현되었다^[7]. 상기 프레임 패킹 기법은 지오메트리 아틀라스를 수평 방향으로 분할 후 텍스처 아틀라스 하단에 삽입하여 아틀라스 개수를 절반으로 감소시켜 복호기 개수의 절감을 유도한다. 이때 텍스처 및 분할된 지오메트리 아틀라스는 region 단위로 구별된다. 그림 2는 MIV의 프레임 패킹 과정을 도시한다. 좌측에 도시된 프레임 패킹이 적용되지 않은 아틀라스에 대한 메타데이터는 V3C 비트스트림의 video parameter set (VPS)에 일부가 저장된다. 텍스처와 지오메트리 아틀라스가 존재하므로 이에 상응하는 선택스 (syntax) 로 vps_attribute_video_present_flag(), vps_geometry_video_present_flag() 의 값을 1로 하여 V3C 비트스트림의 VPS에 저장하면 MIV 복호기가 이를 인식하여 복호화를 진행한다. MIV에서는 색상 정보를 텍스처로 정의하나 선택스의 기반이 되는 V3C 표준은 색상 정보를 어트리뷰트 (attribute) 로 정의하므로 텍스처는

선택스 상에서 어트리뷰트로 표현된다. 프레임 패킹이 적용되었을 경우 텍스처와 지오메트리가 한 아틀라스 내에 저장되므로 상기 두 개의 값은 0이 되며, vps_packed_video_present_flag() 의 값이 1로 활성화되어 프레임 패킹된 영상이 있음을 MIV 복호기에 알릴 수 있다. 상기 값이 1일 경우 그림 2의 우측 표에 도시된 것과 같이 MIV 부호기는 packing_information() 선택스를 추가로 저장하는데, 상기 선택스에는 각 region이 표현하는 영상의 종류, 너비 및 높이, 프레임 패킹 전/후의 좌상단 좌표가 포함된다.

[7]에서 제안되어 채택된 MIV의 프레임 패킹은 V3C 선택스와 호환되도록 작성되었으나, 현존하는 영상 압축 도구를 이용한 실험에는 여러 문제가 존재한다. MIV는 여러 실험을 통해 지오메트리 아틀라스에 낮은 QP를 부여하는 것이 비트율 측면에서 효율적이라는 것을 확인하였고^[16], 이는 MIV CTC에 반영되어 있다. 따라서, 프레임 패킹 적용 시 영상 압축 도구는 같은 아틀라스 내에서 텍스처, 지오메트리 영역에 서로 다른 양자화 매개변수 (quantization parameter; QP) 를 부여할 수 있어야 한다. HEVC 참조 소프트웨어인 HEVC test model (HM)^[17] 은 상기 기능을 지원하지 않으며, HEVC의 오픈 소스 부호기인 Kvazaar^[18]의 경우 코딩 트리 유닛 (coding tree unit; CTU) 단위 QP 할당을 지원하나 실험 결과 비대칭 크기를 가지는 region에 대한 부호화 시 비트스트림 크기가 크게 상승하는 것을 확인하였다. VVC의 참조 소프트웨어인 VVC test model (VTM)^[19] 역시 HM과 마찬가지로 상기 기능을 지원하지

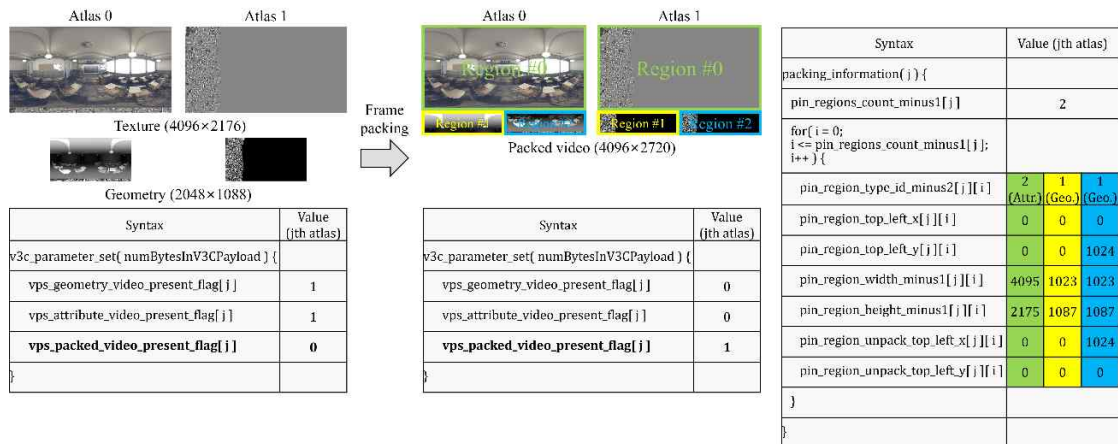


그림 2. MIV 부호기에서의 프레임 패킹 과정
Fig. 2. Frame packing process by the MIV encoder

않고, MIV CTC가 권고하는 VVC의 최적화 소프트웨어인 versatile video encoder (VVenC) 역시 상기 기능을 지원하지 않는다^[20]. 따라서, 프레임 패킹 기법에 대한 추가 실험 및 성능 탐색을 위해 표준 소프트웨어가 지원하는 프레임 패킹 기법이 필요하다. 또한, 기존 MIV가 지원하는 프레임 패킹 기법은 픽셀 수준에서 텍스처와 지오메트리 아틀라스를 병합한 후 부호화를 진행하므로 적응적 복호기 개수 조절을 위해서는 프레임 패킹이 적용되지 않은 비트스트림을 별도로 가지고 있어야 한다. 이는 서버의 연산량 및 저장공간 증가를 의미하므로 프레임 패킹 기법이 널리 사용되지 못하는 문제를 야기할 수 있다. 따라서, 복호기 적응적이지만 부호기 측면에서도 효율적인 프레임 패킹 기법 연구가 필요하며, 이는 MIV 및 V3C 표준을 준수하여야 한다.

III. 적응적 복호기 개수 조절을 위한 Geometry Packing

본 절은 적응적 복호기 개수 조절이 가능한 geometry

packing 기법의 설계 및 최신 V3C 신택스를 고려한 동작 기법에 대해 기술하며, 최신 TMIV 소프트웨어에서의 구현에 대해 설명한다. 그림 3은 TMIV 부호기 상위 다이어그램에서의 geometry packing 실시예를 나타내며, 제안하는 방법이 적용된 구성 요소는 빨간 실선 상자로 강조되었다. 몰입형 영상은 뷰 매개변수, 지오메트리 및 어트리뷰트 (텍스처) 구성 요소, 그리고 엔티티 (entity) 맵으로 구성된다. TMIV 부호기는 부호화 전에 지오메트리 품질 평가 진행 및 입력 영상들을 그룹으로 분할한다. TMIV는 서로 다른 영역을 표현하는 몰입형 영상을 시야각에 따라 그룹으로 분할하는 그룹 기반 부호화를 지원하며^[21,22] 현재 MIV CTC는 오직 하나의 그룹만을 사용할 것을 권고한다. 이후 TMIV 부호기는 뷰 합성 품질을 향상시키기 위해 인페인팅 (inpainting) 이 적용된 배경 영상을 생성하고 이는 아틀라스에 포함된다. 그 후 몰입형 영상의 각 시점은 기본 또는 추가 시점으로 구분되고, TMIV 부호기는 영상 간 중복성을 제거하는 프루닝 (pruning) 및 잔여 영상을 병합하여 아틀라스에 저장하는 패킹 (packing) 과정을 그룹 단위로 진행한다. 파라미터 셋, 뷰 파라미터 셋, 그리고 아틀라스로부터

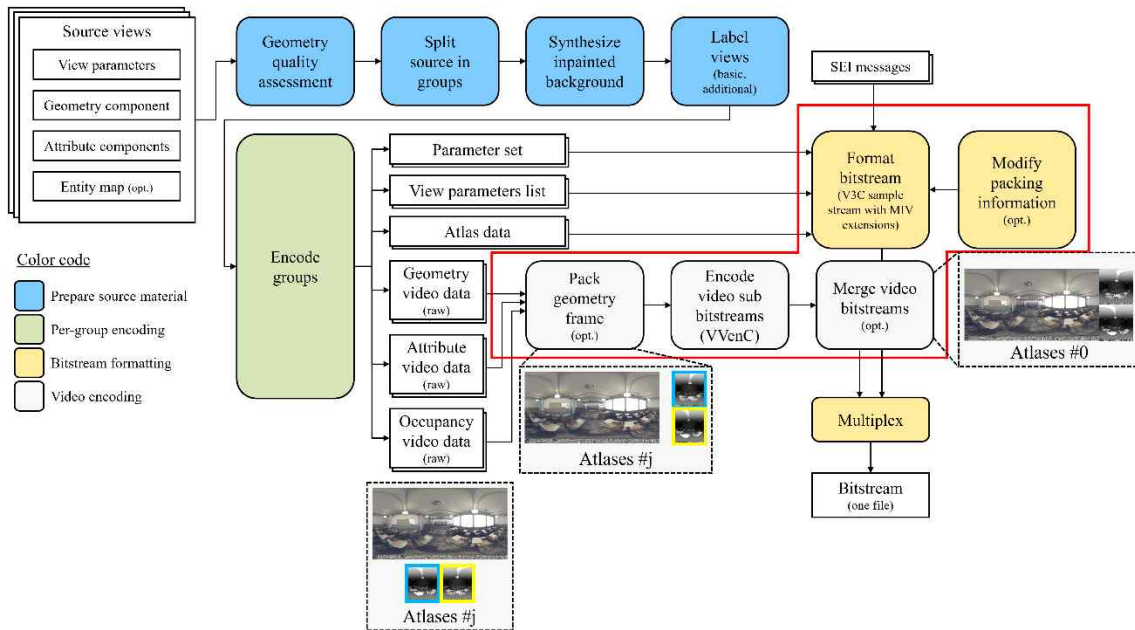


그림 3. TMIV 부호기 상위 다이어그램에서의 geometry packing 실시예. 빨간 실선 상자는 제안하는 시스템 구성 요소를 나타냄
 Fig. 3. Example of geometry packing in high-level block diagram of the TMIV encoder. A solid red box represents the proposed system components

터 원본 시점 영상 복원에 필요한 정보들을 담은 아틀라스 데이터는 MIV 확장형 V3C 샘플 비트스트림으로 출력된다. TMIV 부호기에 의해 출력되는 영상 데이터들은 지오메트리, 어트리뷰트 (텍스처), 그리고 점유 (occupancy) 영상 데이터들로 구성된다. 점유 영상 데이터는 아틀라스 내 유효영역을 구분하기 위해 사용될 수 있으며, MIV CTC는 점유 영상 데이터를 지오메트리 내에 포함시키는 것을 권고하므로 그림 3의 실시예는 지오메트리와 텍스처 아틀라스만 표현하였고, MIV CTC에 따라 지오메트리 아틀라스는 텍스처 아틀라스 대비 2x2 비율로 다운샘플링되어 저장되었다.

제안하는 geometry packing 기법을 적용 시 TMIV 부호기는 지오메트리 아틀라스를 region으로 분할 후 패킹을 진행하여 패킹된 지오메트리 아틀라스를 생성한다. 이 때 패킹된 지오메트리 아틀라스의 높이는 비트스트림 병합을 위해 텍스처 아틀라스의 높이와 같도록 설정된다. 이후 아틀라스들은 VVenC 부호기에 의해 1x1 서브픽처로 부호화되어 아틀라스 수에 대응하는 영상 비트스트림들이 생성된다. MIV CTC를 준수하여 생성된 비트스트림들의 경우 표 1에 도시된 바와 같이 통상 Main 10 프로파일에서 레벨 5.2의 복잡도를 가지는데, 상기 복잡도 이상의 연산능력을 가지고 있거나 단일 복호기만을 보유하여 다수의 복호기 간 동기화 구현이 어려운 장비들을 위해 제안하는 기법은 비트스트림 병합을 진행할 수 있다. 제안하는 기법은 비트스트림 영역에서 텍스처 아틀라스 비트스트림 우측에 패킹되어 부호화된 지오메트리 비트스트림을 위치시켜 비트스트림 병합을 진행한다. 비트스트림 병합은 VTM의 SubpicMergeApp 소프트웨어를 통해 실행 가능하다^[23]. 비트스트림 병합이 완료되면 아틀라스 수만큼의 비트스트림이 생성되고, 이후 영상 복호기를 통해 복호화된 픽처들을 기반으로 TMIV 부호기가 뷰 합성을 진행한다. 아틀라스 비트스트림 병합 시 텍스처와 지오메트리 아틀라스가 한 픽처 내에 위치하므로, 아틀라스 및 패킹 정보들을 담고 있는 V3C 비트스트림 정보 역시 변경되어야 한다. 아틀라스 비트스트림 병합이 진행되면 TMIV 부호기는 V3C 비트스트림 복호화 후 패킹 정보들을 병합된 아틀라스 비트스트림 내 픽처에 맞게 변경하여 변경된 V3C 비트스트림을 생성한다. 이후 아틀라스 비트스트림과 V3C 비트스트림은 멀티플렉싱 (multiplexing) 되어 하나의 비트스트림으로 통합된다. 만약 클라

이언트 측 복호기가 상기 동기화 문제가 없다면 제안하는 기법은 아틀라스 비트스트림 병합 및 V3C 비트스트림 수정을 진행하지 않고 텍스처 및 패킹된 지오메트리 아틀라스와 V3C 비트스트림만을 멀티플렉싱하여 클라이언트 측으로 전송하고, 이후 클라이언트 내 TMIV 복호기는 디멀티플렉싱 (demultiplexing), 영상 복호화, 뷰 합성을 진행하여 사용자 시점에 대응하는 가상 시점 영상을 생성하게 된다.

제안하는 geometry packing 기법은 현존하는 영상 압축 도구에서 문제없이 동작하도록 설계되어 HEVC 구현체인 Kvazaar, VVC 구현체인 VTM, VVenC에서의 구동이 가능함을 확인하였다. 비트스트림 추출기 및 병합기를 모두 제공하는 VTM과 달리 HM은 비트스트림 추출기만을 제공하기에, 성균관대학교 연구팀은 HEVC 타일 병합기를 개발 및 구현하여 제공한다^[24]. HEVC 참조 소프트웨어인 HM은 1x1 타일 부호화를 지원하지 않아 본 논문이 제안하는 비트스트림 병합 기법 사용이 불가하나, HEVC 오픈 소스 부호기인 Kvazaar를 통한 실험은 가능함을 확인하였다. HM의 타일 추출기는 타일 정보가 저장된 supplemental enhance information (SEI) 메시지를 해석하여 타일 추출을 진행하고, HM을 통한 타일 기반 영상 부호화 시 상기 SEI 메시지가 생성되나, Kvazaar를 이용한 영상 부호화 시 상기 SEI 메시지가 생성되지 않아 HM이 제공하는 타일 추출기를 이용한 타일 추출이 불가하다. 따라서 성균관대학교 연구팀은 SEI 메시지를 해석하지 않고 각 타일 및 슬라이스 (slice)를 해석하여 타일을 추출하는 타일 추출기를 제공한다^[25].

제안하는 기법은 복호기 수를 줄이기 위한 프레임 패킹을 픽셀 영역이 아닌 비트스트림 영역에서 진행하여 서비스 제공자 입장에서 기존 MIV CTC가 권고하는 프로파일 및 프레임 패킹 기능을 둘 다 제공 시 각각에 대한 별도의 비트스트림을 생성할 필요가 없으므로 연산량 및 저장 공간 측면에서 효율적이다. 아틀라스 비트스트림 병합 시 추가적인 연산량이 소모되나, 비트스트림 기반 연산이므로 매우 적은 연산량 증가만을 확인하였다.

IV. 최신 TMIV 소프트웨어 내 구현

본 절은 제안하는 geometry packing 기법의 최신 TMIV

소프트웨어 내 구현에 대해 설명한다. 제안하는 기법은 TMIV v14 버전을 기준으로 구현되었다. 그림 3에 도시된 TMIV 부호기는 부호화된 아틀라스 비트스트림과 V3C 비트스트림을 멀티플렉싱하여 하나의 비트스트림을 생성하고, 이는 TMIV 복호기에 직접 입력되어 복호화될 수 있다. 이를 TMIV는 in-band라 정의한다. 반면, 멀티플렉싱 과정을 거치지 않고 아틀라스 비트스트림을 복호화하여 복호화된 픽처와 V3C 비트스트림을 TMIV 복호기에 입력해줄 수 있는데, 이를 TMIV는 out-of-band라 정의한다. 최신 버전

의 TMIV는 out-of-band 메타데이터를 json 형태로 생성하여 아틀라스 구성 요소를 설명하고, 상기 메타데이터는 TMIV 복호기에 입력되어야 한다. 그림 4는 0번째 아틀라스와 out-of-band 메타데이터 예시를 나타낸다. 그림 4에서 알 수 있듯 out-of-band 메타데이터는 아틀라스 픽처 크기, 아틀라스 인덱스, 유닛 종류를 나타낸다. 메타데이터에서의 vuh는 V3C unit header를 의미하고, vuh_unit_type은 vuh의 종류를 나타내며 이는 ISO/IEC 23090-5 V3C 표준에 정의되어 있다. 표 2는 V3C 유닛의 종류를 도시한다.

표 2. V3C 유닛 종류
 Table 2. V3C unit types

vuh_unit_type	Identifier	V3C unit type	Description
0	V3C_VPS	V3C parameter set	V3C level parameters
1	V3C_AD	Atlas data	Atlas information
2	V3C_OVD	Occupancy video data	Occupancy information
3	V3C_GVD	Geometry video data	Geometry information
4	V3C_AVD	Attribute video data	Attribute information
5	V3C_PVD	Packed video data	Packing information
6	V3C_CAD	Common atlas data	Information that is common for atlases in a CVS. Specified in ISO/IEC 23090-12
7...31	V3C_RSVD	Reserved	

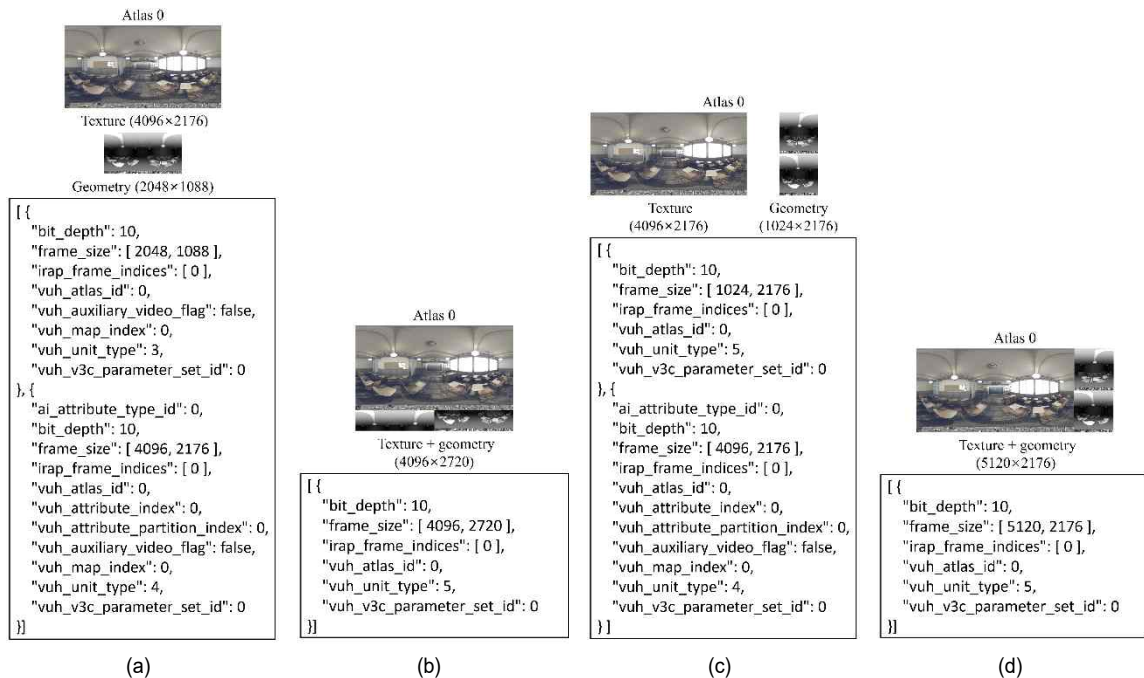


그림 4. 0번째 아틀라스 및 out-of-band 메타데이터 예시. (a) MIV 메인 프로파일, (b) MIV 프레임 패킹, (c) geometry packing, (d) geometry packing + 비트스트림 병합

Fig. 4. Example of 0th atlas and out-of-band metadata. (a) MIV main profile, (b) MIV frame packing, (c) geometry packing, (d) geometry packing + bitstream merging

현재 MIV는 0~6 사이의 값을 사용하며, out-of-band 메타데이터의 vuh_unit_type 값이 3일 경우 해당 구역은 지오메트리 아틀라스를, 4일 경우 텍스처 (어트리뷰트) 아틀라스를, 5일 경우 패킹된 아틀라스를 의미한다. 그림 4(a)는 MIV CTC가 권고하는 MIV 메인 프로파일 결과물을 나타내는데, 텍스처 및 지오메트리 아틀라스가 따로 존재하므로 out-of-band 메타데이터도 각각 존재함을 볼 수 있다. 반면 그림 4(b)는 MIV 프레임 패킹 실행 예시를 나타내고, 텍스처와 지오메트리 아틀라스가 한 픽처 내에 존재하므로 vuh_unit_type은 5이며 픽처 크기 역시 그에 대응하여 작성되었다. 그림 4(c)는 제안하는 geometry packing 기법만을 적용한 예시를 나타낸다. 지오메트리 아틀라스에 대해서만 프레임 패킹이 적용되었으므로 텍스처와 지오메트리 아틀라스 각각에 out-of-band 메타데이터가 정의되고, 두 픽처의 높이는 동일한 것을 확인할 수 있다. 그림 4(d)는 geometry packing 및 비트스트림 병합을 적용한 예시를 나타낸다. 비트스트림 병합 후 복호화된 픽처 내에는 텍스처와 지오메트리가 존재하므로 out-of-band 메타데이터의 vuh_unit_type 값은 5로 패킹된 아틀라스에 대한 설명이 메타데이터에 기록되었다. 상기 비트스트림 병합은 VTM의 SubpicMergeApp을 통해 가능하고, 그림 4(d)의 실시예를 위해 본 논문은 최신 버전의 TMIV 내에 BitstreamMerger라는 프로젝트를 생성 및 구현하였다. 상기 프로젝트는 개별 실행 파일을 생성하고 이를 통해 geometry packing 후 생성된 out-of-band 메타데이터 및 V3C 비트스트림을 입력받아 병합된 아틀라스 비트스트림에 대응하는 메타데이터 및 V3C 비트스트림을 출력하였다. TMIV 복호기는 프레임 패킹된 아틀라스를 입력받으면 V3C 비트스트림의 packing_information() 실택스를 참조하여 프레임 언패킹 작업을 수행 후 뷰 합성을 진행한다. 이 때 패킹된 지오메트리 아틀라스는 V3C_PVD, 패킹되지 않은 텍스처 아틀라스는 V3C_

AVD로 인식되므로 MIV 복호기 단 수정 없이 복호화가 가능하였다. 복호기 단이 수정되지 않았으므로 제안하는 기법은 TMIV의 in-band 및 out-of-band 모두에서 동작이 가능함을 확인하였다. 제안하는 기법은 프레임 패킹이 적용되는 MIV 확장 프로파일에 적용되므로 MIV 메인 프로파일에 대한 연산량 증가를 야기하지 않고, 프레임 패킹 적용 시에도 복호기 수정 없이 실행이 가능하였으며 실행 시간의 미미한 차이만이 관찰되었고 V3C 및 MIV 표준을 준수하였다.

[7]에서 제안하는 프레임 패킹 및 본 논문이 제안하는 geometry packing 기법은 V3C 비트스트림에 packing_information() 실택스를 추가로 저장하여, 프레임 패킹 미 적용 대비 V3C 비트스트림의 크기가 증가한다. V3C 비트스트림은 아틀라스로부터 가상 시점을 생성 시 반드시 필요한 정보를 포함하여 MIV 기반 스트리밍 시 반드시 전송되어야 하므로, 본 논문은 각 패킹 기법에 대한 V3C 비트스트림의 비트율을 산출 및 비교하였다. 표 3은 본 논문에서 실험한 4가지 조건을 나타낸다. A17은 프레임 패킹을 적용하지 않은 MIV mode를 사용하여 17 프레임을 부호화한 조건을 의미한다. MIV 부호기는 MIV mode 및 MIV view mode를 지원하며, MIV mode에서는 시점 간 중복성이 제거되어 높은 압축률을 확보할 수 있으나 가상 시점 합성 시 패치 단위로 연산이 이루어져 복잡도가 높다. MIV view mode는 입력 영상에서 기본 시점만을 추출하여 아틀라스에 저장하여 복잡도는 낮으나 압축률은 MIV mode에 비해 낮다. [26]에 따르면 MIV mode는 MIV view mode 대비 평균 454.29%, 175.48%의 MIV 부호화 및 복호화 실행 시간 증가를 보이나, 몰입형 영상 품질 평가를 위해 고안된 immersive video peak signal-to-noise ratio (IV-PSNR)^[27]에서 TMIV와 HM 병행 사용 시 11.48%의 Bjontegaard-elta rate (BD-rate) 절감을 보여주었다. 따라서, 현재 MIV

표 3. 실험 조건

Table 3. Experimental Conditions

Condition	Description	Profile	Components within an atlas index
A17	MIV mode anchor for 17 frames	MIV main	AVD, GVD
P17	Frame packing in MIV mode for 17 frames	MIV extended	PVD containing AVD and GVD
GP17	Geometry packing in MIV mode for 17 frames	MIV extended	AVD and PVD containing GVD
GPM17	Geometry packing and bitstream merging in MIV mode for 17 frames	MIV extended	PVD containing AVD and GVD

CTC는 MIV mode를 대조군 (anchor) 으로 설정하여 17 프레임에 대해 실험할 것을 권고하며 본 실험에서도 MIV mode를 17 프레임에 대해 적용하였다. A17에서는 별도의 프레임 패킹 기술이 적용되지 않으므로 텍스처를 저장하는 attribute video data (AVD) 와 지오메트리를 저장하는 geometry video data (GVD) 가 따로 존재한다. P17은 [7]이 제안하여 TMIV에 채택된 프레임 패킹 기법을 MIV mode에서 17프레임에 대해 적용한 조건을 의미하며, AVD와 GVD가 포함된, 프레임 패킹된 아틀라스를 저장하는 packed video data (PVD) 형태의 영상만 포함한다. GP17은 본 논문이 제안하는 geometry packing 기법을 MIV mode에서 17 프레임에 대해 적용한 조건을 의미한다. GP17 조건에서는 텍스처에 대한 패킹이 이루어지지 않고, 지오메트리에 대한 패킹만 이루어지므로 동일 아틀라스 인덱스에서 하나의 AVD와 GVD를 포함하는 하나의 PVD가 포함된다. GPM17은 GP17에 비트스트림 병합을 적용하여 AVD와 PVD를 비트스트림 영역에서 합쳤으므로, 동일 아틀라스 인덱스에서 하나의 PVD가 존재하고, 상기 PVD는 AVD와 GVD를 포함한다.

MIV CTC는 16개의 서로 다른 특성을 가진 테스트 시퀀스에 대해 실험할 것을 권장하고, 그 중 9개의 대표 테스트 시퀀스에 대한 실험 결과는 표준화 기고 시 필수로 제출해야 한다. 본 실험은 상기 9개의 테스트 시퀀스인 ClassroomVideo, Museum, Painter, Frog, Kitchen, Chess, Fan, Carpark, Group에 대해 진행되었다. 표 4는 MIV 부호화 후 생성된 V3C 비트스트림의 비트레이트를 나타낸다. GPM17

의 비트레이트는 GP17 조건에서 생성된 V3C 비트스트림을 BitstreamMerger에 입력하여 출력된 V3C 비트스트림을 대상으로 산출되었다. 프레임 패킹이 적용된 P17, GP17, GPM17은 프레임 패킹이 미적용된 A17 대비 packing_information() 신택스가 추가로 저장되어 약 1Kbps가량의 적은 비트레이트 상승이 관찰되었다. 따라서, 프레임 패킹 및 제안하는 geometry packing이 적용되어도 V3C 비트스트림 비트레이트 상승은 작거나 없는 수준이라는 것을 확인하였고, 이는 제안하는 기법이 기존 프레임 패킹 미적용 대비 복잡도 증가 역시 미미한 수준임을 의미한다. 표 5는 각 조건에 대한 MIV 부호화 시간을 나타낸다. 실험은 i7-17000 CPU (8 cores, 16 threads), 64GB memory가 장착된 Windows 10 PC에서 진행되었다. GPM17은 GP17에 비트스트림 병합만 적용하고 아틀라스 자체는 GP17 조건에서 생성된 것을 사용하므로 GPM17에 대한 MIV 부호화 시간은 도시되지 않았다. 비트레이트와 마찬가지로 MIV 부호화 실행 시간 역시 각 모드 별 큰 차이가 없었고, 따라서 제안하는 geometry packing 기법이 기존 기법과 비슷한 복잡도를 가지므로 널리 사용될 수 있음을 확인하였다. 또한 [7]에서 제안하는 프레임 패킹 기법은 픽셀 수준에서의 패킹만을 적용하여 고정된 수의 아틀라스 및 비트스트림을 제공하는 반면, 본 논문이 제안하는 geometry packing 기법은 매우 경량화된 연산을 요구하는 비트스트림 병합을 통해 아틀라스 및 비트스트림 수를 조절할 수 있어 몰입형 영상 스트리밍 서비스 제공 시 적은 자원으로 적응적 스트리밍 서비스를 제공할 수 있다.

표 4. V3C 비트스트림 비트레이트 (Kbps)
 Table 4. Bitrates of V3C bitstreams (Kbps)

Sequence	A17	P17	GP17	GPM17
<i>ClassroomVideo</i>	236.27	237.55	237.14	237.50
<i>Museum</i>	151.51	152.79	152.38	152.73
<i>Painter</i>	264.55	265.83	265.42	265.77
<i>Frog</i>	208.61	209.90	209.49	209.84
<i>Kitchen</i>	274.13	275.42	275.01	275.36
<i>Chess</i>	138.00	139.28	138.87	139.22
<i>Fan</i>	202.85	204.14	203.73	204.08
<i>Carpark</i>	52.18	53.25	52.91	53.21
<i>Group</i>	223.55	224.83	224.42	224.78
<i>Average</i>	194.62	195.88	195.48	195.83

표 5. MIV 부호화 시간 (초)
 Table 5. ntimes of MIV encoding (seconds)

Sequence	A17	P17	GP17	GPM17
<i>ClassroomVideo</i>	4053.62	4134.12	4245.95	-
<i>Museum</i>	4633.15	4696.66	4640.76	-
<i>Painter</i>	1024.71	1084.07	1049.23	-
<i>Frog</i>	853.232	789.99	793.66	-
<i>Kitchen</i>	1811.91	1795.14	1737.92	-
<i>Chess</i>	1792.32	1703.61	1712.62	-
<i>Fan</i>	1057.66	1065.12	975.04	-
<i>Carpark</i>	539.84	501.66	500.65	-
<i>Group</i>	1664.13	1532.05	1554.94	-
<i>Average</i>	1936.73	1922.49	1912.30	-

V. 결 론

본 논문은 복호기 인스턴스 개수를 적응적으로 절감하기 위한 MIV에서의 geometry packing 기법을 제안하고 이를 최신 버전의 TMIV에서 구현하였다. 제안하는 geometry packing 기법은 지오메트리 아틀라스를 수직 방향으로 분할하고 프레임 패킹하여 텍스처 아틀라스와 같은 높이를 가지도록 한 뒤, 개별 아틀라스에 대해 타일 또는 서브픽처 기반 부호화를 진행하였다. 이후 선택적으로 아틀라스 비트스트림 병합을 진행하여 복호기 개수를 절감하고, V3C 비트스트림, out-of-band 메타데이터 수정을 진행하여 최신 TMIV 소프트웨어에서 문제없이 동작하도록 하였다. 만약 클라이언트 단에서 복호기 동기화 문제가 없다면 비트스트림 병합을 진행하지 않은 아틀라스 비트스트림 및 V3C 비트스트림만을 전송하여, 프레임 패킹을 위해 별도의 비트스트림을 생성해야 했던 기존 기법 대비 연산량 및 저장 공간을 절감할 수 있다. 제안하는 기법에 의해 생성된 아틀라스들은 기존 TMIV 복호기에 의해 문제없이 복호화됨을 확인하였다. 향후 연구로는 추가적인 복호기 개수 절감을 위해 서로 다른 인덱스를 가진 아틀라스 간 프레임 패킹 기법을 연구 및 구현할 계획이다.

참 고 문 헌 (References)

- [1] Oculus Quest 2 Specification, [Online]. Available at: <https://developer.oculus.com/resources/oculus-device-specs/>.
- [2] G. J. Sullivan, J. -R. Ohm, W. -J. Han, T. Wiegand. 2012. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649-1668. doi: <https://doi.org/10.1109/TCSVT.2012.2221191>
- [3] B. Bross, Y. -K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, J. -R. Ohm. 2021. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 10, pp. 3736-3764. doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [4] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, L. Yu. 2021. MPEG immersive video coding standard. *Proceedings of the IEEE*, Vol. 109, No. 9, pp. 1521-1536. doi: <https://doi.org/10.1109/JPROC.2021.3062590>
- [5] ISO/IEC JTC 1/SC 29/WG 7. 2021. Text of ISO/IEC DIS 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition. Standard ISO/IEC JTC 1/SC 29 WG 7, MPEG/n0188.
- [6] P. Hanhart, X. Xiu, Y. He, Y. Ye. 2019. 360 video coding based on projection format adaptation and spherical neighboring relationship. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 9, No. 1, pp. 71-83. doi: <https://doi.org/10.1109/JETCAS.2018.2888960>
- [7] B. Salahieh, G. Naf, J. Boyce. 2021. "Frame Packing Implementation in TMIV." Standard ISO/IEC JTC 1/SC 29/WG 4, MPEG/m56827.
- [8] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. "Sub bitstream packing based lightweight tiled streaming for 6 degree of freedom immersive video". *Electronics Letters*, Vol. 57, No. 25, pp. 973-976. doi: <https://doi.org/10.1049/ell2.12329>
- [9] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. "DWS-BEAM: Decoder-Wise Subpicture Bitstream Extracting and Merging for MPEG Immersive Video". *International Conference on Visual Communications and Image Processing 2021 (VCIP2021)*, pp. 1-5. doi: <https://doi.org/10.1109/VCIP53242.2021.9675419>
- [10] J. -B. Jeong, S. Lee, E. -S. Ryu. 2022. "[MIV] Geometry Packing Implementation in TMIV for Frame Packed Video". Standard ISO/IEC JTC 1/SC 29 WG 4, MPEG/m59442.
- [11] B. Salahieh, J. Jung, A. Dziembowski. 2022. Test Model 14 for MPEG immersive video. Standard ISO/IEC JTC 1/SC 29 WG 4, MPEG/n0242.
- [12] M. M. Hannuksela, Y. Yan, X. Huang, H. Li. 2015. Overview of the multiview high efficiency video coding (MV-HEVC) standard. *IEEE International Conference on Image Processing 2015 (ICIP2015)*, pp. 2154-2158. doi: <https://doi.org/10.1109/ICIP.2015.7351182>
- [13] J. Jung, B. Kroon. 2022. Common Test Conditions for MPEG Immersive Video. Standard ISO/IEC JTC 1/SC 29/WG 4, MPEG/n0232.
- [14] J. Fleureau, F. Thudor, T. Tapie, R. Gendrot, G. Briand, R. Doré. 2021. [MIV] MIV / MPI on Android. Standard ISO/IEC JTC 1/SC 29/WG 4, MPEG/m56727.
- [15] J. Boyce, B. Salahieh. 2020. Frame Packing and Extended MIV Profile. Standard ISO/IEC JTC 1/SC 29/WG 11, MPEG/m54491.
- [16] B. Wang, L. Yu, B. Kroon, and J. Jung. 2018. [MPEG-I Visual] Results on depth QPs in CTC of 3DoF+ Video. Standard ISO/IEC JTC 1/SC 29/WG 11, MPEG/m44688.
- [17] HEVC test model (HM), [Online]. Available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.
- [18] M. Viitanen, A. Koivula, A. Lemmetti, A. Y. Outinen, J. Vanne, T. D. Hämäläinen. 2016. Kvazaar: open-source HEVC/H. 265 encoder. *Proceedings of the 24th ACM international conference on Multimedia (ACM MM)*, pp. 1179-1182. doi: <http://dx.doi.org/10.1145/2964284.2973796>
- [19] VVC test model (VTM), [Online]. Available at: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM.
- [20] A. Wieckowski, J. Brandenburg, T. Hinz, C. Bartnik, V. George, G. Hege, C. Helmrich, A. Henkel, C. Lehmann, C. Stoffers, I. Zupancic,

- B. Bross, D. Marpe. 2021. VVenC: An open and optimized VVC encoder implementation. Proceedings of 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1 - 2. doi: <https://doi.org/10.1109/ICMEW53276.2021.9455944>
- [21] B. Salahieh, S. Bhatia, J. Boyce. 2019. Grouping Implementation in TMIV. Standard ISO/IEC JTC 1/SC 29/WG 11, MPEG/m49859.
- [22] S. Lee, J. -B. Jeong, E. -S. Ryu. 2022. Group-based Adaptive Rendering for 6DoF Immersive Video Streaming. Journal of Broadcast Engineering, Vol. 27, No. 2, pp. 216-227. doi: <https://doi.org/10.5909/JBE.2022.27.2.216>
- [23] A. Hallapuro, M. M. Hannuksela. 2020. AHG3/AHG12: Subpicture merging software. Standard ITU-T SG 16 WP3 ISO/IEC JTC 1/SC 29/WG 11, JVETS0162.
- [24] HEVC Tile Merger, [Online]. Available at: <https://gitlab.com/MCSLJeong/hevctilemerger>.
- [25] Kvazaar-compliant HEVC Tile Extractor, [Online]. Available at: <https://gitlab.com/MCSLJeong/hevctileextractor>.
- [26] J. -B. Jeong, S. Lee, E. -S. Ryu. 2022. Performance Analysis of 6DoF Video Streaming Based on MPEG Immersive Video. Journal of Broadcast Engineering (JBE), Vol. 27, No. 5, pp. 773-793. doi: <https://doi.org/10.5909/JBE.2022.27.5.773>
- [27] A. Dziembowski, D. Mieloch, J. Stankowski, A. Grzelka. 2022. IV-PSNR - The objective quality metric for immersive video applications. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 32, No. 11, pp. 7575-7591. doi: <https://doi.org/10.1109/TCSVT.2022.3179575>

저 자 소 개



정 종 범

- 2018년 8월 : 가천대학교 컴퓨터공학과 학사
- 2018년 9월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육학과 석박통합과정
- 2020년 1월 ~ 2020년 3월 : University of California, Santa Barbara 방문연구원
- 2021년 8월 ~ 2022년 1월 : Purdue University 방문연구원
- 2022년 9월 ~ 현재 : 성균관대학교 글로벌융합학부 강사
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준, MPEG immersive video, video-based dynamic mesh coding



이 순 빈

- 2020년 3월 : 가천대학교 컴퓨터공학과 학사
- 2020년 3월 ~ 2022년 3월 : 성균관대학교 컴퓨터교육과 석사
- 2022년 3월 ~ 현재 : 성균관대학교 컴퓨터교육과 박사과정
- ORCID : <https://orcid.org/0000-0002-8951-0335>
- 관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준, MPEG immersive video, Deep learning



류 은 석

- 1999년 8월 : 고려대학교 컴퓨터학과 학사
- 2001년 8월 : 고려대학교 컴퓨터학과 석사
- 2008년 2월 : 고려대학교 컴퓨터학과 박사
- 2008년 3월 ~ 2008년 8월 : 고려대학교 연구교수
- 2008년 9월 ~ 2010년 12월 : 조지아공대 박사후과정
- 2011년 1월 ~ 2014년 2월 : InterDigital Labs Staff Engineer
- 2014년 3월 ~ 2015년 2월 : 삼성전자 수석연구원/파트장
- 2015년 3월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 조교수
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 부교수
- ORCID : <https://orcid.org/0000-0003-4894-6105>
- 관심분야 : 멀티미디어 통신 및 시스템, 비디오 코딩 및 국제 표준, HMD/VR 응용분야