

일반논문 (Regular Paper)

방송공학회논문지 제27권 제5호, 2022년 9월 (JBE Vol.27, No.5, September 2022)

<https://doi.org/10.5909/JBE.2022.27.5.773>

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

MPEG 몰입형 비디오 기반 6DoF 영상 스트리밍 성능 분석

정 중 범^{a)}, 이 순 빈^{a)}, 김 인 애^{a)}, 류 은 석^{a)‡}

Performance Analysis of 6DoF Video Streaming Based on MPEG Immersive Video

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

요 약

다수의 고품질 몰입형 영상 전송을 통해 가상 현실에서 six degrees of freedom(6DoF)를 지원하기 위해 the moving picture experts group (MPEG) immersive video (MIV) 압축 표준이 설립되었다. MIV는 비트율과 연산 복잡도 간 상충관계를 고려하여 1) 시점 간 연관성 제거 또는 2) 대표 시점을 선택하여 전송하는 2가지 압축 방식을 제공한다. 본 논문은 전송한 두 가지 방식에 대해 high-efficiency video coding (HEVC), versatile video coding (VVC) 기반 성능 분석 결과를 입력 영상 위치에 합성한 가상 영상 및 사용자 시점 영상 중심으로 제시한다.

Abstract

The moving picture experts group (MPEG) immersive video (MIV) coding standard has been established to support six degrees of freedom (6DoF) in virtual reality (VR) by transmitting high-quality multiple immersive videos. The MIV exploits two approaches considering tradeoff between bandwidth and computational complexity: 1) eliminating correlation between multi-view videos or 2) selecting representative videos. This paper presents performance analysis on intermediate synthesized views on source view positions and synthesized pose traces using high-efficiency video coding (HEVC) and versatile video coding (VVC) for above-mentioned two approaches.

Keyword : Virtual reality, MIV, 6DoF, HEVC, VVC, Immersive video, 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 the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1F1A1074935). This research was also supported by the SungKyunKwan University and the BK21 FOUR(Graduate School Innovation) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF).

· Manuscript June 29, 2022; Revised August 16, 2022; Accepted August 16, 2022.

1. 서론

최근 오кул러스 퀘스트 2, 밸브 인덱스, HTC 바이브 등의 고품질 영상을 제공하는 head-mounted display (HMD)의 출시에 힘입어, 몰입형 virtual reality (VR)에 대한 관심이 높아지고 있다. 외부 환경과 단절되어 사용자의 field of view (FoV)를 전부 표현하는 HMD의 특성 상 초고품질의 영상이 제공되지 않을 경우 어지럼증 및 불쾌감을 유발할 수 있다. HMD를 통한 영상 시청 시 몰입감 극대화 및 불편감 최소화를 위해서는 최소 12K, 90 frames per second (FPS) 이상의 초고품질 영상이 요구되며 20ms 이하의 motion-to-photon (MTP) 지연 시간이 요구된다. 현재 널리 쓰이는 HMD는 대부분 전방위 360도 영상 수신 및 렌더링을 지원하나, 사용자의 FoV는 약 110도 정도이므로 사용자는 전송된 영상의 일부만을 한 번에 시청 가능하며, 이는 전송된 영상과 렌더링된 영상의 해상도가 같은 종래의 2차원 디스플레이 대비 더 높은 해상도의 영상 전송을 요구한다. 따라서, 영상을 직사각형의 타일 단위로 부호화, 추출, 복호화가 가능하게 하는 motion-constrained tile set (MCTS) 기술 기반으로 부호화하여 사용자 관심 영역 중심으로 스트리밍하는 연구가 제안되었다^{[2][6]}. 한편, 현재 널리 상용화된 HMD를 통해 영상을 시청할 경우 통상 단일 위치에서 취득된 영상을 보게 되므로 사용자의 머리 회전만 지원하고 몸의 움직임은 미지원하는 three degrees of freedom (3DoF)이 사용자에게 부여되고, 제한된 자유도로 인해 사용자의 quality of experience (QoE)가 저하되는 문제가 발생한다. 이를 극복하기 위해 컴퓨터 그래픽 환경 또는 자연 환경에서 다수의 카메라로부터 취득된 몰입형 영상 압축 및 전송 방식을 사용하여 사용자가 가상 현실 환경에서 자유롭게 움직일 수 있도록 하는 six degrees of freedom (6DoF)를 지원할 수 있다. 종래에 널리 쓰이던 영상 압축 도구인 advanced video coding (H.264/AVC), high-efficiency video coding (H.265/HEVC)는 단일 영상 압축을 목표로 개발되어, 상기 몰입형 영상 압축 시 비트율 및 복호기 인스턴스 (instance) 개수가 매우 높아지는 문제가 있다. 따라서 영상 압축 표준화 기관인 moving picture experts group (MPEG)은 기존 영상 압축 표준의 확장 표준으로 multiview video coding (MVC)^[7] 및 multi-view HEVC

(MV-HEVC)^[8] 표준을 개발하여 기존 표준 대비 비트율 절약이 가능했으나, 기존 표준에 대한 종속성, 실시간 가속화가 어렵다는 이유로 널리 사용되지 않는 문제가 있었다. 이 문제를 해결하기 위해 픽셀 영역에서 시점 간 중복성을 제거하여 잔차 영상을 병합하는 방식으로 비트율과 픽셀 수를 모두 절약하는 연구들이 제안되었다^{[9][10]}. MPEG-Immersive (MPEG-I) 표준화 그룹은 영상 압축 도구에 독립적이면서 비트율 및 픽셀 수, 복호기 인스턴스 수 절감을 가능하게 하는 몰입형 영상 압축 개발을 시작하였고, 2019년 1월 call for proposal (CfP)을 발간하여 기술 공모를 진행하였으며^[11], 기고된 기술들을 종합하고 발전시켜^{[12][16]} 2021년 7월 MPEG immersive video (MIV)의 final draft international standard (FDIS)가 발간되었다^[17].

MIV 표준은 색상 정보를 나타내는 텍스처 (texture)와 거리 정보를 나타내는 지오메트리 (geometry) 쌍으로 구성된 다수의 2-D 영상 및 각 영상을 취득한 카메라에 대한 매개변수를 기반으로 몰입형 영상 부호화 및 복호화를 진행한다. MIV는 MIV 모드 (MIV mode), MIV 시점 모드 (MIV view mode) 2가지 압축 방식을 제공한다. MIV 모드는 각 시점 영상을 기본 시점 (basic view) 및 추가 시점 (additional view)으로 분류하고, 픽셀 영역에서 시점 간 중복된 영역을 추가 시점에서 제거하며, 잔차 영상을 직사각형 형태의 패치 (patch)로 추출한 후 아틀라스 (atlas)라 불리는 출력 영상에 저장하는 압축 방식을 제공한다. 반면, MIV 시점 모드는 중복성 제거 과정 없이 입력 영상을 대표하는 영상들을 기본 시점으로 선택하여 아틀라스에 저장한다. 2차원 영상 형태로 저장된 아틀라스는 기존의 2-D 영상 압축 도구를 통해 부호화가 가능하다. 현재 MIV의 기본 압축 방식은 MIV 모드이고 이를 중심으로 성능 평가가 이루어지고 있으나, 영상 복호화 외에도 패치 기반 영상 복원이 필요한 MIV 특성상 저복잡도 6DoF 스트리밍 시스템에서는 MIV 시점 모드 역시 고려되어야 한다.

본 논문은 ISO/IEC 23090 Part 12에서 표준화가 진행중인 MIV 기반 6DoF 몰입형 영상 스트리밍 성능 분석 결과를 소개한다. 그림 1은 MIV 기반 6DoF 몰입형 영상 스트리밍 시스템을 나타낸다. 몰입형 영상이 입력되면 MIV 부호기는 MIV 모드, MIV 시점 모드에 대해 각각 2개의 텍스처 아틀라스 및 상응하는 지오메트리 아틀라스를 생성하고,

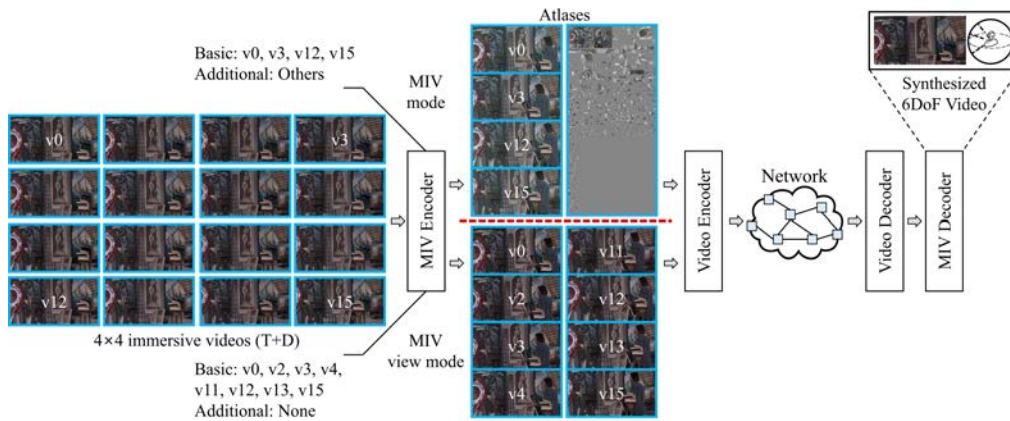


그림 1. MIV 기반 6DoF 몰입형 영상 스트리밍 시스템 구조
 Fig. 1. Conceptual diagram of the MIV-based 6DoF immersive video streaming system

아틀라스에 대한 정보를 포함하는 구문 (syntax) 및 의미론 (semantics) 을 표현하는 MIV 비트스트림이 생성된다. 생성된 아틀라스는 AVC, HEVC 및 차세대 영상 압축 도구인 versatile video coding (VVC)^{[18][19]}에 의해 부호화되어 비트스트림 형태로 저장된다. 클라이언트는 서버로부터 전송 받은 비트스트림을 복호화하고, MIV 복호기를 통해 사용자 시점 영상을 생성하여 몰입형 영상을 화면에 렌더링 (rendering)한다.

본 논문의 구성은 다음과 같다. 2 절에서는 배경 연구로

MIV 표준 기반 스트리밍 및 렌더링 연구를 소개한다. 3 절에서는 테스트 시퀀스 및 성능 평가 지표를, 4 절에서는 실험 결과 및 case study를 설명하고 마지막으로 5 절에서는 본 논문의 결론을 서술한다.

II. 배경 연구

MIV 표준은 기존의 블록 기반 압축 방식을 사용하며 특

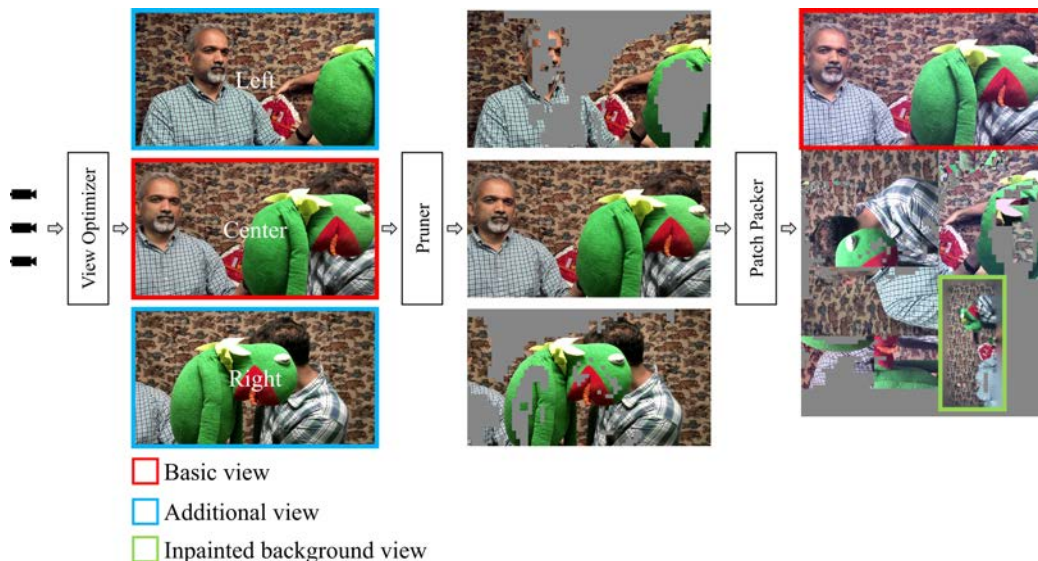


그림 2. MIV 부호기를 통한 몰입형 영상 부호화 과정
 Fig. 2. Immersive video encoding processes by the MIV encoder

정 영상 압축 도구에 종속적인 MVC 및 MV-HEVC의 한계를 극복하기 위해 픽셀 영역에서 와핑 (warping) 기반 몰입형 영상 간 중복성 제거 기법을 채택하였고, 이를 통해 비정형 카메라 배열로부터 취득된 몰입형 영상에 대해서도 높은 압축 효율을 보인다. 그림 2는 몰입형 영상을 MIV 부호기를 통해 MIV 모드로 부호화하는 과정을 나타내며, 이는 크게 view optimizer, pruner, patch packer에 의한 처리 과정을 거친다. MIV 부호기는 동일한 높이에서 좌측, 중앙, 우측으로 정렬된 3개의 카메라로부터 취득된 몰입형 영상을 입력받고, view optimizer는 입력 영상을 기본 및 추가 시점으로 분류한다. 기본 시점은 전체 영상을 가장 잘 나타내도록 선택되며, 단일 기본 시점을 선택할 경우 모든 시점의 중앙에 가까운 위치에, 복수 기본 시점의 경우 기본 시점끼리 겹치는 영역이 거의 없도록 선택된다. MIV 모드에서 기본 시점 외의 모든 시점은 추가 시점으로 분류되며, 이후 pruner를 통해 시점 간 중복된 영역은 오직 추가 시점에서만 제거된다. 상기 제거된 영역은 기본 시점에 의해 표현되고, patch packer를 통해 중복성이 제거된 잔차 영상에서 유효한 영역은 직사각형 형태의 패치로 검출 및 추출되어 아틀라스에 포함된다. 기본 시점은 그 자체로 커다란 하나의 패치로 취급되고, 기본 및 추가 시점 이외에도 MIV 복호기 단에서의 인페인팅 (inpainting) 성능 향상을 위해 모든 시점으로부터 합성된 인페인팅된 배경 뷰 (inpainted background view) 가 생성되어 아틀라스에 포함된다. 현재 MIV 부호기는 아틀라스 생성 시 표 1에 도시된 픽셀 레이트 (pixel rate) 제한 조건을 따라 2개의 텍스처, 지오메트리 아틀라스를 생성하여 총 4개의 영상이 출력된다. 아틀라스 해상도를 최적화하기 위해 각 패치는 회전되어 들어갈 수 있으며, 예시로 그림 2에 녹색 상자로 강조된 인페인팅된 배경 뷰 패치가 90도 회전되어 아틀라스에 삽입된 것을 확인

할 수 있다.

MIV 부호기에 의해 생성된 아틀라스는 2차원 영상의 형태로 저장되므로 이는 종래에 널리 쓰였거나 차세대 기술로 개발되는 2-D 영상 압축 도구를 통해 부호화될 수 있다. 부호화된 몰입형 영상을 수신한 클라이언트는 뷰 합성을 통해 사용자 관심 영역을 복호화할 수 있어야 하므로 아틀라스 내에 포함된 각 패치 정보가 따로 저장 및 전송되어야 한다. 기존에 MPEG에서 표준화가 진행중이던 3차원 포인트 클라우드 (point cloud) 압축을 위한 video-based point cloud coding (V-PCC)는 3차원 포인트 클라우드를 2차원 영상으로 투영 (projection) 하여 2-D 영상 압축 도구를 통해 높은 부호화 효율을 보이며^{[20]-[22]}, V-PCC 역시 MIV와 마찬가지로 패치 및 아틀라스의 개념을 도입하여 전송 측면에서의 구분 및 의미론에서 공통점이 많다. 따라서 상기 아틀라스를 복원하는데 필요한 정보를 MIV 표준에서는 visual volumetric video coding (V3C) / V-PCC 표준에 따라 정의하여 MIV 및 V-PCC는 같은 표준 구분 및 의미론을 공유한다^[23]. 따라서 MIV 기반 스트리밍 시스템은 아틀라스 복원을 위한 MIV 비트스트림을 생성하고 아틀라스를 부호화한 비트스트림과 멀티플렉싱 (multiplexing) 하여 전송한다.

MIV 표준에 의해 부호화 및 전송되는 몰입형 영상은 몰입감 향상을 위해 다수의 고해상도 전방위 360도 영상의 형태를 가질 수 있고, 이는 단일 360도 영상 전송 대비 상당한 비트율 증가를 야기한다. 이에 전송 측면에서의 MIV 표준 최적화를 위해 몰입형 영상에 대해 사용자 시점 기반 타일 스트리밍을 적용한 연구가 제안되었으며, 11.11%의 비트율 절감이 보고되었다^[24]. 또한 다양한 시점 및 패치 정보를 포함하여 반복되는 패턴을 보이는 아틀라스를 효율적으로 부호화하기 위해 screen content coding (SCC) 및 화면

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

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

내에서 블록을 예측하는 *intra block copy* (IBC) 기술을 적용한 연구가 제안되었고, 최대 15.74%의 부호화 성능 향상이 보고되었다^[25]. 한편 MIV 표준화를 진행중인 ISO/IEC SC 29/WG 4 그룹에서는 4개의 아틀라스 비트스트림을 복호화할 때 다수의 복호기 인스턴스 사용으로 인한 지연 및 동기화 문제를 해결하고자 아틀라스를 픽셀, 비트스트림 영역에서 병합하는 프레임 패킹 (*frame packing*) 기술을 논의하였고, 이에 Nokia, Intel, 성균관대학교가 프레임 패킹 기술을 기고하였다^{[26]-[28]}. 상기 프레임 패킹 기술은 복호기의 개수를 2개 또는 1개로 절약 가능하며 성균관대학교 프레임 패킹 기술의 경우 15.78%의 복호화 시간 절감이 보고되었다^{[29][30]}.

멀티플렉싱되어 전송된 비트스트림은 디멀티플렉싱 (*demultiplexing*) 되고, 아틀라스 비트스트림은 2-D 영상 복호기에 의해 복호화된다. 이후 MIV 복호기는 MIV 비트스트림과 복호화된 아틀라스를 입력받아 원본 영상과 동일한 해상도 및 시야각을 가지는 영상 또는 사용자 관심 영역만을 뷰 합성을 통해 생성할 수 있다. 현재 MIV 표준에 채택된 뷰 합성 기술은 *reference view synthesizer* (RVS) 및 *view weighting synthesizer* (VWS) 2가지가 있다. RVS는 Université libre de Bruxelles (ULB) 와 Philips 공동연구팀

에 의해 개발되었고, 삼각형 메쉬 와핑을 통한 뷰 합성을 지원하며 VWS 대비 낮은 복잡도로 실시간화에 유리하다^{[31][32]}. VWS는 한국전자통신연구원 연구팀에 의해 개발되었고, ray 기반으로 각 뷰에 대한 가중치를 부여하여 블렌딩 (*blending*) 하는 방법으로 RVS 대비 향상된 품질을 제공하며 현재 MIV의 표준 뷰 합성 기술로 채택되었다^[33]. 한편, 몰입형 6DoF 스트리밍 시스템 구축을 위해서는 실시간 뷰 합성 기술의 구현이 필수적이다. [34]에 따르면 현재 MIV 표준 뷰 합성 도구 중 하나인 RVS는 OpenGL을 통한 GPU 가속을 통해 2개의 영상을 입력받고 가상 시점 합성시 그림 3과 같이 프레임 당 평균 520.71ms가 소모된다. 이를 가속화하기 위해 ULB 연구팀은 GPU 병렬화 및 Oculus Rift 및 GTX 1080Ti 기반의 테스트베드 상에서 사용자 시점 합성 실험 시 90 FPS를 달성하였으나, 4개의 다운샘플링 (*down-sampling*) 된 1920×1080 (FHD) 영상을 사용했다는 한계점이 있다^[35]. 서울과학기술대학교와 한국항공대학교 공동연구팀은 OpenGL 버퍼와 텍스처 객체 재사용 및 파일 입출력 병렬화를 통해 가상 시점 합성 실험 시 약 26 FPS를 달성하였으나, 2개의 다운샘플링된 2K 영상을 사용하였고, 기존 RVS에서 CPU 영역에서만 구현된 인페인팅 기법을 생략하였다는 한계점이 있다^[36]. 한국전자

| Process | Exc. Time (ms) | Portion (%) |
|-------------------------|----------------|----------------|
| File loading | 83.87 | 17.83% |
| Synthesis shader | 180.50 | 35.23% |
| Texture buffer | 27.93 | 3.79% |
| Texture transfer | 35.30 | 7.24% |
| Create VAO/VBO/EBO | 110.15 | 22.69% |
| Synthesis | 7.11 | 1.51% |
| Blender shader | 45.36 | 8.75% |
| Setup VAO/VBO | 45.20 | 8.72% |
| Blending | 0.16 | 0.04% |
| Result | 210.99 | 38.19% |
| Download | 55.20 | 11.37% |
| Inpaint | 136.90 | 22.55% |
| Save | 18.89 | 4.27% |
| Total | 520.71 | 100.00% |

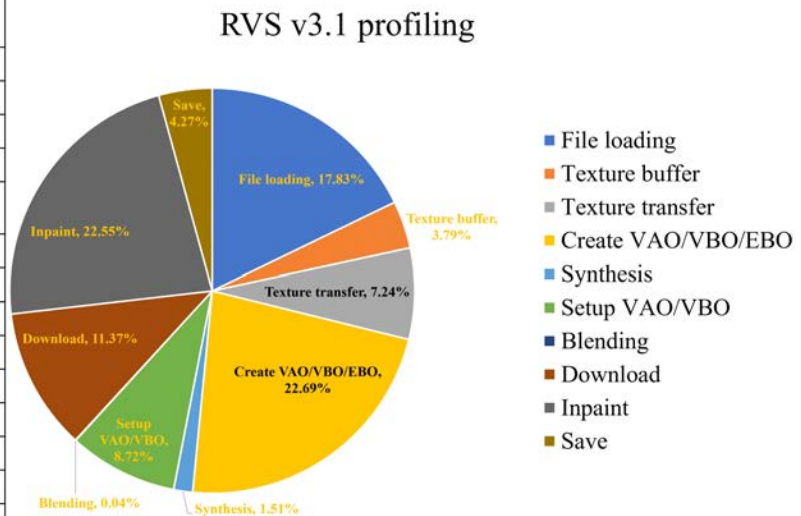


그림 3. OpenGL 가속 기반 RVS v3.1 프로파일링 결과
 Fig. 3. Profiling results of RVS v3.1 based on OpenGL acceleration

통신연구원 연구팀은 VWS에 360도 영상 표현에 사용되는 equirectangular projection (ERP) 영상을 고려한 삼각형 기반 색상화 및 OpenGL 기반 GPU 가속화를 적용하여 기존의 RVS 및 VWS 대비 향상된 뷰 합성 기술을 제공한다³⁷⁾. 손실 블록 압축 알고리즘인 DXT1을 텍스처에 적용하여 적은 화질 열화와 83% 용량 절감을, 16비트 지오메트리를 2개의 8비트 채널로 분할 및 BC4 압축을 적용하여 25% 용량 절감을 이끌어내었으며, 2개의 2K 몰입형 영상 입력 및 2개의 가상 시점 영상 생성 시 Intel Core i7-9800x CPU 및 2개의 RTX 2080Ti GPU로 구성된 테스트베드에서 단일 GPU 사용 시 약 35 FPS, 듀얼 GPU 사용 시 약 65 FPS의 뷰 합성 속도를 달성하였다. ULB, 서울과학기술대학교 / 한국항공대학교, 한국전자통신연구원의 GPU 가속 기반 뷰 합성 기술은 FHD 또는 2K로의 영상 다운샘플링 및 2-4개도의 입력 영상 제한이라는 제약조건을 가져 입력 영상과 가상 시점 간 거리가 멀 경우 홀 (hole) 또는 폐색 (occlusion) 등의 문제가 발생할 수 있으나, 26~90 FPS의 뷰 합성 속도를 제공하고 향후 GPU 기반 가속을 통한 6DoF 영상 합성이 가능함을 증명하였다.

III. 테스트 시퀀스 및 성능 평가 지표

본 논문에서는 MIV 기반 6DoF 몰입형 영상 스트리밍 기법을 평가하기 위해 MIV 공통 실험 조건 (common test conditions; CTC)³⁸⁾을 따라 실험을 진행하였다. 표 2는 본 실험에서 사용된 11개의 몰입형 영상 테스트 시퀀스 정보를 나타낸다. 테스트 시퀀스는 5개의 ERP 영상 및 6개의 2-D perspective 영상으로 구성되며, 8개의 영상은 컴퓨터 그래픽스 환경에서, 3개의 영상은 자연 환경에서 취득되었다. 다양한 카메라 배열에 대한 성능 평가를 진행하기 위해 1-D, 2-D 정형 카메라 배열에서 취득된 5개의 테스트 시퀀스를, 카메라 간 회전 각도가 서로 다르며 불규칙하게 배열된 비정형 카메라 배열에서 취득된 6개의 테스트 시퀀스를 선정하였다. MIV 공통 실험 조건 권고 사항에 따라 MIV 모드는 A17, MIV 시점 모드는 V17로 표기하였다.

표 3은 본 실험에서 사용된 MIV 기반 부·복호화, 뷰 합성, 품질 평가 소프트웨어를 도시한다. MIV 부호화 및 뷰 합성엔 MIV의 참조 소프트웨어인 test model for immersive video (TMIV) v8.0³⁹⁾을, 2-D 영상 부·복호화엔

표 2. MIV 테스트 시퀀스 정보. 약자: A17, 17프레임에 대한 MIV 모드; V17, 17프레임에 대한 MIV 시점 모드
Table 2. Information of the MIV test sequences. Abbreviations: A17, MIV mode for 17 frames; V17, MIV view mode for 17 frames

| Name | Class | Projection | Resolution (source view) (atlas) | No. of source views | No. of basic views (A17 / V17) |
|----------------|-------|-------------|----------------------------------|---------------------|--------------------------------|
| ClassroomVideo | CG-A | ERP | 4096×2048 4096×2176 | 15 | 1 / 2 |
| Museum | CG-B | ERP | 2048×2048 2048×4352 | 24 | 2 / 4 |
| Fan | CG-O | Perspective | 1920×1080 1920×4640 | 15 | 4 / 8 |
| Kitchen | CG-J | Perspective | 1920×1080 1920×4640 | 25 | 4 / 8 |
| Painter | NC-D | Perspective | 2048×1088 2048×4352 | 16 | 4 / 8 |
| Frog | NC-E | Perspective | 1920×1080 1920×4640 | 13 | 4 / 8 |
| Carpark | NC-P | Perspective | 1920×1080 1920×4640 | 9 | 4 / 6 |
| Chess | CG-N | ERP | 2048×2048 2048×4352 | 10 | 2 / 4 |
| Group | CG-R | Perspective | 1920×1080 1920×4640 | 21 | 4 / 8 |
| ChessPieces | CG-Q | ERP | 2048×2048 2048×4352 | 10 | 2 / 4 |
| Hijack | CG-C | ERP | 4096×2048 4096×2176 | 10 | 1 / 2 |

표 3. MIV 압축, 뷰 합성, 품질 평가 소프트웨어
 Table 3. MIV compression, view synthesis, and quality evaluation softwares

| Software | Version |
|---------------------------------------|---------|
| Test model for immersive video (TMIV) | v8.0 |
| HEVC test model (HM) | v16.20 |
| VVC test model (VTM) | v11.0 |
| WS-PSNR | v2.0.1 |
| VMAF | v1.3.14 |
| IV-PSNR | v3.0 |

HEVC의 참조 소프트웨어인 HEVC test model (HM) 16.20^[40]과 VVC의 참조 소프트웨어인 VVC test model (VTM) 11.0^[41]이 사용되었다. 품질 평가 기법으로는 weighted-to-spherically-uniform peak signal-to-noise ratio (WS-PSNR)^[42], video multimethod assessment fusion (VMAF)^[43], immersive video PSNR (IV-PSNR)^[44]이 사용되었다. WS-PSNR은 3-D 구체를 2-D 평면에 투영할 때 위, 아래 극점에서 발생하는 왜곡에 대한 가중치를 적게 부여하기 위해 개발된 360도 ERP 영상을 위한 품질 평가 기법으로, 본 실험에 사용된 ERP 영상에는 WS-PSNR을, 2-D perspective 영상에는 PSNR을 적용하였다. VMAF는 시-공간적 측면에서 support vector machine (SVM) 기반 기법을 사용하여 주관적 평가에 가까운 객관적 품질 점수를 산출하기 위해 개발되었고, 본 실험에서는 4K 영상으로 학습한 v1.3.14 가중치를 사용하였다. 6DoF 몰입형 영상에 대한 뷰 합성 시 한 물체를 표현하기 위해 여러 시점의 정보가

사용되므로 물체의 가장자리에서 아티팩트 (artifact) 가 발생하고, 이는 주관적 화질 관점에서 크게 눈에 띄지 않음에도 불구하고 픽셀 기반 품질 평가 기법인 PSNR의 저하를 야기한다. IV-PSNR은 상기 아티팩트를 무시하고 다시점 영상에서의 색변화를 보정하는 기법을 사용하여 주관적 품질 평가와의 높은 상관관계를 보여주었고^{[44][45]}, 본 실험에서는 IV-PSNR 중심의 객관적 품질 평가를 진행하였다.

$$Q_q = \max(1, [-14.2 + 0.8 \cdot Q_i]) \quad (1)$$

MIV 공통 실험 조건은 5개의 목표 비트레이트 (bitrate) 지점 및 그에 상응하는 양자화 매개변수 (quantization parameter; QP) 를 명시한다. 표 4는 HM 사용 시 MIV 모드, MIV 시점 모드에 대한 텍스처 양자화 매개변수를 나타낸다. 지오메트리 영상은 부호화로 인한 정보 손실이 발생하면 뷰 합성 시 화질 열화를 야기하므로 MIV 공통 실험 조건은 지오메트리에 텍스처 대비 낮은 양자화 매개변수를 부여하여 품질을 향상시킨다. 한 텍스처 양자화 매개변수에 대한 지오메트리 양자화 매개변수는 수식 (1)을 활용하여 구할 수 있으며, 수식 (1)에서 는 지오메트리 양자화 매개변수, 는 텍스처 양자화 매개변수를 의미한다. 표 5는 VTM 사용시 MIV 모드, MIV 시점 모드에 대한 텍스처 양자화 매개변수를 나타내며, 상기 양자화 매개변수는 2021년 5월 제안된 MIV 공통 실험 조건^[46]에서 2-D 영상 부호화 도구로 채택된 versatile video encoder (VVenc)^[47]를 위

표 4. MIV 모드, MIV 시점 모드 텍스처 QP (HM)
 Table 4. MIV and MIV mode QPs for texture (HM)

| Mode | MIV mode (A17) | | | | | MIV view mode (V17) | | | | |
|----------------|----------------|-----|-----|-----|-----|---------------------|-----|-----|-----|-----|
| | 50 | 28 | 16 | 9 | 5 | 50 | 28 | 16 | 9 | 5 |
| Class | QP1 | QP2 | QP3 | QP4 | QP5 | QP1 | QP2 | QP3 | QP4 | QP5 |
| ClassroomVideo | 25 | 27 | 30 | 33 | 38 | 26 | 28 | 30 | 33 | 37 |
| Museum | 21 | 27 | 33 | 37 | 41 | 24 | 30 | 35 | 38 | 42 |
| Fan | 30 | 38 | 45 | 48 | 51 | 31 | 37 | 45 | 48 | 51 |
| Kitchen | 14 | 21 | 27 | 33 | 39 | 18 | 24 | 29 | 34 | 39 |
| Painter | 22 | 28 | 35 | 44 | 51 | 22 | 28 | 34 | 41 | 47 |
| Frog | 30 | 36 | 43 | 47 | 51 | 32 | 37 | 42 | 48 | 51 |
| Carpark | 22 | 26 | 32 | 39 | 47 | 22 | 25 | 31 | 37 | 43 |
| Chess | 11 | 18 | 25 | 31 | 38 | 10 | 17 | 22 | 28 | 33 |
| Group | 24 | 30 | 35 | 40 | 46 | 25 | 30 | 34 | 38 | 42 |
| ChessPieces | 4 | 11 | 18 | 26 | 35 | 3 | 9 | 15 | 22 | 28 |
| Hijack | 16 | 22 | 29 | 38 | 49 | 14 | 18 | 23 | 29 | 35 |

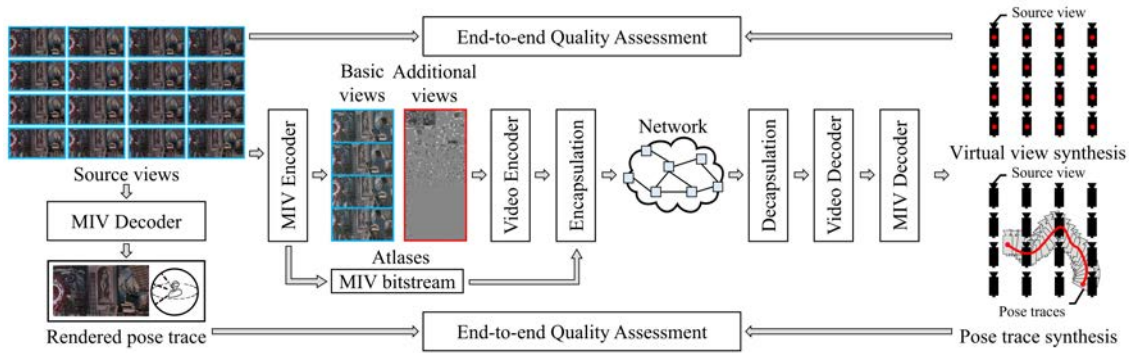


그림 4. MIV 기반 6DoF 몰입형 영상 스트리밍 시스템 평가 프레임워크
 Fig. 4. Evaluation framework of the MIV-based 6DoF immersive video streaming system

표 5. MIV 모드, MIV 시점 모드 텍스처 QP (VTM)
 Table 5. MIV and MIV mode QPs for texture (VTM)

| Mode | MIV mode (A17) | | | | | MIV view mode (V17) | | | | |
|-----------------------|----------------|-----|-----|-----|-----|---------------------|-----|-----|-----|-----|
| | QP1 | QP2 | QP3 | QP4 | QP5 | QP1 | QP2 | QP3 | QP4 | QP5 |
| Target bitrate [Mbps] | 50 | 28 | 16 | 9 | 5 | 50 | 28 | 16 | 9 | 5 |
| ClassroomVideo | 26 | 28 | 33 | 41 | 50 | 27 | 29 | 33 | 41 | 49 |
| Museum | 30 | 38 | 44 | 48 | 51 | 33 | 41 | 46 | 49 | 51 |
| Fan | 30 | 35 | 41 | 46 | 51 | 31 | 34 | 41 | 46 | 51 |
| Kitchen | 17 | 24 | 29 | 35 | 41 | 21 | 27 | 31 | 36 | 41 |
| Painter | 22 | 29 | 38 | 45 | 51 | 22 | 29 | 37 | 42 | 47 |
| Frog | 28 | 32 | 37 | 42 | 46 | 30 | 33 | 36 | 43 | 46 |
| Carpark | 22 | 26 | 33 | 40 | 47 | 22 | 25 | 32 | 38 | 43 |
| Chess | 18 | 25 | 31 | 37 | 45 | 17 | 24 | 28 | 34 | 40 |
| Group | 22 | 28 | 34 | 39 | 46 | 23 | 28 | 33 | 37 | 42 |
| ChessPieces | 4 | 11 | 18 | 26 | 35 | 17 | 22 | 28 | 36 | 43 |
| Hijack | 16 | 22 | 29 | 38 | 49 | 14 | 17 | 22 | 29 | 35 |

한 것으로, 본 실험에서는 VTM을 활용한 영상 부호화에 적용하였다. 그림 4에 도시된 바와 같이 MIV는 입력 영상 위치에 가상 시점을 합성하는 중간 뷰 합성 및 동적인 사용자 시점을 나타내는 포즈 트레이스 (pose trace) 합성을 통해 품질 평가를 진행한다. 합성되는 중간 뷰는 입력 영상과 같은 해상도 및 FoV를 가지며, 포즈 트레이스는 HMD를 통해 6DoF 몰입형 영상을 감상할 때와 같이 전체 영상에서 사용자가 시청 가능한 일부의 영역만을 나타내므로 입력 영상보다 작은 해상도 및 FoV를 가진다. MIV 공통 실험 조건은 각 테스트 시퀀스당 3개의 예시적인 포즈 트레이스를 사용자 위치 좌표 (x, y, z) 및 사용자 회전 각도 (yaw, pitch, roll) 를 통해 정의하였고, 본 실험에서도 각 테스트 시퀀스 별 3개의 포즈 트레이스를 사용하였다. 본 실험은

2개의 intel xeon e5-2687w v4 CPU와 128GB의 메모리가 장착된, Ubuntu 18.04가 설치된 서버에서 진행되었다.

IV. 실험 결과

표 6은 HM 기반 부호화 환경에서 중간 뷰 합성 시 MIV 시점 모드 대비 MIV 모드의 WS-PSNR, VMAF, IV-PSNR에 대한 비트율 절감율을 나타낸다. 평가 지표로는 bjontegaard delta rate (BD-rate) 이 사용되었고, 계산 시 4개의 대역폭 지점을 필요로 하므로 MIV 공통 실험 조건에서는 QP1~QP4 고대역폭 지점을 ‘high’, QP2~QP5 저대역폭 지점을 ‘low’로 정의하였고 본 실험도 상기 조건을 준수하였

표 6. 중간 뷰 합성 시 MIV 시점 모드 + HM 대비 MIV 모드 + HM의 WS-PSNR, VMAF, IV-PSNR BD-rate 성능 (음수 값은 비트율 절감을 나타내며, ‘-’ 기호는 BD-rate 이득이 계산되기에 매우 높음을 의미함)

Table 6. WS-PSNR, VMAF, and IV-PSNR BD-rate performances of the MIV mode + HM compared to the MIV view mode + HM for synthesized intermediate views (Negative value indicates bitrate saving, and ‘-’ symbol indicates the BD-rate gain is too high to be computed)

| | High-BR BD-rate Y-PSNR | Low-BR BD-rate Y-PSNR | High-BR BD-rate VMAF | Low-BR BD-rate VMAF | High-BR BD-rate IV-PSNR | Low-BR BD-rate IV-PSNR |
|----------------|------------------------------|-----------------------------|----------------------------|---------------------------|-------------------------------|------------------------------|
| ClassroomVideo | -11.02% | -5.53% | -65.03% | -19.36% | -27.17% | -14.16% |
| Museum | -54.06% | -41.16% | -19.62% | -14.62% | -65.64% | -49.28% |
| Fan | 28.10% | 39.55% | 54.48% | 58.77% | 28.48% | 39.46% |
| Kitchen | -46.51% | -25.25% | -15.05% | -8.14% | -70.08% | -45.09% |
| Painter | 38.00% | 34.20% | 19.81% | 26.02% | 12.26% | 20.60% |
| Frog | 31.80% | 33.27% | 38.16% | 38.09% | 10.93% | 22.60% |
| Carpark | -18.47% | -8.12% | 44.19% | 20.45% | -27.67% | -15.44% |
| Chess | - | - | -81.47% | -32.12% | - | - |
| Group | 63.18% | 58.03% | 12.37% | 27.96% | 12.61% | 23.84% |
| ChessPieces | - | - | - | -41.50% | - | - |
| Hijack | - | -32.11% | -25.10% | 15.29% | - | -40.42% |
| Average | 2.82% | 4.81% | -3.39% | 6.44% | -11.48% | -5.26% |

다. 표에서 음수 값은 동일 품질 제공 시 해당 비율만큼 비트율이 절감됨을 의미하며, ‘-’ 기호는 비트율 절감폭이 매우 높거나 낮아 계산되지 않고 0.00%의 값으로 표시됨을 의미한다. Low BD-rate 값 대비 high-BD-rate 값이 낮았으므로 MIV 시점 모드 대비 MIV 모드는 고대역폭에서 효율적이었으며, 특히 IV-PSNR은 11.48%의 비트율 절감을 보여주었다. 반면, Y-PSNR에서는 MIV 모드가 MIV 시점 모드 대비 2.82%의 비트율 증가가 있었고, VMAF에 대해서는 3.39%의 비트율 절감이 확인되었다. 이와 같은 결과는 각 품질 평가 기법의 특성 때문인데, Y-PSNR의 경우 픽셀 단위로 품질 차이를 반영한다. MIV 기반 뷰 합성 시 물체의 경계에서 아티팩트가 다수 발생하고, 중복성이 제거된 여러 패치를 통해 뷰 합성을 진행하는 MIV 모드 특성 상 상기 아티팩트가 MIV 시점 모드에 비해 많이 발생한다. Y-PSNR은 상기 아티팩트를 검출하여 MIV 모드에서의 효율이 MIV 시점 모드의 효율보다 낮다는 결과를 도출하였다. 반면, 가중치 계산을 통해 주관적 품질과 유사한 점수를 도출하는 VMAF의 경우 상기 아티팩트가 주관적 화질 측면에서 중요하지 않다고 판단하여 MIV 모드의 품질이 더욱 좋다고 평가하였다. 뷰 합성을 통해 생성된 영상을 평가하기 위해 개발된 IV-PSNR은 상기 아티팩트를 덜 반영하도록 개발되어 VMAF와 마찬가지로 MIV 모드가 더 효율적이라는 결론을 도출하였다. 본 실험에서는 MIV 표준화 활동 중 제안된 IV-PSNR 위주로 실험 결과 분석을 진행하

고, 이에 대한 타당성을 주관적 화질 측면에서 검증하였다.

뷰 합성 아티팩트 이외에도, 일부 시퀀스에서 MIV 시점 모드의 한계점이 확인되었다. 제한된 수의 시점 전송으로 인해 MIV 시점 모드 아틀라스로 표현이 불가능한 중간 뷰 합성 시 정보 부재로 인한 홀 및 폐색의 문제가 발생하여 MIV 시점 모드의 효율 하락이 확인되었다. 실제로 카메라 간 겹치는 영역이 적은 비정형 카메라 배열을 가지는 ClassroomVideo, Museum, Chess, ChessPieces, Hijack 테스트 시퀀스에서는 MIV 모드가, 카메라 간 겹치는 영역이 많은 정형 카메라 배열 기반의 Fan, Painter, Frog 테스트 시퀀스에서는 MIV 시점 모드가 효율적이 확인되었다. Kitchen 테스트 시퀀스에서는 70.08%의 high IV-PSNR BD-rate 절감이 확인되었는데, 상기 시퀀스는 2-D 정형 카메라 배열에서 취득되었으나 영상 내 물체가 많아 다수의 폐색이 존재하였고 이로 인해 MIV 모드의 IV-PSNR BD-rate 절감이 확인되었다. Carpark 테스트 시퀀스는 1-D 정형 카메라 배열에서 취득되었으나 MIV 시점 모드에서 동일 해상도의 타 테스트 시퀀스가 8개의 기본 시점을 선택한 것에 비해 6개의 기본 시점만을 선택하였고, MIV 모드에서 시점 간 중복성 제거를 통해 생성된 패치의 수가 적어 부호화 시 적은 비트율을 소모하였으므로 MIV 모드의 IV-PSNR BD-rate 절감이 확인되었다. Group 테스트 시퀀스는 비정형 카메라 배열로부터 취득되었으나, 카메라 방향이 바깥에서 안쪽을 향하므로 시점 간 변화가 크지 않았

표 7. 중간 뷰 합성 시 MIV 시점 모드 + VTM 대비 MIV 모드 + VTM의 WS-PSNR, VMAF, IV-PSNR BD-rate 성능 (음수 값은 비트율 절감을 나타내며, '-' 기호는 BD-rate 이득이 계산되기에 매우 높음을 의미함)

Table 7. WS-PSNR, VMAF, and IV-PSNR BD-rate performances of the MIV mode + VTM compared to the MIV view mode + VTM for synthesized intermediate views (Negative value indicates bitrate saving, and '-' symbol indicates the BD-rate gain is too high to be computed)

| | High-BR BD-rate Y-PSNR | Low-BR BD-rate Y-PSNR | High-BR BD-rate VMAF | Low-BR BD-rate VMAF | High-BR BD-rate IV-PSNR | Low-BR BD-rate IV-PSNR |
|----------------|------------------------------|-----------------------------|----------------------------|---------------------------|-------------------------------|------------------------------|
| ClassroomVideo | -5.64% | 15.44% | -12.04% | 11.47% | -6.26% | 11.35% |
| Museum | -29.37% | -4.27% | -3.24% | 10.71% | -31.39% | -6.09% |
| Fan | 22.34% | 34.34% | 49.26% | 55.62% | 28.64% | 40.05% |
| Kitchen | -41.90% | -22.05% | -14.66% | -5.46% | -63.55% | -35.91% |
| Painter | 41.31% | 34.84% | 20.77% | 24.78% | 16.11% | 21.69% |
| Frog | 48.83% | 40.12% | 62.79% | 49.46% | 10.91% | 20.40% |
| Carpark | -21.19% | -11.22% | 47.12% | 26.12% | -27.20% | -15.62% |
| Chess | - | - | -44.37% | -0.66% | - | - |
| Group | 101.85% | 72.41% | 17.60% | 34.87% | 12.53% | 23.80% |
| ChessPieces | - | - | -12.17% | 25.38% | - | - |
| Hijack | - | -42.25% | -38.89% | -4.63% | - | -45.54% |
| Average | 10.57% | 10.67% | 6.56% | 20.70% | -5.48% | 1.28% |

고, 해상도가 타 시퀀스 대비 낮아 많은 수의 기본 시점이 MIV 시점 모드에 포함되었으므로 MIV 시점 모드의 효율이 높게 나타났다.

표 7은 VTM 기반 부호화 환경에서 중간 뷰 합성 시 MIV 시점 모드 대비 MIV 모드의 BD-rate 절감율을 나타낸다. 표 6, 7을 비교했을 때, Museum, ChessPieces 테스트 시퀀스가 HM 기반 MIV 모드에서는 높은 BD-rate 절감을 나타내는 것에 대비해 VTM 기반 MIV 모드에서는 일부 IV-PSNR BD-rate 손실을 나타내는 것을 확인할 수 있다. 이는 VVenC에 최적화된 양자화 매개변수를 VTM에 적용하여 MIV 모드와 MIV 시점 모드 간 비트율 차이가 크게 벌어져 BD-rate 산출 시 0.00% 값이 출력되었기 때문이고, 울-왜곡 곡선 상에서는 MIV 모드의 높은 효율이 확인되었다.

그림 5는 HM, VTM 및 MIV 모드 (A17), MIV 시점 모드 (V17) 를 사용하여 중간 뷰 합성 시 IV-PSNR에 대한 울-왜곡 곡선을 도시한다. 그래프의 x축은 대역폭을, y축은 IV-PSNR 값을 나타내고, 그래프가 좌상단에 위치할수록 동일 대역폭에서 높은 품질을 제공한다. 그림 5 (a)는 모든 테스트 시퀀스의 평균에 대한 결과를 도시하며, A17 + VTM, A17 + HM, V17 + VTM, V17 + HM 순으로 스트리밍 효율이 높음을 확인하였다. Chess, ChessPieces, Hijack 테스트 시퀀스의 울-왜곡 곡선은 그림 5 (h), (k), (d)에 도시

되었고, MIV 모드 그래프가 MIV 시점 모드 그래프 대비 좌상단으로 치우쳐 있어 표 6, 7에서 MIV 모드의 BD-rate가 '-' 값으로 나타난 반면 울-왜곡 곡선 상 MIV 모드의 성능이 매우 높음을 보여준다. Fan, Painter, Frog 테스트 시퀀스의 울-왜곡 곡선은 그림 5 (i), (e), (f)에 도시되었고, 상기 시퀀스들은 BD-rate 측면에서 MIV 시점 모드가 더 효율적이었으며 울-왜곡 곡선 역시 이를 뒷받침하는 것을 확인하였다. 그 외의 테스트 시퀀스에서는 그림 5에 도시되어 있듯 MIV 모드가 더 높은 효율을 보여주었다.

그림 6은 MIV 시점 모드의 품질저하가 컸던 Museum, Hijack 테스트 시퀀스에 대해 VTM 및 MIV 모드, MIV 시점 모드를 사용하여 QP1 지점에서 중간 뷰를 합성 및 확대한 그림을 도시한다. QP1 지점에서 Museum 테스트 시퀀스의 시점 v2를 생성했을 때 MIV 모드와 MIV 시점 모드 간 Y-PSNR 차이는 9.44dB로 매우 높으며, 이는 그림 6 (c), (d)에 도시된 바와 같이 MIV 시점 모드 기반 중간 뷰 합성 시 홀이 발생했기 때문이다. Museum 테스트 시퀀스는 각각 180°×180°의 시야각을 가지며 서로의 각도차가 큰 24개의 카메라로부터 취득되었기 때문에 MIV 시점 모드로는 전체 영상을 표현하는 데에 한계가 있었다. 한편 Hijack 테스트 시퀀스의 시점 v9를 생성했을 때 각 모드 간 Y-PSNR 차이는 9.74dB였으며, 그림 6 (g), (h)는 MIV 시점 모드 기반 중간 뷰 합성 시 홀 및 물체의 가장자리에서 아티팩트가

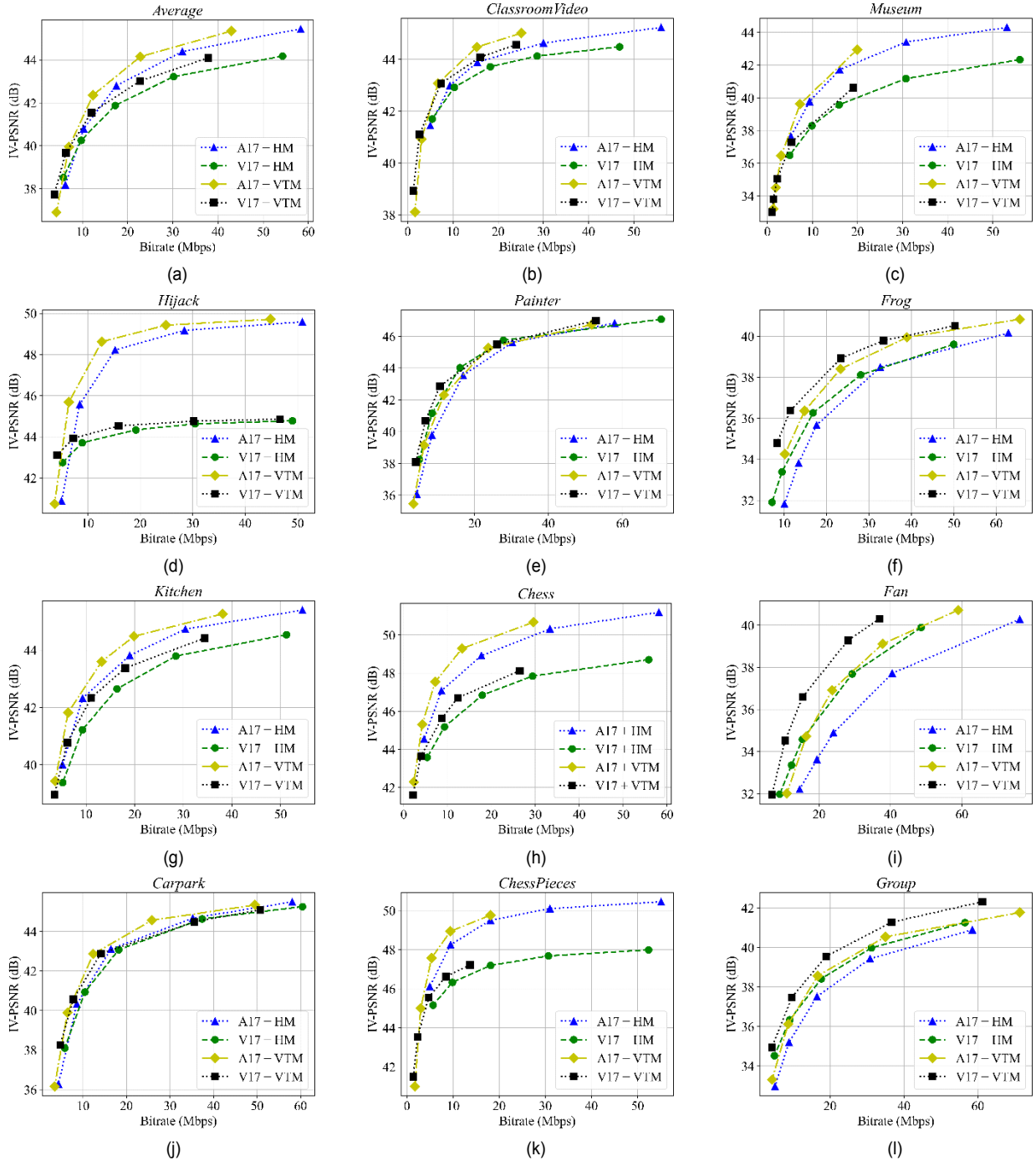


그림 5. 중간 뷰 합성 시 HM, VTM과 MIV 모드 (A17), MIV 시점 모드 (V17) 에 대한 IV-PSNR 율-왜곡 곡선: (a) 평균, (b) ClassroomVideo, (c) Museum, (d) Hijack, (e) Painter, (f) Frog, (g) Kitchen, (h) Chess, (i) Fan, (j) Carpark, (k) ChessPieces, (l) Group

Fig. 5. IV-PSNR RD curves of the MIV mode (A17) and MIV view mode (V17) for the synthesized intermediate views using HM and VTM: (a) average, (b) ClassroomVideo, (c) Museum, (d) Hijack, (e) Painter, (f) Frog, (g) Kitchen, (h) Chess, (i) Fan, (j) Carpark, (k) ChessPieces, (l) Group



그림 6. VTM과 MIV 모드 (A17), MIV 시점 모드 (V17)를 통해 QP1 지점에서 합성된 중간 뷰 확대: (a), (b) Museum view v2, A17 + VTM@20.03Mbps, (c), (d) Museum view v2, V17 + VTM@19.12Mbps, (e), (f) Hijack view v9, A17 + VTM@45.19Mbps, (g), (h) Hijack view v9, V17 + VTM@46.99Mbps
 Fig 6. Enlarged synthesized intermediate views using VTM, MIV mode (A17) and MIV view mode (V17) at QP1: (a), (b) Museum view v2, A17 + VTM@20.03Mbps, (c), (d) Museum view v2, V17 + VTM@19.12Mbps, (e), (f) Hijack view v9, A17 + VTM@45.19Mbps, (g), (h) Hijack view v9, V17 + VTM@46.99Mbps

발생했음을 보여준다. Hijack 테스트 시퀀스는 10개의 카메라가 같은 방향을 바라보나 각 카메라의 위치가 다르고 해상도가 4096×2048로 높아 MIV 시점 모드에서는 오직 2개 시점의 영상만을 아틀라스에 포함할 수 있었으므로 MIV 시점 모드의 효율이 낮게 나타났다.

MIV 부호기는 지오메트리를 기반으로 시점 간 중복성을 추가 시점에서 제거하고, MIV 복호기는 중간 뷰 합성 시 제거된 중복 영역을 기본 시점에서 가져오게 되어 통상 추가 시점 대비 기본 시점의 픽셀 레이트가 높게 나타난다. MIV 스트리밍 시스템 품질 평가 방법에 따르면 합성할 중간 뷰의 위치는 입력 영상의 위치와 동일하게 설정되어 추가 시점 위치에 합성된 중간 뷰보다 기본 시점 위치에 합성된 중간 뷰의 품질이 더 좋게 나타날 확률이 높고, 이는 각 테스트 시퀀스 특성에 따라 부적합한 품질 평가로 이어질 수 있다. 그림 7은 HM을 이용하여 2-D 영상을 부호화하고 QP1 지점에서 중간 뷰 합성 시 MIV 모드, MIV 시점 모드에서 각 뷰 별 IV-PSNR 값 그래프를 도시한다. 수평 방향으로 그려진 파랑, 초록 점선은 각각 MIV 모드, MIV 시점

모드에 대한 IV-PSNR 평균값을 나타내며, ‘□’ 기호와 병기된 뷰는 MIV 모드에서의 기본 시점을, ‘×’ 기호와 병기된 뷰는 MIV 시점 모드에서의 기본 시점을 의미한다. MIV 모드와 MIV 시점 모두 기본 시점으로 표시된 뷰에서 IV-PSNR의 급격한 상승이 있었고, 그 외의 시점에서는 주로 급격한 하락이 일어난 것을 확인하였다. 이는 비슷한 장면을 촬영한 테스트 시퀀스라도 기본 시점의 개수 및 분포에 따라 IV-PSNR 값의 분포 및 평균이 달라질 수 있다는 것을 의미하며, 본 논문에서는 보다 정확한 품질 평가를 위해 사용자 관심 영역을 나타내는 포즈 트레이스 합성 실험을 진행하였고 결과를 정리하였다. 또한 상기 결과는 VTM을 이용한 실험 결과에서도 유사한 추세를 보임을 확인하였다.

표 8은 HM 기반 부호화 환경에서 포즈 트레이스 합성 시 MIV 시점 모드 대비 MIV 모드의 WS-PSNR, VMAF, IV-PSNR BD-rate를 도시한다. Museum, Kitchen, Chess, ChessPieces, Hijack 테스트 시퀀스에서 ‘-’ 값이 도출되어 BD-rate 평균 값이 양수가 되었으나, 포즈 트레이스 합성

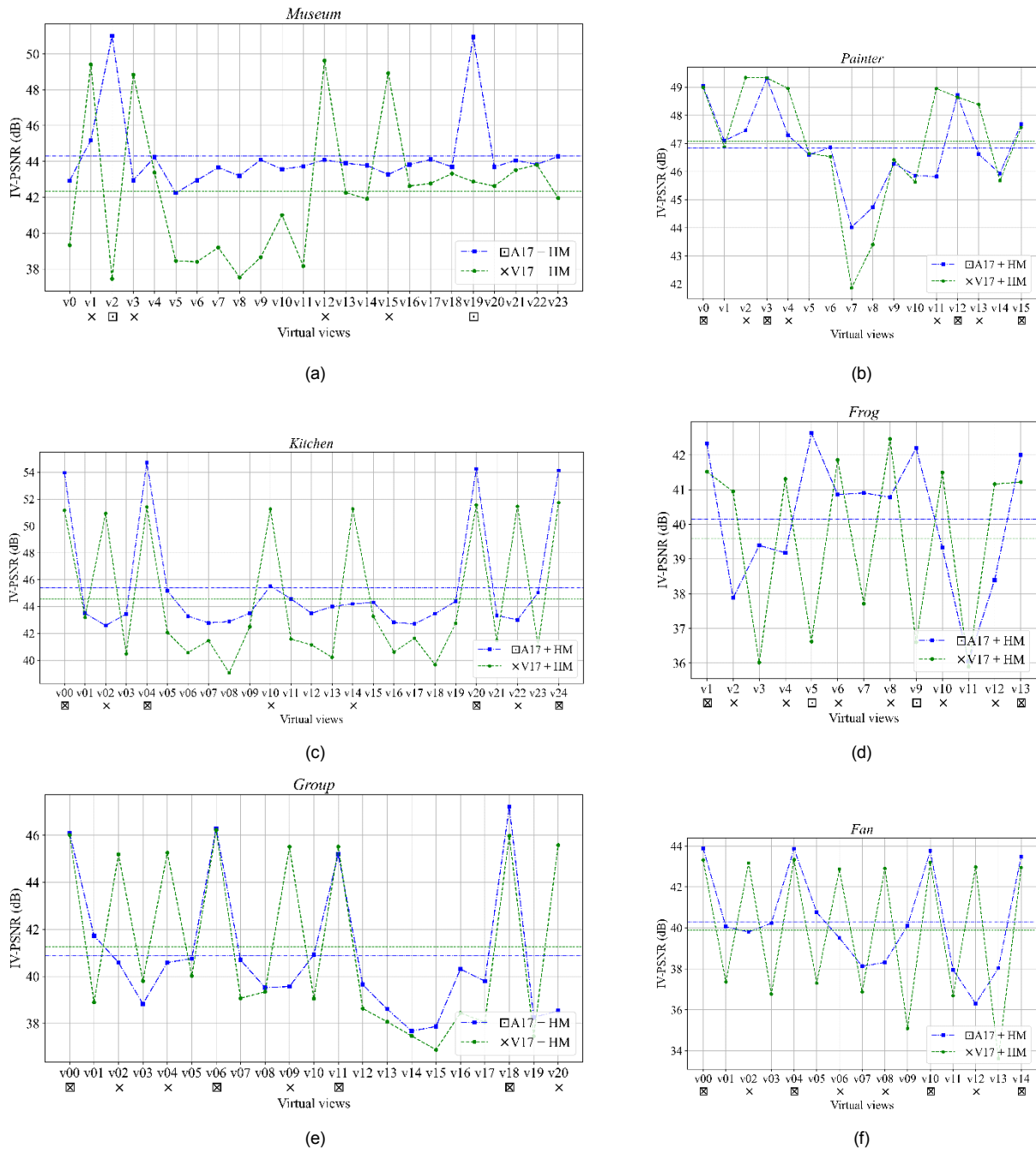


그림 7. QP1 지점에서 중간 뷰 합성 시 HM과 MIV 모드 (A17), MIV 시점 모드 (V17) 에 대한 뷰 별 IV-PSNR: (a) Musuem, (b) Painter, (c) Kitchen, (d) Frog, (e) Group, (f) Fan
 Fig. 7. Per-view IV-PSNR of the MIV mode (A17) and MIV view mode (V17) for the synthesized intermediate views using HM at QP1: (a) Musuem, (b) Painter, (c) Kitchen, (d) Frog, (e) Group, (f) Fan

표 8. 포즈 트레이스 합성 시 MIV 시점 모드 + HM 대비 MIV 모드 + HM의 WS-PSNR, VMAF, IV-PSNR BD-rate 성능 (음수 값은 비트율 절감을 나타내며, '-' 기호는 BD-rate 이득이 계산되기에 매우 높음을 의미함)

Table 8. WS-PSNR, VMAF, and IV-PSNR BD-rate performances of the MIV mode + HM compared to the MIV view mode + HM for synthesized pose traces (Negative value indicates bitrate saving, and '-' symbol indicates the BD-rate gain is too high to be computed)

| | High-BR BD-rate Y-PSNR | Low-BR BD-rate Y-PSNR | High-BR BD-rate VMAF | Low-BR BD-rate VMAF | High-BR BD-rate IV-PSNR | Low-BR BD-rate IV-PSNR |
|-----------------------|------------------------------|-----------------------------|----------------------------|---------------------------|-------------------------------|------------------------------|
| <i>ClassroomVideo</i> | 63.25% | 36.74% | 46.46% | 24.50% | 2.70% | 2.51% |
| <i>Museum</i> | - | - | -38.92% | -26.39% | - | - |
| <i>Fan</i> | 172.93% | 152.78% | 118.94% | 107.42% | 99.61% | 84.77% |
| <i>Kitchen</i> | -70.33% | -37.47% | -40.06% | -18.52% | - | -60.82% |
| <i>Painter</i> | 63.13% | 49.14% | 16.32% | 26.78% | 28.28% | 28.13% |
| <i>Frog</i> | 40.56% | 37.47% | 35.84% | 36.29% | 11.91% | 20.33% |
| <i>Carpark</i> | 174.35% | 56.95% | 140.38% | 46.12% | 96.05% | 40.22% |
| <i>Chess</i> | - | -63.31% | 95.31% | 67.40% | -73.77% | -39.48% |
| <i>Group</i> | 48.69% | 56.33% | 33.18% | 44.92% | 13.10% | 27.41% |
| <i>ChessPieces</i> | - | - | - | 183.70% | - | - |
| <i>Hijack</i> | - | - | -37.53% | 1.17% | - | - |
| <i>Average</i> | 44.78% | 26.24% | 33.63% | 44.85% | 16.17% | 9.37% |

시 울-왜곡 곡선을 나타낸 그림 8에서는 MIV 모드의 높은 효율이 확인되었다. 중간 뷰 합성과 다르게 포즈 트레이스 합성 시 high보다 low에서 MIV 모드의 Y-PSNR, IV-PSNR BD-rate 개선이 확인되었다. 반면, VMAF BD-rate 측정 시 MIV 모드가 high에서 더 높은 효율을 보여주었는데, ChessPieces 테스트 시퀀스에 대한 high VMAF BD-rate가 '-' 값으로 표시된 반면 low VMAF BD-rate가 183.70%라는 높은 양수 값을 기록하였기 때문이다. 표 9는 VTM 기반

부호화 환경에서의 BD-rate를 나타내며, 표 9에서 '-' 값을 보여준 Museum, ChessPieces 테스트 시퀀스에 대해서 높은 BD-rate 절감을 기록하여 MIV 모드가 상기 비정형 카메라 배열로부터 취득된 테스트 시퀀스에 효율적임을 확인하였다. 중간 뷰 합성 시 BD-rate 결과와 마찬가지로 포즈 트레이스 역시 BD-rate 평균값은 양수가 되어 MIV 시점 모드가 비트율 측면에서 효율적으로 보이니, 울-왜곡 곡선을 통해 MIV 모드가 비트율 측면에서 효율적임을 확인하였다.

표 9. 포즈 트레이스 합성 시 MIV 시점 모드 + VTM 대비 MIV 모드 + VTM의 WS-PSNR, VMAF, IV-PSNR BD-rate 성능 (음수 값은 비트율 절감을 나타내며, '-' 기호는 BD-rate 이득이 계산되기에 매우 높음을 의미함)

Table 9. WS-PSNR, VMAF, and IV-PSNR BD-rate performances of the MIV mode + VTM compared to the MIV view mode + VTM for synthesized pose traces (Negative value indicates bitrate saving, and '-' symbol indicates the BD-rate gain is too high to be computed)

| | High-BR BD-rate Y-PSNR | Low-BR BD-rate Y-PSNR | High-BR BD-rate VMAF | Low-BR BD-rate VMAF | High-BR BD-rate IV-PSNR | Low-BR BD-rate IV-PSNR |
|-----------------------|------------------------------|-----------------------------|----------------------------|---------------------------|-------------------------------|------------------------------|
| <i>ClassroomVideo</i> | 24.25% | 31.10% | 11.12% | 19.95% | 5.30% | 17.34% |
| <i>Museum</i> | -89.46% | -57.71% | -13.13% | 8.38% | -76.80% | -40.60% |
| <i>Fan</i> | 175.19% | 154.24% | 130.48% | 111.18% | 108.82% | 89.97% |
| <i>Kitchen</i> | -59.42% | -29.64% | -36.40% | -13.32% | - | -39.21% |
| <i>Painter</i> | 63.82% | 49.14% | 16.11% | 24.53% | 31.80% | 29.02% |
| <i>Frog</i> | 76.24% | 53.25% | 57.43% | 45.84% | 27.20% | 26.31% |
| <i>Carpark</i> | 182.53% | 65.58% | 165.09% | 63.21% | 95.37% | 40.66% |
| <i>Chess</i> | - | -16.71% | 75.16% | 65.86% | -46.50% | -11.54% |
| <i>Group</i> | 82.47% | 72.36% | 52.18% | 55.30% | 8.30% | 24.71% |
| <i>ChessPieces</i> | - | -13.94% | 115.16% | 99.47% | -65.73% | -14.77% |
| <i>Hijack</i> | - | - | -50.88% | -16.04% | - | - |
| <i>Average</i> | 41.42% | 27.97% | 47.48% | 42.21% | 7.98% | 11.08% |

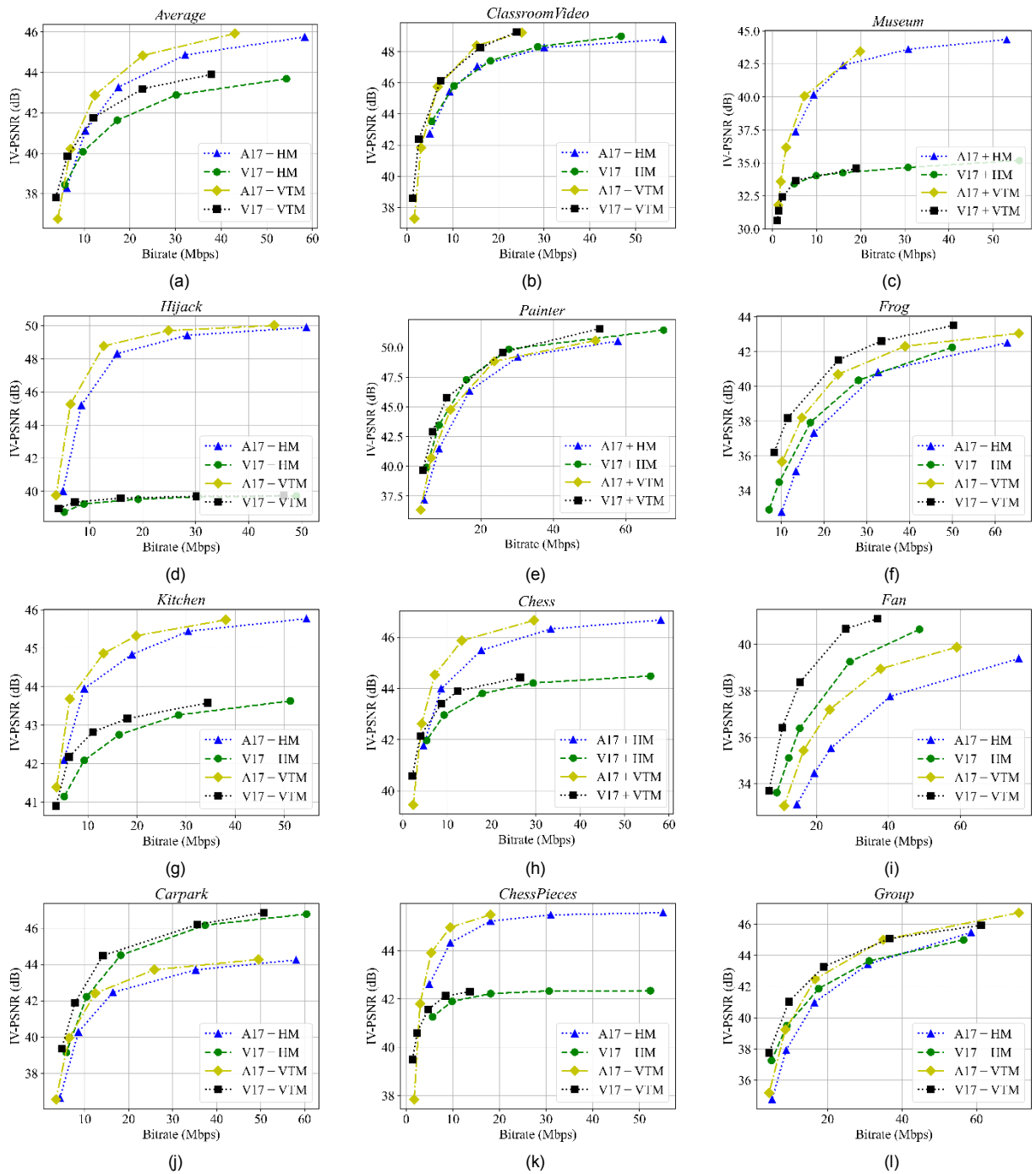


그림 8. 포즈 트레이스 합성 시 HM, VTM과 MIV 모드 (A17), MIV 시점 모드 (V17) 에 대한 IV-PSNR 율-왜곡 곡선: (a) 평균, (b) ClassroomVideo, (c) Museum, (d) Hijack, (e) Painter, (f) Frog, (g) Kitchen, (h) Chess, (i) Fan, (j) Carpark, (k) ChessPieces, (l) Group

Fig. 8. IV-PSNR RD curves of the MIV mode (A17) and MIV view mode (V17) for the synthesized pose traces using HM and VTM: (a) average, (b) ClassroomVideo, (c) Museum, (d) Hijack, (e) Painter, (f) Frog, (g) Kitchen, (h) Chess, (i) Fan, (j) Carpark, (k) ChessPieces, (l) Group

포즈 트레이스 IV-PSNR 율-왜곡 곡선을 표현한 그림 8의 (e), (f), (i), (j), (l)은 각각 Painter, Frog, Fan, Carpark, Group 테스트 시퀀스 결과를 나타내고, 1-D, 2-D 정형 카메라 배열로부터 취득된 상기 테스트 시퀀스들은 모두 MIV 시점 모드에서 높은 효율을 보였다. 특히 Fan, Carpark 테스트 시퀀스의 경우 MIV 모드와 VTM을 사용했을 때 보다 MIV 시점 모드와 HM을 사용했을 때 더욱 높은 효율을 보였다. [48]에 따르면 VTM은 HM 대비 평균 62.66%의 비트율 절감 및 약 8.8배의 부호화 시간 증가를 나타내므로, 상기 특성을 가진 테스트 시퀀스들에 MIV 시점 모드 및 HM을 적용하면 비트율과 부호화 시간을 모두 크게 절감할 수 있다. 그림 9는 Painter, Frog, Fan, Carpark, Group 테스트 시퀀스들에 대해 VTM 기반으로 부호화를 진행하여 MIV 모드, MIV 시점 모드에서 합성된 포즈 트레이스를 나타낸다. MIV 모드 사용 시 높은 양자화 매개변수 사용으로 인한 직사각형 형태의 블로킹 아티팩트 (blocking

artifact), 엣지 아티팩트 (edge artifact), 화질 뭉개짐 현상이 보임에 반해 MIV 시점 모드 사용 시 상기 화질 열화가 상당 부분 제거되었음을 확인하였다.

표 10은 HM과 MIV 및 MIV 시점 모드에서의 실행 시간 비율을 나타낸다. 실험은 3 절에서 언급된 서버와 클라이언트 장비에서 수행되었다. MIV 시점 모드는 각 시점 정보를 이용하여 대표 기본 시점 영상만을 선택하여 아틀라스에 출력하는 반면, MIV 모드는 기본 시점과 추가 시점 간 중복성을 제거하고 패치를 패킹하는 과정을 포함하여 MIV 모드는 MIV 시점 모드 대비 MIV 부호화 시 평균 354.29%의 시간 증가율을 보여주었다. ClassroomVideo 시퀀스에서는 MIV 모드가 245.54%의 시간 증가율을 보여주었는데, 해당 시퀀스가 전방위 360도 영상을 포함하고 시점 간 위치 차이가 크지 않아 시점 간 중복성이 많이 제거되었고, 이에 패킹할 패치의 수가 적어 가장 적은 시간 상승을 보여주었다. 반대로 180도 ERP 카메라 또는 2-D 카메라로부터 취득

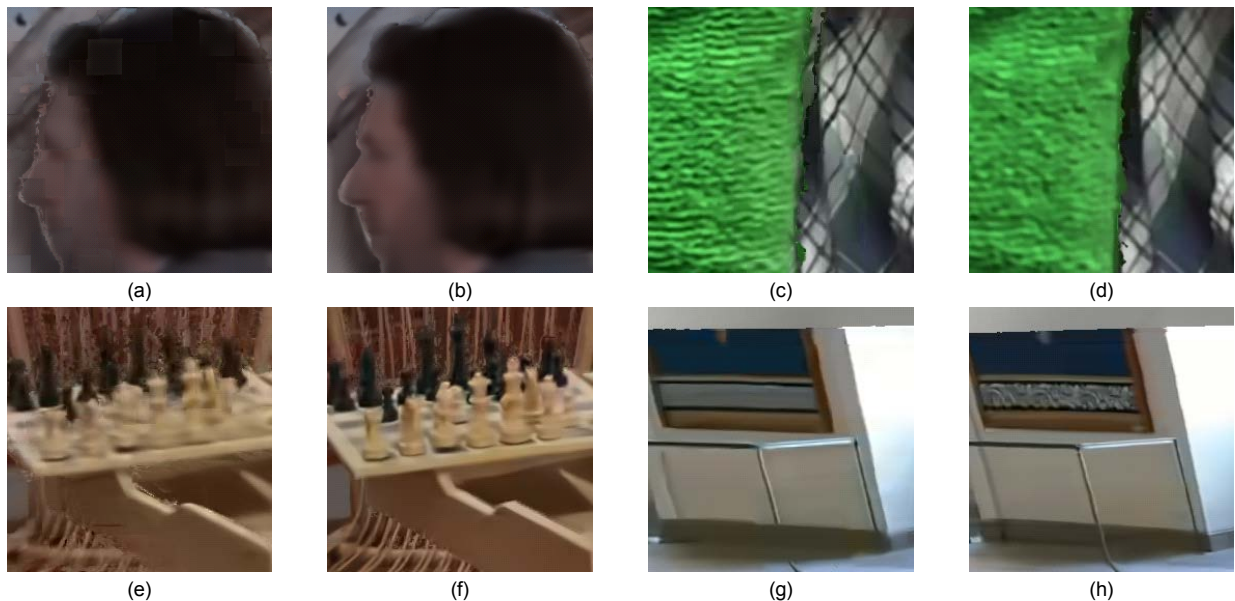


그림 9. VTM과 MIV 모드 (A17), MIV 시점 모드 (V17)를 통해 합성된 포즈 트레이스 확대: (a) Painter pose p02, A17 + VTM@12.06Mbps, (b) Painter pose p02, V17 + VTM@11.07Mbps, (c) Frog pose p03, A17 + VTM@23.55Mbps, (d) Frog pose p03, V17 + VTM@23.60Mbps, (e) Fan pose p03, A17 + VTM@11.19Mbps, (f) Fan pose p03, V17 + VTM@10.68Mbps, (g) Group pose p03, A17 + VTM@4.54Mbps, (h) Group pose p03, V17 + VTM@4.53Mbps

Fig. 9. Enlarged synthesized pose traces using VTM, MIV mode (A17) and MIV view mode (V17): (a) Painter pose p02, A17 + VTM@12.06Mbps, (b) Painter pose p02, V17 + VTM@11.07Mbps, (c) Frog pose p03, A17 + VTM@23.55Mbps, (d) Frog pose p03, V17 + VTM@23.60Mbps, (e) Fan pose p03, A17 + VTM@11.19Mbps, (f) Fan pose p03, V17 + VTM@10.68Mbps, (g) Group pose p03, A17 + VTM@4.54Mbps, (h) Group pose p03, V17 + VTM@4.53Mbps

표 10. MIV 시점 모드 + HM 대비 MIV 모드 + HM의 MIV 부호화, 영상 부호화, 영상 복호화, 중간 뷰 합성, 포즈 트레이스 합성 시 실행 시간 비율 (낮은 값이 시간 절감을 나타냄)

Table 10. Runtime ratios of the MIV mode + HM compared to the MIV view mode + HM for MIV encoding, video encoding, video decoding, intermediate view synthesis, and pose trace synthesis (Low value indicates time saving)

| | MIV encoding | Video encoding | Video decoding | Intermediate view synthesis | Pose trace synthesis |
|-----------------------|--------------|----------------|----------------|-----------------------------|----------------------|
| <i>ClassroomVideo</i> | 345.54% | 89.40% | 99.82% | 261.20% | 264.40% |
| <i>Museum</i> | 457.51% | 102.47% | 104.83% | 293.80% | 340.64% |
| <i>Fan</i> | 502.68% | 107.53% | 103.93% | 133.39% | 129.21% |
| <i>Kitchen</i> | 458.71% | 101.37% | 96.06% | 164.18% | 159.56% |
| <i>Painter</i> | 414.25% | 100.27% | 99.54% | 128.06% | 123.99% |
| <i>Frog</i> | 418.45% | 103.39% | 102.02% | 121.90% | 122.85% |
| <i>Carpark</i> | 347.49% | 97.15% | 97.39% | 117.63% | 113.74% |
| <i>Chess</i> | 557.88% | 100.11% | 100.53% | 165.86% | 175.48% |
| <i>Group</i> | 516.12% | 108.22% | 100.72% | 164.86% | 157.99% |
| <i>ChessPieces</i> | 565.41% | 100.29% | 97.65% | 163.07% | 176.04% |
| <i>Hijack</i> | 413.11% | 105.10% | 104.28% | 216.33% | 188.28% |
| <i>Average</i> | 454.29% | 101.39% | 100.62% | 175.48% | 177.47% |

되었고, 비정형 카메라 배열을 가져 영상 간 중복성이 크지 않은 ChessPieces, Chess, Group 시퀀스들의 경우 각각 465.41%, 457.88%, 416.12%의 시간 증가율을 보였다. HM 을 이용한 영상 부호화 시 MIV 모드가 MIV 시점 모드 대비 1.39%만의 시간 증가율을 보였고, 특히 추가 시점 패치의 수가 적어 아틀라스 내 빈 공간이 많았던 Classroom-Video에서는 10.60%의 영상 부호화 시간 절감이 확인되었다. 영상 복호화 시에는 MIV 모드와 MIV 시점 모드 간 실행 시간 차이가 미미했고, MIV 모드에서 0.62%의 시간 증가율을 보였다. 중간 뷰 합성, 포즈 트레이스 합성 시 패

치의 수가 MIV 시점 모드 대비 많은 MIV 모드가 각각 75.48%, 77.47%의 시간 증가율을 보였다. Carpark 시퀀스의 경우 뷰 수가 9개로 타 시퀀스 대비 적고, 1-D 정형 카메라 배열에서 취득된 점이 작용하여 MIV 모드를 사용하였을 때 중간 뷰 합성, 포즈 트레이스 합성 시 각각 17.63%, 13.74%만큼의, 타 시퀀스 대비 가장 낮은 시간 증가율을 보였다.

표 11은 VTM과 MIV 및 MIV 시점 모드에서의 실행 시간 비율을 나타낸다. MIV 부호화는 HM, VTM으로 영상 부호화 전 이루어지므로 시간 증가율은 앞서 표 10에서 설

표 11. MIV 시점 모드 + VTM 대비 MIV 모드 + VTM의 MIV 부호화, 영상 부호화, 영상 복호화, 중간 뷰 합성, 포즈 트레이스 합성 시 실행 시간 비율 (낮은 값이 시간 절감을 나타냄)

Table 11. Runtime ratios of the MIV mode + VTM compared to the MIV view mode + VTM for MIV encoding, video encoding, video decoding, intermediate view synthesis, and pose trace synthesis (Low value indicates time saving)

| | MIV encoding | Video encoding | Video decoding | Intermediate view synthesis | Pose trace synthesis |
|-----------------------|--------------|----------------|----------------|-----------------------------|----------------------|
| <i>ClassroomVideo</i> | 345.54% | 69.54% | 110.30% | 257.81% | 261.13% |
| <i>Museum</i> | 457.51% | 108.24% | 114.39% | 296.35% | 342.40% |
| <i>Fan</i> | 502.68% | 111.14% | 119.71% | 126.97% | 128.10% |
| <i>Kitchen</i> | 458.71% | 90.41% | 104.79% | 157.73% | 160.56% |
| <i>Painter</i> | 414.25% | 79.57% | 94.99% | 121.10% | 124.05% |
| <i>Frog</i> | 418.45% | 86.74% | 100.81% | 113.88% | 121.40% |
| <i>Carpark</i> | 347.49% | 77.36% | 102.43% | 109.03% | 113.76% |
| <i>Chess</i> | 557.88% | 91.65% | 104.96% | 155.76% | 178.10% |
| <i>Group</i> | 516.12% | 100.64% | 112.37% | 157.17% | 161.18% |
| <i>ChessPieces</i> | 565.41% | 96.10% | 104.91% | 163.46% | 178.30% |
| <i>Hijack</i> | 413.11% | 82.59% | 114.87% | 213.55% | 187.65% |
| <i>Average</i> | 454.29% | 90.36% | 107.68% | 170.25% | 177.88% |

명된 바와 같다. 영상 부호화 시 MIV 모드에서 9.64%의 시간 절감이 확인되었는데, HM 기반 부호화 시 MIV 모드에서 1.39%의 시간 증가가 확인된 바와 대조적이다. 이는 영상을 분할하는 기본 단위인 coding tree unit (CTU)의 크기가 HM에서는 64인데 반해 VTM에서는 128로 면적이 4배 증가하여 빈 공간에 대한 부호화 속도가 증가하였기 때문이다. 반면 영상 복호화 시 MIV 모드는 7.68%의 복호화 시간 증가를 기록하였으나, MIV 부호화만큼의 높은 시간 증가폭을 보여주지는 않았다. 중간 뷰 합성, 포즈 트레이스 합성 시 VTM 기반 결과는 HM과 큰 차이를 보이지 않았고, MIV 모드에서 각각 70.25%, 77.88%의 시간 증가율을 보였다.

V. 결론

본 논문은 MIV 기반 6DoF 몰입형 영상 스트리밍 시스템의 효율을 검증하고 각 테스트 시퀀스 특성에 따른 실험 결과 해석을 소개하였다. 전체 시점에서 일부 시점만을 선택하여 전송하는 MIV 시점 모드 대비 영상 간 중복성을 제거하고 잔차 패치들을 병합하여 전송하는 MIV 모드에서 몰입형 영상 테스트 시퀀스 전반에 걸쳐 BD-rate 절감을 확인하였다. Y-PSNR, VMAF, IV-PSNR 등의 평가 기법 중, 뷰 합성 시 발생하는 아티팩트를 주관적 화질 관점에서 잘 반영하는 IV-PSNR이 6DoF 영상 품질 평가에 유리했다. 1-D, 2-D 정형 카메라 배열로 이루어졌거나 시점 간 차이가 크지 않은 몰입형 영상에 대해서는 MIV 시점 모드가 BD-rate 측면에서 효율적이고, 실제로 주관적 화질에서도 품질 향상이 있었으며, MIV 모드 대비 MIV 부호화 및 MIV 복호화 시 높은 시간 절감을 달성할 수 있었다. 향후 연구로 더 높은 수준의 몰입감을 제공하는, 더 많은 시점을 포함하며 해상도가 높은 몰입형 영상에 대해 MIV 성능 분석을 진행하고, 이를 적응적 모드 선택에 적용할 수 있다면 더 효과적인 몰입형 미디어를 위한 6DoF 영상 스트리밍 시스템을 구축할 수 있을 것이다.

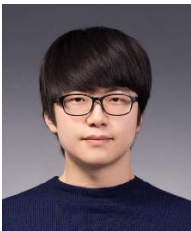
참고 문헌 (References)

- [1] M.-L. Champel, T. Stockhammer, T. Fautier, E. Thomas, R. Koenen. 2016. Quality Requirements for VR. 116th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 116/m39532.\
- [2] T. T. Le, J. B. Jeong, E. -S. Ryu. 2019. Efficient transcoding and encryption for live 360 CCTV system. Applied Sciences, Vol. 9, No. 4. doi: <https://doi.org/10.3390/app9040760>
- [3] J. -B. Jeong, S. Lee, I. Kim, S. Lee, E. -S. Ryu. 2020. Implementing VVC Tile Extractor for 360-degree Video Streaming Using Motion-Constrained Tile Set. Journal of Broadcast Engineering (JBE), Vol. 25, No. 7, pp. 1073-1080. doi: <https://doi.org/10.5909/JBE.2020.25.7.1073>
- [4] T. T. Le, J. -B. Jeong, S. S. Lee, J. Kim, E. -S. Ryu. 2021. An Efficient Viewport-Dependent 360 VR System Based on Adaptive Tiled Streaming. CMC-Computers, Materials & Continua, Vol. 66, No. 3, pp. 2627-2643. doi: <https://doi.org/10.32604/cmc.2021.013399>
- [5] J. -B. Jeong, S. Lee, I. Kim, E. -S. Ryu. 2021. Implementing Viewport Tile Extractor for Viewport-Adaptive 360-Degree Video Tiled Streaming. International Conference on Information Networking 2022 (ICOIN2022), pp. 8-12. doi: <https://doi.org/10.1109/ICOIN50884.2021.9333964>
- [6] S. Lee, D. Jang, J. B. Jeong, E. -S. Ryu. 2019. Motion-constrained tile set based 360-degree video streaming using saliency map prediction. ACM Network and Operating System Support for Digital Audio and Video 2019 (NOSSDAV2019), pp. 20-24. doi: <https://doi.org/10.1145/3304112.3325614>
- [7] A. Vetro, T. Wiegand, G. J. Sullivan. 2011. Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard. Proceedings of the IEEE, Vol. 99, No. 4, pp. 626-642. doi: <https://doi.org/10.1109/JPROC.2010.2098830>
- [8] 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>
- [9] R. Doré, F. Thudor. 2018. Outperforming 3DoF+ Anchors: first evidence. 123th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2018/m43504.
- [10] J. -B. Jeong, S. Lee, D. Jang, E. -S. Ryu. 2019. Towards 3DoF+ 360 Video Streaming System for Immersive Media. IEEE Access, Vol. 7, pp. 136399-136408. doi: <https://doi.org/10.1109/ACCESS.2019.2942771>
- [11] ISO/IEC JTC1/SC29/WG11. 2019. Call for Proposals on 3DoF+ Visual. 125th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG/n18145.
- [12] J. Fleureau, F. Thudor, R. Doré, M. Dmitrichenko, B. Salahieh, J. Boyce. 2019. Description of Technicolor Intel response to MPEG-I 3DoF+ Call for Proposal. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2019/m47445.
- [13] M. Domański, A. Dziembowski, D. Mieloch, O. Stankiewicz, J.

- Stankowski, A. Grzelka, G. Lee, J. Y. Jeong, J. Seo. 2019. Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by Poznań University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI). 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2019/m47407.
- [14] B. Kroon, B. Sonneveldt. 2019. Philips response to CFP on 3DoF+. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2019/m47179.
- [15] V. K. M. Vadakital, K. Roimela, L. Ilola, J. Keränen, M. Pesonen, S. Schwarz, J. Iainema, M. Hannuksela. 2019. Description of Nokia's response to CFP for 3DoF+ visual. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2019/m47372.
- [16] B. Wang, Y. Sun, L. Yu. 2019. Description of Zhejiang University's response to 3DoF+ Visual CFP. 126th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2019/m47684.
- [17] ISO/IEC JTC 1/SC 29/WG 4. 2021. Text of ISO/IEC FDIS 23090-12 MPEG Immersive Video. 135th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/n00111.
- [18] 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>
- [19] Jongseok Lee, Juntaek Park, Hansol Choi, Joohyung Byeon, Donggyu Sim, "Overview of VVC", *Broadcast and Media Magazine*, pp.10-25. Vol. 24, No. 3, Oct. 2019.
- [20] E. S. Jang, M. Preda, K. Mammou, A. M. Tourapis, J. Kim, D. B. Graziosi, S. Rhyu, M. Budagavi. 2019. Video-based point-cloud-compression standard in MPEG: From evidence collection to committee draft [standards in a nutshell]. *IEEE Signal Processing Magazine*, Vol. 36, No. 3, pp. 118-123.
doi: <https://doi.org/10.1109/MSP.2019.2900721>
- [21] D. B. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, A. Tabatabai. 2020. An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, Vol. 9, pp. 1-17.
doi: <https://doi.org/10.1017/ATSIP.2020.12>
- [22] DaeHyeok Gwon, Heeji Han, Haechul Choi, "Video based Point Cloud Compression with Versatile Video Coding" 2020 The Korean Institute of Broadcast and Media Engineers Summer Conference, pp. 375-377, July, 2020.
- [23] 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. 135th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 7, MPEG/n00188.
- [24] J. -B. Jeong, S. Lee, I. -W. Ryu, T. T. Le, E. -S. Ryu. 2020. Towards Viewport-dependent 6dof 360 Video Tiled Streaming for Virtual Reality Systems. *ACM International Conference on Multimedia 2020 (MM 2020)*, pp. 3687 - 3695.
doi: <https://doi.org/10.1145/3394171.3413712>
- [25] S. Lee, J. -B. Jeong, I. Kim, S. Lee, E. -S. Ryu. 2020. Screen Content Coding Analysis to Improve Coding Efficiency for Immersive Video. *Journal of Broadcast Engineering (JBE)*, Vol. 25, No. 6, pp. 911-921.
doi: <https://doi.org/10.5909/JBE.2020.25.6.911>
- [26] L. Kondrad, V. K. M. Vadakital, L. Ilola. 2020. CE-1.3: frame packed video sub-bitstream type. 131th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 11, MPEG/m54274.
- [27] B. Salahieh, G. Naf, J. Boyce. 2021. Frame Packing Implementation in TMIV. 134th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/m56827.
- [28] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. [MIV] Extraction and Merging on Frame Packed Video. 134th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/m56591.
- [29] 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>
- [30] 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>
- [31] ISO/IEC JTC 1/SC 29/WG 11. 2018. Reference View Synthesizer (RVS) manual. 124th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 11, MPEG/n18068.
- [32] S. Facheada, D. Bonatto, A. Schenkel, G. Lafruit. 2018. Depth image based view synthesis with multiple reference views for virtual reality. *The True Vision-Capture, Transmission and Display of 3D Video 2018 (3DTV-CON 2018)*, pp. 1-4.
doi: <https://doi.org/10.1109/3DTV.2018.8478484>
- [33] S. Kwak, J. Yun, J. Jeong, W. -S. Cheong, J. Seo. 2020. [MPEG-I Visual] Ray-based blending weight for 6DoF view synthesis. 131th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 11, MPEG/m54409.
- [34] J. -B. Jeong, S. Lee, E. -S. Ryu. 2021. On OpenGL-based View Synthesizer Optimization for Real-time 6DoF Immersive Video Streaming. *IEEE P3079.1 Contribution (DCN: 3079-21-0057-00-0001)*.
- [35] D. Bonatto, S. Fachada, G. Lafruit. 2020. RaViS: Real-time accelerated View Synthesizer for immersive video 6DoF VR. *Electronic Imaging*, pp. 382-1-382-9.
doi: <https://doi.org/10.2352/ISSN.2470-1173.2020.13.ERVR-381>
- [36] H. Ahn, M. -j. Lee. 2020. MPEG-I RVS Software Speed-up for Real-time Application. *Journal of Broadcast Engineering (JBE)*, Vol. 25, No. 5, pp. 655-664.
doi: <https://doi.org/10.5909/JBE.2020.25.5.655>
- [37] S. Kwak, J. Yun, J. -Y. Jeong, Y. Kim, I. Ihm, W. -S. Cheong, J. Seo. 2021. View synthesis with sparse light field for 6DoF immersive video. *ETRI Journal*, Vol. 44, No. 1, pp. 24-37.
doi: <https://doi.org/10.4218/etrij.2021-0205>
- [38] J. Jung, B. Kroon. 2021. Common Test Conditions for MPEG Immersive Video. 133th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/n00051.
- [39] B. Salahieh, J. Jung, A. Dziembowski, C. Bachhuber. 2021. Test

- Model 8 for MPEG Immersive Video. 133th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/n00050.
- [40] HM reference software, [Online]. Available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.
- [41] VTM reference software, [Online]. Available at: https://vcgit.hhi.fraunhofer.de/jvet/VVCSSoftware_VTM/.
- [42] Y. Sun, A. Lu, L. Yu. 2017. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. IEEE signal processing letters, Vol. 24, No. 9, pp. 1408 - 1412.
doi: <https://doi.org/10.1109/LSP.2017.2720693>
- [43] C. G. Bampis, A. C. Bovik, and Z. Li. 2018. A simple prediction fusion improves data-driven full-reference video quality assessment models. Picture Coding Symposium 2018 (PCS2018), pp. 298 - 302.
doi: <https://doi.org/10.1109/PCS.2018.8456293>
- [44] 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, doi: <https://doi.org/10.1109/TCSVT.2022.3179575>
- [45] J. -B. Jeong, S. Kim, S. Lee, I. Kim, E. -S. Ryu. 2021. Super Metric: Quality Assessment Methods for Immersive Video. Journal of Internet Computing and Services (JICS), Vol. 22, No. 2, pp. 51-58.
doi: <https://doi.org/10.7472/jksii.2021.22.2.51>
- [46] J. Jung, B. Kroon. 2021. Common Test Conditions for MPEG Immersive Video. 134th MPEG meeting of ISO/IEC JTC 1/SC 29/WG 4, MPEG/n00085.
- [47] Versatile video encoder (VVenc), [Online]. Available at: <https://github.com/fraunhoferhhi/vvenc/>.
- [48] F. Bossen, K. Sühring, A. Wiecekowsi, S. Liu. 2021. VVC Complexity and Software Implementation Analysis. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 31, No. 10, pp. 3765-3778.
doi: <https://doi.org/10.1109/TCSVT.2021.3072204>

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
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



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>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준

저 자 소 개



김 인 애

- 2013년 8월 : 배재대학교 외식경영학 학사
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육학과 석사과정
- ORCID : 0000-0003-4263-6448
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



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>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 코딩 및 국제 표준, HMD/VR 응용분야