

특집논문 (Special Paper)

방송공학회논문지 제24권 제6호, 2019년 11월 (JBE Vol. 24, No. 6, November 2019)

<https://doi.org/10.5909/JBE.2019.24.6.1035>

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

6 자유도 전방위 몰입형 비디오의 압축 코덱 개발 및 성능 분석

박 현 수^{a)}, 박 상 효^{a)}, 강 제 원^{a)†}

Toward 6 Degree-of-Freedom Video Coding Technique and Performance Analysis

Hyeonsu Park^{a)}, Sang-hyo Park^{a)}, and Je-Won Kang^{a)†}

요 약

최근 몰입형 비디오의 수요가 점차 늘어남에 따라 국제 표준 단체인 MPEG-I에서 전방위 몰입형 비디오의 처리 기술이 활발하게 개발 중이다. 전방위 몰입형 비디오는 사용자 시점의 자유도가 증가함에 따라 비디오 신호의 크기가 급격히 증가하여 효과적인 압축 기술이 필수적이다. 더욱이 사용자의 움직임에 따른 보다 자유로운 시점 변환을 지원하는 6 자유도 (6-Degree-of_Freedom, 6DoF) 비디오의 압축을 위해서는 보다 우수한 부호화 효율을 제공하는 코덱의 개발이 필요하다. 본 논문에서는 ISO/IEC 23090 Part 7 (Metadata for Immersive Media (Video))에서 진행 중인 몰입형 비디오의 압축 표준 프로젝트의 테스트 모델인 TMIV (Test Model for Immersive Video)에 기준 적용된 High Efficiency Video Coding (HEVC)를 최근 차세대 비디오 압축 표준 개발 중인 Versatile Video Coding (VVC)로 대체하여 성능 분석을 수행하고, VVC의 툴 분석으로부터 디블로킹 필터를 TMIV의 패치 아틀라스에 선택적으로 적용하는 것이 부호화 효율을 증대시킬 수 있음을 보인다. VVC 기반의 6 DoF 비디오 코덱의 성능 평가는 본 논문이 최초로 그에 따른 향후 6DoF 지원 몰입형 비디오 표준 개발 방향을 제시한다. TMIV의 두 가지 작동 모드인 MIV (Metadata for Immersive Video) 모드와 MIV 시점 모드에서 공통 실험 조건에 명시된 일곱 가지 시퀀스에 대해 전체적으로 실험을 진행하였다. 기준 HEVC를 VVC로 대체함으로써 MIV 모드 방식에서 33.8%, MIV 시점 모드에서 30.2%의 Peak Signal-to-Noise Ratio (PSNR) 관점에서의 부호화 성능 향상을 제공하였다. 이외에도 3차원 비디오의 인지 화질 평가를 위하여 사용하는 평가 지표로 IV-PSNR (Immersive Video PSNR)와 MSSIM (Mean Structural Similarity)를 이용하여 성능을 평가하였다.

Abstract

Recently, as the demand for immersive videos increases, efficient video processing techniques for omnidirectional immersive video is actively developed by MPEG-I. While the omnidirectional video provides a larger degree of freedom for a free viewpoint, the size of the video increases significantly. Furthermore, in order to compress 6 degree-of-freedom (6 DoF) videos that support motion parallax, it is required to develop a codec to yield better coding efficiency. In this paper, we develop a 6 DoF codec using Versatile Video Coding (VVC) as the next generation video coding standard. To the authors' best knowledge, this is the first VVC-based 6 DoF video codec toward the future ISO/IEC 23090 Part 7 (Metadata for Immersive Media (Video)) MPEG-I standardization. The experiments were conducted on the seven test video sequences specified in Common Test Condition (CTC) in two operation modes of TMIV (Test Model for Immersive Media) software. It is demonstrated that the proposed codec improves coding performance around 33.8% BD-rate reduction in the MIV (Metadata for Immersive Video) mode and 30.2% BD-rate reduction in the MIV view mode as compared to the state-of-the-art TMIV reference software. We also show the performance comparisons using Immersive Video PSNR (IV-PSNR) and Mean Structural Similarity (MSSIM).

Keyword : 6DoF video, MPEG-I, TMIV software, HEVC, VVC

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

I. 서론

최근 가상 및 증강 환경에서의 새로운 몰입형 비디오 서비스 기술에 대한 수요가 높아지고 있다. 전방위 비디오의 취득 기술이 점차 발전하고 있고 경제적인 Head Mounted Display (HMD)가 잇달아 등장함에 따라 몰입형 비디오는 사용자에게 보다 높은 시점 자유도를 제공함으로써 보다 현장감 있는 경험을 제공하고 있다. 시점 자유도 (Degree of Freedom, DoF)란 3차원 공간에서 물체의 위치 및 움직임을 나타내기 위해 도입된 개념으로, 3 자유도 (3DoF) 비디오는 전방향 VR 영상 시청자의 머리를 중심으로 Yaw, Pitch, Roll과 같이 총 세 가지 방향의 자유도를 제공하는 비디오를 의미한다. 3DoF+는 머리를 중심으로 약간의 움직임 자유도를 제공하며, 더 나아가 6 자유도 비디오 (6 degree-of-freedom, 6DoF)는 시청자의 머리 회전 운동뿐만 아니라 3차원 공간에서의 움직임에 따른 자유로운 시점 변환을 제공한다.

MPEG에서는 2016년 10월 116차 중국 청두 회의에서부터 몰입형 미디어를 위한 차세대 프로젝트인 MPEG-I (Immersive) 표준화를 시작하였다. 특히 ISO/IEC 23090 Part 7 (Metadata for Immersive Media (Video))에서는 전방위 몰입형 비디오의 압축 기술을 개발 중으로, 현재 1단계로 3DoF+ 비디오의 압축 표준을 2020년 초 완료를 목표로 개발 중이다. ISO/IEC 23090 Part 7 (Metadata for Immersive Media (Video))의 두번째 단계이자 최종 목표는 6 DoF 비디오의 압축 기술 개발이다^[1].

6DoF 비디오는 그림 1과 같이 크게 두가지 방식으로 취득할 수 있다. 첫째로, Windowed 6DoF는 다시점 카메라 시스템으로부터 취득한다. 그로 인해 창문 형태의 영역과

같이 사용자가 바라보는 현재 및 이웃의 시점을 평행 이동으로 제한하고 있다. 즉, 시청자가 Windowed 6DoF 비디오를 시청할 때 해당 윈도우 밖을 볼 수 없다. 둘째로, Omnidirectional 6DoF는 360도 비디오를 다시점으로 구성하여 사용자 시점에 맞추어 제한된 공간에서 시청 자유도를 제공한다. 한 예로, 시청자는 HMD를 착용하고 제한된 영역에서 3차원의 전방위 가상 환경을 경험할 수 있다^[2].

그림 1과 같이 Windowed 6DoF 및 Omnidirectional 6DoF 비디오는 그 취득을 위하여 많은 수의 시점을 취득해야 할 뿐만 아니라 전방위 비디오에서 특정 시점을 선택할 수 있도록 초고화질의 해상도를 가져야한다. 그에 따라 전송 및 압축을 위해서는 막대한 양의 비디오 데이터를 처리해야 한다. 효율적인 몰입형 비디오의 처리 및 압축을 위하여 MPEG-I 표준을 중심으로 여러 기술들이 소개가 되었다. 대표적으로 MPEG-I에서는 6DoF 비디오의 압축 코덱 개발 위한 탐색 실험 (EE)을 깊이 정보의 추정에 관한 실험^[5], 중간 시점의 합성에 관한 실험^[6], 시점 변환에 관한 실험^[7], 그리고 핵심 압축 기술 개발에 관한 실험^[8]으로 총 네 가지 실험으로 나누어 진행하였다. 깊이 추정을 위해서는 Depth Estimation Reference Software (DERS)^[9]의 성능을 개선하고 있으며, 중간 시점의 합성을 위해서는 지난 124차 마카오 회의에서 Versatile View Synthesis (VVS)^[10]를 새로운 합성 기술을 위한 공개 소프트웨어로 채택하여 기존 View Synthesis Reference Software(VSRS)를 대체하여 주관적 객관적 성능을 향상시켰다. 시점 변환에서는 전방위 시청 환경 내에서 시청자의 위치 및 시점을 변환하며 주관적 화질 평가를 측정하는 실험이 계속되었다^[7]. 탐색 실험 외에도 시점 간의 색상 톤 편차를 줄여 코딩 효율을 올리기 위한 화질 개선 기술이 제안되었다^[11].

제 120차 마카오 회의에서는 3DoF+ 비디오 압축 표준 제정에 관한 제안 요청에 대응한 다섯 개의 관련 응답을 평가하고 공통적으로 채택된 기술들을 구현한 TMIV (Test Model for Immersive Video) 공개 소프트웨어를 개발하였다^{[12][13]}. 126차 제네바 회의에서 테크니컬러 및 인텔에서는 입력 시점 중 중앙 시점을 설정하여 이외의 시점에서 중앙 시점 간의 차이를 제거한 패치 아틀라스 기술을 제안하였다^[14]. Poznan 대학과 한국전자통신연구원에서는 입력을 기본 계층과 잔차 계층으로 나누어 처리하는 기술을 제안

a) 이화여자대학교 엘텍공과대학 전자전기공학과(Department of Electronic and Electrical Engineering, Ewha W. University)

‡ Corresponding Author : 강제원(Je-Won Kang)

E-mail: jewonk@ewha.ac.kr
Tel: +82-3277-2347

ORCID: <https://orcid.org/0000-0002-1637-9479>

※ This work was supported by IITP grant funded by the Korea government (MSIP)(2018-0-00765 Development of Compression and Transmission Tech. for Ultra High Quality Immersive Videos Supporting 6DoF.

· Manuscript received October 4, 2019; Revised November 8, 2019; Accepted November 8, 2019.

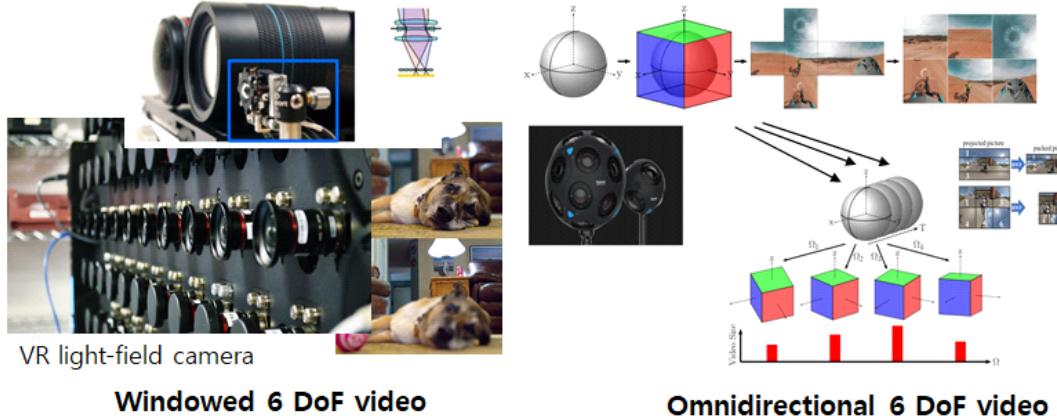


그림 1. Windowed 6DoF (좌) 와 Omnidirectional 6DoF (우) 비디오의 취득 방식 [3][4]
 Fig. 1. Windowed 6DoF (left) and Omnidirectional 6DoF (right) videos [3][4]

하였다^[15]. 노키아에서는 모자이크 기반의 블록 기반 아틀라스 생성 기술을 제안하였고, 중국 제장 대학교에서는 서브 띵쳐 기반의 기술을, 필립스에서는 블록 트리 기반의 기술을 제안하였다^{[16][17][18]}. 아래의 표는 각 기관의 제안 응답 (Call for Response)의 주요 특징과 성능을 정리한 것으로 모든 시점을 HEVC로 압축하는 방식 대비 평균적으로 약 -73.1%에서 -7.3%까지 향상된 부호화 성능을 보였고 기관 별로 보였고, 이 중 성능이 가장 우수한 Technicolor-Intel의 응답을 중심으로 각 기관의 주요 기술을 통합하여 TMIV가 기본 모델로 구현이 되었다. 성능 등 보다 구체적인 내용은 각 기관의 제안 응답에서 확인할 수 있다^[14-18]

표 1. 3DoF+ 다섯 개 기관의 제안 응답의 주요 기술 및 성능
 Table 1. 3DoF+ Main Technique and Performance of Five Organizations Proposed Response

Institute	Core Methods
Technicolor-Intel	Patch Atlas
PUT/ETRI	Layer Separation
Nokia	Mosaic
Zhejiang University	Sub-picture
Philips	Block Tree

몰입형 비디오의 특성상 일반적인 방송 및 스트리밍 서비스에 활용하는 영상보다 훨씬 더 많은 시점을 전달해야

하며, 이에 따라 주어진 네트워크 대역폭에 맞는 고효율 압축기술이 필요하다. 현재 ISO/IEC 23090 Part 7에서는 시점의 색상 및 깊이 영상의 압축을 위하여 현재 High Efficiency Video Coding (HEVC)를 사용하고 있다. 최근, MPEG 과 VCEG이 공동으로 차세대 동영상 압축 기술인 Versatile Video Coding (VVC)을 공동으로 표준화하고 있으며, 최근 보고에 따르면, VVC test model (VTM)이 HEVC test model (HM)보다 임의 접근 (Random-access, RA) 구성 환경에서 평균 33%의 비트율 절감을 보인다^[19]. 특히 VVC는 기존 2차원 비디오뿐만 아니라 전방위 비디오를 포함하는 다양한 비디오 콘텐츠를 포괄적으로 압축할 수 있도록 하는 것을 목표로 개발이 되고 있다.

본 논문에서는 ISO/IEC 23090 Part 7에서 진행 중인 몰입형 비디오의 압축 표준 프로젝트의 테스트 모델인 TMIV에 기존 적용된 HEVC를 VVC로 대체하여 성능 분석을 수행하고, 향후 6DoF 지원 몰입형 비디오 표준의 개발 방향을 제시한다. TMIV의 두 가지 작동 모드와 공통 실험 조건에 명시된 일곱 가지 시퀀스에 대해 전체적으로 실험을 진행하였다. 기존 HEVC를 VVC로 대체함으로써 MIV (Metadata for Immersive Video) 모드에서 33.8%, MIV 시점 모드에서 30.2%의 부호화 성능 향상을 제공하였다. MIV 모드로 인코딩을 하게 될 경우, 일반적인 기본 시점의 영상은 실험 결과와 같이 VVC를 이용하여 효율적인 압축이 가능하다.

그러나 TMIV의 아틀라스 패치 영상의 경우, 일반적인 자연 동영상이나 스크린 콘텐츠의 특성과 다르다. 즉, 영상 하나에 고주파의 많은 텍스쳐 경계가 존재한다. 이러한 아틀라스 패치는 인접 시점과의 연관성을 줄이기 위한 목적으로 생성이 되므로, 인지 화질과는 큰 연관이 없다. 따라서 본 논문에서는 VVC 부호화 기술의 On/Off를 통한 실험으로부터 디블로킹 필터를 TMIV의 패치 아틀라스에 선택적으로 적용하는 것이 부호화 효율을 증대시킬 수 있음을 보인다.

본 논문은 다음과 같이 구성한다. 2장에서는 6DoF 비디오 코덱의 개발을 설명하고, 3장에서는 사용한 6DoF 비디오의 특성 및 성능평가 방식을 설명하며, 4장에서는 성능 평가를 수행한다. 5장에서 결론을 맺는다.

II. 6DoF 비디오 코덱 개발

1. 6DoF 비디오 코덱 개요

제안하는 6DoF 비디오 코덱의 전체적인 구성은 그림 2에서 보이듯이 3DoF+를 목표로 진행 중인 TMIV의 개발적인 구조를 따랐다. 6DoF 비디오 인코더와 디코더는 세 개의 주요 단계로 이루어진다. 인코더의 경우 시점 최적화 (View Optimizer), 아틀라스 생성기(Atlas Constructor), 그리고 비디오 인코더 (Video Encoder)와 메타데이터 컴포서 (Metadata Composer)로 구성된다. 디코더의 경우에 반대 방향으로 비디오 디코더 (Video Decoder)와 메타데이터 파서 (Metadata Parser), 아틀라스 패치 점유 지도 생성기 (Atlas Patch Occupancy Map Generator), 그리고 렌더러 (Renderer)로 구성된다^[13]. 또한, TMIV 소프트웨어와 같이 MIV 모드와 MIV 시점 모드의 두 가지 모드로 작동 가능하다^[20]. MIV 모드에서는 시퀀스의 모든 시점을 기본 시점과 추가 시점으로 구성한 아틀라스를 이용하여 전송하며, MIV 시점 모드에서는 시점을 선택적으로 전송하고 중간 시점을 깊이 정보를 이용하여 생성한다. MIV 시점 모드는 렌더러의 성능을 평가할 수 있는 특징이 있다. 비디오 인코더 및 디코더는 비디오의 텍스쳐 정보와 깊이 정보 모두를 압축하여 전송한다. 본 논문에서는 비디오 인코더 및 디코

더의 개발 코덱을 HEVC 참조 소프트웨어 대신 VVC 참조 소프트웨어로 대체하여 구성하였다.

1.1 MIV 모드

그림 2에서 푸른색으로 표현한 MIV 모드의 워크 플로우에서 보이듯이 전체 시점의 비디오 입력 데이터는 시점 최적화를 통해 기본 시점 (basic view)과 추가 시점 (additional view)로 구분이 된다. 예를 들어 그림 2에서 전체 입력 시점이 v0~v15까지 총 16개의 시점이 있을 때 기본 시점은 v6로 추가 시점은 그 외의 시점으로 표현하였다. 기본 시점은 전체 시점에서 가장 공통의 성분을 갖는 시점이 될 수 있다. 기본 시점 및 추가 시점이 선택이 되게 되면 기본 시점은 보존이 되어 부호기를 거친다. 그러나 추가 시점의 입력 비디오는 아틀라스 생성기를 거쳐 공통의 신호를 제거한 잔차 패치를 비디오에 채운 하나의 아틀라스 비디오로 생성이 된다. 해당 아틀라스 지도 정보는 복호기로 전달되어 복호 과정에서 렌더러가 최종 결과를 출력할 수 있도록 한다^[13]. v0~v15에 해당하는 시점이 기본 시점과 추가 시점의 아틀라스로부터 복원 비디오로 출력된다.

1.2 MIV 시점 모드

MIV 시점 모드에서는 아틀라스의 생성이 없이 기본 시점 만을 압축하여 전송한다. 그러나 MIV 모드와는 다르게 전체 시점의 비디오를 모두 전송하는 것이 아니라 일부의 기본 시점 만을 선택하여 전송하고 복호 측에서 중간 시점을 합성한다. 그림 2의 붉은색으로 표현한 MIV 시점 모드의 워크 플로우에서 선택된 시점들이 시점 최적화와 아틀라스 생성기를 대부분 거치지 않고 처리함을 보인다. 아틀라스 생성기의 패치 구성에서는 잔차 패치를 비디오에 패킹하여 구성하는 것이 아니라 선택된 시점을 각각의 아틀라스에 배치한다. 그에 따라 디코더에서는 아틀라스 패치 점유 지도 생성기에서 아틀라스 당 단일 패치 아이디를 생성하여 렌더러로 출력한다. 렌더러는 메타 데이터 및 복호된 아틀라스, 그리고 깊이 정보를 사용하여 원하는 시점에서의 영상을 출력한다^[13]. 예를 들어, 그림 2에서는 $vn^*(n = 1, 2, 4, 7, 8, 11, 13, 14)$ 는 선택된 시점으로부터 깊이 정보를 함께 이용하여 합성된 시점들이다.

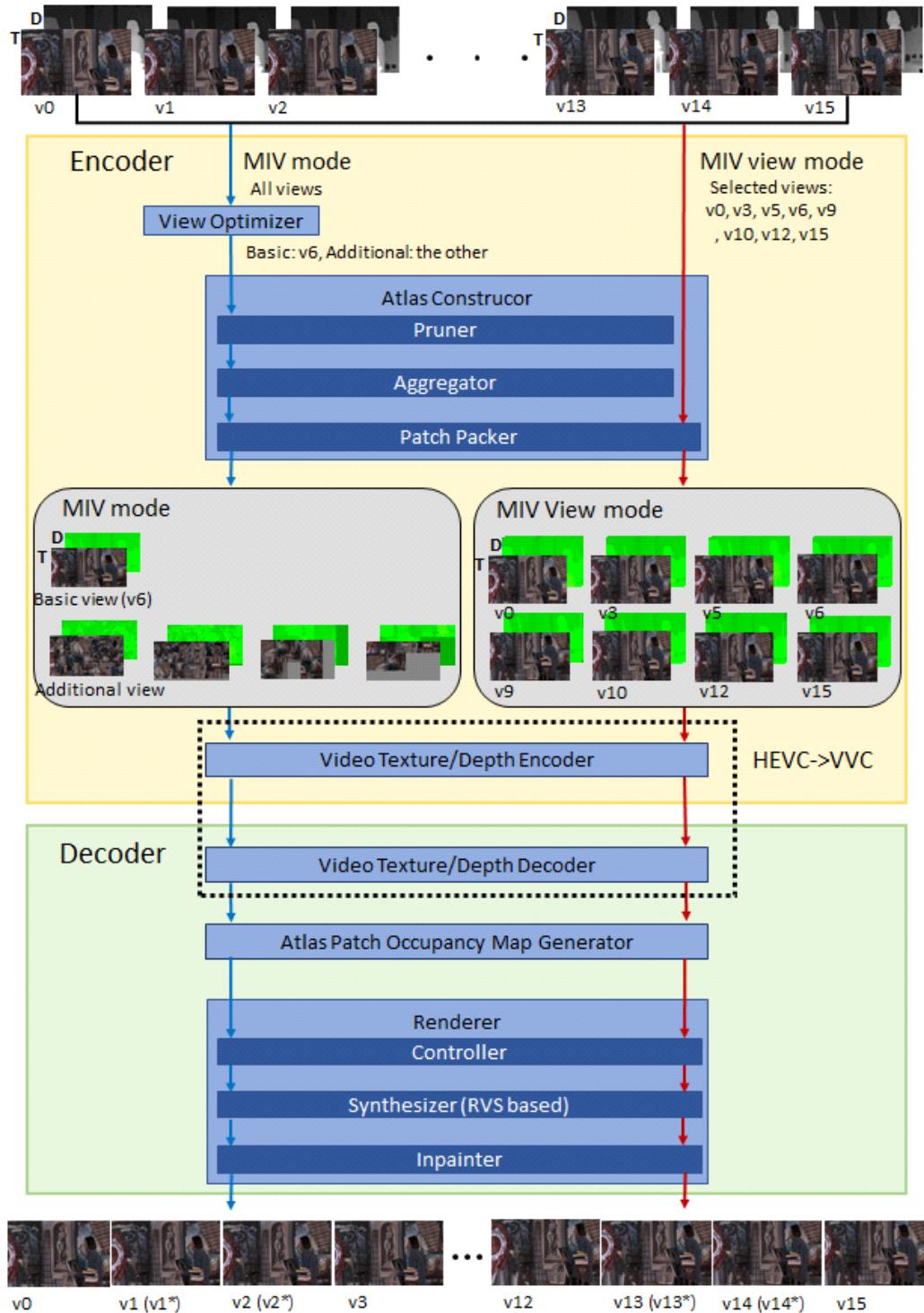


그림 2. TMIV의 부호기 및 복호기의 구성 및 주요 구성요소
 Fig. 2. Architecture and Main Component of TMIV Encoder and Decoder

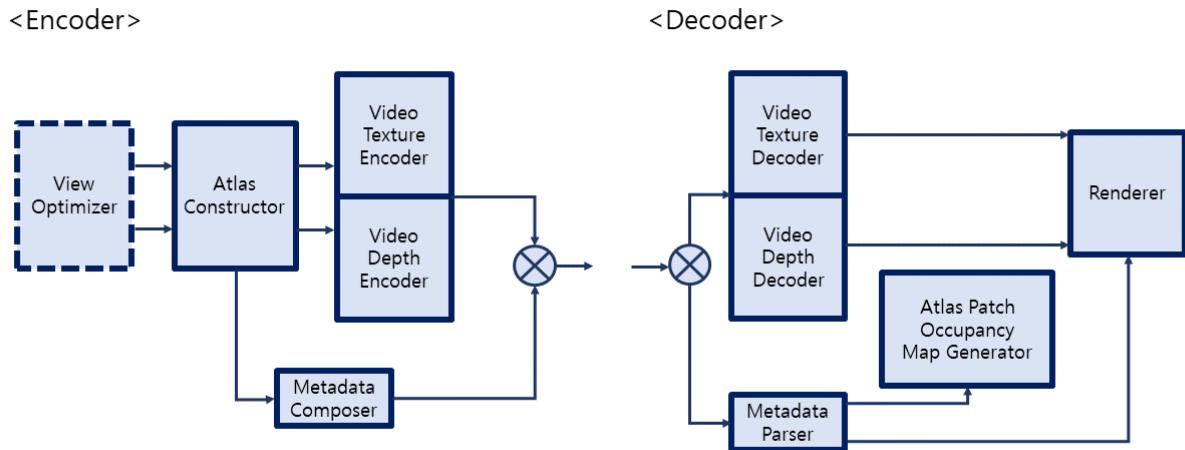


그림 3. TMIV 인코더와 디코더의 블록 다이어그램. 인코더(왼쪽) 디코더(오른쪽)
 Fig. 3. Block diagram of TMIV Encoder. Encoder(Left) Decoder(Right)

2. 6DoF 비디오 코덱의 인코더 및 디코더

그림 3을 보면, TMIV 인코더의 입력은 순서대로 시점 최적화, 아틀라스 생성기를 거치고 비디오 텍스쳐 및 깊이 인코더 (Video Texture/Depth Encoder)를 통해 부호화 된다. 디코더는 반대로 비디오 텍스쳐 및 깊이 정보에 관한 기본 시점 및 아틀라스를 복원한다. 이후 아틀라스 패치 점 유지도 생성기, 렌더러를 통해 최종으로 복원된 출력이 생성된다. 개발 코덱은 TMIV와 유사한 구조를 차용하되, 핵심 비디오 인코더 디코더를 VVC로 대체하여 구성하였다. 다음에서 각 모듈의 역할과 구성에 관하여 설명한다.

2.1 시점 최적화

시점 최적화는 과정에서는 방향 편차, 시야, 거리 및 시야 간 겹침을 고려하여 필요한 기본 시점의 개수를 결정한다. 다음으로 시점 간의 위치와 상호 간의 겹침을 고려하여 기본 시점을 선택한다^[13].

2.2 아틀라스 생성기

기본 시점과 추가 시점의 아틀라스를 그림 4에서 살펴볼 수 있다. 시점 최적화 단계에서 선택된 기본 시점들은 하나의 완전한 단일 아틀라스로 전달된다. 이후 추가 시점들은 기본 시점 기준으로 중복된 부분을 제거한 잔차 패치로 구

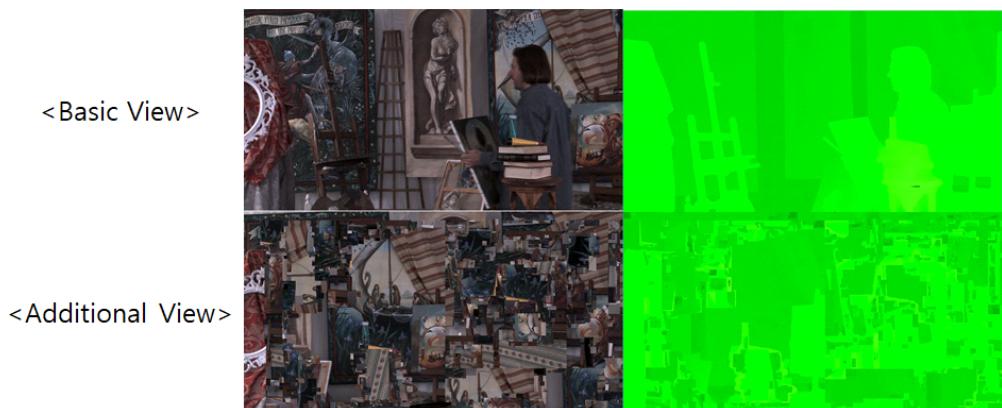


그림 4. TechnicolorPainter 시퀀스의 아틀라스의 예시. 색상영상(왼쪽) 깊이영상(오른쪽)
 Fig. 4. Example of Atlas of TechnicolorPainter Sequence. Texture(Left) Depth(Right)

성한다. 아틀라스 생성 과정 동안에 Pruner는 마스크를 사용하여 기본 시점들을 보존하고 추가 시점들의 중복된 부분을 제거한다. Aggregator에서는 시간적 순서에 따라 비디오 프레임에 사용한 마스크를 업데이트 한다. 이후, Patch packer에서 각 패치 아틀라스들을 패킹하여 최종적으로 아틀라스를 생성한다^{[13][21]}. 그림 4는 각 기본 시점과 추가 시점의 아틀라스 구성 예시를 보인다. 그림에서 보듯이 기본 시점의 텍스쳐 및 깊이 정보는 원본과 같이 구성하되, 추가 시점에 대해서는 텍스쳐 및 깊이 정보 모두 잉여 정보를 블록 패치 형태로 구성한다. 디코더에서는 잔차 블록 패치를 재구성하여 추가 시점의 비디오 프레임을 복원한다.

2.3 비디오 인코더 및 디코더

기존의 TMIV에서 HEVC를 기본 비디오 코덱으로 사용한 반면 개발 6DoF 비디오 코덱에서는 VVC 인코더를 사용하였다. Main10의 임의 접근 구성 (Random Access)이 텍스쳐 영상 및 깊이 영상을 압축하는데 사용된다^[13]. 비디오 인코더의 구체적 구성이 그림 5에 도시 되어있다. MIV 모드에서 전체 10개의 시점에 대해 시점 최적화와 아틀라스 생성기의 과정을 거치면 하나의 기본 시점에 대한 아틀라스와 세 개의 추가 시점에 대한 아틀라스가 생성된다. 이

때, 생성된 아틀라스들은 각각 비디오 부호화 과정을 거쳐 비트스트림으로 구성한다. 제안 코덱에서는 비디오 부호화 과정을 VVC 참조 소프트웨어 (VTM6.0)을 이용하여 대체함으로써 부호화 효율을 증대하였다.

표 2에 비디오 인코더의 주요 부호화 파라미터를 보인다. 여기에 기술한 파라미터는 HEVC와 VVC 모두에 적용하는 공통된 파라미터이며, 그 설정은 모두 HEVC와 VVC 표준화에서 권장하는 공통 실험 조건 (Common Test Condition, CTC) 규격을 따랐다^{[22][23]}.

표 2. Video Encoder의 매개변수

Table 2. Parameter of Video Encoder

Parameter Name	Value	Comment
IntraPeriod	32	Period of I-frame
GOPSsize	16	Number of B-frame plus 1
LambdaFromQPEnable	1	Lambda value adjustment by QP
FEN	1	Fast encoder decision
RateControl	0	Disable rate control
LoopFilterDisable	0	Enable deblocking tiler
SAO	1	Enable sample adaptive offset
TransformSkip	1	Enable transform skip mode
SearchRange	384	Search range for motion estimation
InternalBitDepth	10	Codec operating bit-depth of pixels

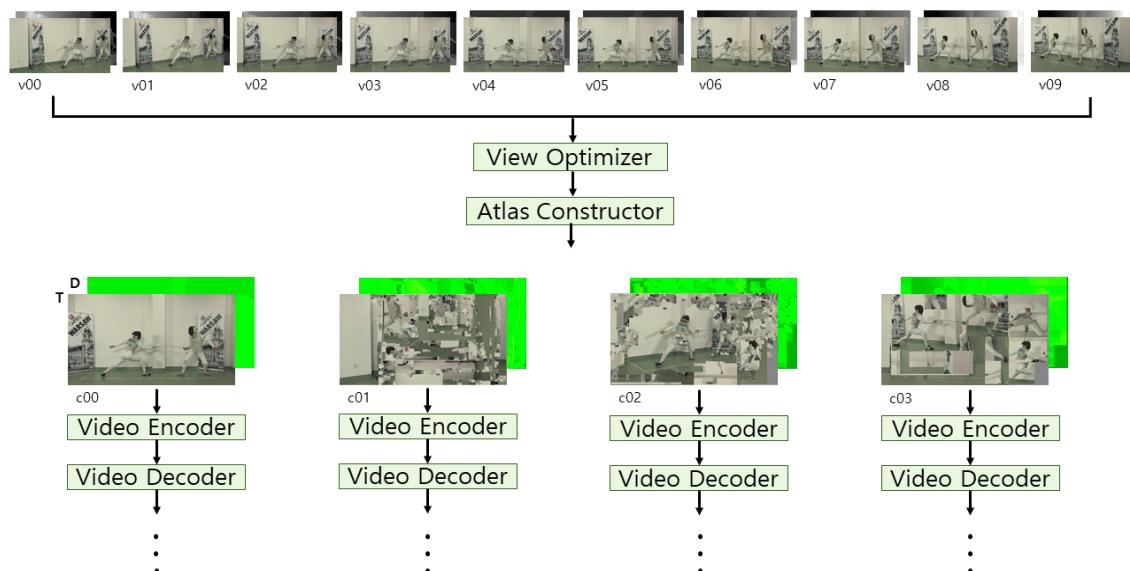


그림 5. TMIV의 알고리듬 구조

Fig. 5. Algorithm architecture of TMIV

최근 VVC에서 HEVC 부호화 기술을 확장, 변형하여 보다 우수하고 다양한 부호화 기술을 개발하였다. 특히 블록 분할 구조가 블록의 보다 세밀한 텍스쳐 정보를 표현할 수 있도록 정교해졌다. 일례로, HEVC에서는 CTU를 픽쳐 단위로 분할할 경우, 쿼드트리 (Quad-tree) 형태로 정사각형 모양으로만 분할할 수 있었다. 그러나 VVC에서는 쿼드트리에 이진 트리 (Binary Tree)와 삼진 트리 (Ternary Tree)를 중첩할 수 있게 한다. 구체적으로는, 루트노드인 CTU 블록에서 항상 쿼드트리로 시작하되, 특정 자식 노드 (Child node)에서부터는 쿼드트리 구조를 종료하고 이진 트리 혹은 삼진 트리를 가질 수 있으며, 그 역 또한 가능하며, 이러한 이진 혹은 삼진 트리 구조에서는 쿼드트리가 더 이상 나타나지 않게 된다. 이진 트리는 가로방향 혹은 세로방향으로 1:1의 크기를 갖는 두 노드의 구조로 분할하며, 삼진 트리는 가로방향 혹은 세로방향으로 1:2:1의 크기를 갖는 세 노드의 구조로 분할한다. 이에 따라, HEVC에서 표현하기 어려운 다양한 블록 모양을 가질 수 있게 되어 코딩 효율이 크게 높아졌다^[24]. 아래 표 3에서 블록 분할에 관한 주요 확장 및 변경 내용을 HEVC와 비교하여 보인다.

표 3. VVC의 확장 매개변수
Table 3. Parameter of Extension in VVC

Parameter Name	HEVC	VVC
CTU Size	64x64	128x128
Max Transform Unit Size	32x32	64x64
Max Transform Skip Size	4x4	32x32
Max Tree Depth	4	4+3
Core Transform Kernel	1	3

그 외에도 압축 성능에 큰 영향을 주는 부호화 기술이 새롭게 추가되었으며, 주요 부호화 기술을 아래와 같이 요약하여 소개한다. 본문의 VTM 실험에서는 아래의 주요 기술을 적용한 상태로 인코딩을 진행하였다. 보다 자세한 알고리듬에 대한 내용은 VTM 알고리듬 서술문서서에 기술되어 있다^[24].

- Chroma Separate Tree: I-frame의 경우, 휘도 성분과 색차 성분에 대한 코딩 트리 구조를 다르게 갖는 구조로서, 색차 성분의 블록 모양의 다양성을 증가시켜 색차 성분 코딩 효율을 크게 높인다.
- Cross-component Linear Model: 휘도 성분의 재구성

된 샘플에 대한 선형 모델을 취하여 색차 성분을 예측함으로써 색차 성분 코딩 효율을 크게 높인다.

- Adaptive Loop Filter: Wiener 필터를 기반으로 코딩 에러를 최소화하는 필터를 루프 내에 적용하여 참조 프레임의 화질을 크게 향상시킨다.
- Affine Motion Estimation/Compensation: 화면 간 예측 기술의 하나로, 기존의 움직임 벡터로 표현하기 힘든 확대/축소/변환 움직임을 Affine 변환 기법으로 예측하게 하여, 보다 다양한 움직임을 효과적으로 표현할 수 있게 한다.
- Adaptive Motion Vector Resolution: HEVC의 경우, 움직임 벡터 차분값(Motion vector difference, 이하 MVD)을 1/4 픽셀 레벨에서만 보낼 수 있었는데, VVC의 경우 정규 화면간 예측인지 Affine 예측인지에 따라 최소 1/16 픽셀레벨에서 최대 4 픽셀 단위로 움직임 벡터를 전송할 수 있게 하여, 압축 효율을 크게 향상시킨다.
- Dependent Quantization: 변환 계수 레벨에 따라 계수 재구성 값의 종류를 더 다양하게 하여, 압축 효율을 향상시킨다.

2.4 렌더러 (Renderer)

렌더러는 컨트롤러 (Controller), 합성 (Synthesizer), 그리고 인페인팅 (Inpainter) 세 가지 단계로 구성되어 주로 중간 시점의 합성을 위하여 사용이 된다^[13]. 합성 시점 소프트웨어로는 TMIV와 같이 RVS를 사용한다^[25]. 기존 MPEG에서 개발하였던 합성 시점 소프트웨어인 VSRS 대비 가장 큰 차이는 보다 우수한 시점 합성 결과를 위해 컨트롤러에서 인접한 뷰 외에도 멀리 있는 뷰 역시 선택 가능하도록

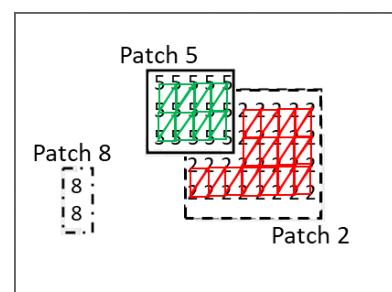


그림 6. 렌더러의 매쉬 형성^[13]
Fig. 6. Mesh Formation of Renderer^[13]

하는 것이다. 합성 알고리즘은 삼각형의 병렬 래스터화가 적용된 RVS 기반의 시점합성 과정을 거친다. 삼각형의 병렬 래스터화가 적용되기 이전에 그림 6와 같이 아틀라스로부터 메쉬(mesh)를 형성한다. 예를 들어, 그림 6이 MIV 모드에서 추가 시점의 한 프레임이라고 생각해보자. 배치된 패치 아틀라스 중 패치2와 패치5와 같이 메쉬를 만들 수 있고 이는 무게 중심 보간을 이용해 래스터화 된다. 이후 인페인터는 누락된 픽셀을 보간하는 역할을 한다.

가지 테스트 시퀀스들에 대하여 실험을 진행하였다. 관련한 설명은 아래 첨부된 표 4에서 찾아볼 수 있다. ClassroomVideo, TechnicolorMuseum, TechnicolorHijack은 Omnidirectional 6DoF 시퀀스이며 모두 컴퓨터로부터 생성되었다. TechnicolorPainter, IntelFrog, PoznanFencing은 Win-

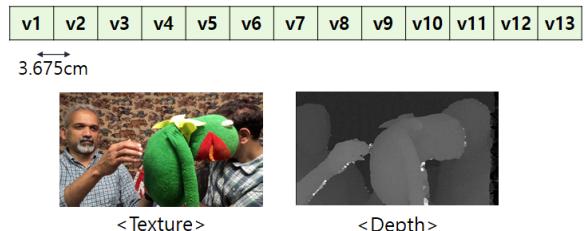


그림 8. IntelFrog 시퀀스의 정보. 색상영상(왼쪽 아래) 깊이영상(오른쪽 아래) 카메라 배열(위) [27]

Fig. 8. Information of IntelFrog sequence. Texture(Left-up) Depth(Left-down) Camera Array(Up) [27]

III. 테스트 6DoF 비디오와 성능 평가 지표

1. Test Materials

성능 평가를 위해 공통 실험 조건 문서에 명시된 일곱

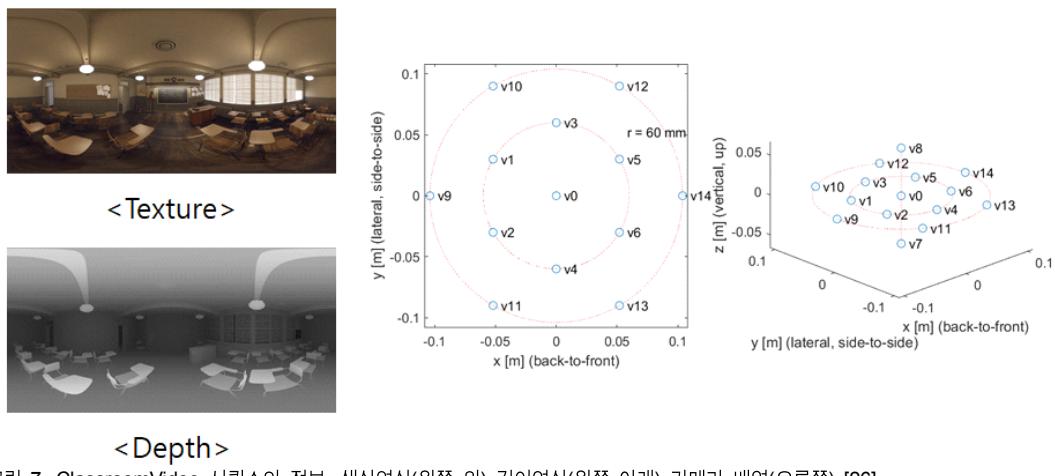


그림 7. ClassroomVideo 시퀀스의 정보. 색상영상(왼쪽 위) 깊이영상(왼쪽 아래) 카메라 배열(오른쪽) [26]
 Fig. 7. Information of ClassroomVideo sequence. Texture(Left-up) Depth(Left-down) Camera Array(Right) [26]

표 4. 실험에 사용한 테스트 비디오의 특성 [20]

Table 4. Properties of test video sequences [20]

Sequence Name	Format	# of source view	Resolution	Class
ClassroomVideo	ERP	15	4096x2048	Computer-generated
TechnicolorMuseum	Semi-ERP	24	2048x2048	Computer-generated
TechnicolorHijack	Semi-ERP	10	4096x4096	Computer-generated
TechnicolorPainter	Perspective	16 (4x4)	2048x1088	Natural contents
IntelFrog	Perspective	13 (13x1)	1920x1080	Natural contents
OrangeKitchen	Perspective	25 (5x5)	1920x1080	Computer-generated
PoznanFencing	Perspective	10 (10x1)	1920x1080	Natural contents

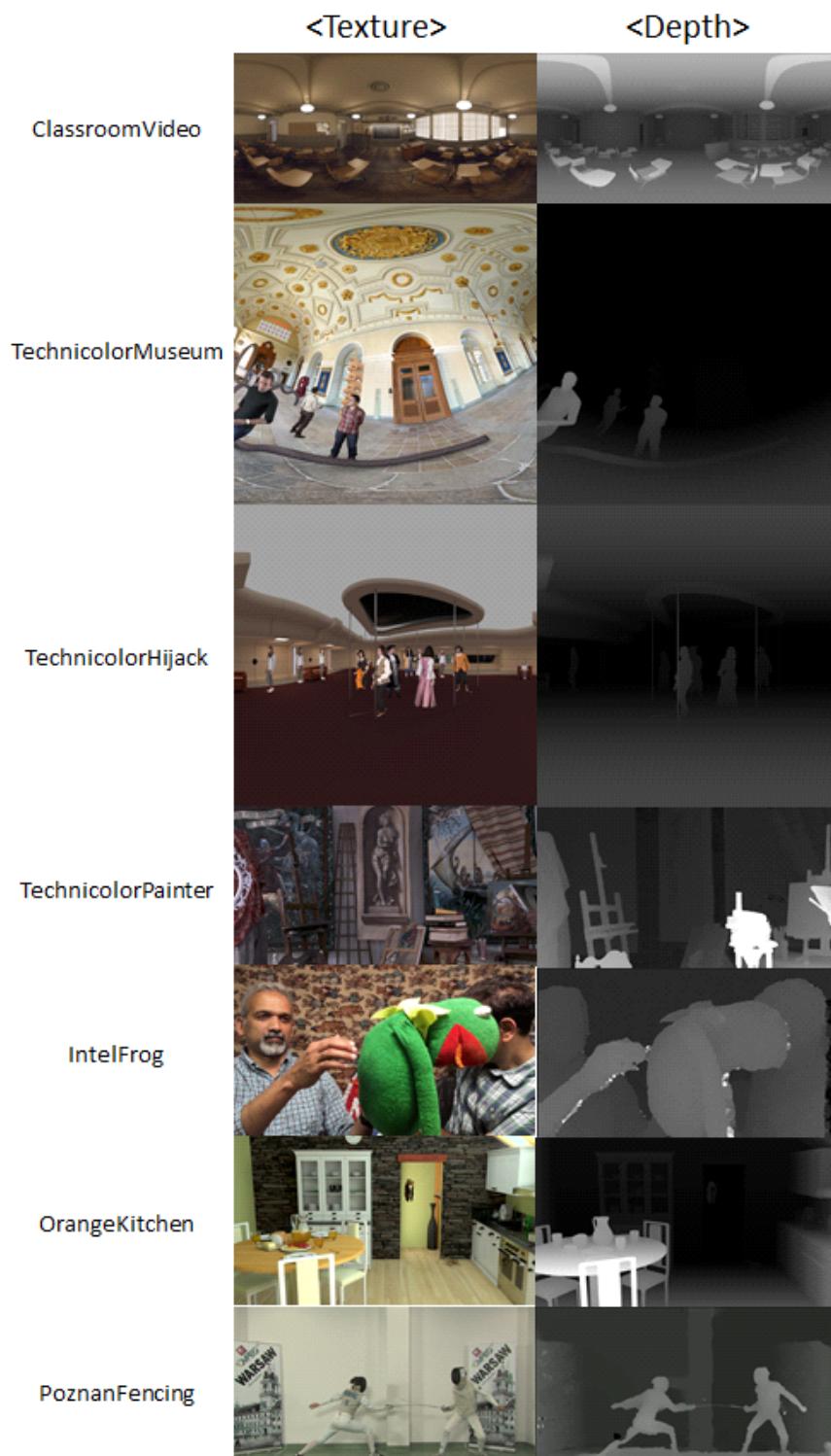


그림 9. 테스트 시퀀스의 정보. 색상영상(왼쪽) 깊이영상(오른쪽)[26][27]

Fig. 9. Information of Test sequence. Texture(Left) Depth(Right) [26][27]

dowed 6DoF 시퀀스이며 자연 영상을 취득하였다. OrangeKitchen은 Windowed 6DoF 시퀀스이며 컴퓨터로부터 생성되었다. Omnidirectional 6DoF 시퀀스는 모두 Equi-Rectangular Projection (ERP) 또는 semi-ERP를 이용하여 프로젝션 하였다. ERP 포맷은 360도 영상을 평면상에 매핑한 것으로, 3차원의 좌표를 경도와 위도로 매핑한다. 대표적으로 ClassroomVideo는 그림 7과 같이 동심원 형태로 카메라를 배열하고 비디오를 취득하였다. Windowed 6DoF 시퀀스인 IntelFrog는 그림 8과 같이 수평 방향으로 카메라가 3.75cm의 간격을 가지고 취득된 영상이다^{[20][26][27]}. TechnicolorPainter는 4x4의 카메라 배열을 이용하여 취득하였고, 이와 유사하게 OrangeKitchen도 다 차원의 카메라 배열을 이용하여 취득하였다. 그림 9에서는 각 시퀀스들의 색상 및 깊이 영상의 예시를 찾아볼 수 있다. 이때, TechnicolorMuseum과 OrangeKitchen의 경우 다른 시퀀스들보다 색상 정보가 많은 것을 알 수 있다. 또한 자연 취득된 깊이 영상의 경우 컴퓨터로 생성된 깊이 대비 잡음이 많이 있음을 알 수 있다.

2. Metrics

몰입형 비디오의 성능을 평가하기 위해 기준 PSNR을 비롯한 많은 평가 척도들이 존재한다. 그 중 WS-PSNR, IV-PSNR을 소개한다. 실험에서는 Windowed 6DoF 영상은 위도에 따른 물체 배경의 변형이 없으므로 PSNR을 사용하고 ERP 영상에 대해서는 WS-PSNR을 사용하였다^[28-29]. 이외에도 IV-PSNR을 이용하여 인지 화질에 관한 성능을 측정하였다.

2.1 WS-PSNR (Weighted-to-Spherically-uniform Peak Signal-to-Noise Ratio)

WS-PSNR은 가장 대표적인 평가 기준이다. 이는 ERP를 비롯한 전방위 비디오의 객관적인 품질을 정확하게 평가하도록 제안된 평가 기준이다. 이 방법은 각 화소 위치에 따른 가중치를 할당하여 구형 표면에서 균일한 가중치로 PSNR을 계산한다. 이와 같은 왜곡을 포함한 객관적 화질 평가를 통해 코덱 성능을 분석을 정확하게 수 있다. 그림 10은 ERP포맷의 WS-PSNR 평가 방식의 예시로, 포맷에서

의 $w(i,j)$ 수식으로 계산된 가중치를 그림으로 도식한 것이다. 밝은 색을 가지는 것은 가중치가 크다는 의미이며, 어두운 영역은 가중치가 작음을 의미한다.

$$W(i,j)_{ERP} = \cos \frac{(j+0.5 - \frac{N}{2})\pi}{N}$$



그림 10. ERP 포맷에서의 가중치 [29]
 Fig. 10. Weight in ERP format [29]

2.2 IV-PSNR (Immersive Video PSNR)

IV-PSNR은 몰입형 비디오와 같은 다시점 영상에 대해 중간시점 합성의 오류의 대부분이 경계 부분에서 일어난다는 특성을 파악하여 PSNR을 블록 단위로 계산하는 평가 기준이다. 기준 PSNR 대비 화소의 이동과 전역적인 색상 편차를 고려한다는 점에서 차이가 있다. 화소의 이동 고려는 물체의 경계의 이동과 같은 재투사(Reprojection) 오류의 영향을, 전역적인 색상 편차 고려는 입력 시점의 위치에 따른 색상 특성에 따른 오류의 영향을 줄인다^[30].

IV. 실험 결과

실험 조건은 TMIV2.0과 MPEG-I 표준에서 사용하는 공통실험조건 (Common Test Condition)을 참고하여, TMIV 인코더-디코더 사이의 비디오 부복호화에서 기준 HEVC를 VVC로 대체하여 성능을 비교하였다. 실험은 HEVC 참조 소프트웨어인 HM 16.16을 이용하여 기준 비교로 사용하였고, 개발 코덱에서는 VVC 참조 소프트웨어인 VTM6.0을 사용하였다. ERP 영상과 같은 Omnidirectional 비디오의 압축을 위하여 360Lib를 ERP 포맷에 맞춰 코덱에 적용을

표 5. MIV mode 실험 결과

Table 5. Test results of MIV mode

Sequence Name	QP	TMIV+HEVC				TMIV+VVC			
		Total kbps	Avg. Y-PSNR	Avg. U-PSNR	Avg. V-PSNR	Total kbps	Avg. Y-PSNR	Avg. U-PSNR	Avg. V-PSNR
ClassroomVideo	22	107717.5	33.97	47.52	49.49	103099.7	34.09	48.72	50.90
	27	24016.3	33.94	46.23	48.14	23968.1	34.11	47.85	49.95
	32	8014.9	33.60	44.77	46.55	7296.1	33.85	46.64	48.58
	37	3580.4	32.96	43.37	45.26	3007.0	33.33	45.26	47.11
	42	2084.7	31.97	42.68	44.59	1722.7	32.59	44.09	45.77
	BD-rate Y-PSNR	High	-51.13%			Low	-44.82%		
TechnicolorMuseum	BD-rate MS-SSIM	High	-35.8%			Low	-33.1%		
	BD-rate IV-PSNR	High	-38.8%			Low	-40.5%		
	22	51200.8	32.73	41.28	42.76	47494.4	33.07	41.58	43.16
	27	31686.7	31.99	40.24	41.69	29846.1	32.48	40.74	42.40
	32	17935.7	30.74	38.78	40.29	17358.4	31.56	39.63	41.38
	37	8818.6	29.32	37.58	39.15	8694.5	30.20	38.38	40.14
TechnicolorHijack	42	3925.9	27.71	37.00	38.55	3896.0	28.54	37.58	39.23
	BD-rate Y-PSNR	High	-33.52%			Low	-35.54%		
	BD-rate MS-SSIM	High	-35.4%			Low	-33.5%		
	BD-rate IV-PSNR	High	-26.0			Low	-26.5%		
	22	43770.1	39.15	51.77	51.78	37477.0	39.42	52.33	52.37
	27	24096.9	38.48	50.58	50.42	20370.3	38.79	51.48	51.37
TechnicolorPainter	32	13697.6	37.63	48.98	48.72	11639.6	38.03	50.23	49.97
	37	7166.4	36.43	47.68	47.23	6106.5	36.97	48.68	48.33
	42	3860.4	35.06	46.72	46.15	3310.9	35.65	47.47	46.94
	BD-rate Y-PSNR	High	-34.50%			Low	-34.61%		
	BD-rate MS-SSIM	High	-34.4%			Low	-35.5%		
	BD-rate IV-PSNR	High	-38.1%			Low	-37.6%		
IntelFrog	22	61271.4	36.19	46.70	46.46	56861.8	36.25	47.29	47.06
	27	35680.3	35.98	46.15	45.73	31342.8	36.03	46.58	46.23
	32	21873.1	35.42	44.90	44.35	18858.0	35.65	45.81	45.22
	37	12462.8	34.43	43.52	42.96	10739.7	34.88	44.65	43.80
	42	7184.9	32.92	42.87	42.16	6374.6	33.60	43.62	42.58
	BD-rate Y-PSNR	High	-27.44%			Low	-29.87%		
OrangeKitchen	BD-rate MS-SSIM	High	•			Low	-29.9%		
	BD-rate IV-PSNR	High	-30.8%			Low	•		
	22	373504.0	27.44	41.54	37.71	384695.9	27.47	41.85	37.88
	27	133595.7	27.12	40.71	37.09	131449.7	27.15	41.24	37.38
	32	62326.9	26.79	39.86	36.49	57277.7	26.90	40.57	36.83
	37	31475.4	26.15	38.91	35.83	27910.2	26.46	39.65	36.12
PoznanFencing	42	16373.5	25.14	38.01	35.05	14795.1	25.67	38.58	35.19
	BD-rate Y-PSNR	High	-21.23%			Low	-31.46%		
	BD-rate MS-SSIM	High	•			Low	-27.0%		
	BD-rate IV-PSNR	High	-14.1%			Low	•		
	22	13761.0	24.96	33.37	31.89	11774.6	25.65	37.48	37.54
	27	8159.2	25.80	35.94	35.24	6907.2	25.53	36.43	35.68
OrangeKitchen	32	4866.0	24.65	32.82	31.34	4132.9	25.03	33.42	32.08
	37	2819.0	24.20	31.27	29.30	2426.3	24.11	31.10	29.26
	42	1745.7	25.16	34.28	33.05	1518.5	24.43	32.55	31.22
	BD-rate Y-PSNR	High	-26.68%			Low	-41.29%		
	BD-rate MS-SSIM	High	-27.3%			Low	-25.8%		
	BD-rate IV-PSNR	High	-28.1%			Low	-30.8%		
PoznanFencing	22	42912.8	28.15	42.71	41.89	39738.0	28.24	42.90	42.04
	27	24666.5	27.98	42.35	41.64	19930.4	28.10	42.60	41.82
	32	16864.6	27.77	41.76	41.16	13359.5	27.87	41.99	41.30
	37	10755.7	27.54	41.42	40.84	9058.6	27.59	41.35	40.75
	42	5814.6	27.28	41.85	41.46	5296.5	27.27	41.32	40.76
	BD-rate Y-PSNR	High	-34.40%			Low	-24.93%		
BD-rate MS-SSIM		High	•			Low	-37.5%		
BD-rate IV-PSNR		High	-33.4%			Low	•		
Average BD-rate PSNR		High	-33.0%			Low	-34.6%		
Average BD-rate MS-SSIM		High	-33.2%			Low	-31.8%		
Average BD-rate IV-PSNR		High	-29.9%			Low	-33.9%		

하였다. 성능평가는 WS-PSNR을 사용하였다. 결과는 TMIV의 두 가지 모드인 MIV 모드와 MIV 시점 모드로 구분하여 비교한다. 공통실험조건에 언급된 양자화 파라메터 (QP) 22, 27, 32, 37, 42에 대하여 모두 실험을 진행하였으며 결과 중 High는 QP 27, 32, 37, 42를 사용하였을 때의 BD-rate이고 Low는 QP 22, 27, 32, 37를 사용할 때의 BD-rate이다. 또한 각 시퀀스 별로 공통실험조건에서 언급된 시작 프레임을 33번째 프레임부터 하여 부호화를 수행하였다. 표5, 6에서 PSNR의 단위는 데시벨 (dB)이다.

일부 시퀀스에 대해 비트율-왜곡 곡선이 컨벡스하지 않은 구간이 발생한 경우 성능에 대해 표를 기입하지 않았다. 이와 같은 이유로 MIV 시점 모드에서 OrangeKitchen의 BD-rate Y-PSNR의 결과와 자연영상의 경우 High BD-rate MS-SSIM과 Low BD-rate의 IV-PSNR은 비트율-왜곡 곡선이 컨벡스한 특성을 보이지 않아 결과에서 제외하였다.

1. Media Immersive Video (MIV) 모드 부호화 성능 평가

실험결과 기준 HEVC 대비 VVC 적용 시 평균적으로 High에서는 33.0%, Low에서는 34.6%의 BD-rate 감소 효율을 제공하였다. 표 5에서 유사한 PSNR을 가질 때 BD-

rate의 감소함을 알 수 있다.

먼저 High일 때의 성능을 살펴보면, Omnidirectional 6DoF 시퀀스의 경우, ClassroomVideo 시퀀스는 물체의 움직임보다는 조명의 변화가 주요 정보인데, 밝기 정보를 담고 있는 Y요소의 성능이 적은 비트로도 높아진 것을 보았을 때, 제안 방법이 적합함을 알 수 있다. Technicolor Museum, TechnicolorPainter 시퀀스는 보다 생상 정보가 많고 물체들의 움직임이 많아 ClassroomVideo 만큼의 성능은 아니지만 평균적인 성능을 보인다. Windowed 6DoF의 경우, Omnidirectional 6DoF 시퀀스들 보다 성능이 떨어지는는데, 이는 Windowed 6DoF의 해상도가 비교적 낮기 때문에 QP가 큰 High에서 영향을 많이 받는 것 같다. Low일 때의 성능을 살펴보면, 컴퓨터로부터 취득된 영상의 성능이 자연으로부터 취득된 영상의 성능보다 우수함을 알 수 있다.

그림 11에서 TechnicolorPainter(위) 시퀀스에서 사람의 얼굴 중앙 부분을 보았을 때 HEVC의 경우 VVC보다 Blocky artifact로 인한 화질 열화가 심함을 알 수 있다. 또한 전체적인 경계 부분이 VVC에서 보존이 잘 되었음을 알 수 있다. IntelFrog(아래) 시퀀스에서 사람의 얼굴의 Block artifact와 셔츠의 블러링으로 인한 화질 열화가 심함을 알 수 있다.



그림 11. MIV mode의 실험 결과. TechnicolorPainter (위), IntelFrog (아래)

Fig. 11. Test results of MIV mode. TechnicolorPainter (Up), IntelFrog (Down)

2. MIV 시점 모드 부호화 성능 평가

표 6. MIV 시점 mode 실험 결과

Table 6. Test results of MIV view mode

Sequence Name	QP	TMIV+HEVC				TMIV+VVC			
		Total kbps	Avg. Y-PSNR	Avg. U-PSNR	Avg. V-PSNR	Total kbps	Avg. Y-PSNR	Avg. U-PSNR	Avg. V-PSNR
ClassroomVideo	22	541287.6	36.59	48.41	50.28	521171.2	36.76	49.91	51.89
	27	125822.6	35.73	46.88	48.73	123775.8	35.92	48.79	50.75
	32	40730.7	35.07	45.23	47.01	37351.7	35.32	47.36	49.23
	37	16633.4	34.15	43.76	45.63	13934.3	34.53	45.85	47.66
	42	8630.7	32.98	43.04	44.95	7117.0	33.56	44.63	46.39
	BD-rate Y-PSNR	High		-33.09%		Low		-37.74%	
TechnicolorMuseum	BD-rate MS-SSIM	High		-33.6%		Low		-30.5%	
	BD-rate IV-PSNR	High		-39.4%		Low		-42.6%	
	22	125575.0	36.33	42.51	43.80	115581.5	36.65	42.97	44.40
	27	78151.5	34.87	40.89	42.28	72878.1	35.37	41.55	43.22
	32	43840.1	32.94	39.05	40.63	42085.1	33.73	40.00	41.85
	37	21169.7	30.78	37.74	39.42	20821.9	31.67	38.57	40.45
TechnicolorHijack	42	9003.6	28.67	37.16	38.83	8956.0	29.51	37.73	39.51
	BD-rate Y-PSNR	High		-23.61%		Low		-27.08%	
	BD-rate MS-SSIM	High		-28.5%		Low		-27.7%	
	BD-rate IV-PSNR	High		-23.7%		Low		-25.1%	
	22	45567.4	42.95	52.38	52.28	39898.8	43.49	53.13	53.01
	27	24570.3	41.53	50.92	50.72	21353.7	42.15	52.03	51.83
TechnicolorPainter	32	13722.5	39.89	49.12	48.90	11994.4	40.72	50.63	50.38
	37	7127.7	38.13	47.68	47.32	6268.0	39.01	48.95	48.60
	42	3875.6	36.26	46.78	46.25	3370.1	37.20	47.62	47.18
	BD-rate Y-PSNR	High		-34.30%		Low		-35.46%	
	BD-rate MS-SSIM	High		-36.4%		Low		-36.9%	
	BD-rate IV-PSNR	High		-32.8%		Low		-34.8%	
IntelFrog	22	64145.9	38.99	47.24	47.03	61615.2	39.11	47.58	47.41
	27	34370.5	38.38	46.33	45.98	30806.1	38.58	46.94	46.63
	32	20379.0	37.40	45.07	44.56	17883.6	37.73	45.95	45.38
	37	11478.2	35.85	43.61	43.01	10043.7	36.43	44.83	43.92
	42	6423.4	33.88	42.97	42.30	5821.2	34.62	43.86	42.79
	BD-rate Y-PSNR	High		-25.16%		Low		-26.76%	
PoznanFencing	BD-rate MS-SSIM	High		•		Low		-25.5%	
	BD-rate IV-PSNR	High		-28.7%		Low		•	
	22	287174.2	32.21	43.63	42.21	289832.2	32.38	44.20	42.72
	27	94760.8	31.11	42.39	41.03	92326.6	31.33	43.37	41.89
	32	41791.7	30.28	41.16	39.80	38960.3	30.67	42.44	40.87
	37	20613.3	29.11	39.93	38.60	18492.0	29.71	41.29	39.51
	42	10558.5	27.48	39.02	37.56	9659.2	28.35	40.22	38.22
PoznanFencing	BD-rate Y-PSNR	High		-27.46%		Low		-34.53%	
	BD-rate MS-SSIM	High		•		Low		-3.7%	
	BD-rate IV-PSNR	High		-36.4%		Low		•	
	22	40870.2	34.16	46.63	45.39	41795.7	34.34	46.92	45.56
	27	20621.5	33.69	45.78	44.78	17587.3	33.83	46.18	44.99
	32	13159.2	33.09	44.59	43.87	10835.0	33.33	45.18	44.21
	37	8065.3	32.21	43.66	43.08	7003.6	32.56	44.06	43.28
	42	4222.7	31.01	42.96	42.54	3978.0	31.45	43.16	42.50
	BD-rate Y-PSNR	High		-28.97%		Low		-28.12%	
PoznanFencing	BD-rate MS-SSIM	High		•		Low		-36.4%	
	BD-rate IV-PSNR	High		-27.6%		Low		•	
	Average BD-rate Y-PSNR	High		-28.8%		Low		-31.6%	
	Average BD-rate MS-SSIM	High		-32.8%		Low		-26.8%	
	Average BD-rate IV-PSNR	High		-31.4%		Low		-34.2%	

실험결과 기존 HEVC 대비 VVC 적용 시 평균적으로 High에서는 28.8%, Low에서는 31.6%의 BD-rate 감소 효율을 제공하였다. 표 6에서 유사한 PSNR을 가질 때 BD-rate이 감소함을 알 수 있다.

먼저 High일 때의 성능을 살펴보면, Omnidirectional 6DoF 시퀀스의 경우, ClassroomVideo 시퀀스는 물체의 움직임보다는 조명의 변화가 주요 정보인데, 밝기 정보를 담고 있는 Y의 성능이 적은 비트로도 높아진 것을 보았을 때, 제안 방법이 우수함을 알 수 있다. TechnicolorPainter 시퀀스는 보다 생상 정보가 많고 물체들의 움직임이 많아 ClassroomVideo 만큼의 성능은 아니지만 평균적인 성능을 보인다. TechnicolorMuseum의 경우 색상 정보가 Omnidirectional 6DoF 중 낮은 편인데, 이는 MIV 시점 모드에서 색상 정보를 MIV 모드만큼 복원하지 못하는 것으로 보인다. Windowed 6DoF의 경우, Omnidirectional 6DoF 시퀀스들 보다 성능이 떨어지는데, 이는 Windowed 6DoF의 해상도가 비교적 낮기 때문에 QP가 큰 High에서 영향을 많이

받는 것 같다. Low일 때, Windowed 6DoF 시퀀스인 IntelFrog의 선택된 시점들의 복원된 성능이 좋은 점을 제외하고 High와 유사하다.

그림 12에서 TechnicolorPainter(위) 시퀀스에서 사람의 얼굴 중 눈이 VVC에서 보다 복원이 잘 되었으며 경계 부분 역시 뚜렷함을 알 수 있다. IntelFrog(아래) 시퀀스에서 경계 부분이 눈에 띄게 뚜렷함을 알 수 있다.

3. TMIV 특성에 맞는 VVC 툴 제어 연구

HEVC나 VVC의 경우, 코딩 블록의 양자화 적용에 따른 블록 경계면을 필터링하는 디블로킹 필터(deblocking filter)가 주관적 화질 향상을 위해 필수이며, 중요한 코딩 툴 중 하나로 여겨진다. 그러나 아틀라스 영상의 경우, 오히려 이러한 경계면을 수정하게 되면 원치 않는 코딩 에러가 될 수 있으며, 추후 아틀라스로 추가 시점 영상들을 생성하려 할 때 그 에러가 치명적으로 영향을 줄 수 있다. 실제로,



그림 12. MIV 시점 mode의 실험 결과. TechnicolorPainter (위), IntelFrog (아래)
 Fig. 12. Test results of MIV view mode. TechnicolorPainter (Up), IntelFrog (Down)

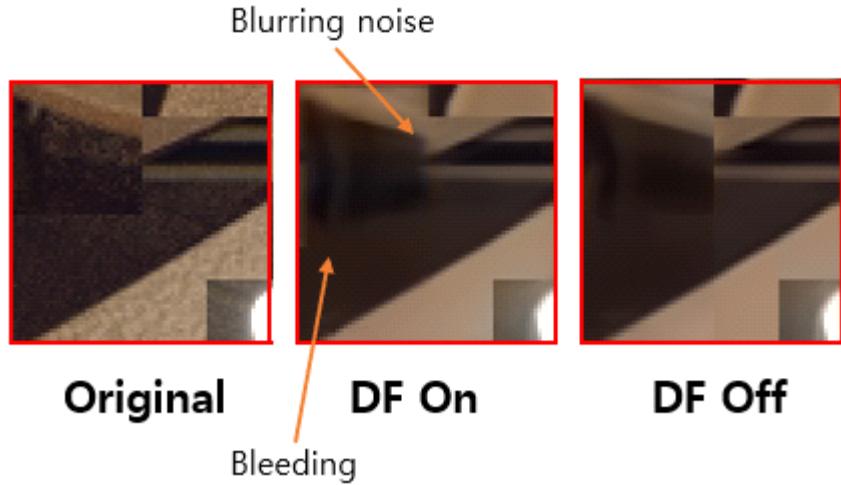


그림 13. 아틀라스 원본 영상(왼쪽)에 디블로킹 필터(DF)를 적용하였을 때(가운데)와 적용하지 않았을 때(오른쪽)의 예시

Fig. 13. Example of original atlas image (Left), encoded atlas image with deblocking filter (Center), and without deblocking filter (Right)

아래 그림 13과 같이, 원 아틀라스 영상에 대하여 VVC의 디블로킹 필터를 적용하였을 때와 적용하지 않았을 때, 주관적으로 화질이 변화하는 것을 볼 수 있다. 디블로킹 필터를 적용하게 되면, 원치 않는 블러링 효과와 픽셀 값이 번지는 bleeding 애러가 발생하는 것을 확인할 수 있다.

이처럼 TMIV의 경우, 기존의 동영상 기반의 압축기술이 오히려 성능을 저해할 수 있다. 6DoF의 방대한 데이터를 효율적으로 압축하기 위해서는 단순히 HEVC에서 VVC로의 적용을 넘어서서, TMIV 영상에 맞는 효율적인 압축 기

술을 개발해야 하며, 또한 기존 코딩 툴 역시 영상 특성에 따라 제어하는 경우가 필요하다고 본다. 이를 관찰하고자 특별히, 디블로킹 필터에 대한 시점 별 VVC 압축 결과를 실험해 보았다. 동일한 VTM 버전에 동일한 TMIV의 MIV 모드 영상들에 대하여 기존 VTM 대비 디블로킹 필터를 사용하지 않게 한 VTM 인코딩의 실험을 진행하였다. 표 7에서 보는 바와 같이, VVC로 TMIV를 압축할 경우, 오히려 디블로킹 필터를 적용하지 않는 경우가 전반적인 압축률 상승이 나타나는 것을 볼 수 있다. 아틀라스의 경우에 약

표 7. VVC로 TMIV 영상을 압축할 때, 디블로킹 필터에 따른 성능 변화

Table 7. Test results of deblocking filter when encoding TMIV by VVC

Sequence Name / View	QP	TMIV+VVC				TMIV+VVC (without deblocking filter)			
		Total kbps	Y-PSNR	U-PSNR	V-PSNR	Total kbps	Y-PSNR	U-PSNR	V-PSNR
ClassroomVideo / Basic view	27	17271.93	35.49	49.03	50.95	17284.87	35.51	49.10	50.99
	32	4116.47	34.73	47.59	49.44	4117.05	34.74	47.63	49.50
	37	1169.73	33.99	46.04	47.91	1172.17	33.99	46.03	47.88
	42	508.67	33.14	44.79	46.44	509.46	33.13	44.70	46.51
BD-rate						-1.18%			
ClassroomVideo / Atlas	27	4614.59	41.47	52.06	53.93	4615.15	41.52	52.47	54.37
	32	1613.33	40.47	49.89	51.48	1611.21	40.56	50.26	51.92
	37	642.61	39.34	47.42	49.09	640.28	39.47	47.75	49.37
	42	355.80	37.95	45.75	47.17	353.24	38.15	45.86	47.28
BD-rate						-7.83%			

7.83%의 압축률 향상이 나타나는 것을 볼 수 있다.

디블로킹 필터와 같이 일반적인 동영상에서 압축 효율을 증진시키던 기술이 아틀라스의 경우 오히려 압축률을 떨어뜨리는 작용을 할 수 있음을 확인하였다. 실험에 포함하지는 않았지만 VVC의 스크린콘텐츠를 위하여 채택된 Intra-block copy 기술이나 HEVC 대비 새로 채택이 된 움직임 예측 기술은 크게 영향을 미치지 않음을 확인하였다. 6DoF 영상에 대한 기준 연구방법을 단순히 적용하기 보다는, 아틀라스와 같은 특별한 경우에 대하여 어떻게 동영상 압축을 진행할 것인지, 어떠한 코덱을 사용하는 것이 효율적인지는 아직 열린 문제이다. 이에 따라, 기존 코딩 툴의 on/off 테스트뿐 아니라, 아틀라스 특성에 특화된 코딩 기법, 적응적 제어 기법 등의 새로운 연구가 필요하다.

V. 결 론

본 논문에서는 국제 표준 단체인 MPEG-I에서 개발 중인 3DoF+ 전방위 몰입형 비디오 처리 기술의 테스트 모델인 TMIV에 기준 적용된 비디오 인코더 HEVC를 VVC로 대체하여 성능 분석을 수행하였다. 실험의 결과, TMIV의 기준 비디오 인코더인 HEVC보다 VVC가 평균적으로 MIV mode의 Low에서 34.6%, Hight에서 33.0% BD-rate를 MIV 시점 mode의 Low에서 31.6%, High에서 28.8%의 BD-rate를 개선하였다. 이는 향후 6DoF 표준의 방향을 제시한다는 점과 VVC 기반의 6 DoF 비디오 코덱의 성능 평가는 본 논문이 최초라는 점에서 의의를 가진다. 후속 연구로는 최종 복원된 영상에 Video Super Resolution 딥 러닝 모델을 적용하여 화질을 개선하는 것이다.

참 고 문 헌 (References)

- [1] J. Lee, "Immersive Video Format Standardization Trends", *Broadcast and Media*, 23(4), pp.31-40, 2018.
- [2] ISO/IEC JTC1/SC29/WG11/M41555, "Proposed architectures for supporting Windowed 6DoF, Omnidirectional 6DoF and 6DoF media", *MPEG 120*, Oct. 2017.
- [3] <https://www.theverge.com/2017/4/19/15345738/facebook-surround-360-video-cameras-f8-conference-2017>
- [4] Yu-Chuan su, Kristen Grauman, "Learning Compressible 360° Video Isomers", In *CVPR*, 2018.
- [5] ISO/IEC JTC1/SC29/WG11/M48278, "[MPEG-I Visual] Proposal of Depth Improvement for Depth Estimation Reference Software (DERS8.0)", *MPEG 127*, July. 2019.
- [6] ISO/IEC JTC1/SC29/WG11/M45995, "[MPEG-I Visual] Results of Exploration Experiments for MPEG-I: Omnidirectional 6DoF - EE synthesis", *MPEG 125*, Jan. 2019.
- [7] ISO/IEC JTC1/SC29/WG11/M47705, "[MPEG-I Visual] 6DoF EE Proposal for ClassroomImage Pose e", *MPEG 126*, Mar. 2019.
- [8] ISO/IEC JTC1/SC29/WG11/N18564, "Exploration Experiments for MPEG-I: 6DoF", *MPEG 127*, July. 2019.
- [9] ISO/IEC JTC1/SC29/WG11/N18450, "Manual of Depth Estimation Reference Software (DERS 8.0)", *MPEG 126*, Mar. 2019.
- [10] ISO/IEC JTC1/SC29/WG11/N18172, "Versatile View Synthesizer (VVS) 2.0 manual", *MPEG 125*, Jan. 2019.
- [11] ISO/IEC JTC1/SC29/WG11/M47170, "[MPEG-I Visual] Proposal of Trimming and Color Matching of Multi-View Sequences", *MPEG 126*, Mar. 2019.
- [12] ISO/IEC JTC1/SC29/WG11/N18145, "Call for Proposals on 3DoF+ Visual", *MPEG 125*, Jan 2019.
- [13] ISO/IEC JTC1/SC29/WG11/N18577, "Test Model 2 for Immersive Video", *MPEG 127*, July 2019.
- [14] ISO/IEC JTC1/SC29/WG11/M47544, "Description of Technicolor Intel response to MPEG-I 3DoF+ Call for Proposal", *MPEG 126*, Mar. 2019.
- [15] ISO/IEC JTC1/SC29/WG11/M47407, "Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by Poznan University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI)", *MPEG 126*, Mar. 2019.
- [16] ISO/IEC JTC1/SC29/WG11/M47372, "Description of Nokia's response to CFP for 3DOF+ visual", *MPEG 126*, Mar. 2019.
- [17] ISO/IEC JTC1/SC29/WG11/M47684, "Description of Zhejiang University's response to 3DoF+ Visual CfP", *MPEG 126*, Mar. 2019.
- [18] ISO/IEC JTC1/SC29/WG11/M47179, "Philips response to CfP on 3DoF", *MPEG 126*, Mar. 2019.
- [19] JVET of ITU-T and ISO/IEC, "Meeting Report of the 15th Meeting of the Joint Video Experts Team (JVET)", Document JVET-O2000, July 2019.
- [20] ISO/IEC JTC1/SC29/WG11/N18563, "Common Test Conditions for Immersive Video", *MPEG 127*, July 2019.
- [21] ISO/IEC JTC1/SC29/WG11/N18576, "Working Draft2 of Metadata for Immersive Video", *MPEG 127*, July 2019.
- [22] JCT-VC of ITU-T and ISO/IEC, "Common test conditions", Document JCTVC-Z1100, January 2017
- [23] JVET of ITU-T and ISO/IEC, "JVET common test conditions and software reference configurations for SDR video", Document JVET-N1010, March 2019.
- [24] JVET of ITU-T and ISO/IEC, "Algorithm description for Versatile Video Coding and Test Model 6 (VTM 6)", Document JVET-O2002-v2, July 2019.
- [25] ISO/IEC JTC1/SC29/WG11/N18068, "Reference View Synthesizer

- (RVS) manual”, MPEG 124, Oct. 2018.
- [26] ISO/IEC JTC1/SC29/WG11/M42415, “3DoF+ test sequence ClassroomVideo”, MPEG 122, April. 2018.
- [27] ISO/IEC JTC1/SC29/WG11/M43748, “Kermit test sequence for Windowed 6DoF Activities”, MPEG 123, July. 2018.
- [28] Y. Sun, A. Lu, L. Yu, “Weighted-to-spherically-uniform quality evaluation for omnidirectional video”, IEEE Signal Process. Lett., 24(9), pp.1408-1412, 2017.
- [29] G. Choi, G. Bang, “MPEG-I Video Standardization Trends”, Broadcast and Media, 23(4), pp.41-57, 2018.
- [30] ISO/IEC JTC1/SC29/WG11/N18709, “Software manual of IV-PSNR for Immersive Video”, MPEG 127, July 2019.

저자 소개

박현수



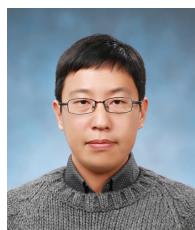
- 2018년 2월 : 이화여자대학교 전자공학과 (학사)
- 2018년 3월 ~ 현재 : 이화여자대학교 전자전기공학과 (硕사)
- ORCID : <https://orcid.org/0000-0001-6996-6821>
- 주관심분야 : Immersive Video, Image/Video Processing, Deep Learning

박상효



- 2011년 2월 : 한양대학교 컴퓨터전공 (학사)
- 2017년 8월 : 한양대학교 컴퓨터-소프트웨어학과 (硕사)
- 2017년 5월 ~ 2018년 2월 : 전자부품연구원 지능형영상처리센터 Post-doc.
- 2018년 3월 ~ 2018년 12월 : 연세대학교 바른ICT연구소 연구원
- 2019년 2월 ~ 현재 : 이화여자대학교 전자전기공학과 박사후연구원
- ORCID : <https://orcid.org/0000-0002-7282-7686>
- 주관심분야 : HEVC, VVC, Encoding/Decoding Complexity, Omnidirectional Video, Deep Learning

강제원



- 2006년 2월 : 서울대학교 전기컴퓨터공학부 (학사)
- 2008년 2월 : 서울대학교 전기컴퓨터공학부 (硕사)
- 2012년 5월 : University of Southern California (硕사)
- 2012년 8월 ~ 2014년 2월 : Qualcomm (Multimedia RnD and Standard team)
- 2014년 3월 ~ 2019년 3월 : 이화여자대학교 전자공학과 조교수
- 2019년 3월 ~ 현재 : 이화여자대학교 전자전기공학과 부교수
- ORCID : <https://orcid.org/0000-0002-1637-9479>
- 주관심분야 : HEVC (High Efficiency Video Coding), 3D Video, Image/Video Processing, Machine Learning