

특집논문 (Special Paper)

방송공학회논문지 제25권 제6호, 2020년 11월 (JBE Vol. 25, No. 6, November 2020)

<https://doi.org/10.5909/JBE.2020.25.6.911>

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

몰입형 비디오 압축을 위한 스크린 콘텐츠 코딩 성능 분석

이순빈^{a)}, 정종범^{a)}, 김인애^{a)}, 이상순^{b)}, 류은석^{a)*}

Screen Content Coding Analysis to Improve Coding Efficiency for Immersive Video

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

요약

최근 MPEG-I (Immersive) 그룹에서는 몰입형 비디오(Immersive Video)에 대한 표준화 프로젝트를 통해 압축 성능 탐색을 진행하고 있다. MIV(MPEG Immersive Video) 표준 기술은 다수의 시점 영상과 깊이 맵을 통한 깊이 맵 기반 이미지 렌더링(DIBR)을 바탕으로 제한적인 6DoF를 제공하고자 하는 기술이다. 현재 MIV에서는 바탕 시점(Basic View)과 각 시점의 고유한 영상 정보를 패치 단위로 모아둔 추가 시점(Additional View)으로 처리하는 모델을 채택하고 있다. MIV에서 생성된 아틀라스는 포함되는 시점의 성격에 따라 다른 영상의 특성을 나타내어 비디오 코덱의 압축 효율에 대한 고찰이 필요하다. 따라서 본 논문에서는 다양한 시점과 패치들이 반복되는 패턴에 착안하여 화면 내 블록 카피(IBC: intra block copy) 등의 압축 기법이 포함된 스크린 콘텐츠 코딩 툴에 대한 성능 비교 분석을 진행하여 복원 영상에서 최대 -15.74% Peak Signal-to-Noise Ratio (PSNR) 관점에서의 부호화 성능 향상을 제공하였다.

Abstract

Recently, MPEG-I (Immersive) has been exploring compression performance through standardization projects for immersive video. The MPEG Immersion Video (MIV) standard technology is intended to provide limited 6DoF based on depth map-based image rendering (DIBR). MIV is a model that processes the Basic View and the residual information into an Additional View, which is a collection of patches. Atlases have the unique characteristics depending on the kind of the view they are included, requiring consideration of the compression efficiency. In this paper, the performance comparison analysis of screen content coding tools such as intra block copy (IBC) is conducted, based on the pattern of various views and patches repetition. It is demonstrated that the proposed method improves coding performance around -15.74% BD-rate reduction in the MIV.

Keyword : Immersive Video, V3C, Screen Content Coding, Intra Block Copy

a) 성균관대학교 컴퓨터교육과(Department of Computer Education, Sungkyunkwan University)

b) 가천대학교 컴퓨터공학과(Department of Computer Engineering, Gachon University)

* Corresponding Author : Eun-Seok Ryu

E-mail: esryu@skku.edu

Tel: +82-2-760-0677

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

※ 이 논문의 연구 결과 중 일부는 “2020년 한국방송·미디어공학회 하계학술대회”에서 발표한 바 있음.

※ 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019R1A2C1010476). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2019R1A2C1010476).

· Manuscript received September 21, 2020; Revised November 11, 2020; Accepted November 12, 2020.

Copyright © 2020 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.”

1. 서론

최근 가상현실(Virtual Reality, VR) 및 증강현실에서의 서비스 기술에 대한 수요가 점차 높아지고 있다. 대표적인 VR 기술로는 360 전방위 영상 서비스가 있으며, 단순한 영상을 넘어 사용자에게 보다 높은 시점 자유도(Degrees of Freedom, DoF)를 제공하려는 기술에 대한 요구사항 또한 점차 확대되고 있다. Head Mounted Display (HMD)를 이용하여 3차원 공간에서 감상자의 머리 방향에 따른 세가지 자유도를 제공하는 비디오를 3DoF로 정의하며, 머리 회전 뿐만 아니라, 위치에 따른 움직임에 대해 제한된 시점 변환을 제공하는 3DoF+, 자유로운 시점 변환을 제공하는 6DoF 비디오 기술로 정의되고 있다.

이와 같은 몰입형 비디오에 대한 수요에 발맞추어, 현재 표준화 단체 MPEG(moving picture experts group)에서는 여러 위치에 다양한 시점으로 존재하는 영상들을 취득하고 처리하여 사용자에게 화면 전환 이동에 따른 운동 시차(motion parallax)를 제공할 수 있는 몰입형 비디오 기술에 관한 표준화가 진행되고 있다. 2019년 3월 MPEG 126차 회의에서 MPEG-I (Immersive)에서는 3DoF+와 6DoF를 제공하는 몰입형 비디오를 지원하는 테스트 모델(Test Model for Immersive Video, TMIV)를 정의하였으며^[1], 해당 기술을 통해 감상자가 HMD를 착용하고 제한된 영역에서 몰입형 비디오에 대한 전방위 가상 시점을 경험할 수 있도록 한다.

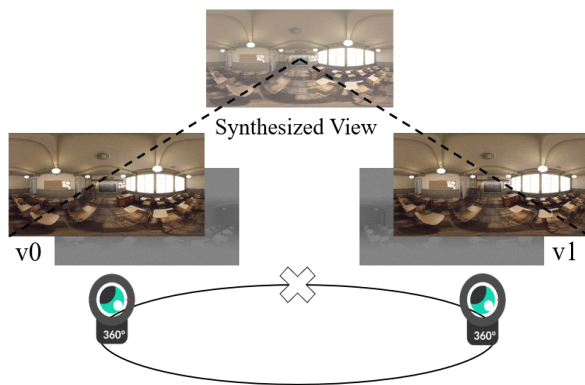


그림 1. 깊이 맵 기반 이미지 렌더링(DIBR)을 통한 중간 시점 생성의 예시
Fig. 1. Example of synthesized view using depth image-based rendering

그림 1은 깊이 맵 기반 이미지 렌더링 기술을 이용하여 중간 시점의 영상을 생성해내는 예시를 보여준다. 이 때 몰입형 비디오는 텍스처(texture) 영상과 깊이 맵(depth map)을 포함하는 다수의 영상으로 이루어진다. 그리고 취득한 깊이 맵을 바탕으로 각 영상 간의 시점과 거리를 계산하게 된다. 이 계산된 시점과 거리를 통해 감상자의 시선과 위치에 대응되는 가상 시점의 영상을 합성해내는 방식으로 이루어지며, 따라서 사용되는 영상의 개수나 위치, 그리고 깊이 맵의 정확도에 따라 합성된 시점의 품질이 매우 달라질 수 있다.

몰입형 비디오 기술 표준화는 다수의 영상 처리를 목적으로 MIV(MPEG Immersive Video)라는 이름으로 성능 탐색을 현재도 지속적으로 수행하고 있다. 6DoF 비디오는 자유로운 사용자의 시선과 움직임에 따른 시점을 지원해야 하므로, 이를 위해 여러 위치에서 취득된 영상들을 동시에 압축 및 전송하여 사용자 시점의 영상을 취득하고 합성하는 과정에서 높은 대역폭과 연산량이 요구된다. 따라서 MPEG-I에서는 6DoF 비디오의 압축 코덱 개발을 위해 여러 요소에 걸쳐 다양한 탐색 실험(Exploration Experiment, EE)을 진행하고 있다^[2-4].

MPEG-I의 테스트 모델인 TMIV에서의 가장 큰 특징은 바탕 시점(Basic View)과 각 시점의 고유한 영상 정보를 패치로 모아둔 추가 시점(Additional View)으로 분리하여 처리하는 모델을 채택하고 있다^[3]. 바탕 시점은 일반적으로 완전한 하나 이상의 시점 영상을 가리키며, 추가 시점의 경우 바탕 시점과의 차분 정보만을 가진 차분 패치의 모음을 의미한다. TMIV 부호화 과정을 통해 생성된 영상을 아틀라스라고 부르며, 아틀라스에는 하나 이상의 바탕 시점과 추가 시점이 포함될 수 있다. 이러한 아틀라스 영상은 처리 방식에 따라 일반적인 자연 동영상과는 매우 다른 특성을 나타내게 된다. 본 논문에서는 이러한 아틀라스 영상의 특성에 착안하여 스크린 콘텐츠 코딩 툴을 적용해 부호화 효율을 높일 수 있음을 보인다. 또한 스크린 콘텐츠 코딩 툴 중 부호화 효율이 가장 높다고 알려졌으며, 자기 유사적인 형태에 대해 높은 부호화 성능을 보이는 툴인 화면 내 블록 카피(intra block copy, IBC)에 주목하여 부호화 효율 분석을 진행하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 몰입형 비디오

오 코덱에 대한 전반적인 내용을 설명하고, 3장에서는 실험에 사용한 시퀀스의 특성과 실험환경, 4장에서는 성능 평가와 결과를 분석한다. 마지막으로 5장에서는 결론을 맺는다.

II. 몰입형 비디오 코덱

1. MIV 표준화 개요

MPEG-I 프로젝트(ISO/IEC 23090)에서는 전방위 미디어 포맷(Part 2 Omnidirectional Media Format, OMAF)을 아우르는 3차원 자유도를 제공하는 3DoF 기술에 대하여 2018년 첫 번째 단계 표준화를 시작하였다. 이후 시청자의 시야 회전 대응뿐만 아니라 3차원 공간에서의 움직임에 대한 시점 변환을 제공하는 6DoF 기술에 대한 논의가 이루어졌으며 이후 125차 회의에서 진행된 제안 요청(Call for Proposals)를 통해 몰입형 비디오(Part 12 MPEG Immersive Video, MIV)를 정의하고 표준화를 시작하였다^[5]. 이후 129차 회의에서 비디오 기반 포인트 클라우드 압축(Part 5 Video-based Point Cloud Compression, V-PCC)과의 유사성을 고려하여 용어, 신택스 구문, 복호화 과정 등 통합을

논의하였고 그 결과 130차 회의에서 공통 명세인 Visual Volumetric Video-based Coding(V3C)을 정의하였다^[6-7]. V3C는 V-PCC와 MIV의 확장 메커니즘을 제공하며, MIV 역시 V3C 단위 비트스트림 구조를 가지게 된다.

3DoF+ 압축 표준 제정은 120차 회의에서 다섯 개의 기관에 대한 제안 응답(Call for Response)에 따라 각 기관의 주요 기술을 통합하여^[8-12] 레퍼런스 소프트웨어 모델인 TMIV가 제정되었다.

그에 따른 TMIV 부호화기의 구성은 그림 2와 같이 나타낼 수 있다. 첫 번째 단계로, 다수 입력 영상들 중 특별한 처리 과정 없이 원본을 그대로 전송할 바탕 시점에 대한 결정이 시점 최적화기(View Optimizer)에서 이루어진다. 이후 카메라 파라미터 등 미리 취득되었던 몰입형 비디오에 대한 시퀀스 파라미터와 함께 아틀라스 생성기(Atlas Constructor)를 거치게 된다. 아틀라스 생성기에서는 첫 번째로 Pruner를 통해 각 시점의 영상들에 대하여 바탕 시점과의 차이를 계산하여, 중복된 부분이 제거된 잔여 영상, 즉 패치들을 생성하게 된다. 다음으로 Aggregator를 통하여 비디오 프레임에 대한 마스크를 시간적으로 고려하여 하나의 단위로 묶게 되고, Patch Packer를 거쳐 각 잔여 패치들을 하나의 영상에 병합하고 마지막으로 추가 시점의 아틀라스 영상을 생성하게 된다. 전체 과정은 텍스처 영상과 깊

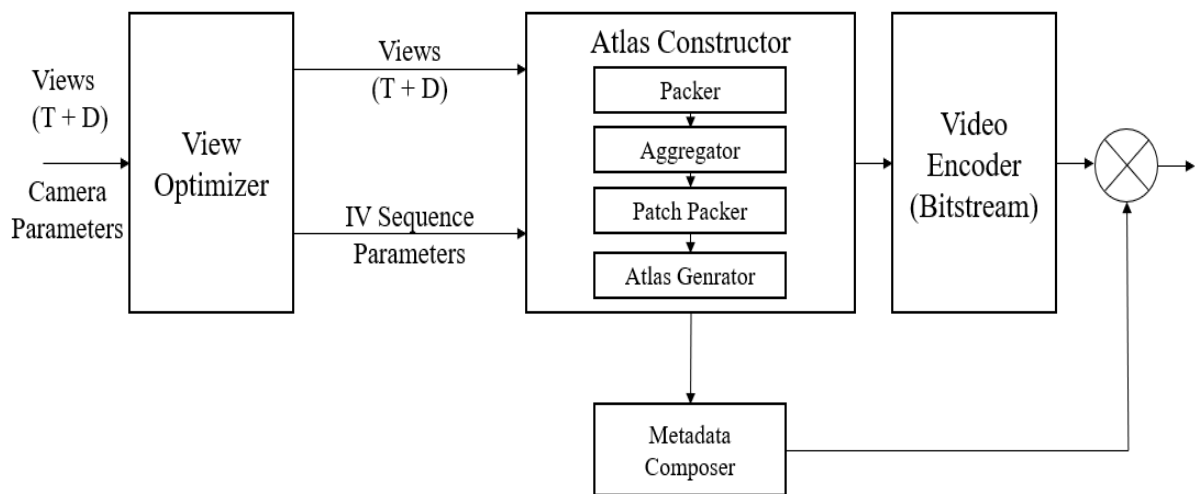


그림 2. TMIV 인코더의 블록 다이어그램 (T: Texture, D: Depth)
 Fig. 2. Block diagram of TMIV Encoder (T: Texture, D: Depth)

이 맵 영상 모두 적용된다. 추가 시점은 영상의 특성에 따라 바탕 시점이 미처 포함할 수 없는 주변 영상, 각 영상의 차이, 바탕 시점에서 볼 때 가려져 보이지 않는 물체 등을 포함한다. 위와 같은 처리를 거쳐 TMIV로 부호화된 아틀라스 영상들을 기존 비디오 코덱인 High Efficiency Video Coding (HEVC)등을 통해 최종적으로 비트스트림 형태로 압축하게 되며, 복원에 필요한 잔여 패치 영상의 위치 정보 등을 담은 메타데이터와 함께 병합하여 전송하는 모델을 기본으로 하게 된다. 이후 현재 131차 회의를 통해 6번째 테스트 모델(Test Model 6)이 구현되었으며^[13], 채택된 중점적인 개선 사항들은 다음과 같다.

- 그룹 기반 인코딩(Group-based encoding)^[14] : TMIV에서 모든 시점을 한번에 고려할 경우, 전경과 배경 구분이 어려워짐에 따라 최종 렌더링 과정에서 배경에 해당하는 패치가 전경 물체에 합성되는 등의 합성 오류를 일으킬 수 있다는 문제가 지적되었다. 따라서 이러한 사항을 개선하기 위해 중심 축을 기준으로 시점을 그룹으로 분할하였을 때 최대 3dB까지의 이득을 얻을 수 있음을 보고하였다. 이와 같은 맥락으로 선택적인 복원을 진행할 수 있는 개체의 위치 정보가 주어졌을 때 개체 별로 분할하여 진행할 수 있는 개체 기반 인코딩(Entity-based encoding) 기법 또한 포함되었다^[15].

- 패치의 시공간적 중복성 제거(Spatio-temporal patch redundancy removal)^[16] : 패치의 형태가 픽처 단위(Group of Pictures, GOP)로 고정될 경우 영상에서 움직임이 일어나면 특정 프레임만이 필요한 정보가 전체 픽처 단위에서 중복되는 문제를 제기하였다. 이에 따라 패치의 시공간적 중복성을 제거하여 전송되는 비트레이트의 양을 효과적으로 감소시켰으며, 깊이 맵이 부정확한 자연 영상에서는 팽창(dilation)을 통해 패치의 정보를 보존하는 것이 더 효율적임을 보였다^[17].
- 바탕 시점 할당기(Basic view allocator)^[18] : 모바일 등의 단말에서 현실적인 물입형 비디오 감상을 위해 최대 LumaPictureSize와 최대 복호화기 개수의 제한을 만족하며 아틀라스를 생성하고자 하는 방법이 제안되었다. 특히 바탕 시점이 갖는 완전한 영상 형태의 정보가 중요하다고 여겨짐에 따라 ‘바탕 시점의 우선 할당, 회전과 분할 금지’등의 규칙과 함께 최대한 바탕 시점을 포함하도록 TMIV의 Packing 방식이 재구성되었다.

2. MIV 부호화 특성

그림 3은 Fan 영상의 TMIV 6.0 부호화 결과를 간소화하여 나타낸다. TMIV에서는 최대 Luma picture size와 복호

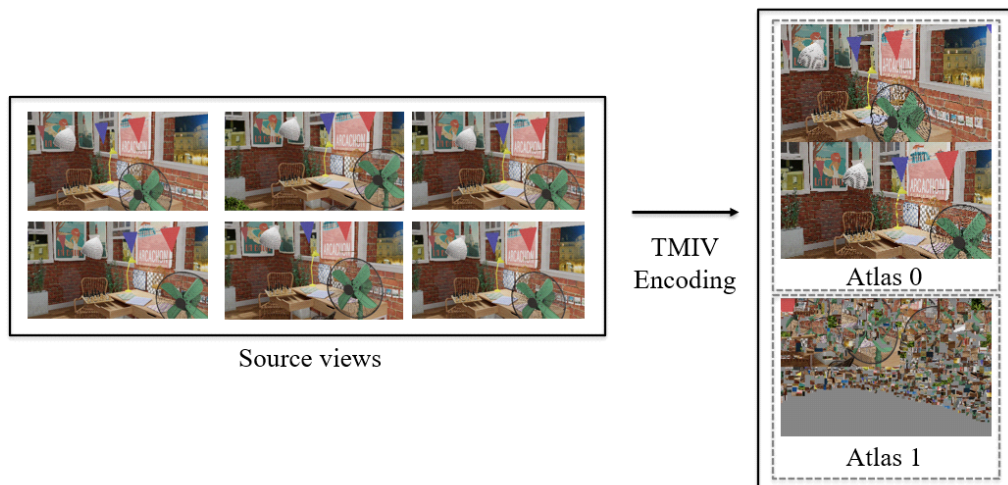


그림 3. Fan 영상의 아틀라스 예시
Fig. 3. Example of TMIV 6.0 encoded views (Fan Sequence)

화기 개수 제한에 따라 아틀라스의 해상도 크기와 개수를 결정한다. 그림 3에서는 2개의 아틀라스로 결정된 예시를 나타내며, 이후 주어진 제한에 만족하도록 바탕 시점의 비율을 계산하여 첫 번째 아틀라스에 바탕 시점을 우선적으로 포함한다. 이 때 바탕 시점이 모두 포함되더라도 아틀라스 내에 빈 공간이 남게 된다면 추가 시점의 잔여 패치가 따라 붙을 수 있다. 아틀라스 내에 남은 공간이 없다면, 다음 번 아틀라스에 대해 작업을 수행하게 된다. 그림 3에서 나타나듯이 바탕 시점이 담긴 첫 번째 아틀라스(Atlas 0)는 다수의 원본 시점 영상의 완전한 형태로 나타나며, 두 번째 아틀라스(Atlas 1)에는 나머지 잔여 패치가 붙는 형태가 된다. 이 때 첫 번째 아틀라스의 경우 담긴 영상들이 서로의 정보를 많은 부분 공유하고 있으며 중복되는 패턴을 많이 포함하고 있는 하나의 영상으로 간주할 수 있다. 또한 두 번째 아틀라스의 경우에도 비슷한 형태의 차분 패치들이 영상 안에 나타나는 어느정도 자기 유사적인 형태를 보인다.

3. 스크린 콘텐츠 코딩

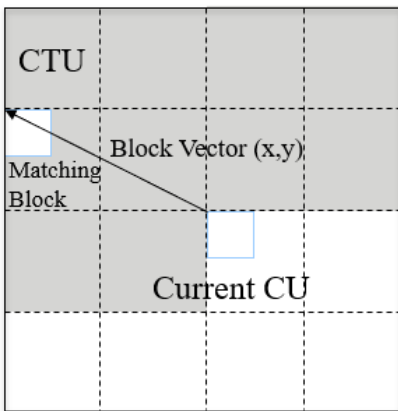


그림 4. 화면 내 블록 카피 기술 개념도.
 Fig. 4. Concept diagram of intra block copy

스크린 콘텐츠 코딩 툴은 기존의 비디오 압축 표준 기술로는 현재 수요가 증가하고 있는 애니메이션, 컴퓨터 그래픽스와 같은 스크린 콘텐츠 영상에 대한 압축을 효율적으로 지원하지 못한다는 문제가 제기됨에 따라 HEVC-SCC에 대한 표준화가 진행되었다. HEVC-SCC에서는 비교적 색상과 모양이 단조로운 특성을 반영한 색 공간 예측 기법,

팔레트 모드 등 효율적인 스크린 콘텐츠 영상 압축을 위한 기술들을 포함한다. 특히 그 중에서도 화면 내 블록 카피(IBC) 기법은 SCC에 채택된 부호화 기술 중 가장 높은 성능 향상을 보여주고 있다. 스크린 콘텐츠 코딩 툴에 포함된 압축 기법으로는 대표적으로 다음과 같은 항목이 있다^[19].

- 1) 화면 내 블록 카피 (IBC): 같은 프레임 내 또는 슬라이스 단위 내의 영역에서, 이미 복원된 샘플들 중 부호화 중인 블록과 가장 유사한 블록을 예측 블록으로 사용하는 기술이다. IBC 기법은 단순한 형태나 그래픽 특성상 텍스처 반복 등이 자주 일어나는 영상들에 대하여 예측성능을 높이기 위한 모드이다. 특히 반복적으로 같은 모양을 나타내는 텍스트와 그래픽이 혼합된 영상에서 IBC의 선택 비율이 상대적으로 높게 나타난다.
- 2) 팔레트 부호화 모드 (Palette Coding mode): 스크린 콘텐츠 영상에서 대표로 빈번하게 등장하는 색상 값을 인덱스로 하는 매핑 테이블을 구성한 후, 나머지 예외적인 색상 값의 경우 양자화와 부호화를 수행하여 전송하는 기술이다.
- 3) 색 공간 변환 기법 (Adaptive Color Space Transform): 원본 영상에서 색 공간을 새로운 공간으로 방법으로 색상 요소 간 중복성을 줄이기 위한 목적으로 4:4:4 동영상 형식을 부호화하는데 효율 향상을 보임이 알려졌다. HEVC-SCC에서는 CU 단위로 부호화 적용 여부를 판단하여 모드 정보를 전송하는 방식으로 변환이 이루어지게 된다. 본 논문에서는 4:2:0 동영상 형식을 이용하여 색 공간 변환 기법의 경우 비활성화하여 실험을 진행하였다.

III. 실험 환경

1. 테스트 6DoF 비디오와 실험 환경

현재 MIV의 공통 실험 조건(Common test conditions,

CTC) 문서에서는 몰입형 비디오의 테스트 조건에 대해 다음과 같은 제한을 고려하고 있다^[20]. 특히 최대 복호화기 개수를 만족시키기 위해 바탕 시점 할당기에서 최대 아틀라스 개수를 2개로 제한하는 것으로 공통 실험 조건 설정 사항에 정의되어 있다. 깊이 맵의 경우 다운샘플링의 여부가 선택적으로 이루어지기 때문에 완전한 형태의 깊이 지도 맵의 경우를 고려하여 최대 아틀라스 개수를 2개로 제한한다. 표 1은 공통 실험 조건 하의 Pixel rate 테스트 조건 제한을 나타낸다. ‘minNonCodedViews’는 최소한으로 포함되어야 하는 바탕 시점의 개수를 나타낸다. 아틀라스 내의 바탕 시점의 비율을 50%(maxBasiViewFraction:0.5)로 설정할 경우 현재 공통 실험 조건의 영상들이 모두 최대 아틀라스 개수 제한을 만족하게 되어, 성능 비교 기준(anchor) 설정으로 지정되어있다. 표 2는 앞서 서술한 바탕 시점 할당기의 설정 값을 나타낸다.

표 1. Pixel rate 테스트 조건 제한
Table 1. Pixel rate test condition constraints

Category	Max luma sample rate	Max luma picture size	Max Decoder Instance
Low pixel rate (HEVC Main 10 Profile Level 5.2)	1,069,547,520 (32 MP, 30fps)	8,912,896 (4096x2048)	4
High pixel rate (HEVC Main 10 Profile Level 6.2)	4,278,190,080 (128 MP, 30fps)	35,651,584 (8192x4096)	4

표 2. 바탕 시점 할당기 설정 값
Table 2. configuration value of basic view allocator

maxBasicViewFraction	0.5
outputAdditionalViews	true
minNonCodedViews	3

본 논문에서는 MIV 영상에서 스크린 콘텐츠 코딩 툴에 대한 성능 분석을 진행하였다. HM 16.16+SCM8.6, MIV 6.0으로 현재 MPEG-I에서 진행중인 CTC에 정의되어 있는 영상에 대하여 실험을 진행하였다^[20]. 표 3은 공통 실험 조건의 각 영상의 정보를 나타낸다. 자연 취득된(NC) 영상의 경우, 컴퓨터로 생성된(CG) 영상에 비해 깊이 맵 영상의 잡음이 심한 편이며, 따라서 최종 합성 영상의 품질에도 영

향을 끼치게 된다. 공통 실험 조건의 영상들은 모두 10비트 영상으로 구성되어 있지만 HEVC에 포함된 SCC 툴의 경우 8비트만을 지원하는 문제로 8비트로 변환하여 부복호화를 시행한 뒤, 이를 다시 10비트로 재변환하여 TMIV 복호화 과정을 거쳐 실험을 진행하였다^[21].

표 3. 테스트 시퀀스의 특성
Table 3. Properties of test video sequences

Sequence name	Class	No. of source view	Resolution	Category
ClassroomVideo	A	15	4096x2048	Computer-generated
Museum	B	24	2048x2048	Computer-generated
Hijack	C	10	4096x4096	Computer-generated
Chess	N	10	2048x2048	Computer-generated
Painter	D	16 (4x4)	2048x1088	Natural contents
Frog	E	13 (13x1)	1920x1080	Natural contents
Fencing	L	10	1920x1080	Natural contents
Carpark	P	9 (9x1)	1920x1088	Natural contents
Street	U	9 (9x1)	1920x1088	Natural contents
Hall	T	9 (9x1)	1920x1088	Natural contents
Group	R	21	1920x1080	Computer-generated
Fan	O	15	2048x2048	Computer-generated
ChessPieces	Q	10	2048x2048	Computer-generated

2. 성능 평가 방법과 지표

몰입형 비디오는 가상의 시점에 대응한다는 영상 특성상 성능 평가 방법이 3가지로 정의되어 있다.

첫 번째는 아틀라스에 대한 압축 성능 평가이다. 원본 시점 영상들로부터 TMIV 부호화를 거친 아틀라스 영상에 대해, 부복호화를 거친 후 성능 평가를 진행한다. 아틀라스에 대한 압축 성능 평가는 원본 시점들을 다시 복원하지 않아도 되어 비교적 간단하게 수행할 수 있다는 장점이 있지만, 아틀라스에 대한 압축 성능 평가만으로는 몰입형 비디오에

대한 복원된 영상의 성능 평가를 진행할 수 없다.

두 번째 방법으로 부복호화를 거친 아틀라스 영상을 다시 TMIV 복호화를 거쳐 각 시점의 복원 영상들과 원본 영상 간의 평균을 비교하는 방법이다. 복원 영상에 대한 품질 평가를 진행함에 따라 전체적인 몰입형 비디오 압축 과정에 대한 평가를 진행할 수 있다.

세 번째는 복원 영상을 통해 시청자의 임의의 시점에 따라 합성된 가상 시점 합성 영상에 대한 품질 평가를 진행하는 방법이다. 이 방법의 경우 원본 시점 영상으로 합성된 가상 시점 합성 영상 또한 합성 오차를 내포하고 있기 때문에 객관적인 품질 평가보다는 주로 주관적 품질 평가를 통해 이루어진다.

몰입형 비디오의 성능 평가 지표로는 전방위 비디오의 객관적 품질 평가를 위해 제안된 Weighted-to-Spherically-uniform Peak Signal-to-Noise Ratio(WS-PSNR), 비디오 품질 평가를 위해 Netflix에서 제안한 혼합 평가 지표인 Video Multimethod Assessment Fusion (VMAF), 그리고 합성 오차에 대한 보정과 렌더링 과정에서 일어나는 색차에 대한 보정을 포함하도록 블록 단위의 PSNR 측정을 사용하는 PSNR for Immersive Video (IV-PSNR) 등이 있다^[22]. 또한 QP 단계를 5단계로 지정하고, 높은 비트레이트

구간(QP1~QP4)과 낮은 비트레이트 구간(QP2~QP5)에 대한 성능 평가를 구분하여 보다 세밀한 평가가 진행될 수 있도록 정의하였다.

IV. 실험 결과 및 분석

표 4는 HEVC 대비 SCC 적용 시 아틀라스에 대한 압축 성능에 대해 BD-rate 이득을 나타낸다. 바탕 시점이 우선적으로 포함되는 첫 번째 아틀라스인 Atlas 0과, 추가 시점의 잔여 패치가 많이 포함될 가능성이 높은 다음 번 아틀라스 Atlas 1로 나누어 측정을 진행하였다.

표 4의 Atlas 0의 텍스처 이득을 살펴보면, CG 영상 class D와 NC 영상 class O를 제외하고는 상대적으로 높지 않은 효율을 보였다. class D 영상은 4x4의 카메라 배열로 촬영된 영상으로, 각 시점의 중복도가 큰 경향의 영상이다. 이는 바탕 시점 포함에 따른 아틀라스의 특성보다는 각 영상의 형태와 취득 카메라 배열에 따라 SCC의 효율이 크게 달라진다는 것을 시사한다. CG 영상과 NC 영상간의 유의미한 이득의 상관관계는 관찰할 수 없었다. 깊이 맵의 이득을 살펴보면, NC 영상에서 약 -9~17% 정도의 이득을 보이는 반면

표 4. HEVC 대비 HEVC-SCC의 아틀라스 BD-rate (Y-PSNR)
 Table 4. Atlas BD-rate HEVC over HEVC-SCC (Y-PSNR)

Atlas 0	SA	SB	SC	SD	SE	SL	SN	SO	SQ	SR	ST	SU	SP
Texture high(1~4)	-1.82%	-1.63%	-0.44%	-20.48%	-3.67%	-0.50%	-0.77%	-12.33%	-2.15%	-0.42%	-1.46%	-3.51%	-3.51%
Texture low(2~5)	-3.70%	-1.74%	-0.61%	-21.21%	-3.27%	-0.36%	-0.89%	-12.41%	-0.51%	-0.94%	-1.42%	-4.97%	-4.97%
Texture avg.	-2.76%	-1.69%	-0.52%	-20.85%	-3.47%	-0.43%	-0.83%	-12.37%	-1.33%	-0.68%	-1.44%	-4.24%	-4.24%
Depth high(1~4)	-34.85%	-19.55%	-24.46%	-12.33%	-8.55%	-14.06%	-	-35.19%	-	-16.92%	-15.09%	-15.00%	-17.78%
Depth low(2~5)	-35.81%	-17.11%	-21.70%	-12.28%	-10.99%	-12.95%	-23.85%	-32.20%	-	-16.34%	-15.15%	-12.84%	-16.39%
Depth avg.	-35.33%	-18.33%	-23.08%	-12.31%	-9.77%	-13.50%	-	-33.69%	-	-16.63%	-15.12%	-13.92%	-17.09%
Atlas 1													
Texture high(1~4)	-9.14%	-1.20%	-5.22%	-2.23%	-1.65%	-2.51%	-3.15%	-4.64%	-4.08%	-2.96%	-2.55%	-2.07%	-2.07%
Texture low(2~5)	-9.97%	-1.63%	-5.91%	-2.14%	-1.35%	-1.79%	-3.77%	-4.75%	-3.82%	-3.68%	-3.52%	-1.90%	-1.90%
Texture avg.	-9.55%	-1.42%	-5.56%	-2.18%	-1.50%	-2.15%	-3.46%	-4.69%	-3.95%	-3.32%	-3.03%	-1.98%	-1.98%
Depth high(1~4)	-46.41%	-27.58%	-35.80%	-18.22%	-14.66%	-19.00%	-	-35.73%	-	-26.72%	-2.55%	-17.32%	-23.28%
Depth low(2~5)	-47.42%	-26.86%	-32.76%	-18.08%	-16.23%	-17.02%	-34.38%	-31.94%	-	-26.71%	-3.52%	-16.23%	-21.22%
Depth avg.	-46.91%	-27.22%	-34.28%	-18.15%	-15.44%	-18.01%	-	-33.83%	-	-26.72%	-3.04%	-16.77%	-22.25%

CG 영상에서는 -16~35%의 이득을 보인다. 이는 자연 취득 영상의 깊이 맵에 잡음이 많은 문제로 인하여 깊이 맵의 예측이 상대적으로 어렵고 불규칙한 형태를 나타내 이에 따라 SCC에 따른 압축 효과를 보지 못하는 것으로 해석된다.

잔여 패치가 포함될 가능성이 높은 Atlas 1의 텍스처 이득의 경우, class A 영상을 제외하면 -1~5%의 효율에 그쳐 자기 유사적인 형태를 지닌 잔여 패치는 SCC 적용에 따른 압축 효과를 크게 볼 수 없음을 보여준다. class A의 경우 하나의 바탕 시점을 가지는 360 전방위 영상으로 Atlas 1에서는 잔여 패치들만이 존재한다.

표 5. 영상 카테고리별 HEVC 대비 HEVC-SCC의 평균 아틀라스 BD-rate (Y-PSNR)
Table 5. Average atlas BD-rate HEVC over HEVC-SCC in Category (Y-PSNR)

	Computer-generated	Natural Content
Atlas 0 (Texture)	-2.88%	-5.77%
Atlas 0 (Depth)	-25.41%	-13.61%
Atlas 1 (Texture)	-4.56%	-2.18%
Atlas 1 (Depth)	-33.79%	-15.61%

표 6은 복원 영상 평균에 대한 BD-rate를 나타낸다. 복호화된 아틀라스를 이용하여 다시 원본 시점 영상들을 복원해내는 시점을 만들고, 원본 시점과의 품질 비교를 진행한다. 이 때 품질 평가는 모든 시점에 대해 평균을 낸 값으로 정의되며, 각각 WS-PSNR, MS-SSIM, VMAF, IV-PSNR 지표를 이용하여 측정된 값의 결과이다. 아틀라스 압축 성능 평가시 Atlas 0에서 높은 이득을 보여주었던 class D 영

표 6. HEVC 대비 HEVC-SCC의 복원 영상 BD-rate
Table 6. Synthesized views BD-rate (HEVC over HEVC-SCC)

Metric	SA	SB	SC	SD	SE	SL	SN	SO	SQ	SR	ST	SU	SP
WS-PSNR Synthesized	-11.71%	-4.32%	-15.74%	-14.08%	-5.62%	-7.21%	-9.99%	-18.02%	-9.65%	-8.23%	-5.92%	-6.80%	-8.35%
MS-SSIM Synthesized	-11.42%	-4.89%	-16.06%	-14.04%	-5.61%	-7.46%	-10.75%	-17.52%	-10.72%	-8.92%	-6.14%	-7.57%	-9.34%
VMAF Synthesized	-12.77%	-4.69%	-15.97%	-14.08%	-5.65%	-7.37%	-10.26%	-17.79%	-10.13%	-8.54%	-6.12%	-7.12%	-9.08%
IV-PSNR Synthesized	-10.28%	-4.42%	-15.29%	-14.53%	-5.89%	-7.33%	-9.76%	-18.07%	-9.47%	-8.12%	-6.15%	-7.22%	-8.55%

상은 상대적으로 복원 영상 평가에서는 가장 높은 효율을 보여주지 않았는데, 이는 CG 영상에서 깊이 맵 압축 효율에 따른 문제로 판단된다. 효율이 가장 높은 class O 영상의 경우, 텍스처와 깊이 맵 모두에서 높은 이득을 보여 전체 비트레이트를 크게 낮추었기 때문으로 해석된다.

아틀라스 평가에서 텍스처 이득이 두드러지지 않았던 영상들의 경우에도 SCC 적용 시 깊이 맵의 이득으로 인해 복원 영상의 BD-rate에서 전체적인 상승이 있는 것을 관찰할 수 있다. 또한 합성 오차를 보정하는 IV-PSNR의 경우 깊이 맵에 잡음이 많은 NC 영상에 대하여 더 높은 이득을 보이는 경향이 있다.

그림 5는 각각 HEVC와 HEVC-SCC를 이용하여 ‘v0’번째 영상을 복원한 결과이다. SCC를 적용한 오른쪽 영상이 같은 비트레이트 대역에서 더 선명하고 높은 화질을 가지는 것을 확인할 수 있다. 또한 VVC와 VVC-SCC를 실험해본 결과, 모든 영상의 텍스처 아틀라스에서 평균 -1% 미만의 이득을 확인하였다. HEVC-SCC에 구현된 IBC 기법은 다르게, VVC에서 현재 채택되어 포함되어 있는 IBC 기법은 하드웨어의 효율적인 구현을 위해 참조 샘플의 범위가 Left 1-CTU 크기로 제한되어 있다^[23]. 따라서 아틀라스의 특징적인 형태에 대한 이득을 IBC 기법으로 얻을 수 없는 것으로 판단된다.

V. 결론

TMIV에서 생성된 아틀라스는 처리 과정에서 자연적인

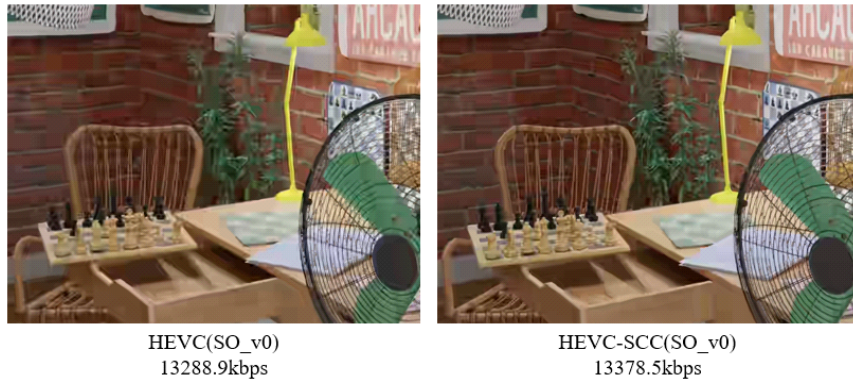


그림 5. SO_v0 영상의 복원 영상 결과
 Fig. 5. The results of synthesized SO_v0 view. HEVC (Left), HEVC-SCC (Right)

영상과는 다른 패턴이 나타나며, 이에 따라 IBC 기법을 포함한 스크린 콘텐츠 코딩 툴 적용에 대한 분석을 진행하였다. 여러 중복된 부분들이 포함된 영상들을 모아놓은 아틀라스에서도 카메라 배열이나 영상 특성에 따라 SCC의 효율이 매우 달라짐을 관찰하였다. 그러나 IBC 기법보다는 일반적으로 깊이 맵 압축 효율에 대한 이득이 뛰어나며, 이로 인해 복원 영상 평가에서는 더 높은 효율 이득을 나타낸다. 현재 TMIV에서는 바탕 시점을 우선적으로 첫 번째 아틀라스에 할당하지만, 분배하는 방법에 따라 IBC 기법의 효율 또한 달라질 수 있다. 몰입형 비디오는 중복성이 있는 각 시점간의 영상이나 잔여 패치들이 반복되는 패턴을 갖고 있거나, 영상 간의 의존성을 가지고 있는 특성으로 해당 부분에 대한 분석이 계속해서 이루어져야 할 것이다^[24-25].

참 고 문 헌 (References)

- [1] ISO/IEC JTC1/SC29/WG11/M41555, "Proposed architectures for supporting Windowed 6DoF, Omnidirectional 6DoF and 6DoF media", MPEG 120, Oct. 2017.
- [2] ISO/IEC JTC1/SC29/WG11/N19491, "Exploration Experiments on Coding for Future MPEG Immersive Video", MPEG 130, Apr. 2020.
- [3] ISO/IEC JTC1/SC29/WG11/N19492, "Exploration Experiments on Processing for Future MPEG Immersive Video", MPEG 131, Jul. 2020.
- [4] ISO/IEC JTC1/SC29/WG11/N19485, "Description of MPEG Immersive Video Core Experiments 1", MPEG 131, Jul. 2020.
- [5] ISO/IEC JTC1/SC29/WG11/N18145, "Call for Proposals on 3DoF+ Visual", MPEG 125, Jan. 2019.
- [6] ISO/IEC JTC1/SC29/WG11/M52994, "Proposed simplifications of MIV", MPEG 130, Apr. 2020.
- [7] ISO/IEC JTC1/SC29/WG11/M49590, "Video-based Point Cloud Coding High Level Syntax: Updates and Unification with the Working Draft on Metadata for Immersive Video", MPEG 127, Jul. 2019.
- [8] ISO/IEC JTC1/SC29/WG11/M47544, "Description of Technicolor Intel response to MPEG-I 3DoF+ Call for Proposal", MPEG 126, Mar. 2019.
- [9] ISO/IEC JTC1/SC29/WG11/M47407, "Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by PUT and ETRI", MPEG 126, Mar. 2019.
- [10] ISO/IEC JTC1/SC29/WG11/M47372, "Description of Nokia's response to CFP for 3DOF+ visual", MPEG 126, Mar. 2019.
- [11] ISO/IEC JTC1/SC29/WG11/M47684, "Description of Zhejiang University's response to 3DoF+ Visual CFP", MPEG 126, Mar. 2019.
- [12] ISO/IEC JTC1/SC29/WG11/M47179, "Philips response to CFP on 3DoF", MPEG 126, Mar. 2019.
- [13] ISO/IEC JTC1/SC29/WG11/N19484, "Test Model 6 for MPEG Immersive Video", MPEG 131, Jul. 2020.
- [14] ISO/IEC JTC1/SC29/WG11/M49406, "Group-Based TMIV", MPEG 127, Jul. 2019.
- [15] ISO/IEC JTC1/SC29/WG11/M52475, "Object-Based Implementation in TMIV", MPEG 129, Jan. 2020.
- [16] ISO/IEC JTC1/SC29/WG11/M51604, "Immersive Video CE3-related: Spatio-temporal patch redundancy removal", MPEG 129, Jan. 2020.
- [17] ISO/IEC JTC1/SC29/WG11/M53701, "CE-3: additional patch dilation in temporal patch redundancy removal", MPEG 130, Apr. 2020.
- [18] ISO/IEC JTC1/SC29/WG11/M54145, "[MIV] Philips CE-2 related response "basic view allocator"", MPEG 131, Jun. 2020.
- [19] "Overview of the Emerging HEVC Screen Content Coding Extension," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 50-62, Jan. 2016.
- [20] ISO/IEC JTC1/SC29/WG11/N19484, "Common Test Conditions for MPEG Immersive Video", MPEG 131, Jul. 2020.
- [21] ISO/IEC JTC1/SC29/WG11/M53427, "[MPEG-I Visual] HEVC-SCC in TMIV", MPEG 130, Apr. 2020.

- [22] ISO/IEC JTC1/SC29/WG11/N19495, "Software manual of IV-PSNR for Immersive Video", MPEG 131, Jul. 2020.
- [23] X. Xu, X. Li and S. Liu, "Intra block copy in Versatile Video Coding with Reference Sample Memory Reuse," Picture Coding Symposium (PCS), Ningbo, China, 2019.
- [24] J. Jeong, S. Lee, D. Jang and E. Ryu, "Towards 3DoF+ 360 Video Streaming System for Immersive Media," in IEEE Access, vol. 7, pp. 136399-136408, 2019.
- [25] J. Jeong, S. Lee, I. Ryu, T. T. Le, and E. Ryu. "Towards Viewport-dependent 6DoF 360 Video Tiled Streaming for Virtual Reality Systems." In Proceedings of the 28th ACM International Conference on Multimedia (MM '20), pp. 3687 - 3695, 2020.

저 자 소 개



이 순 빈

- 2020년 3월 : 가천대학교 컴퓨터공학과 학사
- 2020년 3월 ~ 현재 : 성균관대학교 컴퓨터교육과 석사과정
- ORCID : <https://orcid.org/0000-0002-8951-0335>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



정 종 범

- 2018년 8월 : 가천대학교 컴퓨터공학과 학사
- 2018년 9월 ~ 2019년 8월 : 가천대학교 컴퓨터공학과 석사과정
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 석박통합과정
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



김 인 애

- 2013년 8월 : 배재대학교 외식경영학 학사
- 2019년 9월 ~ 현재 : 성균관대학교 컴퓨터교육과 석사과정
- ORCID : <https://orcid.org/0000-0003-4263-6448>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 압축 표준



이 상 순

- 1982년 2월 : 인하대학교 전자공학과 학사
- 1986년 2월 : 인하대학교 전자계산학과 석사
- 2005년 2월 : 인천대학교 컴퓨터공학과 박사
- 1994년 2월 ~ 현재 : 가천대학교 컴퓨터공학과 부교수
- ORCID : <https://orcid.org/0000-0001-6680-2637>
- 주관심분야 : 컴퓨터네트워크, 시스템소프트웨어, IoT

저 자 소 개



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 조교수
- ORCID : <https://orcid.org/0000-0003-4894-6105>
- 주관심분야 : 멀티미디어 통신 및 시스템, 비디오 코딩 및 국제 표준, HMD/VR 응용분야