

특집논문 (Special Paper)

방송공학회논문지 제25권 제5호, 2020년 9월 (JBE Vol. 25, No. 5, September 2020)

<https://doi.org/10.5909/JBE.2020.25.5.665>

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

효율적인 3DoF+ 비디오 부호화를 위한 작은 블록 제거를 통한 아틀라스 생성 기법

임 성 균^{a)}, 김 현 호^{a)}, 김 재 곤^{a)†}

An Atlas Generation Method with Tiny Blocks Removal for Efficient 3DoF+ Video Coding

Sung-Gyun Lim^{a)}, Hyun-Ho Kim^{a)}, and Jae-Gon Kim^{a)†}

요 약

MPEG-I에서는 최대 6자유도(6DoF)를 가지는 몰입형(immersive) 비디오의 압축 표준화를 진행하고 있다. 360 비디오의 전방위 뷰(view)와 함께 움직임 시차(parallax)를 제공하는 3DoF+ 비디오는 고정된 위치에서 상체 움직임 정도의 제한된 공간에서 획득된 여러 뷰 비디오를 사용하여 임의의 원하는 시점의 뷰를 렌더링한다. MPEG-I 비주얼 그룹에서는 이러한 3DoF+ 비디오의 압축 표준화를 진행하고 있으며, 그 과정에서 TMIV(Test Model for Immersive Video)라는 시험모델을 개발하고 있다. TMIV에서는 다수의 입력 뷰 비디오들의 중복성을 제거하고 남은 텍스처 및 깊이 영상 패치(patch)들을 촘촘히 프레임으로 패킹한 아틀라스(Atlas)를 생성하고 이를 부호화한다. 본 논문은 보다 효율적인 3DoF+ 비디오 부호화를 위해서 작은 크기의 블록들을 제거하는 아틀라스 생성 기법을 제시한다. 제안기법은 TMIV 대비 자연영상과 그래픽 영상에서 각각 0.7%와 1.4%의 BD-rate 비트율 감소의 성능 개선을 보였다.

Abstract

MPEG-I is actively working on standardization on the coding of immersive video which provides up to 6 degree of freedom (6DoF) in terms of viewpoint. 3DoF+ video, which provides motion parallax to omnidirectional view of 360 video, renders a view at any desired viewpoint using multiple view videos acquisitioned in a limited 3D space covered with upper body motion at a fixed position. The MPEG-I visual group is developing a test model called TMIV (Test Model for Immersive Video) in the process of development of the standard for 3DoF+ video coding. In the TMIV, the redundancy between a set of input view videos is removed, and several atlases are generated by packing patches including the remaining texture and depth regions into frames as compact as possible, and coded. This paper presents an atlas generation method that removes small-sized blocks in the atlas for more efficient 3DoF+ video coding. The proposed method shows a performance improvement of BD-rate bit savings of 0.7% and 1.4%, respectively, in natural and graphic sequences compared to TMIV.

Keyword : MPEG-I, 3DoF+, Immersive video, Atlas, TMIV, 360 video

I. 서론

최근 360도 비디오 및 VR(Virtual Reality) 콘텐츠가 확산되고 있으며 360도의 전방위 비디오를 넘어 사용자에게 6DoF(Degree of Freedom)까지 시점의 자유도를 제공함으로써 높은 몰입도를 제공하는 몰입형(immersive) 비디오가 주목받고 있다. MPEG에서는 이러한 몰입형 미디어의 압축 및 전송에 대한 표준 패키지로 MPEG-I (MPEG-Immersive) 표준화를 진행하고 있다. MPEG-I는 2016년 10월 중국 청두에서 열린 116차 회의에서 만들어진 프로젝트 (“Coded Representation of Immersive Media,” ISO/IEC 23090)로 몰입형 미디어를 위한 아키텍처, 전송 포맷, 오디오/비디오 압축, 메타데이터 등을 포함한 표준 패키지이며 이에 대한 표준화가 진행되고 있다^{[1][2]}.

현재 MPEG-I 비디오 그룹에서는 360도 비디오에 상체의 움직임 정도의 제한된 범위에서의 움직임 시차(motion parallax)를 제공하는 3DoF+ 비디오 부호화 표준화를 진행하고 있다. 3DoF+ 비디오는 제한된 범위에서 획득 또는 생성된 10여개 이상의 360 비디오 또는 2D 비디오를 제공하고 이로부터 합성 및 렌더링(rendering)을 통하여 움직임 시차를 제공한다. 따라서 다수의 고화질(4K 또는 8K) 비디오로 구성되는 3DoF+ 비디오는 방대한 대용량 비디오가 되며 이를 효율적으로 전송하기 위한 보다 뛰어난 성능의 압축 기술이 요구된다. MPEG-I 비디오 그룹에서는 이러한 대용량의 3DoF+ 비디오를 효율적으로 압축하기 위한 MIV (MPEG Immersive Video) 표준화가 진행 중이며, 최근 CD(Committee Draft)^[3]를 발간하였으며 표준 개발 과정에서 시험모델인 TMIV(Test Model for Immersive Video)^[4]를 개발하고 있다.

3DoF+ 비디오는 제한된 공간의 서로 다른 뷰 포인트에서 획득 및 생성된 뷰 비디오로 구성되며, 따라서 각 뷰 비

디오 간에 높은 상관성이 존재한다. 3DoF+ 비디오 압축을 위한 MIV의 접근방법은 부호화 이전에 전처리를 통하여 3DoF+의 뷰 비디오 간에 존재하는 높은 중복성을 줄임으로써 화소율 압축(pixel rate reduction)을 하는 것이다. TMIV는 뷰 비디오 간의 중복성을 제거하고 남은 뷰 영상들로 아틀라스(Atlas)를 생성하는 화소율 압축의 전처리 과정과 HEVC(High Efficiency Video Coding)를 이용한 부/복호화 이후 중복성이 제거된 복호화된 비디오로부터 임의의 가상시점을 합성하는 렌더링(rendering)의 후처리 과정을 포함한다.

본 논문에서는, 아틀라스 내의 작은 블록들을 제거하여 3DoF+ 비디오의 부호화 효율을 개선하는 아틀라스를 생성 기법을 제안한다. 제안기법은 복원 및 합성되는 뷰포트(viewport) 영상의 화질에는 영향을 주지 않으면서 3DoF+ 비디오의 부호화 효율을 향상시킨다.

본 논문의 구성은 다음과 같다. 2장에서는 TMIV의 처리과정과 아틀라스 생성 방법에 대해 설명하고, 3장에서 제안기법을 제시한다. 4장의 실험결과에서는 제안기법의 성능검증과 분석 내용을 기술하고 5장에서 결론을 맺는다.

II. Test Model for Immersive Video (TMIV)

그림 1은 TMIV 부호화기의 전체적인 구성도로 3DoF+ 비디오의 부호화 전처리 과정을 보여준다. 먼저, 제한된 공간내의 다수의 위치 위치에서 획득 또는 생성된 소스 뷰 비디오와 해당 뷰의 카메라 파라미터가 입력된다. 이들 소스 뷰들은 입력된 비디오 그대로 부호화 될 기준 뷰(Basic View)와 기준 뷰와 중복되는 영역이 제거될 부가 뷰(Additional View)로 분류된다. 다음으로, Atlas Constructor에서는 입력된 부가 뷰에서 기준 뷰와의 중복되는 영역을 제거하는 가지치기(pruning) 과정을 거친다. 가지치기는 각 뷰의 깊이 정보만을 사용하여 진행되며, 각 뷰에서 중복성이 제거된 화소 위치는 비유효영역(invalid region)이라고 부르고, 중복되지 않아 제거되지 않고 전송되어야 할 화소 위치는 유효영역(valid region)이라고 부른다. 그림 2는 가지치기 과정을 도식화한 것이다.

a) 한국항공대학교 항공전자정보공학부(Korea Aerospace University, School of Electronics and Information Engineering)

‡ Corresponding Author : 김재곤(Jae-Gon Kim)
E-mail: jgkim@kau.ac.kr
Tel: +82-2-300-0414

ORCID: <https://orcid.org/0000-0003-3686-4786>

※ This work was supported by IITP grant funded by Korea Government (MSIT) (No. 2017-0-00486).

· Manuscript received July 10, 2020; Revised August 5, 2020; Accepted August 5, 2020.

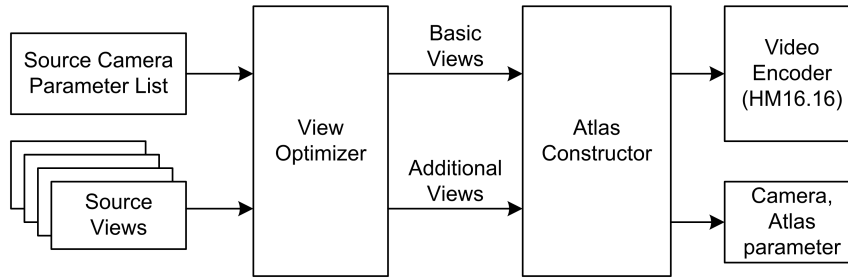


그림 1. TMIV 부호화기 구성도
Fig. 1. Architecture of TMIV encoder

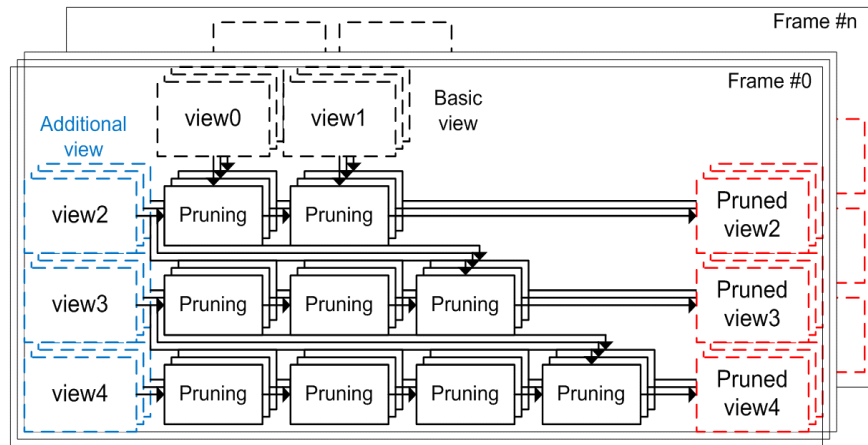


그림 2. Atlas Constructor의 가지치기 과정
Fig. 2. Pruning process in the atlas constructor

가지치기 이후, 각 뷰의 유효 영역들은 클러스터링(clustering) 과정을 통해 다수의 분리된 영역으로 구성되고, 이들 각 영역에 직사각형 형태의 틀을 씌운 것을 패치(patch)라고 한다. 그림 3은 임의의 뷰에서 가지치기 된 결과와 패치

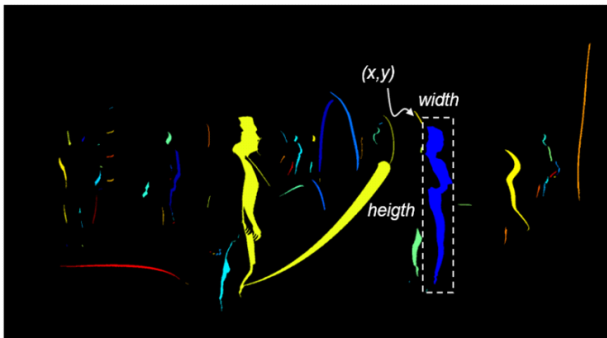


그림 3. 임의의 뷰에서의 가지치기 결과 생성된 패치의 예
Fig. 3. An example of the generated patches as the results of pruning

가 생성된 예시를 보여준다. 동일한 색으로 표현된 각 영역은 하나로 클러스터링 된 영역을 나타내고 이 영역들은 각각 직사각형 형태의 패치로 생성된다. 그림 3의 예시와 같이 직사각형으로 구성된 패치는 내부에 유효영역 뿐만 아니라 비유효영역까지 같이 포함하고 있음을 알 수 있다.

위 과정을 통해 각 뷰마다 생성된 패치들을 효율적으로 전송하기 위해, 생성된 모든 패치들을 아틀라스라고 하는 하나의 프레임에 패킹(packing)한다. 각 패치들은 아틀라스 내의 점유 공간을 최소화하기 위해 서로의 유효영역을 침범하지 않는 범위에서 겹쳐지도록 패킹된다. 패킹된 패치 내의 유효영역에는 해당 패치가 원래 위치했던 뷰에서 대응되는 부분의 텍스처 및 깊이가 채워지며, 비유효영역과 패치가 패킹되지 않은 부분은 중간값이나 0값이 채워지면서 아틀라스 생성이 마무리된다.

III. 제안 기법

앞서 설명한 것처럼 TMIV 4.0 버전에서 아틀라스에 패킹된 각 직사각형 모양의 패치들은 안에 유효영역과 비유효영역을 모두 포함하고 있다. 그러나 전송되는 정보를 최소화하는 측면에서 뷰포트 렌더링에 필수적이지 않은 비유효

영역은 텍스처 영상에서는 중간값으로 채워지고 깊이 영상에서는 0값으로 채워진다. 유효영역은 8x8화소 단위의 격자에 맞춰 남게 되는데 그림 4는 패치의 유효영역이 어떤 식으로 아틀라스에 남게 되는지를 보여준다.

이렇게 하면 아틀라스 부호화에 의해 발생하는 비트율을 크게 감소시킬 수 있다. 하지만 패치 내의 유효 및 비유효영역의 분포가 프레임마다 변동되기 때문에 작은 크기의 유효

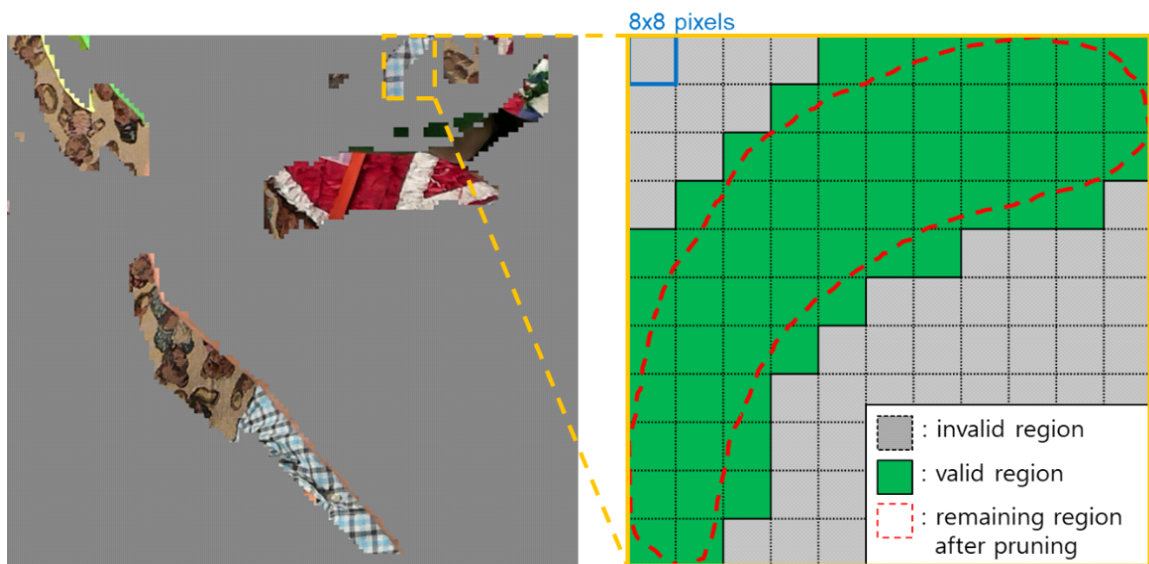


그림 4. 8x8화소 단위의 격자에 맞춰 형성되는 유효영역의 예
Fig. 4. An example of valid region in the form of 8x8 grids

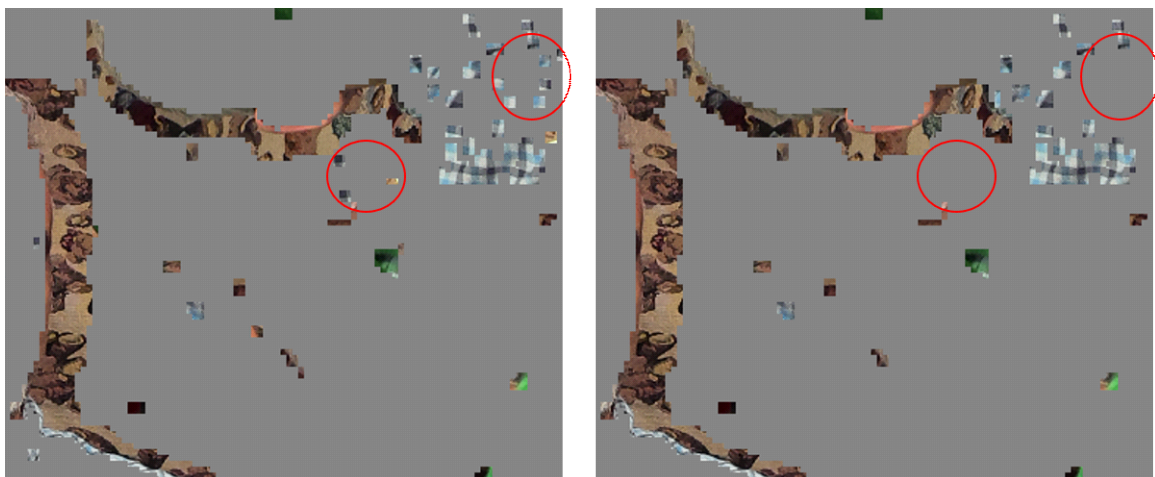


그림 5. 제안기법을 적용한 아틀라스 예(좌: TMIV Anchor, 우: 제안기법)
Fig. 5. A generated atlas with the proposed method (left: TMIV Anchor, right: Proposed method)

효영역에 의해 생기는 블록들이 불규칙적으로 나타났다 사라지는 현상이 생긴다. 이러한 현상은 아틀라스 부호화를 수행할 때, 프레임 간의 유사성을 떨어뜨리며 화면간 예측을 어렵게 해서 부호화 효율을 감소시킬 수 있다.

제안기법은 이런 문제를 해결하기 위해 특정 값보다 크기가 작은 블록들을 아틀라스에서 제거한다. 작은 크기라도 유효영역은 뷰포트 합성에 필요한 영역이므로 제거되면 최종 주관적 화질에 영향을 줄 수 있다. 따라서 제거될 영역의 크기의 적절한 기준을 정하는 것이 중요하다. TMIV에는 패치의 최소 크기를 정해주는 매개변수 ‘MinPatchSize’의 기본값이 16으로 정해져 있다. 즉, 가로 및 세로의 크기가 16화소 이하인 패치는 아틀라스에 패킹되지 않고 제거되며, 뷰포트 렌더링 시에 사용되지 않는다. 이것은 TMIV에서 16x16 화소 이하의 크기로 남은 유효영역은 제거되더라도 렌더링 화질에 큰 영향을 주지 않는다는 것을 보여준다. 따라서, 본 제안 기법에서 제거될 블록의 크기도 ‘MinPatchSize’에 맞추어 16x16 화소로 정했다.

제안기법을 통해 아틀라스 내의 작은 블록들을 제거하면 렌더링 영상의 화질이 약간 저하될 수 있지만 아틀라스 부호화 효율이 향상된다. 그림 5는 TMIV를 통해 생성된 아틀라스의 일부를 예시로 보여준다. 그림 5의 왼쪽은 기존 TMIV 기법으로 생성된 아틀라스이며, 오른쪽은 제안기법이 적용된 아틀라스 생성 결과이다. 그림 5에 붉은색 원으로 표시된 부분을 보면 제안방식을 적용하여 크기가 작은 블록들이 아틀라스 상에서 제거된 것을 볼 수 있다.

IV. 실험결과

제안기법은 TMIV4.0^[4]에 구현되었으며, MPEG-I의 물입형 비디오 공통 테스트 조건(CTC: Common Test Condition)^[5]에 따라 실험하였다. CTC에 따라 3DoF+ 비디오의 아틀라스를 부/복호화하기 위한 핵심 코덱은 HEVC의 시험 모델인 HM16.16^[6]을 사용하였다. CTC에서는 3DoF+ 비디오 특성을 고려한 화질평가를 위하여 360 비디오의 화질평가와 동일하게 WS-PSNR(Weighted Sphere PSNR)을 사용하여 객관적 화질을 측정한다. 또한, 인간의 시각 인식 특성을 고려하여 화질을 평가하는 VMAF(Video Multi-method Assessment Fusion)와 같은 방식도 하여 압축 성능을 측정한다.

표 1은 제안기법을 적용했을 때, 아틀라스만의 부호화에서 비트율이 얼마나 감소하는지를 보여준다. 표 1의 결과로부터 제안한 방법을 통해 평균적으로 1.24%의 비트율을 얻을 수 있음을 확인하였다. 표 2는 제안기법의 단대단(end-to-end) 부호화 실험결과를 BD-rate(Bjontegaard delta bit rate)^[7] 측면에서 나타낸 것이다. 단대단 부호화 성능은 입력 비디오 뷰와 이를 압축 복원한 뷰 비디오를 이용하여 합성한 뷰포트 비디오의 화질을 비교한 것으로 아틀라스의 부호화 성능과 합성 성능이 모두 포함된 결과이다. 또한, BD-rate 값은 동일한 PSNR 화질을 기준으로 Anchor 대비 평균적인 비트율을 나타내므로 음의 값은 Anchor 보다 비트를 절감하여 부호화 성능이 개선되었음을 나타낸

표 1. 제안기법을 적용한 아틀라스 부호화의 비트율 변화(Anchor: TMIV4.0)

Table 1. The bitrate changes in the atlas coding with the proposed method (Anchor: TMIV4.0)

	Bitrate (%)					
	Computer Graphic (CG)			Natural Content (NC)		
	Classroom Video	Technicolor Museum	Orange Kitchen	Technicolor Painter	IntelFrog	Poznan Fencing
QP1	-0.5%	-0.2%	-1.0%	-1.4%	-0.6%	-1.5%
QP2	-1.0%	-0.2%	-1.0%	-1.5%	-0.7%	-1.7%
QP3	-1.7%	-0.3%	-1.2%	-1.7%	-0.9%	-1.9%
QP4	-2.1%	-0.4%	-1.3%	-1.8%	-1.0%	-2.1%
QP5	-2.4%	-0.6%	-1.3%	-2.0%	-1.0%	-2.2%
Average	-1.54%	-0.34%	-1.16%	-1.68%	-0.84%	-1.88%
	-1.24%					

표 2. 제안기법을 적용한 3DoF+ 비디오 단대단 압축 실험결과(Anchor: TMIV4.0)

Table 2. Experimental results on the compression of 3DoF+ videos with the proposed method (Anchor: TMIV4.0)

Test class	Sequences (97 frames)	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
Computer Graphic (CG)	ClassroomVideo	-0.9%	-1.8%	-1.4%	-2.0%	-1.3%	-1.8%
	TechnicolorMuseum	0.0%	-0.1%	-0.1%	-0.4%	-0.2%	-0.3%
	OrangeKitchen	-1.2%	-1.4%	-1.4%	-1.4%	-1.2%	-1.3%
	Average	-0.7%	-1.1%	-1.0%	-1.3%	-0.9%	-1.1%
Natural Content (NC)	TechnicolorPainter	-1.5%	-1.7%	0.0%	0.0%	0.0%	0.0%
	IntelFrog	-0.6%	-0.9%	-0.8%	-1.0%	-0.6%	-0.8%
	PoznanFencing	-0.8%	-1.6%	-0.7%	-1.6%	-0.9%	-1.5%
	Average	-1.0%	-1.4%	-0.5%	-0.9%	-0.5%	-0.8%

다. 표 2의 결과에서 모든 시퀀스에서 부호화 이득이 발생 하는 것을 확인할 수 있다. 앞서 기술했듯이 제안기법은 유효영역을 포함하는 아틀라스 내의 작은 블록들을 제거하면 렌더링 된 영상의 화질을 저하시킬 수 있다는 단점이 있다. 하지만 렌더링 화질에 크게 영향을 주지 않도록 블록이 제거되는 크기의 기준을 적절히 설정해 주었고, 또한 제안기법을 통해 비트율을 크게 감소시킴으로써 BD-rate 측면에서 성능 향상이 있었다.

그림 6은 TMIV를 통해 렌더링한 뷰포트를 예시로 보여 준다. 그림 6의 왼쪽은 기존 TMIV로 실험한 결과이며, 오른쪽은 제안기법이 적용된 TMIV로 실험한 결과이다. 양측의 결과에 육안으로 확인 가능한 차이가 없음을 확인할 수 있다. 이를 통해 제안기법은 렌더링 되는 뷰포트 영상의 화질에 큰 영향을 주지 않는다는 것을 알 수 있다.

V. 결 론

본 논문에서는 3DoF+ 비디오의 부호화 효율을 향상시키기 위한 아틀라스 생성 기법으로 아틀라스 내의 작은 블록 제거 방법을 제안하였다. 실험결과 제안기법은 아틀라스 자체의 부호화 성능을 향상시키고, 입력 3DoF+ 비디오에서 렌더링된 3DoF+ 비디오의 단대단 부호화 성능에서도 BD-rate을 감소시키는 부호화 성능 개선을 보였다. 제안한 아틀라스 생성 방법은 움직임 시차를 지원하는 전방위 비디오인 3DoF+ 비디오의 부호화 효율 향상을 위한 유용한 전처리 기법으로 활용될 수 있을 것으로 기대되며, 추가 성능개선을 위한 확장기법의 연구를 진행할 예정이다.

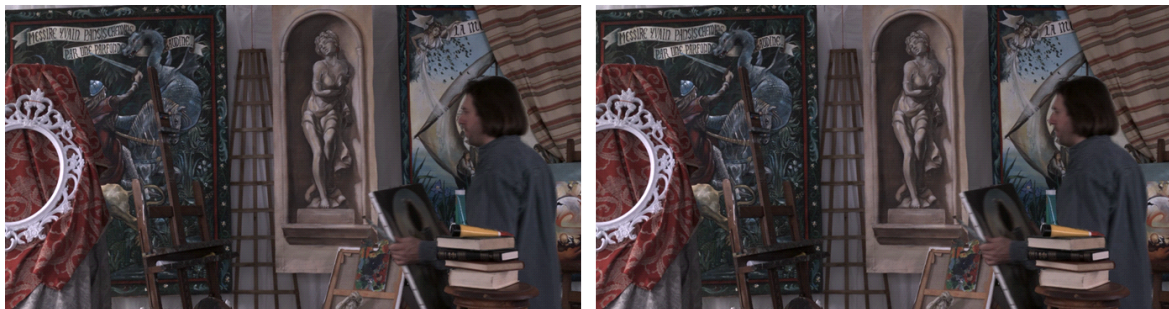


그림 6. 렌더링된 뷰포트 예(좌: TMIV Anchor, 우: 제안기법)

Fig. 6. An example of the rendered viewport (left: TMIV Anchor, right: Proposed method)

참 고 문 헌 (References)

- [1] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," ISO/IEC JTC1/SC29/WG11, N16768, Apr. 2017.
- [2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, "Standardization Status of Immersive Video Coding," IEEE Jour. Emerg. Select. Topics Circuits Syst., vol. 9, no. 1, Mar. 2019.
- [3] J. Boyce, R. Dore, V. Vadakital, "Working Draft 4 of Immersive Video," ISO/IEC JTC1/SC29/WG11, N19001, Jan. 2020.
- [4] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 4 for Immersive Video," ISO/IEC JTC1/SC29/WG11, N19002, Jan. 2020.
- [5] J. Jung, B. Kroon, J. Boyce, "Common Test Conditions for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18997, Jan. 2020.
- [6] HM reference software, [Online]. Available at http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware
- [7] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, "On BD-rate Calculation," JCTVC-F270, Jul. 2011.
- [8] S. Lim, H. Kim, J. Kim, G. Lee, J. Jeong, J. Seo, "CE3-Related: Patch Trimming Using Hole Filling and Tiny Region Removal in Atlas," ISO/IEC JTC1/SC29/WG11, m53834, Apr. 2020.

저 자 소 개



임 성 균

- 2020년 2월 : 한국항공대학교 항공전자 및 정보통신공학부 학사
- 2020년 3월 ~ 현재 : 한국항공대학교 항공전자정보공학과 석사과정
- ORCID : <https://orcid.org/0000-0001-6581-5268>
- 주관심분야 : 영상처리, 비디오 코덱, 360 비디오/VR



김 현 호

- 2018년 2월 : 한국항공대학교 항공전자정보공학과 학사
- 2020년 2월 : 한국항공대학교 항공전자정보공학과 석사
- 2020년 3월 ~ 8월 : 한국항공대학교 항공전자연구소 연구원
- ORCID : <https://orcid.org/0000-0002-8645-6142>
- 주관심분야 : 영상처리, 비디오 코덱, 360 비디오/VR



김 재 곤

- 1990년 2월 : 경북대학교 전자공학과 학사
- 1992년 2월 : KAIST 전기 및 전자공학과 석사
- 2005년 2월 : KAIST 전기 및 전자공학과 박사
- 1992년 3월 ~ 2007년 2월 : 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 9월 ~ 2002년 11월 : Columbia University 연구원
- 2007년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- 2014년 12월 ~ 2016년 1월 : UC San Diego 방문교수
- ORCID : <https://orcid.org/0000-0003-3686-4786>
- 주관심분야 : 비디오 신호처리, 비디오 부호화 표준, Immersive Video, Deep Learning