

몰입형 비디오 부호화의 아틀라스 생성을 위한 홀 채움 기법

임성균, *이광순, 김재곤
 한국항공대학교, *한국전자통신연구원
 sglim@kau.kr, *gslee@etri.re.kr, jgkim@kau.ac.kr

A Method of Hole Filling for Atlas Generation in Immersive Video Coding

Sung-Gyun Lim, *Gwangsoon Lee, and Jae-Gon Kim
 Korea Aerospace University, *ETRI

요 약

MPEG 비디오 그룹은 제한된 3D 공간 내에서 움직임 시차(motion parallax)를 제공하면서 원하는 시점(view)을 렌더링(rendering)하기 위한 표준으로 TMIV(Test Model for Immersive Video)라는 테스트 모델과 함께 효율적인 몰입형 비디오의 부호화를 위한 MIV(MPEG Immersive Video) 표준을 개발하고 있다. 몰입감 있는 시각적 경험을 제공하기 위해서는 많은 수의 시점 비디오가 필요하기 때문에 방대한 양의 비디오를 고효율로 압축하는 것이 불가피하다. TMIV는 여러 개의 입력 시점 비디오를 소수의 아틀라스(atlas) 비디오로 변환하여 부호화되는 화소수를 줄이게 된다. 아틀라스는 선택된 소수의 기본 시점(basic view) 비디오와 기본 시점으로부터 합성할 수 없는 나머지 추가 시점(additional view) 비디오의 영역들을 패치(patch)로 만들어 패킹(packing)한 비디오이다. 본 논문에서는 아틀라스 비디오의 보다 효율적인 부호화를 위해서 패치 내에 생기는 작은 홀(hole)들을 채우는 기법을 제안한다. 제안기법은 기존 TMIV8.0에 비해 1.2%의 BD-rate이 향상된 성능을 보인다.

1. 서론

최근 시청 자유도를 증가시켜 몰입감 있는 시각적 경험을 제공하는 몰입형(immersive) 미디어가 주목을 받고 있으며, MPEG에서는 몰입형 미디어를 압축, 전송하기 위한 표준을 개발 중이다[1], [2]. 몰입형 비디오는 사용자가 고정된 위치에서 전방향을 둘러볼 수 있도록 자유도를 제공하는 전방위 360도 비디오에서 추가적으로 사용자의 움직임에 따른 시점의 변화까지 지원하는 6DoF(Degree of Freedom) 비디오를 포함하고 있다. 이러한 움직임에 의한 시차(motion parallax)를 제공하기 위해서는 다시점 비디오와 함께 그 깊이 정보를 제공하고 이를 기반으로 원하는 위치의 가상 시점을 합성하는 렌더링(rendering) 기법이 사용된다. 따라서, 몰입형 비디오는 제한된 공간의 여러 위치에서 동시에 획득된 텍스처(texture)와 깊이(depth) 비디오들로 구성된다. 이에 따라, 몰입형 비디오는 매우 큰 용량을 가질 수밖에 없기 때문에 고효율 압축이 필수적으로 요구된다. 본 논문에서는 몰입형 비디오의 렌더링 화질을 높이면서 아틀라스 부호화 효율을 높이기 위한 아틀라스 생성 방법을 제안한다.

2. TMIV

MPEG 비디오 그룹에서는 대용량의 몰입형 비디오를 효율적으로 압축 및 전송하기 위한 표준 기술이 개발되고

있으며, 최근 DIS(Draft IS)[3]와 TMIV8.0 버전[4]을 발간했다. TMIV(Test Model for Immersive Video)의 몰입형 비디오 압축을 위한 주요 개념은 입력 시점 비디오들을 부호화 하기 전의 전처리를 통해 높은 상관관계를 가지는 뷰들 간의 중복성을 제거하여 압축할 데이터의 양을 최대한 줄이는 것이다.

그림 1은 TMIV 부호화기[4]의 전체적인 구조를 보여준다. TMIV 부호화기에서는 입력된 서로 다른 시점의 비디오들을 몇 개의 그룹으로 나누고 각 그룹 내에서 다시 기본 시점과 추가 시점으로 분류한다. 각 그룹별로는 기본 시점과의 중복성이 있는 추가 시점의 영역들을 제거하는 프루닝(pruning) 과정이 수행된다. 프루닝 후에 남은 영역들은 직사각형 형태의 틀에 맞춰서 각각 개별적인 패치로 생성된다. 하나의 패치에는 전송되어야 할 잔여 픽셀들을 포함하는 유효 영역(valid region)과 프루닝에 의해 제거된 부분인 비유효 영역(invalid region)이 모두 포함된다.

추가 시점에서 생성된 패치들은 기본 시점 영상과 함께 소수의 아틀라스에 패킹 된다. 패킹되는 패치들은 유효 영역이 서로 겹치지 않도록 하되 패치들이 차지하는 공간이 최소가 되도록 조밀하게 패킹된다. 패킹 공간을 최대한으로 줄이기 위해서 패치를 회전하여 패킹할 수도 있으며, 각 패치의 패킹 위치 및 회전 정보들은 메타데이터(matadata)로 구성되어 아틀라스 부호화 비트스트림(bitstream)과 함께 복호화기로 전송된다. 이런 메타데이터 정보는 TMIV 복호화기에서 아틀라스에 무작위로 패킹된 패치들이 추가 시점들의 원래 위치로 복원될 수 있도록 한다.

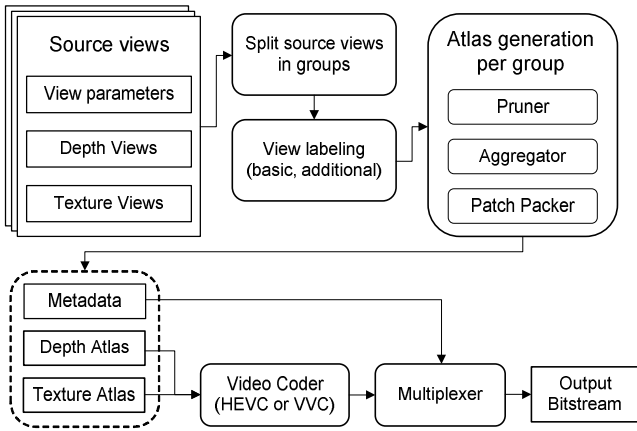


그림 1. TMIV 부호화기의 전체 구조도[4]

3. 제안기법

앞 절에서 설명한대로 각 직사각형 패치에는 유효 영역과 비유효 영역이 모두 포함되며, 패치의 직사각형 경계내 영역들의 유효 여부는 16x16 또는 32x32 크기의 블록 단위로 결정된다. 그리고 유효 픽셀에는 입력 시점 영상의 상응하는 위치로부터 가져온 픽셀값이 포함되지만 비유효 픽셀에는 중간값이나 0 값이 들어간다. 따라서, 한 패치 내의 유효 영역과 비유효 영역의 분포에 따라 블록 모양의 작은 홀(hole)이나 균열이 발생할 수 있다. 이런 현상은 아틀라스의 불연속성을 증가시키며 이로 인하여 부호화 효율을 떨어뜨리는 원인이 된다.

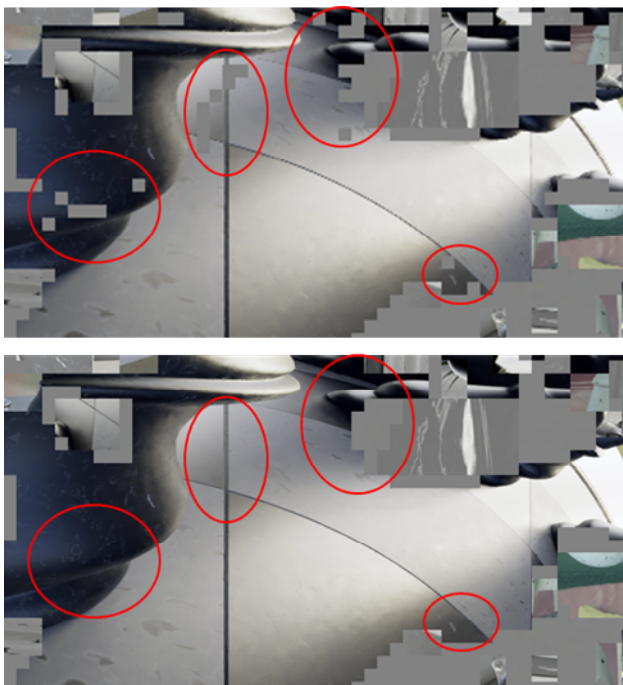


그림 2. 제안 방법을 적용하여 생성한 아틀라스의 일부 예 (위: 기존 TMIV, 아래: 제안기법)

이 문제를 해결하기 위해서 유효 영역에 팽창(dilation)과 침식(erosion)을 적용하여 패치 내부에 발생한 구멍이나 균열을 채우는 홀 채움 기법을 제안한다. 즉, 유효 영역을 일정 범위만큼 팽창시켰다가 동일한 정도로 침식시켜서 작은 크기의 홀들을 제거한다. 하지만, 상관성이 낮은 서로 다른 패치들 사이에 위치한 홀들은 제안기법에 의해 제거되지 않도록 한다. 그림 2 는 제안기법을 사용하여 패치 안의 홀들을 채우는 예를 보여준다[5]. 그림에서 붉은색 원으로 표시한 부분을 보면 제안기법을 적용함으로써 패치 안에 생긴 작은 크기의 홀들이 제거된 것을 확인할 수 있다. 결과적으로 아틀라스 영상의 불연속성이 감소하여 부호화 효율이 향상될 수 있고, 전송되는 유효 영역이 증가함으로써 뷰포트(viewport) 합성의 정확도가 높아져서 화질이 개선될 수 있다.

4. 실험결과

제안기법의 성능을 평가하기 위해서 입력 시점 비디오와 동일한 시점으로 합성된 렌더링 비디오를 비교하여 화질을 측정했다. MIV 공통 테스트 조건(Common Test Condition, CTC)[6]에 따라 화질 측정에는 WS-PSNR[7]과 IV-PSNR[8]을 사용했다. 표 1 은 BD-rate 측면에서 제안기법을 사용한 몰입형 비디오 부호화 성능을 기존 TMIV 의 성능과 비교한 결과이다. 제안기법은 TMIV8.0.1 에 구현하였으며, VVenC[9]를 사용하여 CTC 에 따라 생성된 아틀라스를 압축했다. 기존 TMIV 방식에 비해서 의미 있는 성능 향상이 있는 것을 확인할 수 있으며, 휘도 성분 WS-PSNR 에 대하여 평균 1.2%의 이득을 보여준다.

표 1. 제안기법을 적용한 몰입형 비디오 부호화 성능 실험결과 (Anchor: TMIV8.0)

Sequences (resolution)	WS-PSNR (Y)	IV-PSNR
SA (4096x4096)	-0.4%	-0.5%
SB (2048x2048)	0.4%	0.4%
SO (1920x1080)	-1.6%	-1.0%
SJ (1920x1080)	-1.4%	-0.9%
SD (2048x1088)	-0.1%	-0.1%
SE (1920x1080)	0.1%	0.1%
SP (1920x1080)	-0.4%	-0.3%
SN (2048x2048)	-5.2%	-1.8%
SR (1920x1080)	-2.0%	-0.7%
Average	-1.2%	-0.5%

그림 3 은 ‘Group’ 시퀀스를 사용하여 기존 TMIV8.0 과 제안기법의 주관적 화질을 비교한 것이다. 그림 4 와 같이 TMIV 로 생성된 아틀라스로 렌더링한 시점 영상에서 보이는 시각적 아티팩트(visual artifact)가 제안기법을 적용한 결과에서는 제거된 것을 확인할 수 있다. 이를 통해 제안기법은 표 1 에서 확인한 객관적 화질 뿐만 아니라 주관적 화질 측면에서도 눈에 띄는 개선을 보이는 것을 알 수 있다.



그림 3. 'Group' 시퀀스에서 렌더링 된 시점 영상의 예
(좌: 기존 TMIV, 우: 제안기법)

- [6] J. Jung, B. Kroon, "Common Test Conditions for MPEG Immersive Video," ISO/IEC JTC1/SC29/WG04, N00051, Jan. 2021.
- [7] "WS-PSNR Software Manual," ISO/IEC JTC1/SC29/WG11, N18069, Oct. 2019.
- [8] A. Dziembowski, "Software manual of IV-PSNR for Immersive Video," ISO/IEC JTC1/SC29/WG04, N00013, Oct. 2020.
- [9] VVenC software, [Online]. Available at: <https://github.com/fraunhoferhhi/vvenc/tree/v0.2.0.0>

Acknowledgement

이 논문은 2018 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00207, 이머시브 미디어 전문연구실).

5. 결론

본 논문에서는 몰입형 비디오 부호화 효율 개선을 위하여 아틀라스 생성에서의 홀 채움 기법을 제시하였다. 본 논문의 홀 채움 기법은 아틀라스의 불연속성을 줄임으로써 부호화 효율을 높이고 시점 합성에 필요한 유효 픽셀을 추가적으로 전송함으로써 렌더링 화질도 개선시키는 효과를 보여준다. 실험결과에서 제안기법으로 휘도성분의 WS-PSNR 에서 BD-rate 1.2%의 부호화 효율 개선과 두드러진 주관적인 화질 개선을 얻을 수 있음을 확인할 수 있었다.

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 J. Emerg. Select. Topics Circuits Syst., vol. 9, no. 1, pp. 5-17, Mar. 2019.
- [3] "Text of ISO/IEC DIS 23090-12 MPEG Immersive Video," ISO/IEC JTC1/SC29/WG04, N00049, Jan. 2021.
- [4] B. Salahieh, J. Jung, A. Dziembowski, and C. Bachhuber (Eds.), "Test Model 8 for Immersive Video," ISO/IEC JTC1/SC29/WG04, N00050, Jan. 2021.
- [5] S. Lim, H. Kim, and J. Kim, "[MIV] CE3.1 Patch Trimming," ISO/IEC JTC1/SC29/WG04, m56002, Jan. 2021.