

3DoF+ 비디오 부호화를 위한 깊이 매핑 기법

박지훈, 이준성, 박도현, 김재곤

한국항공대학교

{pss6861, juns2014, dhpark}@kau.kr, jgkim@kau.ac.kr

A Depth Mapping Method for 3DoF+ Video Coding

Ji-Hun Park, Jun-Sung Lee, Dohyeon Park, and Jae-Gon Kim
Korea Aerospace University

요 약

3DoF+ 비디오 부호화 표준을 개발하고 있는 MPEG-I 비주얼 그룹은 표준화 과정에서 참조 SW 코덱인 TMIV(Test Model for Immersive Video)를 개발하고 있다. TMIV 는 제한된 공간에서 동시에 여러 위치에서 획득한 뷰(view)의 텍스처(texture) 비디오와 깊이(depth) 비디오를 효율적으로 압축하여 임의 시점의 뷰 렌더링(rendering)을 제공한다. TMIV 에서 수행되는 깊이 비디오의 비트 심도 스케일링 및 압축은 깊이 정보의 손실을 발생하며 이는 렌더링(rendering)된 임의 시점 비디오의 화질 저하를 야기한다. 본 논문에서는 보다 효율적인 깊이 비디오 압축을 위한 히스토그램 등화(histogram equalization) 기반의 구간별(piece-wise) 깊이 매핑 기법을 제안한다. 실험결과 제안기법은 자연 영상(natural sequence)의 End-to-End 부호화 성능에서 평균적으로 3.1%의 비트율 절감이 있음을 확인하였다.

1. 서론

VR(Virtual Reality)/360 비디오 서비스가 확산되고 더 나아가 사용자에게 6DoF(Degree of Freedom)까지의 자유도를 제공할 수 있는 몰입형(immersive) 미디어가 부각되고 있다. 이에 따라 방대한 양의 새로운 형태의 몰입형 비디오를 효율적으로 제공하기 위한 몰입형 비디오 압축 표준이 요구되었다. 비디오 압축 전문가 그룹인 MPEG(Moving Picture Experts Group)에서는 3DoF 보다 확장된 3DoF+ 비디오와 최종적으로 6DoF 를 제공하는 몰입형 미디어를 압축하기 위한 표준화를 진행하고 있다. 현재 MPEG-I 비주얼 그룹에서는 3DoF+ 비디오를 효율적으로 압축하기 위한 표준화를 진행하고 있으며, 표준화 과정에서 표준 기술을 검증하기 위한 참조 SW 코덱인 TMIV(Test Model for Immersive Video)를 개발, 배포하고 있다[1].

몰입형 미디어는 자유도에 따라서 3DoF, 3DoF+, 6DoF 의 단계로 나눌 수 있다. 3DoF 는 3D 의 자유도를 가지는 것으로, 사용자가 고정된 자리에서 x, y, z 축 방향으로 머리를 회전하면서 전방위 뷰를 소비할 수 있는 상태를 말한다. 3DoF+는 3DoF 에서 조금 더 나아가서 고정된 위치에서 상체의 제한적인 움직임을 허용한 것이다. 6DoF 는 사용자의 모든 시점 및 움직임이 허용되는 자유도이다. 3DoF+를 제공하는 비디오의 경우, 제한된 범위 내 여러 시점에서 획득된 뷰(view) 비디오 및 해당 뷰의 깊이(depth) 비디오로 구성된다. TMIV 는 복수의 입력 비디오의 중복성을 제거함으로써 압축 성능을 향상하며 압축된 깊이 비디오를 이용하여 제한된 범위 내 임의의 시점에서 뷰 비디오를 렌더링(rendering) 한다. 본 논문에서는 TMIV 및 비디오 코덱에서 깊이 뷰 비디오를 효율적으로 압축하여 가상 시점의 뷰 렌더링의 성능을 향상하기 위한 깊이 비디오 매핑 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 TMIV 에 대하여 살펴본 후, 3 절에서는 본 논문에서 제안한 깊이 매핑 기법을 제안한다. 4 절에서는 제안 기법의 성능을 실험을 통해서 확인한다. 마지막으로 5 절에서는 본 논문에 대한 결론을 맺는다.

2. 3DoF+ 비디오 압축: TMIV

그림 1 은 3DoF+ 비디오 압축/복원하기 위한 TMIV 의 전반적인 구성도이다. 입력으로 다양한 시점의 텍스처(texture) 및 깊이 비디오가 입력된다. 입력 비디오는 전처리 과정을 거쳐서 기준이 되는 기준뷰(basic view)와 나머지 각 뷰에 대해서 이 기준 뷰와 중복되지 않는 부분의 영상 조각 패치(patch)를 모아둔 아틀라스(atlas)를 생성한다. 생성된 아틀라스는 HEVC(High-Efficiency Video Coding) 코덱을 통해 부호화 된다. 3DoF+ 비디오의 부호화를 위해서 아틀라스 비트스트림 뿐만 아니라 멀티뷰 복원에 필요한 메타데이터(metadata) 정보도 함께 전송된다. TMIV 디코더는 비디오 코덱을 통해서 복원된 아틀라스와 전송된 메타데이터를 이용하여 입력된 멀티뷰의 텍스처 및 깊이 비디오를 복원하며, 복원된 다양한 시점의 깊이 비디오를 이용하여 임의의 시점 및 방향에 대한 뷰를 렌더링한다.

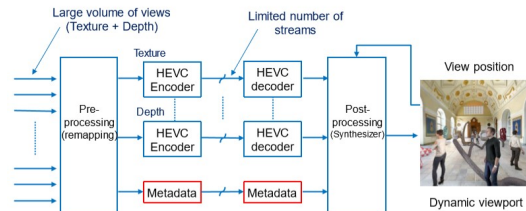


그림 1. TMIV 의 3DoF+ 비디오 압축/복원 구성도

3. 제안하는 깊이 매핑 기법

TMIV 인코더에서는 텍스처 비디오 및 깊이 비디오에 대해서 각각 기준 뷰와 패치를 포함하는 아틀라스를 생성한다. 이때 깊이 영상은 16 비트 심도에서 10 비트 비디오 부호화를 위해 10 비트의 심도로 변환된다. 이 과정에서 깊이 영상의 손실이

발생하며, 생성된 아틀라스를 HEVC 로 부호화하는 과정에서 추가적인 손실이 발생한다. 본 논문에서는 표준 마무리 단계인 차세대 비디오 부호화 표준 VVC(Versatile Video Coding)의 LMCS(Luma Mapping with Chroma Scaling)[2] 기술에 기반하여 TMIV 부/복호화 과정에서 발생하는 깊이 영상의 손실을 줄여 깊이 영상의 화질을 최대한 유지하기 위한 깊이 매핑 기법을 제안한다.

LMCS 는 휘도신호의 동적 범위에 대한 코드워드(codeword)를 휘도신호 값의 빈도수와 지역적 분산 값을 기반한 구간 선형모델을 통해 재분배한다. 제안하는 깊이 매핑 기법은 LMCS 의 구간별 선형모델 매핑 기법을 참고하여 깊이 영상을 효율적인 압축 및 보존이 가능한 동적 범위로 변환한다. TMIV 의 아틀라스 생성 및 HEVC 비디오 부호화는 Intra-Period 단위로 수행되기 때문에 깊이 매핑 또한 Intra-Period 단위로 수행되며 깊이 아틀라스의 구간별 누적 빈도수에 대한 히스토그램 등화(equalization)를 통해 변환될 깊이 아틀라스의 구간별 코드워드를 유도한다. 또한, 매핑된 코드워드 구간의 최소 코드워드의 개수를 유지하여 매핑으로 인한 원본 깊이 영상의 손실을 줄였으며, 최소 코드워드의 개수는 구간별로 다르게 정의한다. 일반적으로 영상 렌더링에 있어 렌즈와 가까운 객체에 대한 깊이 정보의 중요도가 크기 때문에 최소 코드워드를 점진적으로 시차(disparity)가 큰 구간에 많이 할당한다. 최소 코드워드 할당 후 나머지 영역에 앞서 구한 구간별 빈도수에 따라 구간별로 추가 할당을 하여 최종적인 매핑된 깊이 영역의 코드워드를 정의한다.

$$v_{map} = (b_2 - b_1) / (a_2 - a_1) \times (v_{org} - a_1) + b_1 \quad (1)$$

$$v_{org} = (a_2 - a_1) / (b_2 - b_1) \times (v_{map} - b_1) + a_1 \quad (2)$$

매핑된 코드워드를 구한 후, 식 (1)을 통해 깊이 영상을 매핑한다. 여기에서 v_{map} 및 v_{org} 는 매핑이 완료된 깊이 값과 매핑 전 원본 깊이의 값을 의미한다. a_1 과 a_2 는 입력 깊이 값(v_{org})을 포함하는 원본 구간의 첫번째와 마지막 값을 의미하고 b_1 , b_2 는 해당 매핑 전 구간에 대한 매핑 후 구간의 첫번째와 마지막 값을 의미한다. 또한 디코더에서 임의의 시점에 대한 렌더링을 수행하기 위해서는 매핑된 복원 영상의 역매핑이 필수적이다. 역매핑은 식 (2)를 통해 수행한다.

$$BinCW = OrgCW + DeltaCW \quad (3)$$

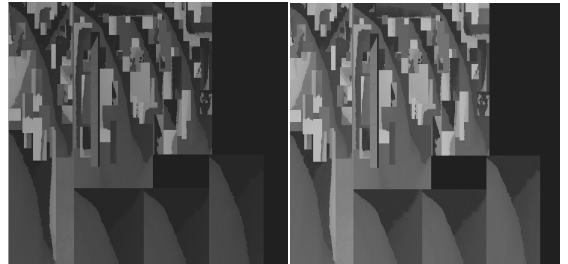
$$DeltaCW = (1 - 2 \times delta_sign_cw_flag) \times delta_abs_cw \quad (4)$$

디코더에서 깊이 역매핑을 수행하기 위해서는 b_1 , b_2 값이 필요하므로 인코더에서 구한 재분배된 코드워드에 대한 정보를 메타데이터로 시그널링 한다. 시그널링 기법은 LMCS 에서 휘도 매핑 구간을 전송하는 선택스(syntax) 구조를 참고하였으며, 식 (3)의 BinCW 는 재분배 후의 각 구간에 대한 코드워드를 의미하고 최대 15 비트의 크기를 가진다. 이에 대한 비트스트림 오버헤드를 줄이고자 BinCW 를 직접 전송하지 않고 재분배 전에 각 구간에 동일하게 할당된 BinCW 와의 차이 값인 DeltaCW 를 전송한다. 또한, 식 (4)와 같이 DeltaCW 는 부호 정보와 절댓값을 분리하여 전송한다.

4. 실험결과

본 논문의 제안 기법은 TMIV-2.01 에 구현하였다. 아틀라스 부호화를 위한 비디오 코덱은 HM16.16 을 사용하였으며, 실험에

사용된 QP 는 22, 27, 32, 37, 42 로 성능 평가를 진행하였다[3]. 테스트 시퀀스는 그래픽 영상(ClassroomVideo), 자연 영상(IntelFrog)를 사용하였으며 각 시퀀스는 16 비트의 깊이 영상을 포함한다. 실험결과 End-to-End 부호화 성능은 압축된 아틀라스와 메타데이터의 비트율과 렌더링된 합성영상의 화질을 다양한 화질 측정 방법(WS-PSNR, VMAF, MS-SSIM, IV-PSNR)으로 확인하였다. 실험결과, 그래픽 영상에서 0.8%, 자연영상에서 3.1%의 BD-rate(WS-PSNR) 부호화 성능 향상이 있음을 확인하였다.



(a) 매핑 전 깊이 아틀라스 (b) 매핑 후 깊이 아틀라스
그림 2. 매핑 전, 후 깊이 아틀라스 예

표 1. 제안 방식을 적용한 3DoF+ 비디오 압축 실험결과

Sequence	WS-PSNR	VMAF	MS-SSIM	IV-PSNR
ClassroomVideo	-0.8%	-0.3%	2.0%	1.9%
IntelFrog	-3.1%	-5.3%	-6.8%	1.5%

5. 결론

본 논문에서는 3DoF+ 비디오의 효율적인 부호화를 위한 깊이 매핑 기법을 제안하였다. 제안 기법은 3DoF+ 비디오 부호화 표준의 시퀀스모델인 TMIV 에 적용하여 깊이 아틀라스를 매핑함으로써 End-to-End 부호화 성능 향상 및 렌더링 화질을 향상시켰다. 실험결과 제안 기법은 자연 영상에서 3.1%의 비트율을 절감의 부호화 이득이 있음을 확인하였다.

감사의 글

이 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00486)

참고문헌

- [1] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18577, July 2019.
- [2] B. Bross, J. Chen, S. Liu, Y. Wang, "Versatile Video Coding (Draft8)," JVET of ITUT-T SC16 WP3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-Q2001, Jan. 2020.
- [3] J. Jung, B. Kroon, J. Boyce, "Common Test Conditions for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18443, Mar. 2019.