

실감형 미디어를 위한 다시점 360 비디오 압축 시스템 구현

정중범, 이순빈, 장동민, 류일웅, Tuan Thanh Le, 유재성, 류은석
가천대학교

uof4949@gc.gachon.ac.kr, soon0698@gc.gachon.ac.kr, dogzz9445@gc.gachon.ac.kr,
dlfdnd96@gc.gachon.ac.kr, tuanlt@gc.gachon.ac.kr, poopoo96@gc.gachon.ac.kr,
esryu@gachon.ac.kr

Implementing Multi-view 360 Video Compression System for Immersive Media

Jong-Beom Jeong, Soonbin Lee, Dongmin Jang, Il-Woong Ryu, Tuan Thanh Le,
Jaesung Ryu and Eun-Seok Ryu
Gachon University

요 약

본 논문에서는 사용자 시점에 대응하는 고품질 360 비디오 제공을 위해 다시점 360 비디오 중복성 제거 기법을 적용하고 잔여 비디오를 하나의 영상으로 병합하여 압축 후 전송하는 시스템을 구현한다. 사용자 움직임 적응적 360 비디오 스트리밍을 지원하는 three degrees of freedom plus (3DoF+)를 위한 시스템은 다시점에서 촬영된 다수의 고품질 360 비디오 전송을 요구한다. 이에 대한 방안으로 다시점 비디오 간 중복성 제거를 위한 3D warping 을 기반으로 하는 뷰 간 중복성 제거 기술과 비디오 복원에 필요한 타일들만 추출 및 병합해주는 잔여 뷰 병합 기술에 대한 구현 내용을 설명한다. 제안된 시스템을 기반으로 다시점 360 비디오 전송을 수행하면, 기존 high-efficiency video coding (HEVC)을 사용하여 전송했을 때 대비 최대 20.14%의 BD-rate 감소가 가능함을 확인하였다.

1. 서론

최근 virtual reality (VR) 시장이 급격히 증가하고 있다. 그에 따라 머리에 장착 가능한 영상 제공 장치인 head-mounted display (HMD) 와 몰입형 콘텐츠를 취득할 수 있는 360 카메라가 시장에 보급되고 있다. 따라서 몰입형 미디어를 제공하기 위하여 효율적인 360 비디오 스트리밍 기술이 요구되고 있다. 그러나 현재 고품질 가상 현실 비디오를 제공하기 위한 컴퓨터 연산능력과 대역폭은 한계가 존재한다. 예를 들어 가상 현실에서 사용자가 이질감을 느끼지 않을 정도의 몰입형 경험을 제공하려면 ultra high-definition (UHD) 이상의 초고화질 360 영상과 낮은 움직임 지연이 필요하다[1].

이에 moving picture experts group (MPEG) 은 몰입형 미디어를 표준화하기 위해 MPEG-I 서브그룹을 설립하였고, 가상현실 비디오 표준화 단계를 사용자의 자유도에 따라 3 degrees of freedom (DoF), 3DoF+, 그리고 6DoF 로 분류하였고[2], 2021 년까지 표준화가 완료될 예정이다. 현재 상용화된 가상 현실 기술은 대부분 3DoF 를 지원하는데, 이는 사용자에게 제한적인 몰입형 콘텐츠를 제공한다. 따라서 의자에 앉은 사용자의 움직임에 대응하는 영상을 제공하는 3DoF+의 필요성이 대두되었다. 사용자의 머리 움직임에 따른 가상 영상을 제공하기 위해 여러 개의 고품질 360 비디오들이 전송되어야 하는데, 다수의 360 영상을 down-sampling 하여 압축하고 전송하여 복호화 후 up-sampling 하여 비트레이트를

절감하는 연구가 제안되었다[3].

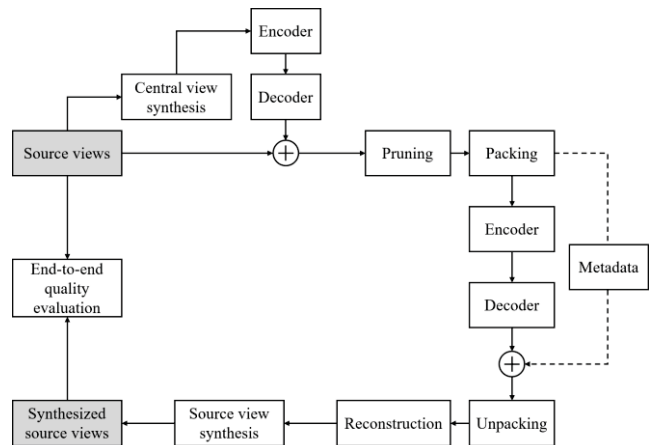


그림 1. 제안하는 다시점 360 비디오 압축 시스템 구조

일반적으로 영상 전송에는 표준화된 advanced video coding (AVC) 또는 high-efficiency video coding (HEVC) 압축기술을 사용하여 영상 압축 및 전송처리를 하고 있다. 해당 기술은 개별 영상에 대한 압축을 전제로 하며, 다수의 UHD 급 영상을 전송하려면 막대한 양의 연산량과 대역폭이 요구되는데 이에 대응하는 시스템이 없다. 이를 해결하기 위해 MPEG-I 는 3DoF+ 시스템에 대응하는 call for proposals (CfP) 를 제안하였다[4]. 요구되는 3DoF+ 시스템은 기존 HEVC 를 사용하여 비디오를 압축 및 전송하되, 비트레이트를 절감하기

위한 pre · post processing 을 허용하며 관련 메타데이터는 MPEG-I part 7 에서 표준화가 진행될 예정이다[5].

본 논문은 3DoF+ CIP 에 대응하는 다시점 360 비디오 영상 전송을 위한 시스템을 제안한다. 다수의 영상 간 중복성을 제거하는 프루닝 기술과 중복성이 제거된 영상들을 하나의 영상으로 병합하는 패킹 기술을 설명한다. 그림 1 은 본 논문이 제시하는 프루닝과 패킹을 이용한 3DoF+ 360 비디오 압축 시스템을 보여준다.

본 논문의 구성은 다음과 같다. 2 절에서는 관련 연구로 뷰 합성 기술과 다시점 비디오 코딩 기술을 설명하고 3 절에서는 제안하는 프루닝과 패킹 기술에서의 이슈와 구현 내용을 설명한다. 4 절에서는 제안하는 시스템의 결과를 설명하고 마지막으로 5 절에서는 본 논문의 결론을 서술한다.

2. 관련 연구

2.1 View Synthesis

몰입형 미디어를 위해서는 사용자의 움직임에 대응하는 영상을 제공할 필요가 있다. 그러나 360 비디오를 취득할 수 있는 카메라의 개수와 위치는 제한되어 있다. 이를 극복하기 위해 3DoF+와 6DoF 는 이미 취득된 영상을 기반으로 사용자의 뷰포트에 해당하는 가상 영상을 제공하는 기술을 포함하는데, 이를 뷰 합성이라 한다. 뷰 합성을 위해서는 텍스처와 맵스 맵이 필요하다. 텍스처는 카메라가 촬영한 물체의 색상 정보를 포함하고, 맵스 맵은 카메라로부터 물체까지의 거리에 대한 정보를 포함한다. 두 정보 이외에도 카메라 메타데이터 등 여러 정보가 뷰 합성 도구에 요구되기에, MPEG-I 에서는 뷰 합성 도구 표준화를 위해 3DoF+에는 reference view synthesizer (RVS) [6]를, 6DoF 에는 versatile view synthesizer (VVS) [7]를 참조 뷰 합성 도구로 채택하였다. 그림 2 는 RVS 의 뷰 합성 구조도를 나타낸다. RVS 는 입력된 비디오들의 맵스 맵을 타겟 뷰포인트에 매핑시키고, 그것을 기반으로 텍스처를 매핑시킨다. 텍스처가 제대로 매핑되지 않은 부분은 입력 비디오에서 표현할 수 없는 영역이므로 이를 제거하고, 제거된 홀 영역을 채우는 작업을 실행한다. 이후 각 입력 비디오에서 도출된 결과 비디오들을 하나로 합치는 블렌딩 작업 후 몇몇 잡음이 포함된 영역을 처리하기 위해 인페인팅을 실행하면 최종 가상 뷰가 출력된다.

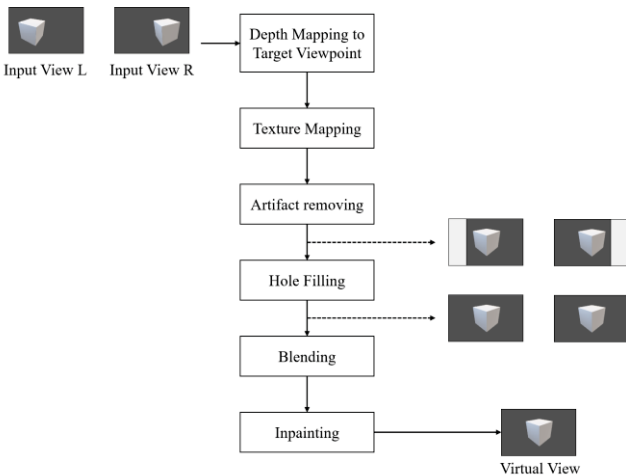


그림 2. RVS 의 뷰 합성 구조도

2.2 Multi-view Video Coding

단일 시점 영상만을 제공하여 제한적인 자유도를 보장하는 기존 VR 서비스의 단점을 극복하기 위해 다시점 영상의 필요성이 대두되었다. 사용자가 몰입감을 느끼게 하기 위해 2.1 절에서 언급된 바와 같이 여러 위치에서 동시에 취득된 다시점 영상을 이용하여 가상 영상을 제공할 수 있다. 이를 위해 다시점 비디오 코딩 기술이 개발되었는데, MPEG 의 경우 free viewpoint television (FTV)을 위한 3D 비디오 시스템[8]을 제안하였다. 해당 시스템은 다시점 비디오 취득, 압축, 전송, 복호화 등의 내용을 포함한다. 다시점 비디오는 여러 영상들을 포함하기에, 이를 효율적으로 압축하기 위한 방법이 요구되었다. 이에 공간적 · 시간적 중복성 외에 뷰 간 중복성을 제거하는 multi-view video coding (MVC) [9][10] 이 제안되었다. 그림 3 은 MVC 의 다시점 비디오 압축 구조를 나타낸다. 시간의 흐름에 따른 프레임 간 차이만을 저장하는 기존 압축 기술에 뷰 간 계층적 압축 구조가 포함되어 있다. 여러 뷰들 중 하나의 뷰를 기본 뷰로 설정하고 나머지 뷰들은 기본 뷰를 참조하여 압축되는 확장 뷰로 정의되어 다시점 비디오 압축이 이루어진다.

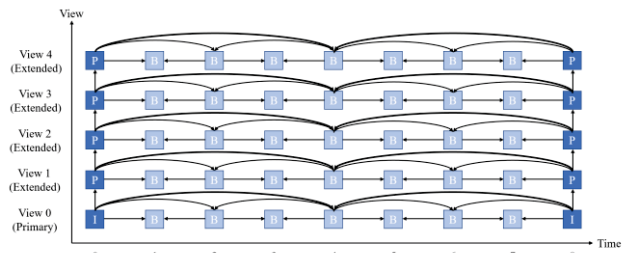


그림 3. MVC 의 비디오 압축 구조도

3. 영상 위치 기반 다시점 비디오 압축 시스템

본 논문은 3DoF+에 대응하는 영상 위치 기반 다시점 비디오 압축 시스템을 제안한다. 제안하는 시스템의 주요 모듈은 2 가지로 나뉘어진다. 3.1 절에서는 외평 기반 뷰 간 중복성 제거 기법에 대해 설명하고, 3.2 절에서는 영역 확장 기반 잔여 뷰 병합 기법에 대해 설명한다.

3.1 외평 기반 뷰 간 중복성 제거 기법

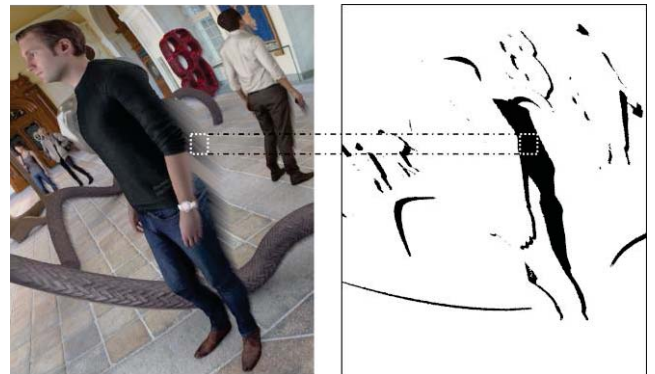


그림 4. 외평 기반 뷰 간 중복성 제거 예시

3DoF+에 대응하는 서비스를 제공하기 위해서는 다시점에서 취득된 영상 전송이 요구되는데, 2.2 절에서 살펴본 바와 같이 뷰 간 중복성을 제거하면 압축을 좀 더 효율적으로 할 수 있다. 중복성 제거 시 기준이 되는 영상을 본 논문에서는

중양 뷰라고 정의하고, 중양 뷰는 취득된 다시점 비디오들의 한가운데 위치한다. 만약 그 위치에 중양 뷰가 없을 경우 제안하는 시스템은 RVS 를 이용하여 중양 뷰를 합성한다.

중복되는 영역을 판단하기 위해 본 논문에서는 RVS 의 와핑 기술을 이용하였는데, 그림 4 는 와핑 기반 뷰 간 중복성 제거 예시를 나타낸다. 그림 4 의 왼쪽 그림은 중양 뷰에서 다시점 비디오 중 하나의 위치로 와핑된 이미지를 나타낸다. 흰색 점선으로 표시된 사각형 박스를 보면 픽셀들이 왜곡된 것을 볼 수 있는데 이 영역은 중양 뷰가 표현할 수 없고 오직 해당 위치의 비디오에 의해서만 표현될 수 있다. 때문에 본 논문이 제안하는 시스템은 왜곡된 부분을 탐색한 후 이를 그림 4 의 오른쪽 그림과 같이 이진 마스크(binary mask)로 나타낸다. 검정색 영역은 해당 위치의 비디오에서 텍스처를 가져올 영역을 나타내고, 하얀색 영역은 임의의 값으로 채워도 무방한 영역을 나타낸다. 마스크가 생성되면 검정색 영역에 해당하는 부분만 텍스처를 매핑하여 중복성이 제거된 뷰를 출력한다. 이후 중복성이 제거된 잔여 뷰를 인코딩 후 전송 시 원본을 인코딩하여 전송할 때보다 더 적은 대역폭을 요구한다.

3.2 영역 확장 기반 잔여 뷰 병합 기법

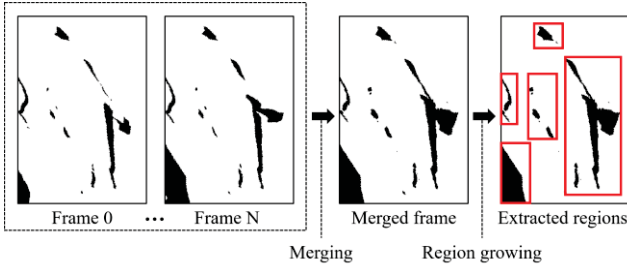


그림 5. 프레임 병합 및 영역 확장·추출 예시

3.1 절에서 소개된 뷰 간 중복성 제거 기술의 적용으로 잔여 뷰 압축 후 전송 시 필요한 대역폭이 감소하였으나, 보낼 영상의 수는 감소하지 않아 다수의 디코더를 요구하여 클라이언트에 부담이 된다. 이를 해결하기 위해 본 논문에서는 잔여 뷰 병합 기법을 제안한다. 해당 기술은 영역 확장(region growing) [11]을 사용하여 잔여 뷰에서 복원 시 필요한 부분을 블록 단위로 추출하고, 하나의 영상으로 병합한다.

프레임마다 영역 확장 후 추출하여 하나의 영상으로 병합하면 같은 영역이라도 프레임마다 크기가 달라지므로 병합된 영상에서의 위치가 변경될 수 있다. 이는 프레임 간 압축의 효율성을 떨어뜨리고 비트레이트의 증가를 야기한다. 이를 해결하기 위해 본 논문에서는 이진 마스크를 intra period 단위로 병합한다. 그림 5 는 이진 마스크의 프레임이 병합되는 과정을 나타낸다. 병합된 프레임에서 영역을 추출하기 위해 영역 확장을 시행하는데, 그림 5 의 오른쪽 그림에서 빨간색 박스는 추출된 영역을 나타낸다. 이진 마스크에서 확보된 영역에 해당하는 부분을 텍스처에서 추출하여 이를 높이에 따른 내림차순으로 정렬하고 하나의 영상으로 병합하고 압축 및 전송이 이루어진다. 이렇게 잔여 뷰를 하나로 합침으로서 비트레이트 절약과 필요한 디코더 수의 감소를 유도할 수 있다.

4. 구현 내용 및 결과

본 논문에서는 다시점 비디오 압축 성능을 실험하기 위해 3DoF+를 위한 테스트 시퀀스인 ClassroomVideo [12],

TechnicolorMuseum [13]을 사용하였으며 해당 시퀀스들은 random access (RA) 방식으로 인코딩되었다. 실험에는 2 개의 intel xeon e5-2687w v4 CPU 와 128GB 의 메모리가 장착된 서버가 사용되었고, ubuntu 16.04 환경에서 진행되었다. 실험에는 3DoF+ CfP 에서 권장하는 프로그램들이 사용되었다. 인코딩과 디코딩에는 HEVC test model (HM) 16.16 버전이 사용되었고 뷰 합성에는 OpenCV 3.4.2 [14] 버전과 연계된 RVS 3.1 버전이 사용되었으며 품질 평가에는 ERP 영상을 위한 평가지표인 weighted-to-spherically-uniform peak signal-to-noise ratio (WS-PSNR) 2.0 버전이 사용되었다.

표 1 은 제안하는 시스템을 사용하여 다시점 비디오를 전송할 때 절약되는 BD-rate 와 pixel-rate 결과 값을 보여준다. 대조군으로는 HEVC encoder 만을 사용하여 인코딩된 비디오를 사용하였고, 평균 12.43%의 BD-rate 절약률과 82.73%의 pixel-rate 절약률을 보여주었다.

Name	BD-rate saving	Pixel-rate saving
ClassroomVideo	20.14%	78.75%
TechnicolorMuseum	4.72%	86.72%
Average	12.43%	82.73%

표1. 제안하는 시스템을 사용하여 다시점 비디오를 전송할 때의 BD-rate, pixel-rate 절약률

5. 결론

본 논문은 몰입형 미디어를 위한 다시점 360 비디오 압축 시스템을 제안한다. RVS 를 활용하여 다시점 비디오 간 중복성을 제거하고 필요한 영역만 추출하여 하나의 영상으로 병합하여 압축 및 전송한다. 제안하는 시스템을 사용하여 3DoF+를 위한 다시점 영상을 전송했을 때 평균 12.43%의 BD-rate 와 82.73%의 pixel-rate 절약률을 보여주었다. 제안하는 시스템은 MPEG-I 에서 현재 표준화 단계에 있는 3DoF+와 6DoF 에도 적용이 가능하다.

Acknowledgement

이 논문은 2018 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2018-0-00765, 6DoF 지원 초고화질 몰입형 비디오의 압축 및 전송 핵심 기술 개발).

참고문헌

- [1] Mary-Luc Champel, Thomas Stockhammer, Thierry Fautier, Emmanuel Thomas, Rob Koenen. 2016. Quality Requirements for VR. 116th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 116/m39532.
- [2] Xin Wang, Lulin Chen, Shuai Zhao, and Shawmin Lei. 2017. From OMAF for 3DoF VR to MPEG-I Media Format for 3DoF+, Windowed 6DoF and 6DoF VR. 119th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG119/m44197.
- [3] Jong-Beom Jeong, Dongmin Jang, Jangwoo Son, and

- Eun-Seok Ryu. 2018. 3DoF+ 360 Video Location-Based Asymmetric Down-Sampling for View Synthesis to Immersive VR Video Streaming. *Sensors* 18 ,9(2018), 3148.
- [4] ISO/IEC JTC1/SC29/WG11. 2019. Call for Proposals on 3DoF+ Visual. 125th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG/n18145.
- [5] ISO/IEC JTC1/SC29/WG11. 2018. 3DoF+ Software Platform description. 125th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG2018/n18070.
- [6] ISO/IEC JTC1/SC29/WG11. 2018. Reference View Synthesizer (RVS) manual. 124th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG124/n18068.
- [7] Patrick Boissonade, Joel Jung, and P.Nikitin. 2018. [MPEG-IVisual] View synthesis algorithm for windowed-6DoF. 124th MPEG meeting of ISO/IECJTC1/SC29/ WG11, MPEG2018/m44788.
- [8] ISO/IECJTC1/SC29/WG11.2008. Overview of 3D Video Coding. 84th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG2008/n9784.
- [9] Philipp Merkle, Aljoscha Smolic, Karsten Muller, and Thomas Wiegand. 2007. Multi-view video plus depth representation and coding. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, Vol. 1. IEEE, 1- 201.
- [10] Karsten Müller, Heiko Schwarz, Detlev Marpe, Christian Bartnik, Sebastian Bosse, Heribert Brust, Tobias Hinz, Haricharan Lakshman, Philipp Merkle, FranzHunn Rhee, etal. 2013. 3D high-efficiency video coding for multi-view video and depth data. *IEEE Trans. Image Processing* 22, 9(2013), 3366- 3378.
- [11] M Mary Synthuja Jain Preetha, L Padma Suresh, and M John Bosco. 2012. Image segmentation using seeded region growing. In *2012 International Conference on Computing, Electronics and Electrical Technologies (ICCEET)*. IEEE, 576- 583.
- [12] Bart Kroon. 2018. 3DoF+ test sequence ClassroomVideo. 122th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2018/m42415.
- [13] Renaud Dore. 2018. Technicolor 3DoF+ test materials. 122th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2018/m42349.
- [14] 2018. Open computer vision. [Online]. <https://github.com/opencv/opencv/releases/tag/3.4.2>