

VVC 에서의 움직임 제한 타일 셋 기반 타일 추출기 구현

정종범¹⁾, 이순빈¹⁾, 류일웅²⁾, 김성빈¹⁾, 김인에¹⁾, 류은석¹⁾

1) 성균관대학교 컴퓨터교육과, 2) 가천대학교 컴퓨터공학과
uof4949@skku.edu, soonbinlee@skku.edu, dlfdnd96@gc.gachon.ac.kr,
beencooke@skku.edu, inaelk@skku.edu, esryu@skku.edu

Implementing Motion-constrained Tile Set Based Tile Extractor on VVC

Jong-Beom Jeong, Soonbin Lee, Il-Woong Ryu, Sungbin Kim, Inae Kim,
Eun-Seok Ryu

- 1) Department of Computer Education, Sungkyunkwan University,
2) Department of Computer Engineering, Gachon University

요 약

최근 몰입형 가상 현실을 제공하기 위한 360 도 영상 전송 기술이 활발히 연구되고 있다. 그러나 현재 가상현실 기기가 가지는 연산 능력 및 대역폭으로는 고화질의 360 도 영상을 전송 및 재생하기에 한계가 있다. 해당 문제점을 극복하기 위해 본 논문에서는 사용자 시점의 고화질 360 도 영상 제공을 위해 사용자 시점 타일을 추출하는 움직임 제한 타일 셋 기반 타일 추출기를 구현한다. Versatile video coding (VVC) 기반 타일 인코더를 이용해 360 도 영상에 대한 비트스트림을 생성한 후, 사용자 시점에 해당하는 타일들을 선택한다. 이후 선택된 타일들은 제안하는 타일 추출기에 의해 추출되고 전송된다. 또한, 전체 360 도 영상에 대한 저화질 비트스트림을 전송하여 갑작스러운 사용자 시점 변경에 대응한다. 제안된 타일 추출기를 기반으로 360 도 영상 전송을 수행하면, 기존 VVC 기반 시스템 대비 대비 평균 24.81%의 bjontegaard delta rate (BD-rate) 감소가 가능함을 확인하였다.

1. 서론

최근 가상 현실 기술이 주목을 받으면서, 관련 시장이 급격히 증가하고 있다. 대표적으로 머리장착형 영상장치인 head-mounted display (HMD)의 보급이 증가하고 있고, 모바일 디바이스도 이를 지원하기 위한 기능이 탑재되고 있다. 이에 따라 몰입감있는 가상 현실 미디어를 제공하기 위한 효율적인 360 도 영상 스트리밍 기술이 요구되고 있다. 그러나 현재 시장에 보급된 가상 현실 영상을 제공하는 디바이스의 연산능력과 대역폭은 한계가 존재한다. 영상 압축 표준을 진행하는 moving picture experts group (MPEG)에서 제안된 보고서[1]에 의하면, 가상 현실에서 사용자가 어지러움을 느끼지 않을 정도의 영상을 제공하려면 12K 이상, 90 fps 이상의 초고화질 360 도 영상과 낮은 전송 지연이 필요하다고 한다. 이에 MPEG 과 ITU-T video coding experts group (VCEG) 은 차세대 영상 압축 코덱 표준화를 위해 joint video coding experts team (JVET)을 설립 후, 360 도 영상의 효율적 전송을 위해 관련 실험 조건 및 평가 조건을 확립하였고[2], 이후 다양한 360 도 영상 스트리밍 방법이 제안되었다. 360 도 영상을 적응적으로 다운샘플링하는 연구[3], 360 도 영상을 타일로 분할하여 인코딩한 후 압축된 비트스트림 레벨에서 사용자 시점 타일만 추출 및 전송하는 연구[4][5], 복수 영상 전송 시 360 도 영상 간 중복성을 제거하는 연구[6] 등이 제안되었다.

사용자 시점의 360 도 영상 렌더링 시 영상의 전체를 사용하는 것이 아닌 영상의 일부만 HMD 에 디스플레이 되기 때문에, 기존에 널리 사용되던 high-efficiency video coding (HEVC) 기반 타일 추출기 및 스트리밍 기법이 제안된 바 있다. 본 논문은 최근 표준화가 진행중인 versatile video coding (VVC) 에서 동작하는 사용자 시점 기반 타일 스트리밍 시스템을 위한 타일 추출기를 제안한다. 이를 위해 360 도 영상의 일부를 타일의 형태로 추출가능하게 하는 motion-constrained tile set (MCTS)을 VVC 인코더에서 동작가능하도록 하고, 해당 인코더를 사용하여 360 도 영상에 대한 고화질 비트스트림을 생성한다. 이후 추출기는 비트스트림의 파라미터 셋을 복호화하여 타일에 대한 정보를 저장하고, 해당 정보를 기반으로 사용자 시점 타일들에 대한 추출을 진행한 후 전송한다. 동시에 사용자 시점이 갑작스럽게 변경될 때를 대비하여 전체 영상을 저화질로 인코딩하여 전송한다. 그림 1 은 본 논문에서 제안하는 타일 추출기의 구조도를 나타낸다.

본 논문의 구성은 다음과 같다. 2 절에서는 관련 연구로 HEVC 기반 타일 스트리밍을 소개한다. 3 절에서는 제안하는 사용자 시점 타일 선택 및 타일 추출기 설계 및 구현 내용을 설명한다. 4 절에서는 제안하는 타일 추출기의 실험 결과를 설명하고 마지막으로 5 절에서는 본 논문의 결론을 맺는다.

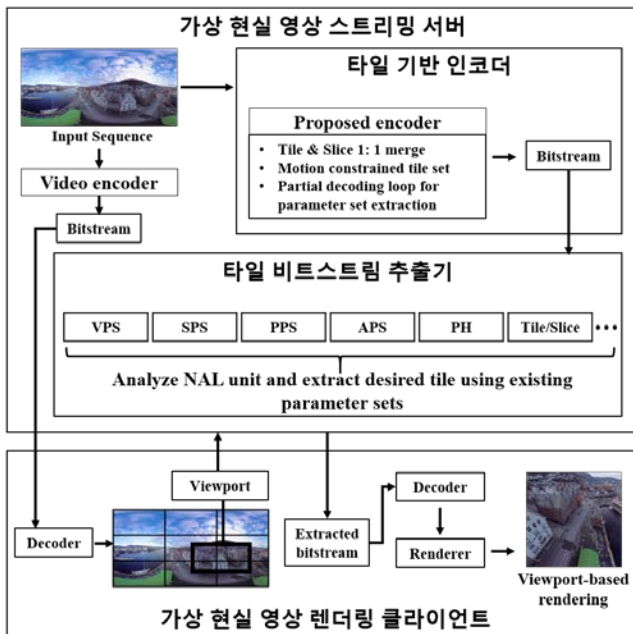


그림 1. 제안하는 VVC 타일 추출기 기반 360도 영상 스트리밍 구조도

2. 관련 연구

HMD 에서 가상 현실 영상을 재생하게 되면 서버에서 전송받은 360 도 영상의 일부만을 화면에 표현한다. 따라서, 가상 현실 영상 스트리밍의 경우 사용자가 바라보고 있는 영역에 해당하는 영상만 전송하면 대역폭을 절감할 수 있다. 전체 360 도 영상을 단순히 분할해서 인코딩한 후 전송해도 대역폭 절감이 가능하지만 서버가 감당해야 할 연산량이 급격히 증가하고, 다루어야 할 비트스트림의 개수가 늘어나는 문제가 있다. 때문에, HEVC 에서는 하나의 비트스트림에서 전체 영상의 일부에 해당하는 직사각형 영역, 즉 타일을 추출하는 기술이 구현되었다[7]. HEVC 에서 하나의 픽처는 다수의 슬라이스로 구성되는데, 슬라이스 간에는 의존성이 없어 비트스트림을 패킷으로 분할하여 전송 시 일부 패킷이 손상되더라도 다른 슬라이스는 디코딩이 가능하다. 이를 타일로 확장시켜 하나의 슬라이스가 하나의 타일을 가지도록 하고, 해당 타일은 직사각형의 형태를 취하도록 하면 전체 픽처의 일부만을 타일의 형태로 추출할 수 있다. 추출 후 디코딩 시 의존성 문제를 없애기 위해 프레임 간 인코딩 시 움직임 벡터의 범위를 타일 내부로 제한하여 각 타일의 독립적 추출 및 디코딩이 가능하도록 하는 기술이 MCTS 인데, 이를 활용하여 타일의 추출이 가능하다. [7]은 사용자의 field of view (FoV) 만을 전송 시 360 도 영상 전체를 전송하는 기존 기법에 비해 평균 35.42%의 bjontegaard delta rate (BD-rate) 절약을 보임을 증명하였다. 본 연구는 기존 연구와 다르게 최근 표준화가 진행중인 영상 압축 코덱인 VVC 에서의 MCTS 활성화 및 타일 비트스트림 추출 과정을 상세하게 기술한다.

3. 움직임 제한 타일 셋 기반 타일 추출기

본 절은 VVC 에서의 MCTS 기반 타일 추출기의 설계 및 동작 과정에 대해 기술한다. 본 논문에서 제안하는 추출기는 VVC test model (VTM) 7.3 에 기반하여 설계되었다. 해당

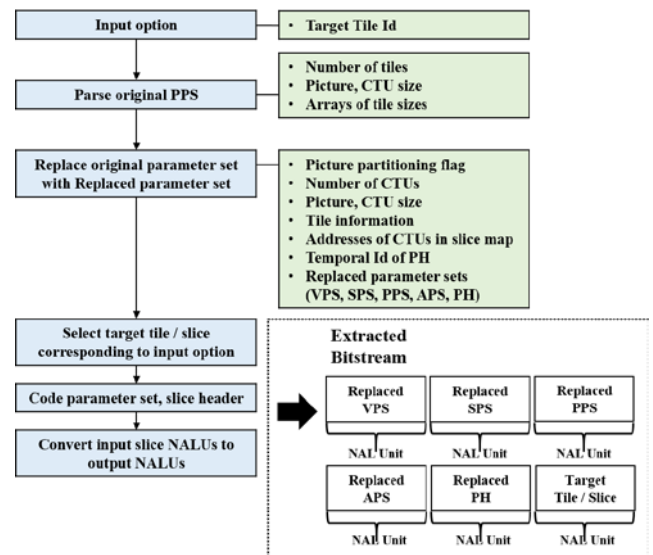


그림 2. 제안하는 VVC 타일 추출기 기능 흐름도

버전의 인코더에서 MCTS 기능을 활성화하려면 EnablePicPartitioning, MCTSEncConstraint 옵션에 1 을 할당해야 한다. 동시에 타일의 너비, 높이를 명시하고 슬라이스 간 루프 필터를 비활성화하여야 한다. 본 논문이 제안하는 타일 추출기는 HEVC 에서의 타일 추출기와 같이 타일과 슬라이스가 1 대 1 로 대응됨을 전제로 한다.

이후 타일 추출기는 인코딩된 비트스트림과 추출하고자하는 타일 인덱스를 입력 받아 목표 타일을 추출하여 비트스트림의 형태로 저장한다. 비트스트림은 network abstraction layer (NAL) unit 의 형태로 저장되는데, 추출기는 타일 추출 시 필요한 정보들을 NAL unit 의 헤더에서 가져온다. HEVC 타일 추출기의 경우 video parameter set (VPS), sequence parameter set (SPS), picture parameter set (PPS) 에 해당하는 NAL unit 에서 정보를 가져온다. MCTS 가 적용된 HEVC 인코더는 타일들에 대한 VPS, SPS, PPS 를 extraction information sets (EIS) supplemental enhancement information (SEI) message 의 형태로 가공하고, NAL unit 으로 저장한다. 이를 HEVC 타일 추출기가 복호화하여 타일 추출 시 사용하게 된다.

한편, VVC 에서는 SEI message 를 사용하려면 HEVC_SEI flag 를 활성화시켜야 하는데, VTM 7.3 기준으로 해당 flag 는 비활성화 되어있어 VVC 타일 추출기는 HEVC 타일 추출기와는 다른 추출 방법이 필요하다. 따라서 본 논문에서는 타일들의 coding tree unit (CTU) 주소를 검사하는 방식을 채택하였다. 이를 통해 EIS SEI message 를 저장할 필요가 없으므로 비트스트림의 크기를 더욱 줄일 수 있다는 장점이 있다. VVC 에는 VPS, SPS, PPS 외에도 adaption parameter set (APS), picture header (PH) 등의 NAL unit 이 존재하여 타일 추출기가 인코딩된 비트스트림을 읽어들 때 상기 NAL unit 을 읽어들 수 있도록 하였다. 제안하는 추출기는 언급된 5 개의 parameter set 들을 읽어오고 저장하며, 전체 픽처 크기, 타일 개수 및 크기, CTU 크기 등을 읽어온다. 이후 추출기는 추출하고자 하는 타일의 CTU 단위 시작 주소를 계산하고, 타일을 포함하는 슬라이스 NAL unit 을 읽었을 때 해당 타일의 시작 주소가 목표 타일과 일치하면 추출한다. 이후 추출기는 저장된 5 개의 parameter set 에서 타일 관련 옵션을 비활성화하고, 슬라이스의 시작 주소를 수정하여 인코딩한 후 출력 비트스트림에 저장한다. 추출된 슬라이스는 입력 NAL

unit 의 형태이기에, 추출기는 출력 NAL unit 으로 변환하여 출력 비트스트림에 저장한다. 상기 과정을 반복하여 목표 타일 비트스트림이 출력되고, 이는 VVC 디코더로 복호화할 수 있다.

4. 실험 결과

본 논문에서는 제안하는 추출기 효율을 측정하기 위해 JVET 에서 정의한 360 도 영상에 대한 실험 조건인 common test conditions (CTC)를 준수하였다. 실험에는 CTC 에 정의된 테스트 시퀀스인 AerialCity, DrivingInCity, DrivingInCountry 를 사용하였으며 해당 시퀀스들은 VTM 7.3 인코더에 의해 random access (RA) 방식으로 인코딩되었다. 양자화 파라미터 (quantization parameter, QP) 로는 고화질 영상 인코딩 시 22, 27, 32, 37 이 사용되었고, 저화질 영상 인코딩 시 42 가 사용되었다. 본 실험에서는 사용자의 field of view (FoV)를 90 도로 설정하였고, 이에 따라 테스트 시퀀스는 3×6 배열로 분할되어 한 프레임 당 18 개의 타일을 가지도록 인코딩되었다. 테스트 시퀀스는 모두 3840×1920 크기의 영상이고, 각 타일은 640×640 의 크기를 가진다. 360 도 영상에서의 사용자 시점은 360 library [8]에 기초하여 fixed viewport 옵션으로 결정되었고, 사용자 시점 타일을 추출하기 위해 omnidirectional media format (OMAF) [9]에 기반하여 사용자 시점 타일을 선택하였다. 사용자 시점 타일은 AerialCity, DrivingInCity, DrivingInCountry 에 각각 6, 9, 6 개가 선택되었고, 평균 7 개의 타일이 추출 및 전송되었다. 즉, 전체 영상의 약 38.88%에 해당하는 영역이 고화질 타일의 형태로 전송되었고, 전체 영상에 대해서는 저화질 영상의 형태로 전송되었다.

실험에는 2 개의 intel xeon e5-2687w v4 CPU 와 128GB 의 메모리가 장착된 서버가 사용되었고, ubuntu 18.04 환경에서 진행되었다. 인코딩과 디코딩에는 VTM 7.3 이 사용되었고 본 논문에서 제안하는 추출기 또한 해당 버전을 사용하여 개발되었다. 품질 평가에는 peak signal-to-noise ratio (PSNR), video multimethod assessment fusion (VMAF) [10], multi-scale structural similarity (MS-SSIM) [11], immersive video PSNR (IV-PSNR) [12] 가 사용되었다.

표 1 은 제안하는 타일 추출기를 사용하여 고화질 사용자 시점 타일과 저화질 전체 영상을 전송했을 때 고화질 전체 영상 전송 대비 BD-rate 절약을 보여준다. 제안하는 타일 추출기 사용 시 기존 시스템 대비 Y-PSNR, VMAF [10], MS-SSIM, IV-PSNR 에 대해 평균 24.81%, 12.85%, 18.54%, 30.86%의 BD-rate 절약을 가능하다. 그림 2 는 제안하는 타일 추출기 및 기존 시스템에 대한 rate-distortion curve 를 보여준다. 가로축은 대역폭, 세로축은 품질을 나타내는데, 모든 대역폭 영역에서 제안하는 타일 추출기 사용 시 동일 대역폭에서 기존 시스템보다 더 좋은 품질의 영상을 제공할 수 있음을 알 수 있다.

Sequence name	Y-PSNR	VMAF	MS-SSIM	IV-PSNR
AerialCity	-21.01%	-3.07%	-15.61%	-32.04%
DrivingInCity	-24.98%	-16.70%	-18.13%	-31.18%
DrivingInCountry	-28.45%	-18.79%	-21.88%	-29.37%
Average	-24.81%	-12.85%	-18.54%	-30.86%

표 1. 제안하는 타일 추출기의 기존 시스템 대비 BD-rate 절약을

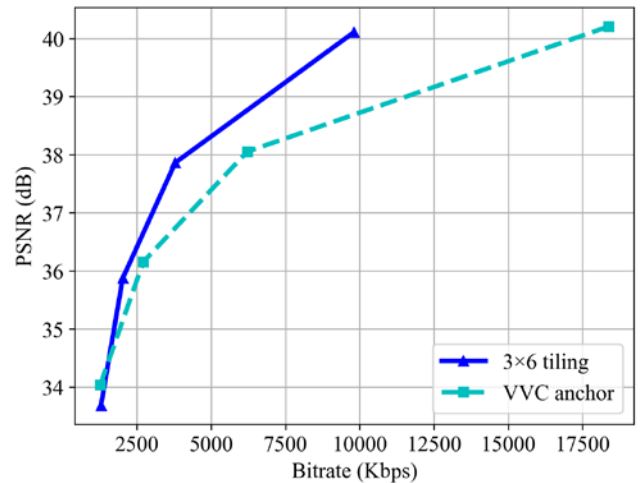


그림 3. 제안하는 타일 추출기 및 기존 시스템 RD-curve

5. 결론

본 논문은 VVC 에서의 움직임 제한 타일 셋 기반 타일 추출기를 제안한다. 제안하는 타일 추출기는 각 타일에 대한 주소 검사를 통해 목표 타일을 추출하여 HEVC 타일 추출기와 달리 EIS SEI message 와 같은 추가적인 metadata 를 요구하지 않는다. 추출된 고화질의 타일은 전체 영상에 대한 저화질 영상과 같이 전송되어 사용자 시점 변경에 대응함과 동시에 고화질의 영상을 사용자에게 제공한다. 제안하는 타일 추출기를 사용하였을 때 평균 24.81%의 BD-rate 절약을 확인하였다. 향후 연구로는 다양한 타일 크기에 대한 스트리밍 효율을 측정하여 타일 기반 스트리밍에 대한 trade-off 를 도출할 계획이다.

Acknowledgement

이 논문은 2020 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). 이 논문은 또한 2020 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2019R1A2C1010476).

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] Jill Boyce, Elena Alshina, Adeel Abbas, Yan Ye. 2017. JVET common test conditions and evaluation procedures for 360° video. 118th MPEG meeting of ISO/IEC JTC1/SC29/WG11, ISO/IEC JTC1/SC29/WG11/n16891.
- [3] Jong-Beom Jeong, Dongmin Jang, Jangwoo Son, 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] Jang–Woo Son, Dongmin Jang, Eun–Seok Ryu. 2018. Implementing Motion–Constrained Tile and Viewport Extraction for VR Streaming. In *Proceedings of the 28th ACM Network and Operating System Support for Digital Audio and Video*, pp. 61–66.
- [5] Jangwoo Son, Eun–Seok Ryu. 2018. Tile–Based 360–Degree Video Streaming for Mobile Virtual Reality in Cyber Physical System. *Computers & Electrical Engineering*, 72, 361–368.
- [6] Jong–Beom Jeong, Soonbin Lee, Dongmin Jang, Eun–Seok Ryu. 2019. Towards 3DoF+ 360 Video Streaming System for Immersive Media. *IEEE Access*, 7, pp. 136399–136408.
- [7] Alireza Zare, Alireza Aminlou, Miska Hannuksela, Moncef Gabbouj. 2016. HEVC–compliant tile–based streaming of panoramic video for virtual reality applications. In *Proceedings of the 24th ACM international conference on Multimedia*, pp. 601–605.
- [8] 360lib 5.1. 2017. Retrieved from https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-5.1
- [9] 오세진. 2017. 360 미디어를 위한 MPEG Omnidirectional Media Format (OMAF) 표준 기술. *방송공학회논문지*, 22(5), pp. 600–607.
- [10] Christos Bampis, Alan Bovik, Zhi Li. 2018. A Simple Prediction Fusion Improves Data–driven Full–Reference Video Quality Assessment Models. In *2018 Picture Coding Symposium (PCS)*, pp. 298–302.
- [11] Zhou Wang, Eero Simoncelli, Alan Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The Thrity–Seventh Asilomar Conference on Signals, Systems & Computers*, Vol. 2, pp. 1398–1402.
- [12] Adrian Dziembowski. 2019. Software manual of IV–PSNR for Immersive Video. 128th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG127/n18709.