

사용자 시점 기반 3DoF+ 스트리밍을 위한 타일 그룹 분할 방법

이순빈¹, 정중범², 김성빈¹, 류은석^{2*}가천대학교¹, 성균관대학교²soon0698@gc.gachon.ac.kr¹, uof4949@skku.edu²,holyvin@gc.gachon.ac.kr¹, esryu@skku.edu^{2*}

HEVC Tile Grouping for Viewport-dependent 3DoF+ Streaming

Soonbin Lee¹, Jong-Beom Jeong², Sungbin Kim¹, Eun-Seok Ryu^{2*}Gachon University¹, Sungkyunkwan University²

요 약

최근 MPEG-Immersive 표준화 그룹에서는 여러 위치에 존재하는 다수의 360 영상을 취득하고 처리하여 사용자에게 제한적인 화면 전환 이동에 따른 운동 시차(motion parallax)를 제공하는 3 degrees of freedom plus (3DoF+) 기술에 관한 표준화가 진행되고 있다. 이러한 3DoF+ 표준화 기술에서의 압축은 기존의 HEVC 코덱을 사용하나, 현재 존재하는 360 영상 사용자 시점 기반 스트리밍 기법을 3DoF+ 스트리밍에 적용하기 위해서는 기존의 360 영상과 다르게 영상 서로 간의 참조과정으로 인한 의존성으로 문제가 존재한다. 본 논문에서는 해당 문제점을 서술하고, MCTS(Motion-Constrained Tile Set)를 사용하여 보다 적은 디코더와 사용자 시점에 맞게 효율적으로 영상을 전송할 수 있는 MCTS 기반 타일 그룹 분할 방법을 제안한다.

1. 서론

360 영상과 virtual reality (VR) 기술 등에 대한 관심이 증대됨에 따라, 보다 몰입할 수 있는 실감형 미디어 압축 기술에 대한 논의가 활발하게 이루어지고 있다. 이에 따라 moving picture experts group (MPEG)에서는 몰입형 미디어를 표준화하기 위해 MPEG-I (Immersive) 서브그룹을 결성하여 가상현실 비디오 표준화를 각각 3DoF, 3DoF+, 6DoF 로 분류하여 표준화를 진행중에 있다 [1].

이러한 여러 분류들 중 3DoF+ 표준화 기술은 여러 위치에 존재하는 다수의 360 영상을 취득하고, 여러 감상 시점에 맞는 중간 뷰 합성을 진행하여 사용자에게 제한적인 움직임의 운동 시차를 제공하고자 하는 기술이며, 현재 ISO/IEC 23090 Part 7 (Metadata for Immersive Media (Video))에서는 첫 단계로 3DoF+ 비디오의 압축 표준을 2020 년 초 완료를 목표로 하고 있다 [2]. 현재 3DoF+ 표준화에서는 기존의 HEVC 비디오 코덱을 사용하여 평가를 진행중에 있으며, 향후 Versatile Video Coding (VVC) 등의 차세대 코덱을 도입할 것으로 보인다.

또한 전방위 몰입형 미디어의 특성에 따라 사용자는 감상하는 영상에서 동시에 제한된 시야각(FoV, Field of View)으로 콘텐츠를 감상하게 된다. 이에 따라 사용자의 관심 영역(RoI, Region of Interest)만을 전송하거나, 해당하는 영역만을 고화질로 압축하여 전송하는 사용자 시점(Viewport) 기반 스트리밍에 대한 연구가 기존의 360 영상 콘텐츠에 대해 진행되온 바 있다 [3,4].

HEVC 를 사용하는 360 영상의 사용자 시점 기반 스트리밍은 보다 적은 코덱의 오버헤드, 독립적인 영역 전송을 위해 기본적인 병렬 프로세스 단위인 타일(Tile)을 단위로 나누어지게 된다. 이 때, 임의의 영역으로 나뉘어진 타일들이 독립적으로 전송되기 위해서는 압축 과정에서 타일 간의 참조

의존성을 제한해야 할 필요가 있다. 이에 따라 각 타일의 공간 및 시간적 참조가 타일 내부를 벗어나지 않도록 인코딩하는 기법인 움직임 참조 제한 타일(MCTS, Motion Constrained Tile Set) 기법이 제안되고 [5], 현재 널리 연구되고 있다 [6,7,8].

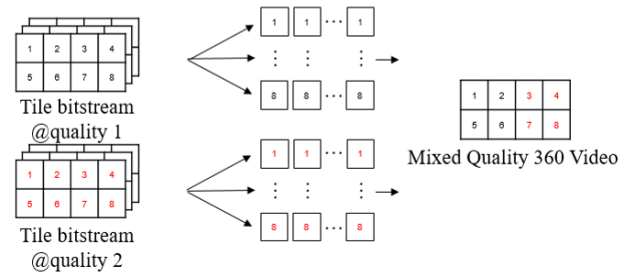


그림 1. 타일 기반 360 영상 스트리밍 개념도.

그림 1 은 다양한 품질 별로 압축된 타일 비트스트림에서 사용자 시점 또는 여러 요구사항에 맞도록 타일들을 독립적으로 송수신하여 스트리밍을 진행하는 개념도를 나타낸다. 본 논문에서는 기존의 360 영상 사용자 시점 기반 스트리밍 기법을 3DoF+ 기술에 적용할 때 나타나는 문제점에 관하여 서술하고, 방향을 고찰한다.

2. 관련 연구

3DoF+ 표준화 동향에서는 다섯 개의 기관에서 제안한 압축 표준 제정 기술에 대하여 평가하고 해당 내용들을 채택한 기술을 토대로 TMIV(Test Model for Immersive Video) 라는 공개 테스트 모델 소프트웨어를 개발하였다 [9,10]. 기본이 되는 모델은 여러 입력 360 영상 중 정보를 가장 많이 담고 있는 시점의 영상을 계산해 기본 시점(Basic View)으로

간주하고, 나머지 시점의 영상들은 기본 시점과의 차이만을 계산하여 패치(Patch) 형태로 하나의 큰 영상 (통상 4K 해상도) 안에 모아 결과적으로 기본 시점과 차분 패치들을 모아놓은 영상들만을 압축하여 전송하는 모델이 된다 [11].

이 때 이 패치들이 모인 이미지들을 아틀라스(Atlas)라고 부르게 되며, 시점 간의 차이를 계산하여 다른 정보를 담고 있는 패치들만을 구하는 과정을 프루닝(Pruning)이라고 한다. 현재 TMIV 에서 구현되어 있는 프루닝 과정은 최대의 압축 효율을 위하여 계층적으로 여러 영상 간을 비교하여 과정을 진행하게 되어 있다.

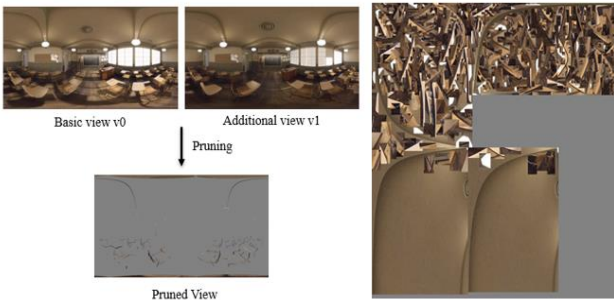


그림 2. (좌) 프루닝 과정의 예시, (우) 여러 차분 패치들이 모인 하나의 아틀라스 이미지.

현재 TMIV 에서 생성된 아틀라스에서 사용자 시점 기반 스트리밍을 적용하기 위해서는 영상들은 서로 간 참조하는 정보가 최대한 적어야 한다. 즉, 임의의 독립적인 영역들의 영상만을 전송해 최종 렌더링 영상을 복원하기 위해서는 다른 영역의 영상들과의 의존성이 없어야 한다.

따라서 현재 TMIV 에서는 (1) 생성된 아틀라스의 패치들을 구분해주어 나눌 필요가 있으며, (2) 프루닝 과정에서 생기는 영상들 간의 의존성을 제거해야 한다.

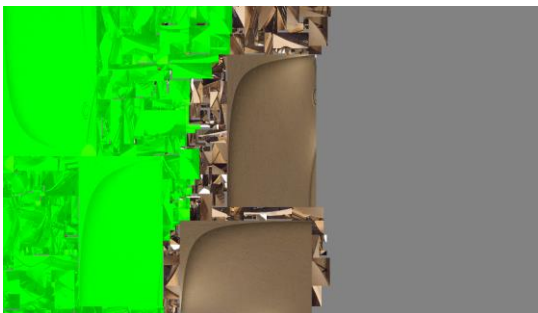


그림 3. 아틀라스에서 특정 카메라의 패치들을 구분한 결과

첫번째 방법의 경우, 단순히 카메라 순으로 패치를 정렬하여 구분하는 방법을 생각해볼 수 있다. 이 경우 현재 TMIV 에서 패치들을 조합할 때, 아틀라스의 크기를 최대한 줄이기 위하여 고안된 PatchInPatch 옵션으로 패치 간 빈 공간에 다시 작은 패치들이 붙을 수 있다.

따라서 해당 옵션의 적용으로 인한 다른 필요없는 영역의 시점들이 포함될 수 있으며 프루닝에 대한 의존성이 여전히 해결되지 않기 때문에, 시점들을 TMIV 전처리 단계에서부터 나누어 각각 따로 전처리 과정을 거치는 그룹화 방법을 생각해볼 수 있다.

이러한 영상의 분할 영역에 맞게 그룹(Group)화하려는 시도가 이미 제안되었으나, 해당 방법은 깊이 맵(depth map)과

카메라 파라미터 부정합으로 인해 나타나는 최종 렌더링 영상의 화질의 열화를 개선하기 위해 지역적 연속성 정보(Local Coherent)를 잘 참조해줄 수 있도록 하는 목적으로 진행되었다 [12]. 그림 3 은 영상들을 특정 그룹별로 나누어 프루닝 과정을 독립적으로 각각 따로 시행하는 개념도를 나타낸다.

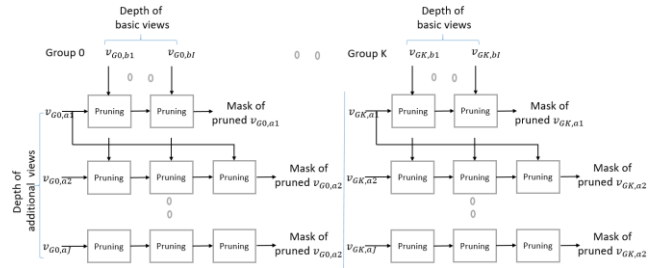


그림 4. Intel 이 제안한 Group-based TMIV 의 프루닝 과정 개념도. [12]

3. MCTS 기반 타일 그룹 분할

위 관련 연구에서 살펴보았듯이 독립적인 영역만을 전송하기 위해서 타일 그룹 간에 의존성이 있는 처리 과정이 있어서는 안된다. 따라서 사용자 시점 기반 3DoF+ 스트리밍 방법은 다른 영상을 참조하는 프루닝, 렌더링 과정에서 얻는 이득이 감소하게 된다.

또한 각 타일들의 크기와 개수, 위치 등 타일에 관한 메타 데이터들이 전송과정에서 함께 포함되어야 하며, 패치들에 대한 메타데이터도 독립적으로 다루어져야 한다.

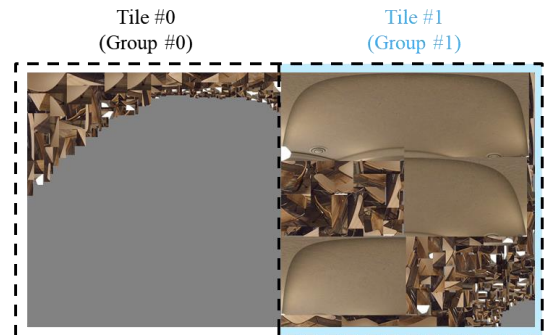


그림 5. 본 논문에서 제안하는 타일 그룹 분할 기반 패킹 구현 결과. (Tile #0 : v0, v3, v4, v5, v6, v12, v13, v14) ,

(Tile #1 : v1, v2, v7, v8, v9, v10, v11)

그러나 MCTS 기반의 타일 분할 방법을 사용하게 되면 사용자 시점에 따라 전송하게 될 영상의 대역폭을 크게 줄일 수 있게 되며 경우에 따라 보다 적은 수의 디코더의 개수가 필요한 장점을 지닌다. 그림 5 는 타일 그룹을 나누어 독립적으로 패치들을 할당한 결과이다. 서로 간의 어떠한 정보도 참조하고 있지 않기 때문에 기본 시점과 함께 사용자 시점에 따라 독립적으로 전송하여 최종 영상을 복원할 수 있게 된다.

특히 MCTS 기반의 타일 분할 방법은 사용자 시점의 시선을 계산하여 기본 시점 또한 타일로 나누어 이후 필요한 시점과 아틀라스 부분만을 추출(Extract)하고 병합(Merge)하여 필요한 환경에 맞추어 유연한 스트리밍을 가능하게 할 수 있다는 장점이 있다.

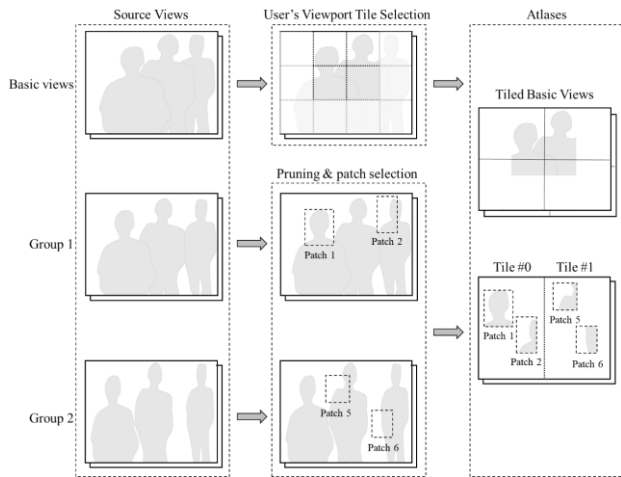


그림 6. 본 논문에서 제안하는 타일 그룹 분할 개념도.

4. 결론 및 향후 연구

현재 해당하는 제안 방법을 구현 및 실험 결과를 작성중이며, 3DoF+ 표준화에서는 보다 높은 압축 효율 또는 화질 개선 효과를 얻기 위하여 영상들 간의 정보를 참조하는 여러 기법들을 사용하고 있다. 독립적인 영역을 전송하는 사용자 시점 기반의 효율적인 스트리밍을 구현하기 위해서는 이러한 영상 간의 참조 정보들을 고려해야 하며, 그에 따른 트레이드오프(Trade-off)를 감안할 필요가 있다.

효율적인 3DoF+ 스트리밍 시스템을 위해 HEVC의 타일, 추후 VVC의 서브픽처(Sub-picture) 등 유연성(Versatility)을 위한 기술들을 적용하여 연구를 확장해나갈 계획이다.

Acknowledgement

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

참고 문헌

[1] “From OMAF for 3DoF VR to MPEG-I Media Format for 3DoF+, Windowed 6DoF and 6DoF VR.” 2017. 119th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG119/m44197.

[2] “Working Draft 1 of Metadata for Immersive Media (Video).” 2019. 126th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG126/w18464.

[3] P. Stefano, V. Swaminathan, M. Hosseini and F. D. Turck. “An HTTP/2-Based Adaptive Streaming Framework for 360° Virtual Reality Videos.” ACM Multimedia (2017).

[4] C. Ozcinar, A. De Abreu, and A. Smolic. 2017. Viewport-aware adaptive 360° video streaming using tiles for virtual reality. In 2017 IEEE International Conference on Image Processing (ICIP). 2174–2178.

[5] R. Skupin, Y. Sanchez, K. Suehring, T. Schierl, E. Ryu, and J. Son. 2017. Temporal MCTS Coding Constraints Implementation. 120th meeting of ISO/IECJTC1/SC29/WG11, MPEG 120/m41626.

[6] J. Son, D. Jang, and E. S. Ryu, “Implementing

Motion-Constrained Tile and Viewport Extraction for VR Streaming,” ACM Network and Operating System Support for Digital Audio and Video 2018 (NOSSDAV2018), Amsterdam, Netherlands, pp. 61–66, 2018.

[7] 장동민, 손장우, 정중범, 류은석, “사용자 시점 기반 360 영상을 위한 렌더러 구현”, 2018, 방송미디어공학회 제 23 권 6 호.

[8] S. Lee, D. Jang, J. Jeong, and E. S. Ryu, “Motion-constrained Tile Set based 360-Degree Video Streaming Using Saliency Map Prediction”, ACM Network and Operating System Support for Digital Audio and Video 2019 (NOSSDAV2019), Jun. 21, 2019.

[9] “Call for Proposals on 3DoF+ Visual.” 2019. 125th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG125/n18145.

[10] “Working Draft2 of Metadata for Immersive Video.” 2019. 127th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG127/n18576.

[11] “Test Model 2 for Immersive Video.” 2019. 127th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG127/n18577.

[12] “Group-Based TMIV.” 2019. 127th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG127/m49406.