

Saliency Map 을 이용한 360 도 비디오 타일 캐싱 기법

류영일^o

류은석

성균관대학교 컴퓨터교육과

Tiled 360-Degree Video Caching Method using Saliency Map

Yeongil Ryu^o

Eun-Seok Ryu

Department of Computer Education

Sungkyunkwan University

yeongilryu@skku.edu

esryu@skku.edu

요 약

최근 몰입형 360 도 비디오 기술은 멀티미디어 분야의 차세대 핵심 기술로서 주목받고 있다. 몰입형 360 도 영상은 콘텐츠 특성상 4K 및 8K 와 같은 매우 높은 품질의 해상도가 요구된다. 본 논문은 초고해상도의 몰입형 360 도 영상을 제한된 네트워크 대역폭에서 효율적으로 전송하기 위하여, 타일(Tile) 기반 360 도 영상 스트리밍 시스템을 위한 엣지 캐싱(Edge caching) 기법을 제안한다. 캐싱 시뮬레이션 실험을 통하여, 제안된 기법이 73%-83%의 캐시 적중률(Cache hit-rate)을 보임을 확인했으며, 타일 기반 360 도 비디오 스트리밍 시스템의 MTP 지연(Motion-to-Photon latency) 저감기술로써 활용될 수 있음을 확인하였다.

1. 서론

최근 가상현실(Virtual Reality, VR) 및 증강현실(Augmented Reality, AR) 어플리케이션들을 위한 몰입형 360 도 비디오 기술은 멀티미디어 분야의 차세대 핵심 기술로서 주목받고 있다. 이러한 흐름에 맞춰 국제 표준화 단체인 MPEG(Moving Picture Experts Group)은 2016 년부터 3DoF(3 Degrees of Freedom), 3DoF+, 6DoF 및 라이트필드 등의 몰입형 미디어를 위한 표준을 마련하기 위해, MPEG-I(Coded Representation of Immersive Media, ISO/IEC 23090) 프로젝트를 시작하였으며, 2021 년 완료 목표로 표준화 작업을 진행하고 있다. 이렇게 VR/AR 관련 기술이 고도화되면서, 이를 활용한 어플리케이션의 QoE(Quality of Experience) 향상 또한 중요한 연구주제로서 대두되고 있다. VR/AR 어플리케이션 QoE 에 직접적 영향을 미치는 요소들은 여러가지가 존재하지만, 사용자의 콘텐츠 몰입도 및 멀미감 등에 직접적 영향을 미치는 요소로서 영상 해상도(Resolution)과 MTP 지연이 있다.

몰입형 360 도 영상은 일반적으로 사용자의 눈과 가까이 위치하는 HMD(Head-Mounted Display)를 통하여 출력되며, 또한 360 도 공간의 정보를 표현하기 때문에 지금까지 주를 이루던 영상 콘텐츠들과 달리 4K 및 8K 와 같은 매우 높은 품질의 해상도가 요구된다. 이러한 초고해상도 영상을 제한된 네트워크 대역폭에서 효율적으로 전송하기 위하여, 영상을 여러 개의 타일로 분할해 360 도 영상 내에

서 사용자가 현재 바라보고 있는 영역의 비디오 데이터만을 스트리밍 할 수 있도록 한 타일 기반 360 도 영상 스트리밍 기법이 소개되었다 [1]. 하지만 타일 기반 360 도 비디오 스트리밍 시스템의 경우, 서버가 현재 프레임의 영상 전체를 전송하는 것이 아니기 때문에, 사용자가 현재 전송된 영역 바깥 쪽으로 시선방향을 이동시켰을 때 사용자의 움직임에 빠르게 대응하지 못하고 높은 MTP 지연을 발생시킬 수 있다. MTP 지연은 사용자의 움직임이 일어난 시점부터 그 움직임에 대응하는 영상이 디스플레이에 출력되기까지의 지연시간을 말한다.

본 논문은 타일 기반 360 도 비디오 스트리밍 시스템에서 발생하는 MTP 지연을 저감 시키기 위하여, 사용자가 영상 내에서 각 영역을 바라볼 확률을 픽셀(Pixel) 단위로 매핑(Mapping)한 Saliency map 을 활용해 사용자가 미래에 바라볼 영상 내의 영역을 예측하여 캐싱하는 기법을 소개한다.

2. Saliency Map 기반 360 도 영상 타일 캐싱

인간의 시각 시스템은 영상 내 모든 오브젝트에 균등하게 주의를 기울이는 것이 아니라, 특정 오브젝트에 더 높은 주의를 기울이게 되는 특성을 가지고 있다. Saliency map 은 이러한 인간의 시각 시스템 특성에 기반하여, 사용자가 영상 내에서 시각적 관심을 가질 가능성을 픽셀 단위로 매핑하여 나타낸다 [2].

본 논문이 제안하는 캐싱 기법은 앞에서 설명한

Saliency map의 특성, 그리고 사용자가 바라보는 영역의 비디오 데이터만을 전송하는 타일 기반 360도 비디오 스트리밍 기술에 기반한다. 그림 1은 타일링된 360도 영상을 엣지 캐시를 이용해 스트리밍하는 네트워크 구조를 보여준다. 이때, 제한된 캐시 저장용량에서 최대의 성능을 얻기 위하여, 본 논문은 여러 개의 타일로 분할된 영상의 Saliency map에 기반해 각 타일의 중요도를 판단한다. 그리고 엣지 캐시의 가용 저장용량에 따라 우선순위가 높은 타일들부터 순차적으로 캐싱한다. 이때, 타일의 우선순위는 각 타일에 포함된 픽셀들의 Saliency map 평균 값을 활용하여 부여한다. 즉, 제안된 기법은 각각 타일들이 포함하고 있는 픽셀들의 평균 Saliency map 값이 높을수록, 해당 타일에게 높은 캐싱 우선순위를 부여한다. 이를 통하여 결과적으로 향상된 캐시 적중률을 얻을 수 있도록 한다.

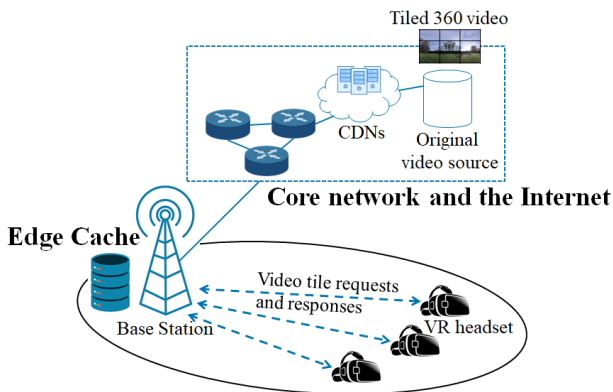


그림 1. 엣지 캐시를 활용한 네트워크 아키텍처

3. 실험 환경 구축 및 실험 결과

본 논문은 제안하는 캐싱 기법의 성능 및 Feasibility를 확인하기 위하여, 10개의 테스트 시퀀스에 대한 사용자들의 시선정보를 제공하는 데이터셋을 사용하였다 [3]. 이 데이터셋은 유튜브에 공개되어 있는 10개의 360도 비디오를 대상으로 각 영상의 모션정보 및 Saliency map을 제공한다. 또한, 각 테스트 시퀀스를 시청한 사용자 50명의 프레임별 시선방향(Yaw, Pitch, Roll) 및 사용자 뷰포트(Viewport)에 포함된 타일 번호 등을 CSV 데이터 포맷으로 제공한다. 여기서 각 타일들은 3840×1920 해상도의 영상을 200등분하여 192×192 크기를 갖는다.

제안하는 캐싱 기법의 Feasibility를 확인하기 위하여, 본 논문은 캐시 용량에 따른 캐시 적중률을 시뮬레이션 하여 실험을 수행하였다. 각 타일의 캐싱 우선순위를 결정하기 위하여, 데이터셋 테스트 시퀀스들의 타일 별 Saliency map 평균 값을 계산하였고, 이를 내림차순으로 정렬하여 가용한 캐시 용량만큼 순차적으로 캐싱 하였다. 가용 캐시 용량은 전체 비디오 데이터 량의 15%부터 50%까지 증가시켜 가며 실험을 진행하였다. 각 가용 캐시 용량에서의 캐시 적중률은 사용한 데이터셋에서 제공하는

50명 사용자의 각 프레임 별 뷰포트 내에 포함된 타일 리스트와 비교하여 계산하였다.

표 1은 4개 테스트 시퀀스의 캐싱 시뮬레이션 결과를 보여준다. 본 논문은 해당 실험을 통하여, 가용 캐시 용량이 전체 비디오 데이터 량의 50%일 때, 약 73%-83%의 캐시 적중률을 보임을 확인하였다. 그리고 이를 통하여 Saliency map을 활용한 캐싱 기법이 타일 기반 360도 비디오 스트리밍 시스템의 MTP 지연 저감기술로써 활용될 수 있음을 확인하였다.

표 1. 가용 캐시 용량 별 평균 캐시 적중률

Test sequence	Average cache hit-rate		
	Capacity 15%	Capacity 30%	Capacity 50%
Hog Rider	33.52	59.83	82.79
Pac-man	37.74	61.65	82.69
Peris Panel	28.39	50.55	74.98
Driving with	27.14	49.29	73.01

4. 결론

본 논문은 엣지 캐시를 활용한 네트워크에서의 타일 기반 360도 비디오 스트리밍 시스템을 위한 캐싱 기법을 제안한다. 제안된 캐싱 기법은 영상 내 각 타일의 평균 Saliency map 값을 구하고, 이를 기반으로 각 타일의 캐싱 우선순위를 부여한다. 그리고 가용 캐시 용량에 따라 우선순위에 기반해 순차적으로 타일들을 캐싱한다.

제안된 캐싱 기법을 시뮬레이션 한 실험을 통하여 가용 캐시 용량 50%에서 약 73%-83%의 캐시 적중률을 확인하였다. 향후 연구로는 본 논문의 실험에 타일 기반 비디오 스트리밍 시스템의 특성을 반영하여 더 높은 캐시 적중률을 보일 수 있도록 고도화해 나갈 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019R1A2C1010476).

5. 참고문헌

- [1] J. Son, D. Jang, and E. S. Ryu, "Implementing motion-constrained tile and viewport extraction for VR streaming," in *Proc. the 28th ACM SIGMM Workshop on NOSSDAV*, pp. 61-66, 2018.
- [2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE TPAMI*, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [3] W.-C. Lo, C.-L. Fan, J. Lee, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "360° Video Viewing Dataset in Head-Mounted Virtual Reality," in *Proc. ACM MMSys'17*, pp. 211-216, 2017.