

특집논문 (Special Paper)

방송공학회논문지 제24권 제6호, 2019년 11월 (JBE Vol. 24, No. 6, November 2019)

<https://doi.org/10.5909/JBE.2019.24.6.1002>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

360VR 콘텐츠의 음원위치정보를 활용한 시점예측 전송기법

정 은 영^{a)}, 김 동 호^{a)†}

Efficient Transmission Scheme with Viewport Prediction of 360VR Content using Sound Location Information

Eunyoung Jeong^{a)} and Dong Ho Kim^{a)†}

요 약

360VR 콘텐츠는 시청자의 시점변화에 따른 즉각적인 반응이 필요하고 고화질의 영상이 제공되어야 한다. 따라서 한정된 대역폭에서 360VR 시청자의 만족도를 보장하는 효율적인 전송기술이 필요하다. 그 일환으로 사용자의 시점을 예측하고 시점에 해당하는 영역과 해당하지 않는 영역에 다른 비트율을 할당하여 전체 대역폭 소모를 감소시키는 연구들이 소개되고 있다. 본 논문에서는 시점 예측의 정확도 향상을 목표로 기존 시각인지 정보만 활용했던 방식에 추가적으로 청각인지 정보인 360VR 콘텐츠의 음원위치정보를 활용한 시점 예측을 제안한다. 또한, 향상된 시점예측 방식을 이용하여 비트율을 효율적으로 할당함으로써 개선된 성능을 제공하는 전송 방식을 제안한다. 성능 분석 결과 제안한 시점 예측방식은 기존 방식 대비 시점 예측의 정확도가 향상되었으며, 이를 바탕으로 제안한 전송 방식은 제한된 대역폭 내에서 사용자의 시점에 해당하는 타일에 고품질의 영상을 제공할 수 있음을 확인하였다.

Abstract

360VR content requires short latency, such as immediate response to viewers' viewport changes and high quality video delivery. It is necessary to consider efficient transmission that guarantees the QoE(Quality of Experience) of the 360VR contents with limited bandwidth. Several research has been introduced to reduce overall bandwidth consumption by predicting a user's viewport and allocating different bit rates to the area corresponding to the viewport. In this paper, we propose novel viewport prediction scheme that uses sound source location information of 360VR contents as auditory recognition information along with visual recognition information. Also, we propose efficient transmission algorithm by allocating a bit rate properly based on improved viewport prediction. The proposed scheme improves the accuracy of the viewport prediction and provides high quality videos to tiles corresponding to the user's viewpoint within the limited bandwidth.

Keyword : 360 Virtual Reality, 360 Video Transmission, Predictive Transmission, Viewport Prediction, MPEG-DASH

a) 서울과학기술대학교 IT미디어공학과(Seoul National University of Science and Technology, Dept. of IT Media Engineering)

† Corresponding Author : 김동호(Dong Ho Kim)

E-mail: dongho.kim@seoultech.ac.kr

Tel: +82-2-970-6417

ORCID: <https://orcid.org/0000-0001-9136-8932/>

· Manuscript received September 26, 2019; Revised November 7, 2019; Accepted November 7, 2019.

I. 서론

360VR(Virtual Reality) 콘텐츠는 일반적인 영상에 비해 많은 정보를 표출할 수 있으며 높은 몰입감을 제공할 수 있다. 유튜브(YouTube)에서 실시한 360VR 콘텐츠 관련 조사에 따르면 360VR 콘텐츠는 선형적인 일반 영상에 비해 상호작용은 41%, 전체 분량 시청률은 46%, 클릭률은 5배 더 높으며 반복 시청 횟수 및 공유 비율도 높은 것으로 나타났다^[1]. 이러한 장점 때문에 오락/엔터테인먼트, 서비스업, 교육/미디어, 산업 등 다양한 분야에서 360VR 콘텐츠를 활용하고 있다^[2].

최근에는 통신기술의 발전 및 모바일 기기의 성능향상으로 모바일 네트워크를 통한 360VR 콘텐츠 소비가 증가하고 있는 추세이다^[3]. 360VR 콘텐츠는 일반 영상에 비해 용량이 크고, 화질에 따라 사용자의 몰입도가 좌우되며 시청자의 시점(viewport)변화에 따른 즉각적인 반응을 요구하는 특성을 갖는다. 따라서 한정된 대역폭을 갖는 통신시스템에서 360VR 콘텐츠를 전송할 때 사용자의 만족도를 보장할 수 있는 효율적인 전송기술 연구가 필요하다.

360VR 콘텐츠를 시청할 때 시청자가 바라보는 창, 즉 시점(viewport)은 전체 360VR 영상의 일부분에 불과하다. 하지만 대개 360°로 구성된 공간 전체의 영상이 전송되어야 하기 때문에 실제로 시청되지 않는 영역을 전송하여 불필요한 요구대역폭이 발생한다. 이러한 문제를 해결하고자 360VR 콘텐츠의 효율적 전송 방안에 관한 여러 연구가 진행되어 왔으며, 시청자의 시점 예측 연구가 대표적이다. 대표적인 시점 예측 연구는 시청자의 관심정보(ROI, Region Of Interest)를 기반으로 시점 예측을 수행하고 시점에 해당하는 영역과 해당하지 않는 영역에 다른 비트율을 할당하여 요구대역폭을 낮추거나 한정된 대역폭 내에서 시청자의 사용경험 성능(QoE, Quality Of Experience)을 개선하는 연구이다^[4]. 그러나 시청자의 시점 예측에 실패할 시에는 시청자가 낮은 비트율이 할당된 영역을 불가피하게 시청하게 되므로 시청자의 경험 만족도 보장이 저하되는 한계가 있다.

본 논문에서는 이러한 한계를 극복하기 위하여 기존의 시각인지(visual recognition) 정보와 더불어 청각인지(auditory recognition) 정보인 360VR 콘텐츠의 음원위치 정보

를 활용하는 시점 예측 방식을 제안한다. 또한 개선된 시점 예측 결과와 현재 시청자의 시점 등을 고려하여 타일별로 적절한 비트율을 할당함으로써 개선된 전송성능을 제공하는 전송 방식을 제안한다.

II. 배경 연구

이번 장에서는 배경 연구로 멀티미디어 전송기술인 MPEG-DASH(Dynamic Adaptive Streaming over HTTP)와 웹 기반 전송환경을 구현하는데 활용한 WebVR API(Application Programming Interface)에 대해 설명한다.

1. MPEG-DASH

MPEG-DASH 방식은 HTTP(Hypertext Transfer Protocol)기반의 적응적 스트리밍 방식으로 서버와 클라이언트의 네트워크 환경과 수신 버퍼의 상황에 따라 다른 비트율을 할당하여 전송함으로써 끊김 없는 서비스 품질을 제공한다. DASH 기술에서 HTTP 서버는 미디어 스트림을 하나 혹은 다수의 구간(Period)으로 나누고 각 구간(period)을 구성하는 영상, 음향, 자막 등의 적응세트(adaptation set)별로 해상도, 음질, 자막 종류 등의 품질을 달리 제공할 수 있다. 대표적인 적응세트인 비디오 혹은 오디오의 품질은 비트율 표현(representation)을 포함한다. 표현은 하나 혹은 그 이상의 세그먼트(segment)로 구성되며 HTTP 프로토콜을 사용하여 각 세그먼트를 다운로드 할 수 있는 서버 내의 주소 정보가 XML(Extensible Markup Language)로 기술되어 있다^[5]. MPD(Media Presentation Description)의 계층적 모델은 그림 1과 같다.

HTTP 서버와 DASH 클라이언트 간 통신은 HTTP 프로토콜을 이용하며 클라이언트의 요청과 서버의 응답으로 이루어진다. 클라이언트는 우선 멀티미디어 파일의 MPD 파일을 요청하고 이를 다운로드 받아 파싱하여 서버 내 세그먼트의 정보를 확인한다. MPD 파일의 정보를 통해 클라이언트의 네트워크 상황 및 재생 버퍼 요구 조건에 최적인 표현 세트를 선택한다. 이후 클라이언트는 타임 라인을 만들고 적절한 미디어 세그먼트를 요청하고 다운로드 받아

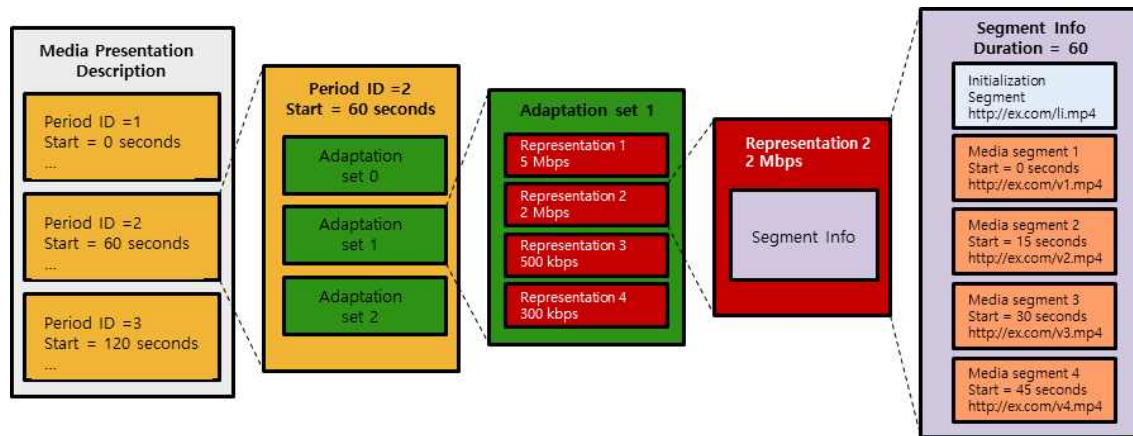


그림 1. MPD(Media Presentation Description)의 계층적 모델

Fig. 1. Hierarchical Model of MPD(Media Presentation Description)

재생한다. 재생이 종료될 때까지 계속해서 클라이언트의 네트워크 상황을 확인하고 그를 기반으로 적응적 표현 세트를 선택하고 위의 과정을 반복한다.

2. WebVR API

HTML5 기반의 웹 브라우저를 통해 몰입형 360VR 콘텐츠를 제공하기 위하여 3D 렌더링 자바스크립트 API인 WebVR API를 활용하였다. WebVR API는 2014년 Mozilla 팀에서 고안되었으며 2016년 Mozilla VR 팀과 구글 크롬 팀이 WebVR API ver 1.0 release를 발표하고 이를

관리하고 있다^[6]. HTML5의 특징에 맞게 별도의 응용프로그램을 플러그인 형태로 설치하지 않고도 사용자가 360VR 콘텐츠를 시청할 수 있는 환경을 제공한다. 본 논문에서는 WebVR API의 프레임워크인 A-FRAME를 이용하여 360VR 콘텐츠 환경을 구성하였다.

III. 본 론

이번 장에서는 360VR 콘텐츠의 음원을 활용한 시점 예측 기법을 제안하고 이를 바탕으로 설계한 효율적 전송 시

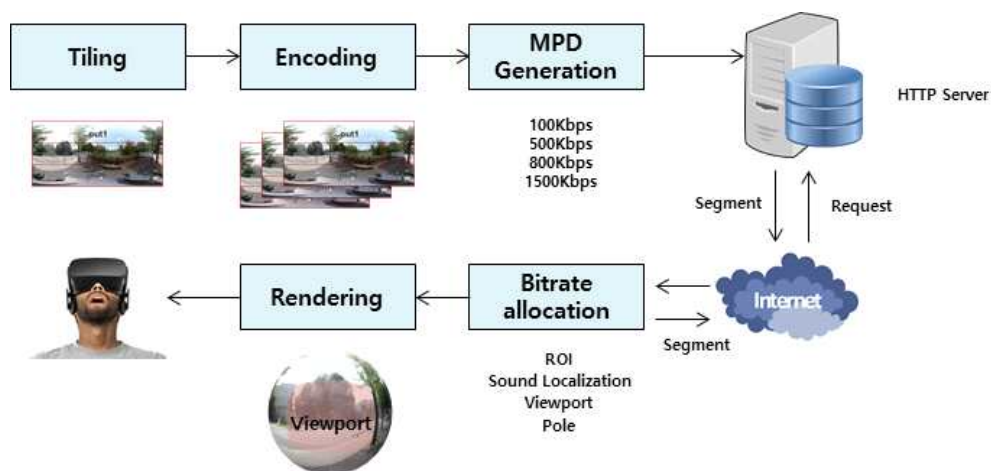


그림 2. 제안하는 시점예측 및 전송기법의 시스템모델

Fig. 2. System model of proposed viewport prediction and transmission scheme

스텝 및 웹 기반 구현 방법에 대해 기술한다. 그림 2는 제안하는 시점예측 및 전송 기법의 구성 블록을 나타낸다. 먼저 영상 타일별로 부분적 비트율을 할당하기 위해 하나의 ERP 포맷 영상을 6개의 타일로 분할한 후 각 타일을 여러 품질의 영상으로 인코딩한다. MPEG-DASH 기반으로 전송하기 위해 해당 타일의 세그먼트의 정보를 MPD로 생성하여 미디어 서버에 저장한 후 시청자의 요구에 따라 관련 정보를 제공한다. 제안 방법은 기존 방법에서 고려하는 시청자 관심정보(ROI)와 더불어 음원 위치정보, 현재 시청자의 시점, 극/적도 영역 여부 등을 추가적으로 고려하여 비트율을 차등 할당한다. 이 후 시청자의 가용 대역폭 상황에 따라 영상 타일별로 영상품질을 다르게 선택하여 전송한다.

1. 부분적 비트율 할당을 위한 영상 분할 및 타일링 기법

1.1 부분적 비트율 할당을 위한 영상 분할

타일의 위치에 따라 다른 비트율을 할당하기 위해 먼저 실험영상을 바탕으로 사용자의 시청 빈도수를 실험하였다. 실험에 사용된 영상은 대체로 조용한 교내 호수 주변에서 약 30초 길이로 촬영된 영상이며, 호수를 중심으로 자동차가 지나가는 상황을 표현한다. 주변은 대체로 조용한 환경

이며 자동차가 지나갈 때 자동차의 엔진 소리에 따라 음원 위치 정보가 이동하게 된다. 실험영상은 그림 3과 같이 두 개의 극 영역과 네 개의 적도 영역으로 분할하였다^[7]. 이러한 분할은 CMP(Cube Map Projection)와 비교하여 극 영역을 축소시키고 적도 영역을 세분화할 시, 불필요한 전송 대역폭을 줄여 효율적인 전송을 할 수 있는 확장성을 가진다. 극 영역과 적도 영역의 시청 빈도수를 확인하기 위하여 시청자의 피치각이 60° 이상 혹은 120° 이하인 경우를 적도, 그 외의 $0^\circ \sim 60^\circ$ 또는 $120^\circ \sim 180^\circ$ 경우를 극 지역으로 기록하도록 정의하였다. 30초 가량의 영상에 대해 시청자의 시청 빈도를 1초당 50회씩 측정하였으며, 각각 10 명의 참여자를 통해 실험하여 시청 빈도수를 평균 낸 결과는 그림 4와 같다. 사전 실험결과에 따르면 특별한 ROI나 음원 정보가 극 지역에 위치한 경우를 제외하고는 적도 영역에 비해 극 영역의 시청 빈도수가 대체로 낮다는 것을 확인할 수 있다. 가용대역폭이 제한된 상황에서 적도 영역과 극 영역의 타일에 다른 비트율을 할당해야 한다면 극 영역에 상대적으로 낮은 비트율을 할당하는 것이 바람직하다는 것을 의미한다. 제안 방법에서는 각 영상의 코덱을 웹 기반 재생이 가능한 H.264으로 변환하고 {100, 500, 800, 1500} Kbps의 해상도로 인코딩한 후 3 초 단위의 세그먼트로 나누고 이를 기반으로 생성한 MPD와 함께 HTTP 서버에 저장한다.

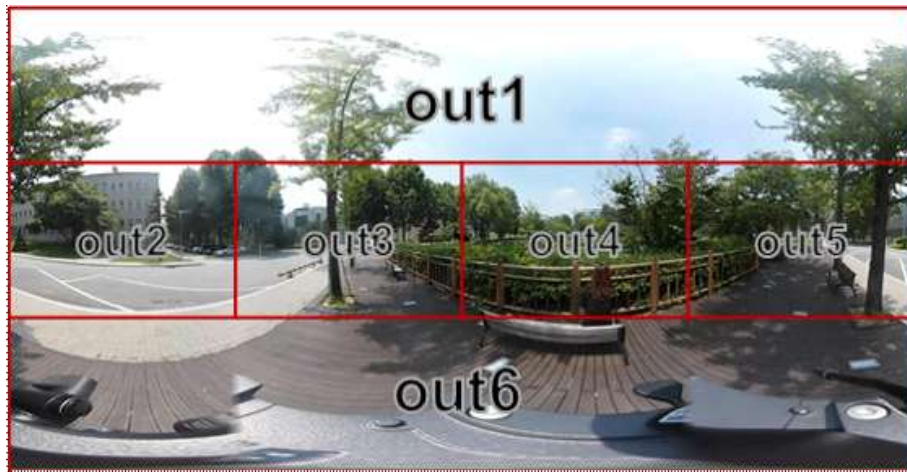


그림 3. 부분적 비트율 할당을 위한 영상 분할
Fig. 3. Cropping the ERP video for partial bit-rate allocation

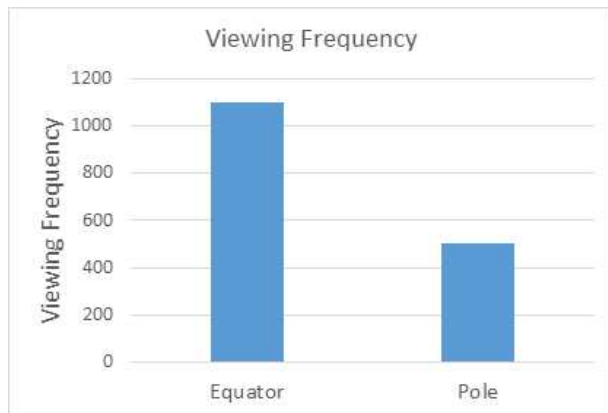


그림 4. 시청 빈도수 실험 결과 (Equator vs Pole)
Fig. 4. Experiment Result of viewing frequency

1.2 시각인지 및 청각인지정보 기반의 영상 타일 세트 지정
제안한 방법에서는 시점 예측의 정확도를 향상하는 목적으로 시각인지정보(visual cognitive information)인 사용자 관심정보(ROI)와 청각인지정보(auditory cognitive information)인 음원 위치정보를 함께 고려한다. 또한 사용자의 현재 시점과 극 영역 타일여부를 함께 고려하여 비트율 할당을 수행함으로써 소음 대역폭감소를 도모한다. 각 고려사항 별 영상 타일 세트 지정에 대해 순차적으로 기술한다.

360VR 콘텐츠 내에서 사용자가 관심을 갖는 ROI 객체는 대화하는 두 명의 사람으로 가정하였으며 ROI 객체가 위치한 영상 타일 세트를 `tile_r` 로 지정하였다. 360VR 콘텐츠 내 ROI 객체 검출은 keras 모델로 구현된 YOLO(You Only Look Once) 알고리즘을 활용하였다. 시점 예측의 정확도를 향상시키기 위해 360VR 콘텐츠 내 음원 위치정보를 추가적으로 고려하였다. 이는 시청자가 ROI 객체가 있을 때 시선을 해당 객체에 설정할 확률적 가능성이 높다는 점과 유사하게 음원이 존재할 때 시선을 해당 위치로 변경 또는 설정할 확률적 가능성이 높다는 점에 착안한 것이다. 특히 특정 객체를 바라보는 상황에서 다른 각도에 소리가 나는 특별한 이벤트가 발생할 경우 특히 그 소리의 절대 크기 또는 소리의 변화값 등이 클 때 해당 방향으로 시점을 옮길 가능성이 높을 것으로 예상된다.

제안 방법의 동작구현을 위해 음원이 위치한 영상 타일 세트를 `tile_s` 로 지정하였다. 시청자는 360°공간 내 음원의 위치를 3차원 좌표값 (x, y, z)로 전달받고, 자신의 방위각 및

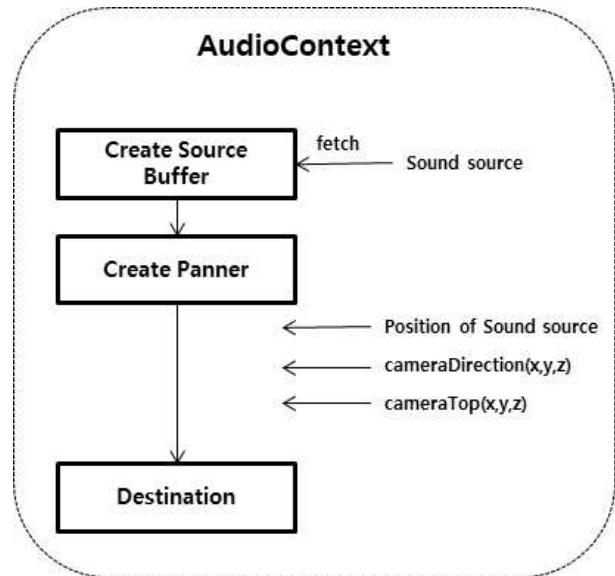


그림 5. 웹 기반 입체음향 구현 동작 흐름
Fig. 5. Web-based stereophonic implementation flow

고도각 정보를 HRTF(Head Related Transfer Function)에 입력하여 입체음향을 제공받는다. 본 논문에서는 WebAudio API를 활용하여 웹 기반으로 입체음향을 구현하였고 해당 동작과정을 그림 5에 나타내었다. WebAudio API는 AudioContext 객체를 생성하면서 시작되며, 음원의 볼륨 크기를 조절하는 GainNode, 음원에 페닝 효과를 적용하고 조절하는 PannerNode 등 여러 모듈로 구성되어 있다. 음원을 바탕으로 버퍼에 저장한 후 PannerNode 모듈을 이용하여 음원의 위치와 시청자의 위치 및 방향 등을 이용하여 공간음향을 구현하였다.

360VR 콘텐츠의 시점 예측에 실패할 경우, 서버가 예측한 다음 시점으로 시청자가 실제로는 시점을 변경하지 않고 현 시점을 그대로 유지하므로 시청경험 만족도가 저하되는 문제점이 있다. 제안하는 방법에서는 이런 상황에서도 시청자의 경험 만족도를 보장하기 위해 시청자가 현재 바라보고 있는 시점에 해당하는 타일에도 상대적으로 높은 비트율을 할당함으로써 서버가 시점 예측에 실패하더라도 시청품질의 저하가 없게 유지할 수 있다. 제안 방법의 동작 구현을 검증하기 위해 시청자의 현 시점에 해당하는 영상 타일 세트를 `tile_v` 로 지정하였다.

시청자의 시점(viewport)이 어느 영상 타일을 바라보고

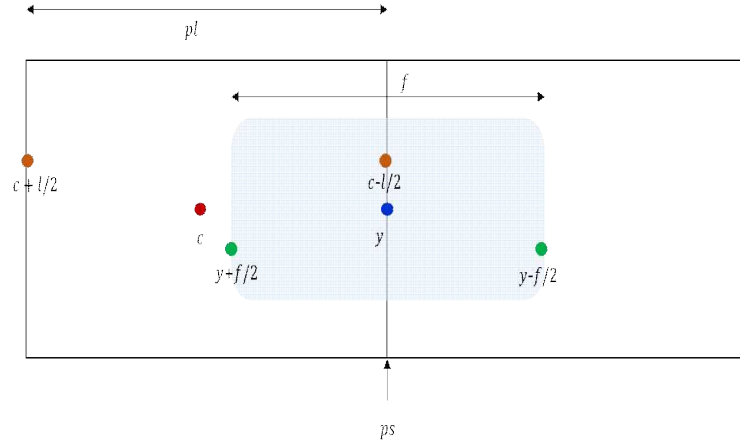


그림 6. 시청자 현재 시점에 해당하는 영상 타일정보 획득 예시

Fig. 6. Example of acquiring tile information corresponding to user's current viewport

있는지 인식하기 위해 영상 타일의 지오메트리(i.e. phi-start, phi-length) 정보와 시청자의 오리엔테이션(i.e. Yaw) 정보를 활용한다. 우선 시청자의 피치각이 $0^\circ \sim 60^\circ$ 이거나 $120^\circ \sim 180^\circ$ 인 경우 각각 극 영역에 위치한 타일($i=1$ or 6)을 바라보고 있다고 판단하며, 피치각이 $60^\circ \sim 120^\circ$ 인 경우 시청자가 적도 영역에 위치한 타일($i \in \{2,3,4,5\}$)을 바라보고 있다고 판단한다. 적도 영역의 타일($i \in \{2,3,4,5\}$) 중 어느 타일을 시청자가 바라보는지는 시청자의 요각(Yaw)을 이용하여 인식할 수 있다. 그림 6에 나타난 바와 같이 두 개의 영상 타일이 있다고 할 때 시청자의 시점(viewport)을 나타내는 하늘색 박스의 가운데 포인트는 현재 시청자의 요각(Yaw)을 나타내며 빨간색 점은 각 영상 타일의 중앙점을 나타낸다. 각 영상 타일의 가로길이는 90° 를 가정하였고, 시청자 시점의 FoV(Field of View)는 80° 를 가정하였다. i 번째 타일의 중앙점 c_i 는 식 (1)로 표현할 수 있다. 이때, ps_i 는 i 번째 영상 타일의 phi-start, pl_i 는 i 번째 영상 타일의 phi-length를 의미한다.

$$c_i = ps_i + \frac{pl_i}{2}, \text{ s.t. } i \in \{2,3,4,5\} \quad (1)$$

시청자의 요각(Yaw)을 y 라 하고, FoV를 f 라 할 때, 시청자의 시선이 바라보는 영상타일 세트를 다음 식 (2)로 판단할 수 있으며 해당 i 번째 영상 타일세트(tile_v)로 지정하였다.

$$|y - c_i| \leq \frac{f}{2} - \frac{pl_i}{2} \rightarrow \text{viewport} \in i\text{-th Tile } i \in \{2,3,4,5\} \quad (2)$$

특히 ROI 또는 음원정보가 두 개 이상의 영상타일 세트에 걸쳐질 경우, 즉 시청자의 시점이 두 개의 영상타일 세트에 걸쳐질 경우 해당 타일 모두를 tile_r 또는 tile_s로 지정하였다.

마지막으로 앞 실험에서 확인한 바와 같이 극 영역($i \in \{1,6\}$)의 시청 빈도수가 적도 영역에 비해 대체로 낮은 점을 비트율 할당 시에 고려하기 위해 극 영역에 해당하는 영상 타일 세트를 tile_p로 지정하였다. 가용 대역폭이 작은 경우에는 tile_p에 해당하는 영상 타일 세트에 낮은 가중치를 할당한다.

2. 클라이언트의 가용대역폭에 따른 타일 세트 별 비트율 할당

본 논문에서는 클라이언트의 가용대역폭 상황에 따라 영상 타일 세트 tile_r, tile_s, tile_v, tile_p에 대해 차별화된 비트율을 할당함으로써 360VR 콘텐츠를 효율적으로 전송하는 방법을 제안한다. 예를 들어, 가용대역폭이 여유 있는 상황(i.e. 50Mbps 이상)이라면 ROI 객체가 들어있는 tile_r, 음원이 위치한 tile_s, 현재 시점을 나타내는 tile_v를 포함하여 극 영역 tile_p까지 모두 높은 비트율(1500kbps)을 할당한다. 반면에 가용대역폭이 열악한 상황(i.e. 5Mbps 이

표 1. 제안 방식의 가용대역폭에 따른 구분 기준

Table 1. Classification according to available bandwidth

BW condition	tile_r	tile_s	tile_v	tile_p	other tiles	available BW
Fairly good	r_3	r_3	r_3	r_3	r_0	$\geq 50\text{Mbps}$
Good	r_3	r_3	r_3	r_2	r_0	$\geq 25\text{Mbps}$
Bad	r_3	r_3	r_2	r_1	r_0	$\geq 10\text{Mbps}$
Very bad	r_3	r_3	r_2	r_0	r_0	$\leq 5\text{Mbps}$

하)이라면 ROI 객체가 들어있는 tile_r, 음원이 위치한 tile_s에 해당하는 영상 타일만을 높은 비트율로 할당하고, 현재 시점을 나타내는 tile_v는 중간 비트율, 극지방을 포함하여 나머지 다른 타일은 최소 비트율(100kbps)만을 할당하는 방법을 적용하였다. 제안한 방법에서 사용한 할당 비트율은 $\{r_0, r_1, r_2, r_3\} = \{100, 500, 800, 1500\} \text{kbps}$ 이며 가용대역폭에 따른 비트율 할당 방법을 표 1에 제시하였다. 클라이언트의 가용 대역폭은 기준 크기를 갖는 이미지를 다운로드 하는 데 걸린 총 시간으로 나누어 계산하고 이를 기준으로 대역폭 상황을 추정하였다. 위의 가용 대역폭 추정 및 비트율 할당과정을 일정 주기 간격으로 반복하여 대역폭 변화에 적응적으로 대응한다.

그림 7은 제안 방식의 영상 타일 세트 별 비트율 할당 알고리즘을 일반화한 동작 흐름도이다. 일반화된 비트율 할당방법에서는 먼저 클라이언트의 가용대역폭에 따라 채

널 상황을 구분하고, 이에 따라 영상 타일 세트 별로 가중치 α_i 를 지정한 후 기준비트율 r_0 에 가중치를 두어 비트율 $r_i = \alpha_i r_0$ 을 할당한다. 그리고 타일 별로 할당된 비트율의 전체 합 과 가용대역폭을 비교하여 전체 합이 가용대역폭 이내인 것을 확인한 후 다운로드를 요청한다.

$$R = \sum_{i=1}^6 \alpha_i r_0 \leq BW_{available} \quad (3)$$

만약 비트율의 전체 합 R 이 가용대역폭을 초과하는 경우 각 타일별로 우선순위를 지정하여 낮은 우선순위를 갖는 타일의 가중치 α_i 를 먼저 낮춘 후 식 (3)의 만족 여부 확인을 반복한다. 타일별 우선 순위는 ROI와 음원위치가 존재하는 tile_r과 tile_s가 가장 우선순위가 높으며, tile_v, tile_p 그리고 이외의 타일 순으로 설정하였다. 위의 과정을 일정 주기 별로 반복하여 대역폭이 가변적인 상황 혹은 필요 대역폭이 가용 대역폭보다 큰 경우에 대처한다.

3. 웹 기반 360°렌더링

웹 기반 360°렌더링 재생을 위해 WebVR의 프레임워크인 A-FRAME의 <a-sky>태그를 활용하여 구현 및 실험하였다. 영상 별 지오메트리에 사용된 변수의 정의를 표 2에 나타내었으며 각 영상 별로 지오메트리(phi-start, phi-length,

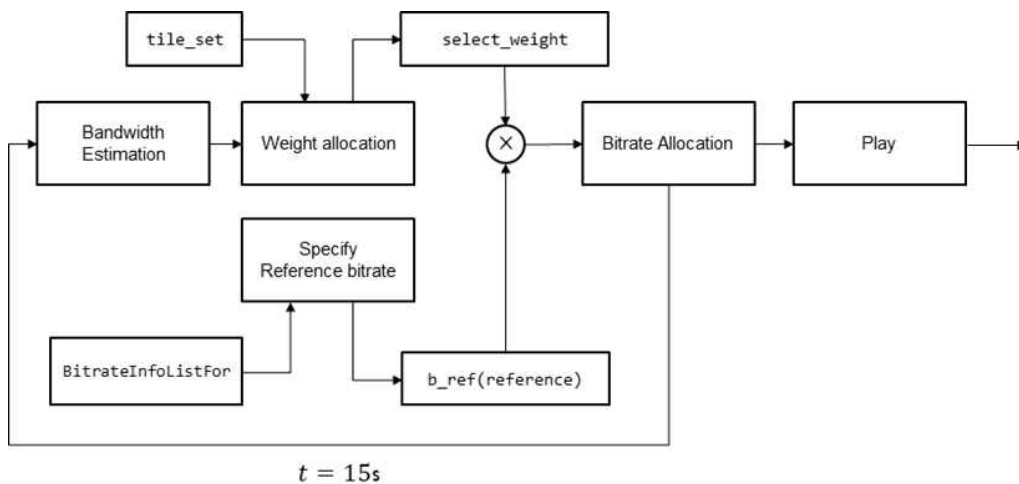


그림 7. 영상 타일 세트 별 비트율 할당 알고리즘의 동작흐름

Fig. 7. Work flow of bitrate allocation algorithms per tile set

표 2. 영상 별 지오메트리
Table 2. Geometry of each video

Variable	Definition of Variable
phi-start	x-axis starting point in spherical coordinates of the image
phi-length	x-axis length in spherical coordinates of the image
theta-start	y-axis starting point in spherical coordinates of the image
theta-length	y-axis length in spherical coordinates of the image
src	File source of the image

theta-start, theta-length) 값을 지정하고 MPD를 이용하여 파일 경로 변수인 src를 연결하여 구현하였다. 그림 8에 여섯 개의 영상을 MPEG-DASH 기반으로 3차원 렌더링한 결과를 나타내었다.

IV. 실험 결과 및 성능 분석

본 논문에서 제안한 방식의 성능 검증을 위해 음원 위치에 따른 시점 예측의 정확도와 전송 효율성 개선 여부를 확인하였다.

1. 음원 위치에 따른 시청자 간 머리 방향 차이의 평균 성능

본 논문에서는 360VR 콘텐츠의 음원위치정보가 시청자 머리 방향에 영향을 주는지 여부를 확인하기 위해 머리 방향 차이의 평균(ADHD: Average Difference of Head Direction)을 성능으로 제시하였다. 실험 환경 시나리오는 화면 내 움직이는 물체의 개수, 즉 시청자가 관심을 가질만한 정보(ROI)의 개수가 매우 적은 경우(건물 앞), 적은 경우(교내 호수), 매우 많은 경우(교내 운동장)로 준비하였다. 각 시나리오의 시퀀스는 ROI가 있으나 입체음향이 구현되지 않은 경우(ROI with no spatial sound), ROI가 있고 입체음향이 구현된 경우(ROI with spatial sound)로 구성하였으며 시나리오에 대해 시퀀스 별 머리 방향 일치도 비교를 진행한 결과는 그림 9와 같다.

머리 방향의 평균차이 성능은 식 (4)와 같이 시청자의 요(Yaw), 피치(Pitch), 롤(Roll) 측정데이터를 기준영상과 비교하여 차이값의 평균으로 도출하였다. 이 값이 작을수록 시청자 간 머리 방향의 일치도가 높으며 시점 예측의 정확도가 높은 것으로 해석할 수 있다.

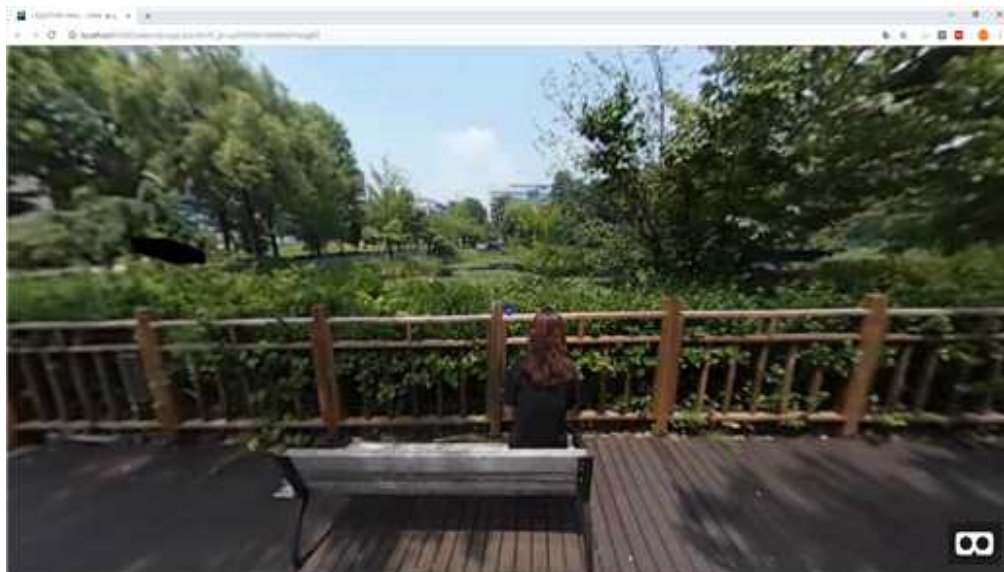


그림 8. 분할 영상의 웹 기반 3차원 렌더링 결과
Fig. 8. Web-based 3D rendering of cropped videos

$$ADHD = \frac{1}{N} \left(\sum_{n=1}^N \sum_{k=1}^3 |u_n^k - u_{ref}^k| \right) \quad (4)$$

그림 9에 나타난 바와 같이 시청자 관심정보가 존재하는 동일한 시퀀스에 대해 음원의 위치에 따른 입체음향 구현 유무에 따라 기준 데이터와 시청자 데이터 간 차이의 평균을 비교한 결과, 세 실험환경 시나리오 모두 음원의 위치에 따른 입체 음향이 구현되었을 때 시청자의 머리 방향이 기준 값에 가까워진다는 것을 확인하였다. 특히 ROI의 개수가 작은 경우인 (a)와 (b)의 경우 음원 위치에 따라 시청자의 시점이 기준영상과 유사하게 된

다는 점을 확인할 수 있었다. 이는 시청자가 시각적으로 집중할 가능성이 적은 경우에는 청각인지정보에 상대적으로 영향을 많이 받게 됨을 의미한다. 따라서 시점 예측 시 360VR 콘텐츠 내 음원의 위치정보를 추가적으로 활용한 시점예측은 기존의 시청자 관심정보 기반의 시점예측보다 예측 정확도를 향상시킬 수 있다는 점을 확인하였다.

2. 총 전송 데이터양 및 시점 내 영상 타일의 비트율

동일한 가용대역폭(5Mbps) 환경에서 제안한 방식이 기

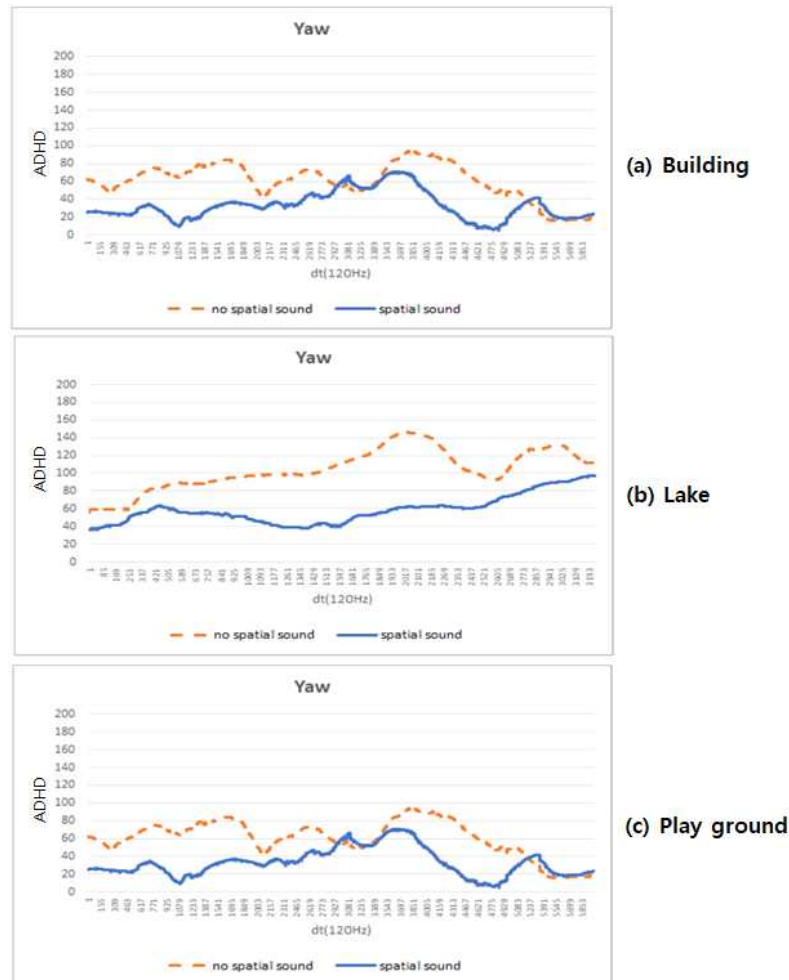


그림 9. 시청자 간 머리방향 차이의 평균 실험 결과

Fig. 9. Experiment result of average difference of head direction

존 방식과 비교하여 개선된 영상품질을 제공할 수 있다는 점을 확인하기 위하여 성능지표로 시청자 시점에 할당된 비트율과 총 전송 데이터율을 비교하였다. 기존 방식과의 성능 비교를 위해 MPEG-DASH 알고리즘 전송기법 (6-DASH)^[5], 사용자 시점 인지 전송기법^[7]을 고려하였고 제안한 전송 방법과 성능비교를 수행하였다. 기존 6-DASH 방식^[6]은 6개의 타일에 동일하게 800kbps의 비트율을 할당하는 방식이며, 사용자 시점 인지 전송방식^[7]은 사용자의 시점을 ROI 기반으로 인지하고 해당 타일에 높은 비트율(1500kbps)를 할당하고 나머지 타일에 상대적으로 낮은 비트율을 할당하는 방식이다. 성능 비교를 위한 실험 결과를 그림 10 및 표 3에 나타내었다.

표 3. 식 (3)의 인덱스

Table 3. Index of equation (3)

Variable	Definition of Variable
k	(Yaw, Pitch, Roll) data index
u_n^k	k -th data of n -th user
N	number of total users of experiment

성능 비교 결과 제안한 방식은 기존 MPEG-DASH 알고리즘 전송에 비해 시청자가 바라보는 시점(viewport)에 높은 비트율(1500kbps)이 할당되어 대체로 높은 화질의 영상을 제공받을 수 있다는 것을 확인하였다. 또한, 가용대역폭 내에서 소모되는 총 전송데이터 전송률이 6-DASH 방식의 4.64Mbps보다 3.26Mbps로 감소하여 가용대역폭을 효율적으로 사용할 수 있음을 확인하였다. 또한, 제안한 방식은 사용자 시점 인지 전송방식^[7]과 비교하더라도 시점 내 할당된 비트율은 1500kbps로 유사한 화질을 제공할 수 있으며 총 전송데이터 전송률을 3.87Mbps에서 3.26Mbps로 감소시켜 가용대역폭을 효율적으로 사용할 수 있음을 확인하였다.

표 4. 총 전송 데이터양 및 시점 내 영상 타일 비트율

Table 4. Total amount of transmitted data and bitrate of tiles in viewport

Scheme	Available BW	BW Consumption	bitrate of tiles in viewport
6-DASH	5Mbps	4.64Mbps	800Kbps
Viewport-Aware	5Mbps	3.87Mbps	1500Kbps
Proposed	5Mbps	3.26Mbps	1500Kbps

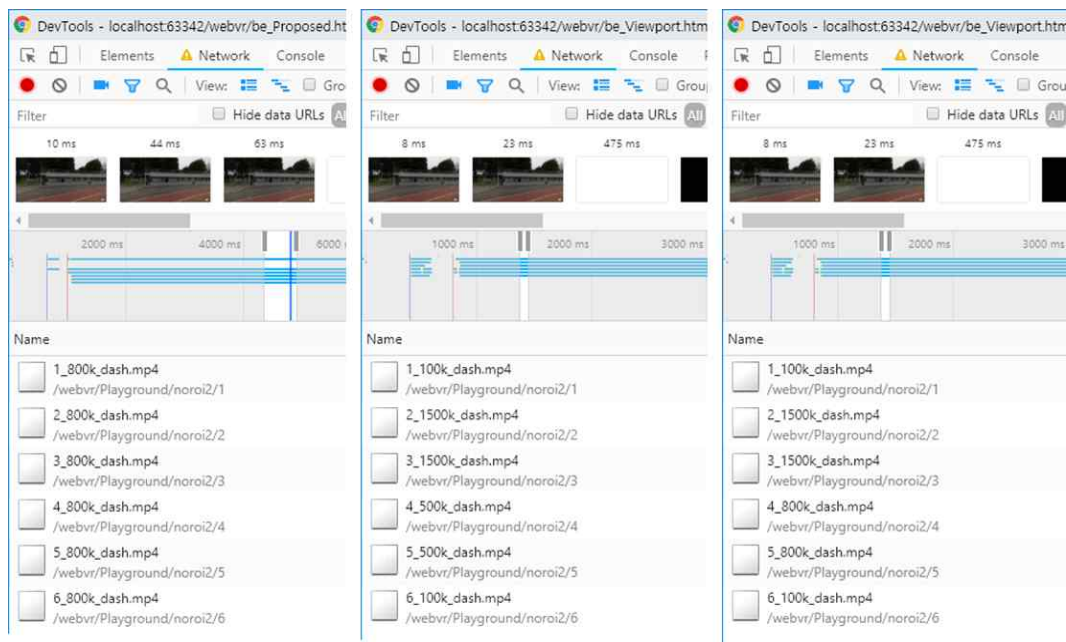


그림 10. 총 전송 데이터양 및 시점 내 영상 타일 비트율

Fig. 10. Total amount of transmitted data and bitrate of tiles in viewport

V. 결 론

본 논문에서는 360VR 콘텐츠에서 음원 위치정보를 기반으로 시점예측을 수행하는 방식과 이를 활용한 효율적인 전송기법을 제안하였다. 본 연구에서는 시각인지정보를 바탕으로 시점을 예측하는 기존 방식에 추가적으로 청각인지정보를 활용하는 새로운 방안을 제시하였으며 기존 시점과의 차이 성능을 비교하여 기존 방식 대비 시점 예측의 정확도가 향상되었음을 확인하였다. 특히 사용자가 시각적으로 집중할 가능성이 적은 경우에는 청각인지정보인 음원의 위치정보에 상대적으로 영향을 많이 받게 된다는 점도 확인하였다.

또한 본 연구에서는 향상된 시점 예측방식과 더불어 예측에 실패할 경우 사용자의 시청경험 만족도 보장을 위해 사용자의 현재 시점 정보를 고려하고 극 영역과 적도 영역을 차별화하여 비트율을 할당하는 방식을 제안하였다. 360VR 콘텐츠 전송 시 제안한 방식을 통해 시점 예측의 정확도가 개선되므로 보다 정확히 비트율을 할당할 수 있게 되어 소모되는 총 전송데이터 전송률을 줄일 수 있어 효율적인 전송이 가능하게 됨을 확인하였다.

참 고 문 헌 (References)

- [1] Jim Habig, "Is 360 video worth it?," July 2016 [Online]. Available: <https://www.thinkwithgoogle.com/advertising-channels/video/360-video-advertising/>
- [2] K. J. Lee, W. S. Jeong, "An Analysis of the Economic Effects for the Immersive Media Industry," J. KICS. vol. 36, no. 7, pp. 795-805, Jul. 2011.
- [3] "VR Industry Trends Survey", Jan. 2017 [Online]. Available: <https://sketchfab.com/trends/q2-2017>
- [4] Y. Ban, L. Xie, "CUB360: Exploiting Cross-Users Behaviors for Viewport Prediction in 360 Video Adaptive Streaming," in Proc. of the IEEE International Conference on Multimedia and Expo, July. 2018.
- [5] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," IEEE Multimedia. vol. 18, no.4, pp. 62-67, Nov. 2011.
- [6] "WebVR API" [Online]. Available: <https://webvr.info/>
- [7] C. Ozcinar, A. D. Abreu, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in Proc. of the IEEE International Conference on Image Processing, Feb. 2017.
- [8] E. Jeong. (2019). A Study on Transmission Optimization Using Viewport Prediction Based on Sound Source Location Information in 360VR Contents. Seoul National University of Science and Technology, Seoul, Korea.
- [9] J. Brutlag, "Speed Matters," June 2009 [Online]. Available: <https://ai.googleblog.com/2009/06/speed-matters.html>

저 자 소 개



정 은 영

- 2016년 2월 : 서울과학기술대학교 전자IT미디어공학과 학사
- 2019년 2월 : 서울과학기술대학교 미디어IT공학과 공학석사
- 2019년 2월 ~ 현재 : LG유플러스
- ORCID : <https://orcid.org/0000-0002-9405-2842>
- 주관심분야 : 멀티미디어통신, 실감미디어 전송기술



김 동 호

- 1997년 2월 : 연세대학교 전자공학과 학사
- 1999년 2월 : 한국과학기술원(KAIST) 전기 및 전자공학과 공학석사
- 2004년 8월 : 한국과학기술원(KAIST) 전기 및 전자공학과 공학박사
- 2004년 9월 ~ 2007년 2월 : 삼성종합기술원 및 삼성전자 책임연구원
- 2007년 3월 ~ 현재 : 서울과학기술대학교 전자IT미디어공학과 교수
- ORCID : <https://orcid.org/0000-0001-9136-8932>
- 주관심분야 : 멀티미디어통신, 무선이동통신, 실감미디어 전송기술